The SAGE Encyclopedia of
SOCIAL SCIENCE RESEARCH METHODS

The SAGE Encyclopedia of SOCIAL SCIENCE RESEARCH METHODS

VOLUME 1

MICHAEL S. LEWIS-BECK
Department of Political Science, University of Iowa

ALAN BRYMAN
Department of Social Sciences, Loughborough University

TIM FUTING LIAO
Department of Sociology, University of Essex and University of Illinois, Urbana-Champaign

EDITORS

A SAGE Reference Publication

SAGE Publications
International Educational and Professional Publisher
Thousand Oaks ▪ London ▪ New Delhi

For information:

SAGE Publications, Inc.
2455, Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
6, Bonhill Street
London EC2A 4PU
United Kingdom

SAGE Publications India Pvt. Ltd.
B-42, Panchsheel Enclave
Post Box 4109
New Delhi 110 017 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Lewis-Beck, Michael S.
The SAGE encyclopedia of social science research methods/Michael S.
Lewis-Beck, Alan Bryman, Tim Futing Liao.
p. cm.
Includes bibliographical references and index.
ISBN 0–7619–2363–2 (cloth)
1. Social sciences—Research—Encyclopedias. 2. Social sciences—Methodology—Encyclopedias.
I. Bryman, Alan. II. Liao, Tim Futing. III. Title.
H62.L456 2004
300.!72—dc22

2003015882

03 04 05 06 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor:	Chris Rojek
Publisher:	Rolf A. Janke
Developmental Editor:	Eileen Haddox
Editorial Assistant:	Sara Tauber
Production Editor:	Melanie Birdsall
Copy Editors:	Gillian Dickens, Liann Lech, A. J. Sobczak
Typesetter:	C&M Digitals (P) Ltd.
Proofreader:	Mattson Publishing Services, LLC
Indexer:	Sheila Bodell
Cover Designer:	Ravi Balasuriya

Contents

List of Entries

The Appendix, Bibliography, appears at the end of this volume as well as in Volume 3.

Reader's Guide

Effects Coefficient
Endogenous Variable
Exogenous Variable
Independent Variable
Path Analysis
Structural Equation Modeling

DISCOURSE/CONVERSATION ANALYSIS

Accounts
Conversation Analysis
Critical Discourse Analysis
Deviant Case Analysis
Discourse Analysis
Foucauldian Discourse Analysis
Interpretative Repertoire
Proof Procedure

ECONOMETRICS

ARIMA
Cointegration
Durbin-Watson Statistic
Econometrics
Fixed Effects Model
Mixed-Effects Model
Panel
Panel Data Analysis
Random-Effects Model
Selection Bias
Serial Correlation (Regression)
Time-Series Cross-Section (TSCS) Models
Time-Series Data (Analysis/Design)
Tobit Analysis

EPISTEMOLOGY

Constructionism, Social
Epistemology
Idealism
Interpretivism
Laws in Social Science
Logical Positivism
Methodological Holism
Naturalism
Objectivism
Positivism

ETHNOGRAPHY

Autoethnography
Case Study
Creative Analytical Practice (CAP) Ethnography
Critical Ethnography
Ethnographic Content Analysis
Ethnographic Realism
Ethnographic Tales
Ethnography
Participant Observation

EVALUATION

Applied Qualitative Research
Applied Research
Evaluation Research
Experiment
Heuristic Inquiry
Impact Assessment
Qualitative Evaluation
Randomized Control Trial

EVENT HISTORY ANALYSIS

Censoring and Truncation
Event History Analysis
Hazard Rate
Survival Analysis
Transition Rate

EXPERIMENTAL DESIGN

Experiment
Experimenter Expectancy Effect
External Validity
Field Experimentation
Hawthorne Effect
Internal Validity
Laboratory Experiment
Milgram Experiments
Quasi-Experiment

FACTOR ANALYSIS AND RELATED TECHNIQUES

Cluster Analysis
Commonality Analysis
Confirmatory Factor Analysis
Correspondence Analysis
Eigenvalue
Exploratory Factor Analysis
Factor Analysis
Oblique Rotation
Principal Components Analysis
Rotated Factor

FEMINIST METHODOLOGY

GENERALIZED LINEAR MODELS

HISTORICAL/COMPARATIVE

INTERVIEWING IN QUALITATIVE RESEARCH

LATENT VARIABLE MODEL

LIFE HISTORY/BIOGRAPHY

LOG-LINEAR MODELS (CATEGORICAL DEPENDENT VARIABLES)

LONGITUDINAL ANALYSIS

MATHEMATICS AND FORMAL MODELS

MEASUREMENT LEVEL

Attribute
Binary
Categorical
Continuous Variable
Dichotomous Variables
Discrete
Interval
Level of Measurement
Metric Variable
Nominal Variable
Ordinal Measure

MEASUREMENT TESTING AND CLASSIFICATION

Conceptualization, Operationalization, and Measurement
Generalizability Theory
Item Response Theory
Likert Scale
Multiple-Indicator Measures
Summated Rating Scale

MULTILEVEL ANALYSIS

Contextual Effects
Dependent Observations
Fixed Effects Model
Mixed-Effects Model
Multilevel Analysis
Nonparametric Random-Effects Model
Random-Coefficient Model
Random-Effects Model

MULTIPLE REGRESSION

Adjusted R-Squared
Best Linear Unbiased Estimator
Beta
Generalized Least Squares
Heteroskedasticity
Interaction Effect
Misspecification
Multicollinearity
Multiple Regression Analysis
Nonadditive
R-Squared
Regression
Regression Diagnostics
Specification
Standard Error of the Estimate

QUALITATIVE DATA ANALYSIS

Analytic Induction
CAQDAS
Constant Comparison
Grounded Theory
In Vivo Coding
Memos, Memoing
Negative Case
Qualitative Content Analysis

SAMPLING IN QUALITATIVE RESEARCH

Purposive Sampling
Sampling in Qualitative Research
Snowball Sampling
Theoretical Sampling

SAMPLING IN SURVEYS

Multistage Sampling
Quota Sampling
Random Sampling
Representative Sample
Sampling
Sampling Error
Stratified Sampling
Systematic Sampling

SCALING

Attitude Measurement
Bipolar Scale
Dimension
Dual Scaling
Guttman Scaling
Index
Likert Scale
Multidimensional Scaling (MDS)
Optimal Scaling
Scale
Scaling
Semantic Differential Scale
Thurstone Scaling

SIGNIFICANCE TESTING

Alpha, Significance Level of a Test
Confidence Interval

SIMPLE REGRESSION

SURVEY DESIGN

TIME SERIES

Contributors

Abdi, Hervé
University of Texas,
Dallas

Adair, John G.
University of Manitoba,
Canada

Aditya, Ram N.
Florida International University,
Miami

Adler, Peter
University of Denver,
Colorado

Adler, Patricia A.
University of Colorado,
Boulder

Aiken, Leona S.
Arizona State University,
Tempe

Albritton, Robert B.
University of Mississippi

Aldridge, Alan
University of Nottingham,
United Kingdom

Allison, Paul D.
University of Pennsylvania,
Philadelphia

Altheide, David L.
Arizona State University,
Tempe

Altman, Micah
Harvard University,
Massachusetts

Alvesson, Mats
University of Lund,
Sweden

Amoureux, Jacque L.
University of Iowa, Iowa City

Andersen, Robert
McMaster University,
Canada

Anderson, Carolyn J.
University of Illinois,
Urbana-Champaign

Angle, John
US Department of Agriculture,
Washington, DC

Angrosino, Michael V.
University of South Florida,
Tampa

Anselin, Luc
University of Illinois,
Urbana-Champaign

Aronow, Edward
Montclair State University,
New Jersey

Ashmore, Malcolm
Loughborough University,
United Kingdom

Atkinson, Robert
University of Southern Maine,
Gorham

Atkinson, Rowland
University of Glasgow,
United Kingdom

Bacher, Johann
Universität Erlangen-Nürnberg,
Germany

Baird, Chardie L.
Florida State University,
Tallahassee

Bakeman, Roger
Georgia State University,
Atlanta

Bakker, Ryan
University of Florida,
Gainesville

Baltagi, Badi H.
Texas A&M University

Barr, Robert
University of Manchester,
United Kingdom

Bartunek, Jean M.
Boston College,
Massachusetts

Batson, C. Daniel
University of Kansas,
Lawrence

Beck, Nathaniel
University of California, San Diego;
New York University

Becker, Saul
Loughborough University,
United Kingdom

Becker, Mark P.
University of Minnesota,
Minneapolis

Benton, Ted
University of Essex,
United Kingdom

Berk, Richard A.
University of California,
Los Angeles

Billig, Michael
Loughborough University,
United Kingdom

Bishop, George F.
University of Cincinnati,
Ohio

Black, Thomas R.
University of Surrey,
United Kingdom

Blaikie, Norman
Formerly at the University of Science,
Malaysia and the RMIT University,
Australia

Blascovich, Jim
University of California,
Santa Barbara

Boehmke, Frederick J.
University of Iowa,
Iowa City

Boklund-Lagopoulos, Karin
Aristotle University of Thessaloniki,
Greece

Bollen, Kenneth A.
University of North Carolina,
Chapel Hill

Bornat, Joanna
The Open University,
United Kingdom

Bornstein, Robert F.
Gettysburg College,
Pennsylvania

Bottorff, Joan L.
University of British Columbia,
Canada

Bourque, Linda B.
University of California, Los Angeles

Box-Steffensmeier, Janet M.
Ohio State University, Columbus

Brady, Ivan
State University of New York, Oswego

Brennan, Robert L.
University of Iowa, Iowa City

Brewer, Marilynn B.
Ohio State University, Columbus

Brick, J. Michael
Westat, Rockville, Maryland

Brookfield, Stephen D.
University of St. Thomas, Minneapolis-St. Paul

Brown, Courtney
Emory University, Atlanta

Brown, Steven R.
Kent State University

Bryman, Alan
Loughborough University, United Kingdom

Buck, Nicholas H.
University of Essex, United Kingdom

Burkhart, Ross E.
Boise State University, Idaho

Burr, Vivien
University of Huddersfield, United Kingdom

Byrne, David
University of Durham, United Kingdom

Cameron, Amna
North Carolina State University, Raleigh

Carless, Sally A.
Monash University, Australia

Carmines, Edward G.
Indiana University, Bloomington

Chapin, Tim
Florida State University, Tallahassee

Charlton, Tony
University of Gloucestershire, United Kingdom

Charlton, John R. H.
Office for National Statistics, United Kingdom

Charmaz, Kathy
Sonoma State University, California

Chartrand, Tanya L.
Duke University, North Carolina

Chell, Elizabeth
University of Southampton, United Kingdom

Chen, Peter Y.
Colorado State University, Fort Collins

Christ, Sharon L.
University of North Carolina, Chapel Hill

Ciambrone, Desirée
Rhode Island College, Providence

Civettini, Andrew J.
University of Iowa, Iowa City

Clark, M. H.
The University of Memphis, Tennessee

Clarke, Harold D.
University of Texas, Dallas

Clegg, Chris
University of Sheffield, United Kingdom

Clinton, Joshua D.
Princeton University, New Jersey

Coffey, Amanda
Cardiff University, United Kingdom

Cole, Michael
University of California, San Diego

Cooper, Harris M.
Duke University, North Carolina

Copp, Martha A.
East Tennessee State University, Johnson City

Corbin, Juliet M.
The University of Alberta, Canada

Corter, James E.
Columbia University, New York

Corti, Louise
University of Essex, United Kingdom

Cortina, Jose M.
George Mason University, Virginia

Couper, Mick P.
University of Michigan, Ann Arbor

Coxon, Anthony P. M.
University of Edinburgh, United Kingdom

Craib, Ian
Late of University of Essex, United Kingdom

Cramer, Duncan
Loughborough University, United Kingdom

Crano, William D.
Claremont Graduate University, California

Croll, Paul
The University of Reading, United Kingdom

Croon, Marcel A.
Tilburg University, The Netherlands

Currivan, Douglas B.
University of Massachusetts, Boston

Dale, Angela
University of Manchester, United Kingdom

De Boef, Suzanna
Pennsylvania State University, University Park

de Vaus, David A.
La Trobe University, Australia

Denzin, Norman K.
University of Illinois, Urbana-Champaign

Dixon-Woods, Mary
University of Leicester, United Kingdom

Doreian, Patrick
London School of Economics,
United Kingdom

Dovidio, John F.
Colgate University,
New York

Dressler, William W.
The University of Alabama,
Tuscaloosa

Drew, Paul
University of York,
United Kingdom

Easterby-Smith, Mark
Lancaster University,
United Kingdom

Edwards, Derek
Loughborough University,
United Kingdom

Edwards, W. Sherman
Westat, Rockville,
Maryland

Eisenhardt, Kathleen M.
Stanford University,
California

Eliason, Scott R.
University of Minnesota,
Minneapolis

Ellis, Charles H.
Ohio State University,
Columbus

Essed, Philomena
University of Amsterdam,
The Netherlands

Fairclough, Norman
Lancaster University,
United Kingdom

Fetterman, David M.
Stanford University,
California

Filmer, Paul
Goldsmiths' College,
United Kingdom

Finkel, Steven E.
University of Virginia,
Charlottesville

Flint, John
University of Glasgow,
United Kingdom

Foley, Hugh J.
Skidmore College,
New York

Fontana, Andrea
University of Nevada,
Las Vegas

Fowler, Jr., Floyd J.
University of Massachusetts,
Boston

Fox, James Alan
Northeastern University,
Massachusetts

Fox, John
McMaster University,
Canada

Franzosi, Roberto
University of Reading,
United Kingdom

Frederick, Richard I.
US Medical Center for Federal Prisoners,
Missouri

Freedman, David A.
University of California,
Berkeley

Frey, James H.
University of Nevada,
Las Vegas

Friendly, Michael
York University,
Canada

Fuller, Duncan
University of Northumbria at Newcastle, United Kingdom

Furbee, N. Louanna
University of Missouri, Columbia

Gallagher, Patricia M.
University of Massachusetts, Boston

Gallmeier, Charles P.
Indiana University Northwest, Gary

Garson, G. David
North Carolina State University, Raleigh

Gephart, Jr., Robert P.
University of Alberta, Canada

Gergen, Kenneth J.
Swarthmore College, Pennsylvania

Gerhardt, Uta
Universität Heidelberg, Germany

Gershuny, Jonathan
University of Essex, United Kingdom

Gibbons, Jean D.
University of Alabama

Gibbs, Graham R.
University of Huddersfield, United Kingdom

Gibson, Nancy
University of Alberta, Canada

Gilbert, Nigel
University of Surrey, United Kingdom

Gill, Jeff
University of Florida, Gainesville

Glasgow, Garrett
University of California, Santa Barbara

Glenn, Norval D.
University of Texas, Austin

Goldstone, Jack
University of California, Davis

Green, Donald P.
Yale University, Connecticut

Greenacre, Michael
Universitat Pomepu Fabra, Spain

Greenland, Sander
University of California, Los Angeles

Greenwood, Davydd J.
Cornell University, New York

Griffith, James W.
Northwestern University, Illinois

Gross, Alan L.
City University of New York

Gschwend, Thomas
University of Mannheim, Germany

Guba, Egon G.
Indiana University

Gubrium, Jaber F.
University of Missouri, Columbia

Gujarati, Damodar N.
United States Military Academy, West Point

Guo, Guang
University of North Carolina, Chapel Hill

Guy, Will
University of Bristol, United Kingdom

Hagenaars, Jacques A.
Tilburg University, The Netherlands

Hagle, Timothy M.
University of Iowa, Iowa City

Hammersley, Martyn
The Open University, United Kingdom

Han, Shin-Kap
University of Illinois, Urbana-Champaign

Hancock, Gregory R.
University of Maryland, College Park

Handcock, Mark S.
University of Washington, Seattle

Hansen, Peter Reinhard
Brown University, Rhode Island

Hardy, Melissa A.
Pennsylvania State University, State College

Hargens, Lowell L.
University of Washington, Seattle

Harper, Douglas
Duquesne University, Pennsylvania

Harré, Rom
University of Oxford, United Kingdom

Headland, Thomas N.
SIL International, Dallas, Texas

Henry, Gary T.
Georgia State University, Atlanta

Herron, Michael C.
Dartmouth College, New Hampshire

Hessling, Robert M.
University of Wisconsin, Milwaukee

Hine, Christine M.
University of Surrey United Kingdom

Hipp, John R.
University of North Carolina, Chapel Hill

Hocutt, Max
The University of Alabama, Tuscaloosa

Hollway, Wendy
The Open University, United Kingdom

Holman, Darryl J.
University of Washington, Seattle

Holman, David J.
University of Sheffield, United Kingdom

Holstein, James A.
Marquette University, Wisconsin

Homan, Roger
University of Brighton, United Kingdom

Horrace, William C.
Syracuse University,
New York

House, Ernest R.
University of Colorado,
Boulder

Hoyle, Rick H.
Duke University,
North Carolina

Hughes, Rhidian
King's College London,
United Kingdom

Hundley, Vanora
University of Aberdeen,
United Kingdom

Iversen, Gudmund R.
Swarthmore College,
Pennsylvania

Jaccard, James J.
State University of New York,
Albany

Jackman, Simon
Stanford University,
California

Jacoby, William G.
Michigan State University,
East Lansing

Jasso, Guillermina
New York University

Jefferis, Valerie E.
Ohio State University,
Columbus

Jefferson, Tony
Keele University,
United Kingdom

Jenkins, Richard
University of Sheffield,
United Kingdom

Johnson, Jeffrey C.
East Carolina University,
Greenville

Johnson, Paul E.
University of Kansas,
Lawrence

Jones, Michael Owen
University of California,
Los Angeles

Kanuha, Valli Kalei
University of Hawaii,
Honolulu

Kaplan, David W.
University of Delaware,
Newark

Karuntzos, Georgia T.
RTI International, Research Triangle Park,
North Carolina

Katz, Jonathan N.
California Institute of Technology,
Pasadena

Kedar, Orit
University of Michigan,
Ann Arbor

Kennedy, Peter E.
Simon Fraser University,
Canada

Kenny, David A.
University of Connecticut,
Storrs

Kimball, David C.
University of Missouri,
St. Louis

King, Jean A.
University of Minnesota, Minneapolis

Kirk, Roger E.
Baylor University,
Texas

Knoke, David H.
University of Minnesota, Minneapolis

Kolen, Michael J.
University of Iowa, Iowa City

Korczynski, Marek
Loughborough University, United Kingdom

Kovtun, Mikhail
Duke University, North Carolina

Krauss, Autumn D.
Colorado State University, Fort Collins

Kroonenberg, Pieter M.
Leiden University, The Netherlands

Krueger, Richard A.
University of Minnesota, St. Paul

Kuhn, Randall
University of Colorado, Boulder

Kuzel, Anton J.
Virginia Commonwealth University, Richmond

Lagopoulos, Alexandros
Aristotle University of Thessaloniki, Greece

Langeheine, Rolf
Formerly at the University of Kiel, Germany

Laurie, Heather
University of Essex, United Kingdom

Lauriola, Marco
University of Rome, "La Sapienza" Italy

Lavrakas, Paul J.
Nielsen Media Research, New York

Le Voi, Martin
The Open University, United Kingdom

Leahey, Erin
University of Arizona, Tucson

Ledolter, Johannes
University of Iowa, Iowa City

Lee, Raymond M.
Royal Holloway University of London, United Kingdom

Lee (Zhen Li), Jane
University of Florida, Gainesville

Levin, Irwin P.
University of Iowa, Iowa City

Lewis-Beck, Michael S.
University of Iowa, Iowa City

Liao, Tim Futing
University of Essex, United Kingdom; University of Illinois, Urbana-Champaign

Lincoln, Yvonna S.
Texas A&M University

Little, Daniel
University of Michigan, Dearborn

Little, Jani S.
University of Colorado, Boulder

Liu, Cong
Illinois State University, Normal

Lockyer, Sharon
De Montfort University,
United Kingdom

Long, Karen J.
University of Michigan,
Ann Arbor

Lowe, Will
University of Bath,
United Kingdom

Lykken, David T.
University of Minnesota,
Minneapolis

Lynch, Michael
Cornell University,
New York

Lynn, Peter
University of Essex,
United Kingdom

Madsen, Douglas
University of Iowa,
Iowa City

Magidson, Jay
Statistical Innovations,
Belmont,
Massachusetts

Mahoney, Christine
Pennsylvania State University,
University Park

Mahoney, James
Brown University,
Rhode Island

Maines, David R.
Oakland University,
Michigan

Manton, Kenneth G.
Duke University,
North Carolina

Marsden, Peter V.
Harvard University,
Massachuesetts

Marsh, Lawrence C.
University of Notre Dame,
Indiana

Mason, Jennifer
University of Leeds,
United Kingdom

Mathiowetz, Nancy
University of Wisconsin,
Milwaukee

Maxim, Paul S.
University of Western Ontario,
Canada

May, Tim
University of Salford,
United Kingdom

Maynard, Mary
University of York,
United Kingdom

McDowall, David
State University of New York,
Albany

McElhanon, Kenneth A.
Graduate Institute of Applied Linguistics,
Texas

McGraw, Kenneth O.
University of Mississippi

McGuigan, Jim
Loughborough University,
United Kingdom

McIver, John P.
University of Colorado,
Boulder

McKelvie, Stuart J.
Bishop's University,
Canada

Meadows, Lynn M.
University of Calgary, Canada

Menard, Scott
University of Colorado, Boulder

Merrett, Mike C.
University of Essex, United Kingdom

Mertens, Donna M.
Gallaudet University, Washington, DC

Miller, Arthur G.
Miami University, Ohio

Millward, Lynne
University of Surrey, United Kingdom

Moffett, Kenneth W.
University of Iowa, Iowa City

Mooney, Christopher Z.
University of Illinois, Springfield

Moreno, Erika
University of Iowa, Iowa City

Morris, Martina
University of Washington, Seattle

Morse, Janice M.
University of Alberta, Canada

Moyé, Lemuel A.
University of Texas, Houston

Mueller, Charles W.
University of Iowa, Iowa City

Mueller, Daniel J.
Indiana University, Bloomington

Mueller, Ralph O.
The George Washington University, Washington, DC

Mukhopadhyay, Kausiki
University of Denver, Colorado

Mulaik, Stanley A.
Georgia Institute of Technology, Atlanta

Nagler, Jonathan
New York University

Nash, Chris
University of Technology, Australia

Nathan, Laura
Mills College, California

Nishisato, Shizuhiko
University of Toronto, Canada

Noblit, George W.
University of North Carolina, Chapel Hill

Norpoth, Helmut
State University of New York, Stony Brook

Noymer, Andrew
University of California, Berkeley

Oakley, Ann
University of London, United Kingdom

Olsen, Randall J.
Ohio State University, Columbus

Olsen, Wendy K.
University of Manchester, United Kingdom

Ondercin, Heather L.
Pennsylvania State University, University Park

Outhwaite, R. William
University of Sussex, United Kingdom

Page, Stewart
University of Windsor, Canada

Palmer, Harvey D.
University of Mississippi

Paluck, Elizabeth Levy
Yale University, Connecticut

Pampel, Fred
University of Colorado, Boulder

Parcero, Osiris J.
Bristol University, United Kingdom

Park, Sung Ho
Texas A&M University

Parry, Ken W.
Griffith University, Australia

Pattison, Philippa
University of Melbourne, Australia

Patton, Michael Quinn
The Union Institute & University, Utilization-Focused Evaluation, Minnesota

Paul, Pallab
University of Denver, Colorado

Pavlichev, Alexei
North Carolina State University, Raleigh

Paxton, Pamela
Ohio State University, Columbus

Peng, Chao-Ying Joanne
Indiana University, Bloomington

Phillips, Nelson
University of Cambridge, United Kingdom

Phua, VoonChin
City University of New York

Pickering, Michael J.
Loughborough University, United Kingdom

Pink, Sarah
Loughborough University, United Kingdom

Platt, Lucinda
University of Essex, United Kingdom

Platt, Jennifer
University of Sussex, United Kingdom

Plummer, Ken
University of Essex, United Kingdom

Poland, Blake D.
University of Toronto, Canada

Pollins, Brian M.
The Mershon Center, Columbus, Ohio

Post, Alison E.
Harvard University, Massachusetts

Potter, Jonathan A.
Loughborough University, United Kingdom

Prosch, Bernhard
Universität Erlangen-Nürnberg, Germany

Pye, Paul E.
University of Teesside, United Kingdom

Quinn, Kevin M.
Harvard University, Massachusetts

Ragin, Charles C.
University of Arizona, Tucson

Rässler, Susanne
Universität Erlangen-Nürenberg, Germany

Redlawsk, David P.
University of Iowa, Iowa City

Reis, Harry T.
University of Rochester, New York

Richardson, Laurel
Ohio State University, Columbus

Riessman, Catherine K.
Boston University, Massachusetts

Ritzer, George
University of Maryland, College Park

Robinson, Dawn T.
University of Iowa, Iowa City

Rodgers, Joseph Lee
University of Oklahoma, Norman

Rodríguez, Germán
Princeton University, New Jersey

Roncek, Dennis W.
University of Nebraska, Omaha

Rosenthal, Robert
University of California, Riverside

Rosnow, Ralph L.
Temple University, Pennsylvania

Rothstein, Hannah R.
Baruch College, New York

Rubin, Donald B.
Harvard University, Massachusetts

Rudas, Tamás
Eotvos Lorand University, Hungary

Sandelowski, Margarete
University of North Carolina, Chapel Hill

Santos, Filipe M.
INSEAD, France

Sapsford, Roger
University of Teesside, United Kingdom

Sasaki, Masamichi
Hyogo Kyoiku University, Japan

Schafer, William D.
University of Maryland, College Park

Schelin, Shannon Howle
North Carolina State University, Raleigh

Schenker, Nathaniel
National Center for Health Statistics, Maryland

Schmidt, Tara J.
University of Wisconsin,
Milwaukee

Schrodt, Philip A.
University of Kansas,
Lawrence

Scott, Steven L.
University of Southern California,
Los Angeles

Scott, John
University of Essex,
United Kingdom

Seale, Clive
Brunel University,
United Kingdom

Shadish, William R.
University of California,
Merced

Shaffir, William
McMaster University,
Canada

Shannon, Megan L.
University of Iowa,
Iowa City

Shipan, Charles R.
University of Iowa,
Iowa City

Shu, Xiaoling
University of California,
Davis

Sieber, Joan E.
California State University,
Hayward

Sijtsma, Klaas
Tilburg University,
The Netherlands

Silver, N. Clayton
University of Nevada,
Las Vegas

Slagter, Tracy Hoffmann
University of Iowa,
Iowa City

Smith, Peter W. F.
University of Southampton,
United Kingdom

Smith, John K.
University of Northern Iowa,
Cedar Falls

Smithson, Michael
Australian National University,
Australia

Snijders, Tom A. B.
University of Groningen,
The Netherlands

Spector, Paul E.
University of South Florida,
Tampa

Standing, Lionel G.
Bishop's University,
Canada

Stark, Philip B.
University of California,
Berkeley

Stevens, Gillian
University of Illinois,
Urbana-Champaign

Stillman, Todd
University of Maryland,
College Park

Stokes, S. Lynne
Southern Methodist University,
Texas

Stovel, Katherine
University of Washington,
Seattle

Strohmetz, David B.
Monmouth University,
New Jersey

Stuart, Elizabeth A.
Harvard University,
Massachusetts

Sutton, Alex
University of Leicester,
United Kingdom

Sweeney, Kevin J.
Ohio State University,
Columbus

Swicegood, C. Gray
University of Illinois,
Urbana-Champaign

Tacq, Jacques
Catholic University of Brussels,
Belgium

Tcherni, Maria
Northeastern University,
Massachusetts

Thomas, Jim
Northern Illinois University,
DeKalb

Thompson, Bruce
Texas A&M University

Thorne, Sally E.
University of British Columbia,
Canada

Tien, Charles
City University of New York

Ting, Kwok-fai
The Chinese University of Hong Kong

Toothaker, Larry E.
University of Oklahoma,
Norman

Tracy, Karen
University of Colorado,
Boulder

Traugott, Michael W.
University of Michigan,
Ann Arbor

Traxel, Nicole M.
University of Wisconsin,
Milwaukee

Tsuji, Ryuhei
Meiji Gakuin University,
Japan

Tuchman, Gaye
University of Connecticut,
Storrs

Tuma, Nancy Brandon
Stanford University,
California

Vallet, Louis-André
LASMAS-Institut du Longitudinal (CNRS) and Laboratoire de Sociologie Quantitative (CREST),
France

van de Pol, Frank J. R.
Statistics Netherlands,
The Netherlands

van der Ark, L. Andries
Tilburg University,
The Netherlands

van Manen, Max
University of Alberta,
Canada

van Teijlingen, Edwin R.
University of Aberdeen,
United Kingdom

Vann, Irvin B.
North Carolina State University,
Raleigh

Vermunt, Jeroen K.
Tilburg University,
The Netherlands

von Hippel, Paul T.
Ohio State University,
Columbus

Walker, Robert
University of Nottingham,
United Kingdom

Walsh, Susan
University of Sheffield,
United Kingdom

Wand, Jonathan
Harvard University,
Massachusetts

Wang, Cheng-Lung
Florida State University,
Tallahassee

Warren, Carol A. B.
University of Kansas,
Lawrence

Wasserman, Stanley
University of Illinois,
Urbana-Champaign

Weisberg, Herbert F.
Ohio State University,
Columbus

Weisfeld, Carol C.
University of Detroit Mercy,
Michigan

Wengraf, Tom
Middlesex University,
United Kingdom

Whitten, Guy
Texas A&M University

Wilcox, Rand R.
University of Southern California,
Los Angeles

Williams, Malcolm
University of Plymouth,
United Kingdom

Wood, B. Dan
Texas A&M University

Woods, James
West Virginia University,
Morgantown

Xie, Yu
University of Michigan,
Ann Arbor

Zimmerman, Marc
University of Houston,
Texas

Zorn, Christopher
National Science Foundation,
Arlington, Virginia

About the Editors

Michael S. Lewis-Beck is F. Wendell Miller Distinguished Professor of Political Science and Chair, at the University of Iowa. He has edited the SAGE green monograph series *Quantitative Applications in the Social Sciences* (QASS) since 1988. Also, he is Data Editor for the journal *French Politics.* He is past Editor of the *American Journal of Political Science.* Professor Lewis-Beck has authored or coauthored more than 110 books, articles, or chapters. His publications in methods include *Applied Regression: An Introduction* and *Data Analysis: An Introduction.* His publications in substantive areas of political science include *Economics and Elections: The Major Western Democracies, Forecasting Elections, How France Votes,* and *The French Voter: Before and After the 2002 Elections.* Besides the University of Iowa, he has taught quantitative methods courses at the Inter-university Consortium for Political and Social Research (ICPSR, University of Michigan); the European Consortium for Political Research (ECPR, University of Essex, England); the TARKI Summer School on Statistical Models (University of Budapest, Hungary); and the Ecole d'Eté de Lille (University of Lille, France). Also, Professor Lewis-Beck has held visiting appointments at the Institut d'Etudes Politiques (Sciences Po) in Paris; the Harvard University Department of Government; the University of Paris I (Sorbonne); the Catholic University in Lima, Peru; and the National Institute for Development Administration in Guatemala City. He speaks French and Spanish. His current research interests include election forecasting, and the modeling of coup behavior in Latin America.

Alan Bryman is Professor of Social Research in the Department of Social Sciences, Loughborough University, England. His main research interests lie in research methodology, leadership studies, organizational analysis, the process of Disneyization, human-animal relations, and theme parks. He is author or coauthor of many books, including *Quantity and Quality in Social Research* (Routledge, 1988), *Charisma and Leadership in Organizations* (SAGE, 1992), *Disney and His Worlds* (Routledge, 1995), *Mediating Social Science* (SAGE, 1998), *Quantitative Data Analysis With SPSS Release 10 for Windows: A Guide for Social Scientists* (Routledge, 2001), *Social Research Methods* (Oxford University Press, 2001; rev. ed., 2004), *Business Research Methods* (Oxford University Press, 2003), and *Disneyization of Society* (SAGE, forthcoming). He is also coeditor of *The Handbook of Data Analysis* (SAGE, forthcoming). He is the editor of the *Understanding Social Research* series for Open University Press and is on the editorial board of *Leadership Quarterly.*

Tim Futing Liao is Professor of Sociology and Statistics at the University of Illinois, Urbana-Champaign, and currently (2001–2003) teaches research methods in the Department of Sociology at the University of Essex, UK. His research focuses on methodology, demography, and comparative and historical family studies. He has authored *Interpreting Probability Models* (SAGE, 1994), *Statistical Group Comparison* (Wiley, 2002), and research articles in both methods and disciplinary journals. Professor Liao served two terms (1992–1996 & 1996–2000) as Deputy Editor of *The Sociological Quarterly,* and he is on the Editorial Board of *Sociological Methods & Research* and the Steering Committee of the ESRC Oxford Spring School in Quantitative Methods for Social Research. His substantive interest led him to a year as visiting research fellow with the Cambridge Group for the History of Population and Social Structure, and his methodological expertise led him to teach summer schools at Peking University, Chinese Academy of Social Sciences, and the University of Essex. Having a catholic interest in methodology as well as in life, he has lived in or visited more than 20 nations, has done research on five societies, has held formal academic appointments in three educational systems, and speaks three and three-quarter languages.

Preface

These volumes comprise an encyclopedia of social science research methods, the first of its kind. Uniqueness explains, at least partly, why we undertook the project. It has never been done before. We also believe that such an encyclopedia is needed. What is an encyclopedia? In ancient Greek, the word signifies "all-encompassing education." This reference set provides readers with an all-encompassing education in the ways of social science researchers. Why is this needed? No one researcher, expert or not, knows everything. He or she occasionally must look up some methodological point, for teaching as well as for research purposes. And it is not always obvious where to go, or what to look up when you get there. This encyclopedia brings together, in one place, authoritative essays on virtually all social science methods topics, both quantitative and qualitative.

A survey researcher may want to learn about longitudinal analysis in order to make sense of panel data. A student of household division of labor may want to learn the use of time diaries. A regression expert might bone up on multidimensional scaling, so as to be better able to aggregate certain measures. A deconstructionist might want to explore Foucauldian discourse analysis. An experimentalist may wish further understanding of the collinearity problem that nonexperimentalists routinely face. A feminist scholar may seek to know more about the use of films in research. A political scientist could have need of a refresher on experimental design. An anthropologist could raise a question about ethnographic realism. Perhaps a psychologist seeks to comprehend a particular measure of association. A philosopher might want to read more about the laws of social science. A sociologist may desire a review of the different approaches to evaluation research. Even in areas where one is highly expert, much can be gained. For example, in the examination of interaction effects, there are many entries from different perspectives, treating the little-used as well as the much-used techniques. These entries are written by acknowledged leading scholars. One would have to be "more expert" than the experts not to benefit from reading them.

Besides practicing researchers and social statisticians, the encyclopedia has appeal for more general readers. Who might these people be? First, there are students, graduate and undergraduate, in the social science classes in the universities of the world. All the major branches—anthropology, communications, economics, education, geography, political science, psychology, public health, public policy, sociology, urban planning—have their students reading articles and books that require an appreciation of method for their full comprehension. The typical undergraduate needs to know how to read a contingency table, interpret a correlation coefficient, or critique a survey. There are entries on these and other such topics that are completely accessible to this student population. These sorts of entries require *no special knowledge of mathematics or statistics* to be understood, nor does the student need to have actually done research in order to grasp the essentials.

Of course, graduate students in these fields are obliged to go beyond simple reading comprehension, to actual application of these techniques. Many of the pieces make the "how-to" of the method clear, such as some of the entries on regression analysis or interviewing techniques. Then, there are more advanced entries that will challenge, but still inform, the typical graduate student. For example, the pieces on Bayesian analysis, or generalized linear modeling, will certainly allow them to sharpen their academic teeth.

In addition to researchers and students, this encyclopedia will be helpful to college-educated readers who want to understand more about social science methodology because of their work (e.g., they are journalists or city managers) or because of their pleasure (e.g., they follow social science reporting in the media). Many entries are written in ordinary English, with no special math or stat requirements, and so will be accessible to such readers. Moreover, many of the more difficult entries are understandable in at least an introductory way, because of instructions to that effect given to our contributors. Having said that, it should be emphasized that none of the essays is "dumbed down." Instead, they are intelligently written by recognized experts. There is nothing "cookbook" or condescending about them. In this domain of style, the only essential restriction placed on the authors was that of length, for we urged them to present their material as concisely as possible.

Our instructions to authors provided four different lengths for the entries—2,500 words (A-level), 1,000 words (B-level), 500 words (C-level), and 50 words (D-level). Of course, certain individual entries—whether A, B, C, or D level—do not conform exactly to these word counts, which are rather more guidelines than rigid standards. Length of entry was finally dictated by the importance, breadth, or, in some cases, complexity of the topic. In general, the longer the entry, the more central the topic. Examples of A-level entries are Analysis of Variance, Bayesian Inference, Case Study, and Regression. B-level topics are also important, but do not demand quite so much space to explain the concept clearly. B-level topics include Historical Method, Inferential Statistics, Misspecification, and Mixed Design. C-level entries usually treat a rather specific method or issue, such as Curvilinearity, Narrative Interview, Prisoner's Dilemma, and Record-Check Studies. The A, B, and C entries all end with references to guide the reader further. D-level entries, on the other hand, are brief and mainly definitional. Examples of D-level entries are Gini Coefficient, Nominal, Secondary Data, and Yates's Correction.

Altogether, there are about 1,000 entries, covering quantitative and qualitative methods, as well as the connections between them. The entries are cross-referenced to each other, as appropriate. For example, take this entry—Discrete (see Categorical, Nominal, Attribute). The number of entries, "1,000," is not magic, nor was it a goal, but it is the number to which we kept returning in our search for a master list of entries. At the beginning of the project, we aimed to come up with an exhaustive list of all possible methods terms. We consulted written materials—subfield dictionaries, statistics encyclopedias, methods journals, publisher flyers—and individuals—members of our distinguished international board of scholars, editorial boards of methodology journals, colleagues at our home institutions and elsewhere, and, last but not least, our graduate students. As the process unfolded, some entries were culled, some consolidated, some added. We have included every methodological term we could think of that someone reading social science research results in a book, a journal, or a newspaper might come across. In the spirit of the first encyclopedia, *l'Encyclopédie,* by Diderot and d'Alembert in 1751, we did try "to examine everything" (*Il faut tout examiner*) that concerned our enterprise. Although something is undoubtedly "not examined," it remains for our readers to tell us what it is.

Once the entry list was basically set, we began contacting potential contributors. We were heartened by the extent of cooperation we received from the social science community. Senior and junior scholars alike gave us their best, and did so promptly. Remarkably, with a handful of exceptions, everyone met their deadlines. This is one measure, we believe, of the value these busy researchers place on this encyclopedia. As can be seen from the bylines, they compose a varied mix from the relevant disciplines, holding important university and research posts around the world.

Besides the contributors, the board, and the colleagues and students of our home institutions, there are others who are responsible for the success of the encyclopedia. Chris Rojek, senior editor at SAGE London, has to be credited with the initial idea of the project. Without that, it would not have happened. On the US side, C. Deborah Laughton, former senior editor at SAGE California, was a dedicated and tireless supporter of the encyclopedia idea from the very beginning. Without her enthusiasm and insight, the final product would have lacked much. Rolf Janke, vice president at SAGE, was a central source of organization strength and encouragement from the home office. Also at SAGE, Vincent Burns, technical editor, was vital in helping piece together the Web site, the database, and the massive manuscript data files. As well, we should mention Melanie Birdsall, production editor, who patiently shepherded us through the alphabet soup of page proofs. Finally, to save the best for last, we must thank Eileen Haddox, our developmental editor. Eileen made it work by doing the detail work. She sent out the contracts, reminded the authors, logged in the contributions, and sent text and corrections back and forth, in the end making the whole corpus ready for the printer. Always, she labored with efficiency and good cheer. In those realms, she set standards that we, the editors of the encyclopedia, tried to achieve. Sometimes we did.

Michael S. Lewis-Beck
Alan Bryman
Tim Futing Liao

Introduction

Social science methodology can be said to have entered our consciousness by the late 19th century when Emile Durkheim penned *The Rules of Sociological Method,* laying the foundation for the systematic understanding and analysis of social phenomena. From that time, it has seen continuous growth characterized by an exponential development over the last few decades of the 20th century. The development came in terms of both the width and the depth of our methodological know-how. The broad range of methods applicable in the wide circle of social science disciplines, and the sophistication and level in the advancement of some analytical approaches and techniques, would have been unthinkable merely a few score years ago. FOCUS GROUP and THICK DESCRIPTION, for example, have virtually become lingua franca among many social science researchers, regardless of their disciplinary orientation. The methods for dealing with MISSING DATA and NONRESPONSE, to take other examples, have advanced so much that specialized workshops on such issues are a perennial favorite among analysts of SURVEY DATA.

It is only natural, at the beginning of a new century, for us to take stock of the entire spectrum of our social science methodological knowledge, even though it is impossible and impractical to include every method that has ever been used in the social sciences. The cardinal aim of this encyclopedia is to provide our readers—be they students, academics, or applied researchers—with an introduction to a vast array of research methods by giving an account of their purposes, principles, developments, and applications. The approximately 1,000 entries, many of which are extensive treatments of the topics and contain recent developments, can be of great use to the novice or the experienced researcher alike.

To accomplish this goal, we offer two major types of entries: Some contain only a definition of no longer than a paragraph or two. These give the reader a quick explanation of a methodological term. True to the encyclopedic form, many other entries are topical treatments or essays that discuss—at varying lengths, often with examples and sometimes with graphics—the nature, the history, the application, and the implication of using a certain method. Most of these entries also give suggested readings and references for the reader to pursue a topic further. These are part and parcel of the encyclopedia and are invaluable for those who would like to delve into the wonder world of research methods. To help provide a more complete explanation than is often achieved within the scope of a single article, we employ small capital letters, such as those appearing in the first paragraph of this introduction, that refer the reader to related terms that are explained elsewhere in the encyclopedia.

With such a variety of specialized essays to write, we are fortunate to have been able to count on the support of our board members and authors, who contributed many a coherent introduction to a method with definitiveness and thoroughness, often with great flair as well. Sometimes, topics are treated in such a novel way that they are not only pleasurable but also thought-provoking to read. For instance, entries such as the essay on ECONOMETRICS by Professor Damodar Gujarati are a pleasant surprise. Rather than merely introducing the topic with the types of methods and models that econometricians use and nothing else, Gujarati takes us on a journey from the ordinary to the extraordinary. He begins with three quotations that illustrate the broad scope of econometrics; here the simple, usual approach of using quotations accomplishes the seemingly undoable task of defining the terrain on which econometricians work and play. He then walks us twice through the research process, from economic theory to data and models to analysis, once in principle and the second time with an example. Such a process is what many of us preach every day but seldom think of when writing an essay for an encyclopedia. Gujarati uses the ordinary process of going about economic research to achieve an extraordinary, profound impact—an impact that will leave a reader thinking about, instead of just the methods and models, the fundamental purpose of econometrics. Entries like this give us knowledge and food for thought.

The diversity of our entries is also great. To take one of many possible contrasts, some of our entries deal with very philosophical issues, such as POSTSTRUCTURALISM, that might appear to be out of step with a set of volumes concerned with methods of research, whereas others discuss advanced statistical techniques that might similarly be viewed as not part of social science research methodology. However, we have taken the view that both are necessary. On the one hand, all researchers need to be aware of the EPISTEMOLOGICAL issues that influence both the nature of RESEARCH QUESTIONS and the ASSUMPTIONS that underpin aspects of the research process; on the other hand, we all need to be aware of the full panoply of ways of analyzing quantitative data, so that the most appropriate and robust techniques can be applied in different situations. It is only when we are knowledgeable about the choices available to us—whether epistemological, statistical, or whatever—that we can completely develop our craft as social researchers.

Examples of excellent treatment of a whole host of topics abound in the following pages and volumes. By assembling into one encyclopedia entries of varied origin that serve different research purposes, we, the editors, hope that readers will come to appreciate the rich heritage of our social science methodology and, more importantly, will be able to benefit from this immense source of methodological expertise in advancing their research.

Michael S. Lewis-Beck
Alan Bryman
Tim Futing Liao

A

ABDUCTION

Abduction is the logic used to construct descriptions and explanations that are grounded in the everyday activities of, as well as in the language and meanings used by, social actors. Abduction refers to the process of moving from the way social actors describe their way of life to technical, social scientific descriptions of that social life. It has two stages: (a) describing these activities and meanings and (b) deriving categories and concepts that can form the basis of an understanding or an explanation of the problem at hand. Abduction is associated with INTERPRETIVISM.

The logic of abduction is used to discover why people do what they do by uncovering largely tacit, mutual knowledge and the symbolic meanings, motives, and rules that provide the orientations for their actions. Mutual knowledge is background knowledge that is largely unarticulated, constantly used and modified by social actors as they interact with each other, and produced and reproduced by them in the course of their lives together. It is the everyday beliefs and practices—the mundane and taken for granted—that have to be grasped and articulated by the social researcher to provide an understanding of these actions.

The concept of abduction has had limited use in philosophy and social science. Two writers, the philosopher C. S. Peirce (1931, 1934) and the sociologist David Willer (1967), have used it but with different meanings to that presented here; the latter is derived from Blaikie (1993). Although not using the concept of abduction, the logic was devised by Alfred Schütz (1963) and used by Jack Douglas, John Rex, and, particularly, Anthony Giddens (1976) (for a review of these writers, see Blaikie, 1993, pp. 163–168, 176–193).

What Schütz, Douglas, Rex, and Giddens have in common is their belief that social science accounts of the social world must be derived, at least initially, from the accounts that social actors can give of the aspect of their world of interest to the social scientist. They differ, however, in what is done with these accounts. Some argue that reporting everyday accounts is all that is possible or necessary to understand social life. Others are prepared to turn these accounts into social scientific descriptions of the way of life of a particular social group (community or society), but they insist on keeping these descriptions tied closely to social actors' language. Such descriptions lend themselves to two other possibilities. The first is to bring some existing THEORY or perspective to bear on them, thus providing a social scientific interpretation or critique of that way of life. The second is to generate some kind of explanation using as ingredients the IDEAL TYPES that are derived from everyday accounts. There are disagreements on both of these latter possibilities—in the first case, about whether criticism of another way of life is legitimate and, in the second case, about what else can or should be added to the everyday account to form a theory.

Interpretive social scientists use some version of abduction with a variety of research methods. However, they rarely articulate their logic. Appropriate methods for using abductive logic are still being developed.

—Norman Blaikie

REFERENCES

Blaikie, N. (1993). *Approaches to social enquiry.* Cambridge, UK: Polity.

Giddens, A. (1976). *New rules of sociological method.* London: Hutchinson.

Peirce, C. S. (1931). *Collected papers* (Vols. 1–2; C. Hartshorne & P. Weiss, Eds.). Cambridge, MA: Harvard University Press.

Peirce, C. S. (1934). *Collected papers* (Vols. 5–6, C. Hartshorne & P. Weiss, Eds.). Cambridge, MA: Harvard University Press.

Schütz, A. (1963). Concept and theory formation in the social sciences. In M. A. Natanson (Ed.), *Philosophy of the social sciences* (pp. 231–249). New York: Random House.

Willer, D. (1967). *Scientific sociology: Theory and method.* Englewood Cliffs, NJ: Prentice Hall.

ACCESS

Access refers to a researcher seeking entry to an environment in which primary research data may be unearthed or generated. Therefore, for researchers interested in analyzing primary data, whether qualitative or quantitative in nature, access is an inevitable stage in the research process. The ability to secure access is a necessary part of the researcher's craft. Despite this, the process of securing access is rarely systematically examined, with researchers often adopting ad hoc, opportunistic, and intuitive approaches to the negotiation of access. When examining the range of tactics that have been used by researchers in their attempts to secure access, it is useful to begin by considering what researchers are seeking access into.

ACCESS TO WHAT, AND VIA WHOM?

Access is often considered in relation to what have been termed *closed and open/public settings* (Bell, 1969). Closed settings are formally constituted settings, such as firms, hospitals, and churches, and their archives. Open/public settings are informal in nature. Research into street corner life or local music scenes is an example of research in open/public settings. A key step in securing access into both closed and open settings involves negotiation with *gatekeepers.* Gatekeepers are the people who, metaphorically, have the ability to open or close the gate to the researcher seeking access into the setting. A preliminary step in seeking access lies in identifying the gatekeepers. In formal, closed settings, the gatekeeper will tend to be a person in a senior position, such as a senior executive. Gatekeepers in informal, open settings are less easy to clearly identify but tend to have status and respect within the setting. Some researchers go directly to the gatekeeper in their attempts to secure access. Others approach the gatekeeper indirectly by first obtaining a "champion" or "sponsor" in the setting who can help persuade the gatekeeper to look favorably on the researcher's request for access.

OBTAINING ACCESS AS TRUST FORMATION AND NEGOTIATION

The ability of researchers to inspire sponsors, central gatekeepers, or dispersed individuals to trust them is central to the process of securing access. Obtaining access can be seen as a specific example of the wider social problem of trust creation. We can, therefore, usefully turn to the literature on trust creation to allow us a better and more systematic understanding of the process of securing access. Korczynski (2000) delineates four main bases of trust that may underlie an exchange. The most common tactics adopted by researchers in negotiating access can be mapped against these categories:

- *Trust based on personal relations.* It is common for researchers to use friends and contacts who know them as trustworthy as a way to obtain access. These friends and contacts may act as a conduit to sponsors, may be sponsors, or may actually be gatekeepers in the research settings.
- *Trust based on knowledge of the other party's internal norms.* Researchers may try to generate trust by giving signals to the gatekeeper or individual that they are "decent" people. For instance, seeking access to respondents by putting a notice in a newspaper or magazine associated with a particular moral, social, or political outlook implicitly invites the reader to associate the researcher with that newspaper's or magazine's norms and values.
- *Trust based on an incentive or governance structure surrounding the exchange.* Researchers can help generate trust by highlighting that it is in the interests of the researchers not to betray such trust. For instance, researchers may give gatekeepers the right to veto the publication of research findings (although such an approach leaves the researcher in a potentially precarious position).

• *Trust based on systems or institutions.* Researchers can seek to generate trust by invoking trust that gatekeepers or other individuals might have in institutions to which the researchers are affiliated. For instance, researchers may stress the quality of a university at which they are employed and highlight relevant research that has been produced by that university in the recent past.

The process of obtaining access should also be seen as a process of negotiation (Horn, 1996). Researchers typically want wide and deep access into a setting, whereas gatekeepers and other concerned individuals may be concerned about potential costs of granting such access. These potential conflicts of interest are typically informally aired in face-to-face meetings. In these implicit negotiations, researchers often offer the gatekeepers or individuals a benefit that would accrue from access being granted. In formal settings, this might take the form of a written report of the findings.

Trust generating and negotiating tend to be highly context-specific processes. This may go some way to explaining why research access tends to be discussed with an emphasis on context-specific tactics used by researchers. Few attempts have been made to construct a more abstract understanding of the process of securing research access.

ACCESS AS AN ONGOING PROCESS

For many qualitative researchers, the process of negotiating access is an ever-present part of the research process. For instance, researchers who wish to individually interview members of an informal group usually must negotiate access with each person in turn and cannot rely on cooperation flowing easily from an informal gatekeeper's implicit sanctioning of the research. Moreover, in formal settings, although respondents may feel the obligation to be interviewed by a researcher, they can still be reticent in their answers. In each case, the researcher needs to generate trust within the ongoing process of negotiating access.

ACCESS AND REACTIVITY

REACTIVITY refers to the presence of the researcher affecting the research environment. The way in which access is negotiated may affect reactivity significantly. For instance, in organizational research, if trust is only generated with high-level gatekeepers and not with lower-level staff who are included in the study, the mode of securing access is likely to significantly bias the research results. As Burgess (1984) notes, "Access influences the reliability and validity of the data that the researcher subsequently obtains" (p. 45).

—Marek Korczynski

REFERENCES

Bell, C. (1969). A note on participant observation. *Sociology, 3*, 417–418.

Burgess, R. (1984). *In the field.* London: Allen & Unwin.

Horn, R. (1996). Negotiating research access to organizations. *Psychologist, 9*(12), 551–554.

Korczynski, M. (2000). The political economy of trust. *Journal of Management Studies, 37*(1), 1–22.

ACCOUNTS

An account is a piece of talk, a way of explaining the self or close others to another person. In an influential essay, sociologists Scott and Lyman (1968) highlighted how talk is the fundamental medium through which people manage relations with others: Offering accounts is how people remedy a problem with another or, more generally, make reasonable any action. Accounts include two main types: (a) accounts *for* action and (b) accounts *of* action.

An account *for* action is a statement offered to explain inappropriate conduct. The inappropriate behavior may be a small act, such as forgetting to make a promised telephone call, being late, or making an insensitive joke, or it may be a large transgression such as taking money without asking, having an affair, or lying about an important matter. Accounting is also done by one person for another, as when an employee seeks to excuse his boss's rude behavior by appealing to its unusualness. Accounts for questionable conduct are typically in the form of excuses or justifications. In giving an excuse, communicators accept that the act they did was wrong but deny that they had full responsibility. A student who tells her teacher, "I'm sorry this paper is late. My hard disk crashed yesterday," is using an excuse. In justifications, in contrast, a speaker accepts full responsibility for an act but denies that the act was wrong. An example would be a teen who responded to his parent's reprimand about being late by saying, "I'm old enough to make my own decisions. You have no right to tell me what to do." Excuses and justifications are alternative ways of accounting, but they may also be paired. A person may say to his partner, "I'm sorry I hollered. I've been under so much pressure lately, but what do you expect when I'm in the middle of a conference call?" This person is simultaneously excusing and justifying, as well as offering an apology (acknowledging an act's inappropriateness and expressing sorrow).

The second and broader type of account is an account *of* action. An account of action (reason giving) is a pervasive feature of communicative life. Whenever people give reasons for what they are doing, they are accounting. Accounting is occurring when a woman tells a friend why she married Tim, a coworker mentions why he is no longer eating red meat, or a husband explains to his wife why they should visit Portugal rather than Greece. Accounts *of* action are often in the form of descriptively detailed stories. One nonobvious feature of accounts, both *for* and *of* action, is what they tell us about a group or culture. When people give reasons for their actions (or thoughts), they are treating the actions as choices. Of note is that people do not account for what they regard as natural or inevitable. Accounts, then, are a window into a culture. A close look at who offers them, about what they are accounting for, and under which circumstances makes visible a culture's beliefs about what counts as reasonable conduct.

—Karen Tracy

REFERENCES

Buttny, R., & Morris, G. H. (2001). Accounting. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 285–301). Chichester, England: Wiley.

Scott, M. B., & Lyman, S. (1968). Accounts. *American Sociological Review, 33,* 46–62.

Tracy, K. (2002). *Everyday talk: Building and reflecting identities.* New York: Guilford.

ACTION RESEARCH

Action research is a strategy for addressing research issues in partnership with local people. Defining characteristics of action research are collaboration, mutual education, and action for change. This approach increases the validity of research by recognizing contextual factors within the research environment that are often overlooked with more structured approaches. Action researchers are sensitive to culture, gender, economic status, ability, and other factors that may influence research partners, results, and research communities.

Ethics is a dominant theme in action research; researchers create an intentional ethical stance that has implications for process, philosophical framework, and evaluation procedures. Ethical principles are usually negotiated at the outset by all partners to identify common values that will guide the research process. These can include consultation throughout the research process, the valuing of various kinds of knowledge, mutual respect, and negotiation of the research questions, methods, dissemination, and follow-on action.

Action research can encompass the entire research project: identification and recruitment of the research team, definition of the research question, selection of data collection methods, analysis and interpretation of data, dissemination of results, evaluation of results and process, and design of follow-on programs, services, or policy changes. Although the degree of collaboration and community involvement varies from one research project to another, action research can be "active" in many ways depending on the needs or wishes of participants. All partners, including community members, play an active role in the research process; the research process itself is a training venue that builds research capacity for all partners; local knowledge and research skills are shared within the team; and the research outcome is a program or policy. Ideally, action research is transformative for all participants.

HISTORY

Action research has its roots in the work of Kurt Lewin (1935/1959) and the subsequent discourse in sociology, anthropology, education, and organizational theory; it is also based on the development theories of the 1970s, which assume the equality and complementarity of the knowledge, skills, and experience of all partners. As the human rights movement has challenged traditional scientific research models, new models for collaborative research have emerged. These models have various names: action research, PARTICIPATORY ACTION RESEARCH (PAR), collaborative research, and emancipatory or empowerment or community-based research, with considerable overlap among them. All models share a commitment to effecting positive social change by expanding the traditional research paradigm to include the voices of those most affected by the research; thus, "research subjects" can become active participants in the research process. Most models include an element of capacity building or training that contributes to the empowerment of the community, the appropriateness of the dissemination process, and the sustainability of action research results.

CASE EXAMPLE: THE CANADIAN TUBERCULOSIS PROJECT

The Canadian Tuberculosis (TB) Project was a 3-year project that addressed the sociocultural factors influencing the prevention and treatment of TB among the most affected people: foreign-born and aboriginal populations. Initiated by a nurse at the TB clinic in the province of Alberta, the need for this research was echoed by leaders within the affected groups. Beginning with the establishment of a community advisory committee (CAC), a set of guiding principles was negotiated. This process contributed to team building and the establishment of trust among the various players, including community, government, and academic partners. The CAC then worked with the administrative staff to recruit two representatives from each of the four aboriginal and six foreign-born communities. These 20 community research associates were trained in action research principles, interview techniques, and qualitative data analysis. They conducted interviews within four groups in each of their communities: those with active TB, those on prophylaxis, those who refused prophylaxis, and those with a more distant family history of TB in their country of origin or on aboriginal reserves. The community research associates, the CAC, and the academic staff then analyzed the interview data using both manual and electronic techniques. The multimethod and multi-investigator analysis strategy maximized the validity of the data and the research findings.

It became very clear that the action research process was valuable to the collaborative team, as members learned to respect each other's ways of knowing about the transmission and treatment of TB. Furthermore, the result of the study confirmed that although the health care system is very effective in treating active tuberculosis, there are few prevention programs in high-risk communities. In addition, health professionals frequently miss TB in their screening because it is not common in the general population in Western countries. The results were reviewed by the CAC and the community research associates to ensure consistency with the original ethical principles. This group then recommended the dissemination plan and the follow-on actions, including an information video, a community education nurse position, and TB fact sheets in their various languages.

LIMITATIONS TO ACTION RESEARCH

The action research approach has been criticized for potential lack of rigor and investigator bias. However, all research can suffer from these weaknesses, and only vigilance ensures high quality of the process and outcomes, regardless of the approach or methodology. Other challenges associated with action research include having longer timelines associated with these projects, sustaining commitment and funding over long periods of time, maintaining consistency with multiple and/or inexperienced researchers, educating funders to support this approach, sustaining a strong theoretical framework, and reducing participant turnover.

EVALUATION

The complexity of action research necessitates a carefully documented and analyzed audit trail. Inclusion of the full range of stakeholders in both formative and process evaluations can enhance the validity of the data and the evaluation process itself (Reason & Bradbury, 2001). These evaluations are most effective when integrated as a routine activity throughout the project. The rigor of the research is also enhanced when the methodology and project goals are measured by their relevance to the research context (Gibson, Gibson, & Macaulay, 2001). It is an important challenge for evaluation to accommodate the dynamic nature of action research while measuring outcomes related to initial goals and evaluating the consistency of adherence to negotiated ethical standards.

OUTCOMES AND ACTION

The process of the action research experience, including mutual learning and capacity building through training, is as valuable as the more tangible research results. Hence, dissemination strategies should include innovative mechanisms such as posters that can be used by community members and academics; use of community-accessible media such as radio, television, and newsletters; town meetings; and academic articles (which are reviewed by all partners prior to publication to prevent unauthorized release or unintentional misrepresentation of community or stigmatization). Because of the educational goal of action research, talks or posters may be presented in multiple venues or languages and can serve as training vehicles themselves, as well as dissemination strategies. The action research process can create effective alliances, building bridges across timeworn chasms between local communities and academics, government, and organizations.

—Nancy Gibson

REFERENCES

Cooke, B., & Kothari, U. (Eds.). (2001). *Participation: The new tyranny?* London: Zed.

Gibson, N., Gibson, G., & Macaulay, A. C. (2001). Community-based research: Negotiating agendas and evaluating outcomes. In J. Morse, J. Swanson, & A. J. Kuzel (Eds.), *The nature of qualitative evidence* (pp. 160–182). Thousand Oaks, CA: Sage.

Lewin, K. (1959). *A dynamic theory of personality: Selected papers* (D. K. Adams & K. E. Zener, Trans.). New York: McGraw-Hill. (Original work published 1935.)

Reason, P., & Bradbury, H. (Eds.). (2001). *Action Research: Participative inquiry and practice.* London: Sage.

Stringer, E. T. (1996). *Action Research: A handbook for practitioners.* Thousand Oaks, CA: Sage.

ACTIVE INTERVIEW

The term *active interview* does not denote a particular type of interview as much as it suggests a specific orientation to the interview process. It signals that all interviews are unavoidably active, interactionally and conversationally. The dynamic, communicative contingencies of the interview—any interview—literally incite respondents' opinions. Every interview is an occasion for constructing, not merely discovering or conveying, information (Holstein & Gubrium, 1995).

INTERVIEWING is always conversational, varying from highly structured, standardized, survey interviews to free-flowing informal exchanges. The narratives produced may be as truncated as FORCED-CHOICE survey responses or as elaborate as oral life histories, but they are all activated and constructed in situ, through interview participants' talk.

Although most researchers acknowledge the interactional character of the interview, the technical literature on interviewing tends to focus on keeping that interaction in check. Guides to interviewing, especially those oriented to standardized surveys, are primarily concerned with maximizing the flow of valid, reliable information while minimizing distortions of what respondents know. From this perspective, the standardized interview conversation is a pipeline for transmitting knowledge, whereas nonstandardized interaction represents a potential source of contamination that should be minimized.

In contrast, Charles Briggs (1986) argues that interviews fundamentally, not incidentally, shape the form and content of what is said. Aaron Cicourel (1974) goes further, maintaining that interviews virtually impose particular ways of understanding and communicating reality on subjects' responses. The point is that interviewers are deeply and unavoidably implicated in creating the meanings that ostensibly reside within respondents. Indeed, both parties to the interview are necessarily and ineluctably *active.* Meaning is not merely elicited by apt questioning, nor simply transported through respondent replies; it is communicatively assembled in the interview encounter. Respondents are not so much repositories of experiential information as they are constructors of knowledge in collaboration with interviewers (Holstein & Gubrium, 1995).

Conceiving of the interview as active leads to important questions about the very possibility of collecting and analyzing information in the manner the traditional approach presupposes (see Gubrium & Holstein, 1997). It also prompts researchers to attend more to the ways in which knowledge is assembled within the interview than is usually the case in traditional approaches. In other words, understanding *how* the meaning-making process unfolds in the interview is as critical as apprehending *what* is substantively asked and conveyed. The *hows* of interviewing, of course, refer to the interactional, narrative procedures of knowledge production, not merely to interview techniques. The *whats* pertain to the issues guiding the interview, the content of questions, and the substantive information communicated by the respondent. A dual interest in the *hows* and *whats* of meaning construction goes hand in hand with an appreciation for the constitutive activeness of the interview process.

—James A. Holstein and Jaber F. Gubrium

REFERENCES

Briggs, C. (1986). *Learning how to ask.* Cambridge, UK: Cambridge University Press.

Cicourel, A. V. (1974). *Theory and method in a critique of Argentine fertility.* New York: John Wiley.

Gubrium, J. F., & Holstein, J. A. (1997). *The new language of qualitative method.* New York: Oxford University Press.

Holstein, J. A., & Gubrium, J. F. (1995). *The active interview.* Thousand Oaks, CA: Sage.

ADJUSTED *R*-SQUARED. *See* GOODNESS-OF-FIT MEASURES; MULTIPLE CORRELATION; *R*-SQUARED

ADJUSTMENT

Adjustment is the process by which the anticipated effect of a variable in which there is no primary interest is quantitatively removed from the relationship of primary interest. There are several well-accepted procedures by which this is accomplished.

The need for adjustment derives from the observation that relationships between variables that capture the complete attention of an investigator also have relationships with other variables in which there is no direct interest. The relationship between ethnicity and death from acquired immunodeficiency syndrome (AIDS) may be of the greatest interest to the researcher, but it must be acknowledged that death rates from this disease are also influenced by several other characteristics of these patients (e.g., education, acculturation, and access to health care). A fundamental question that confronts the researcher is whether it is truly ethnicity that is related to AIDS deaths or whether ethnicity plays no real role in the explanation of AIDS deaths at all. In the second circumstance, ethnicity serves merely as a surrogate for combinations of these other variables that, among themselves, truly determine the cumulative mortality rates. Thus, it becomes important to examine the relationship between these other explanatory variables (or covariates) and that of ethnicity and death from AIDS.

Three analytic methods are used to adjust for the effect of a variable: (a) CONTROL, (b) stratification, and (c) REGRESSION analysis. In the control method, the variable whose effect is to be adjusted for (the adjustor) is not allowed to vary in an important fashion. This is easily accomplished for DICHOTOMOUS VARIABLES. As an example, if it is anticipated that the relationship between prison sentence duration and age at time of sentencing is affected by gender, then the effect is removed by studying the relationship in only men. The effect of gender has been neutralized by fixing it at one and only one value.

The process of stratification adds a layer of complication to implementing the control procedure in the process of variable adjustment. Specifically, the stratification procedure uses the technique of control repetitively, analyzing the relationship between the variables of primary interest on a background in which the adjustor is fixed at one level and then fixed at another level. This process works very well when there are a relatively small number of adjustor variable levels. For example, as in the previous example, the analysis of prison sentence duration and age can be evaluated in (a) only men and (b) only women. This evaluation permits an examination of the role that gender plays in modifying the relationship between sentence duration and age. Statistical procedures permit an evaluation of the relationship between the variables of interest within each of the strata, as well as a global assessment of the relationship between the two variables of interest across all strata. If the adjusting variable is not dichotomous but continuous, then it can be divided into different strata in which the strata definitions are based on a range of values for the adjustor.

The procedure of adjustment through regression analysis is somewhat more complex but, when correctly implemented, can handle complicated adjustment operations satisfactorily. Regression analysis itself is an important tool in examining the strength of association between variables that the investigator believes may be linked in a relationship. However, this regression model must be carefully engineered to be accurate. We begin with a straightforward model in regression analysis, the simple straight-line model. Assume we have collected n pairs of observations $(x_i, y_i), i = 1$ to n. The investigator's goal is to demonstrate that individuals with different values of x will have different values of y in a pattern that is best predicted by a straight-line relationship. Write the simple linear regression model for regressing y on x as follows:

$$E[y_i] = \beta_0 + \beta_1 x_i. \quad (1)$$

For example, y could be the prison sentence length of a prisoner convicted of a crime, and x could be the age of the prisoner at the time of his or her sentencing. However, the investigator believes that the socioeconomic status (SES) of the prisoner's family also plays a role in determining the prisoner's sentence duration. Holding SES constant in this setting would not be a useful approach because it is a continuous variable with wide variability. Consider instead the following model in which prison sentence duration is regressed on both age and the SES simultaneously:

$$E[\text{sentence duration}] = \beta_0 + \beta_1(\text{age}) + \beta_2(\text{SES}). \quad (2)$$

The commonly used LEAST SQUARES estimates for the parameters β_0, β_1, and β_2 are readily obtained from many available software packages. The estimate of β_1, the parameter of interest, is obtained by carrying out three regressions. First, the dependent variable sentence duration is regressed on the adjustor SES. What is left over from this regression is the part of the prison sentence duration variable that is unexplained by SES. This unexplained portion is the prison sentence duration RESIDUAL. Next, the variable age is regressed on the adjustor SES, and the age residual is computed. Finally, the prison sentence duration residual is regressed on the age residual. The result of this final regression produces the estimate of β_1 in equation (2).

These computations are carried out automatically by most regression software. The model produces an estimate of the relationship between sentence duration and age in which the influence of SES is "removed." Specifically, the regression procedure isolated and identified the relationship between sentence duration and the SES adjustor and then removed the influence of the adjustor, repeating this procedure for age. Finally, it regressed the residual of prison sentence duration (i.e., what was left of sentence duration after removing the SES influence) on the residual of age. Thus, in the case of regression analysis, adjustment has a specific meaning that is different from the definition used when one either controls or stratifies the adjustor. In the case of regression analysis, adjustment means that we isolate, identify, and remove the influences of the adjustor variable. When the explanatory variables are themselves discrete, ANALYSIS OF VARIANCE (ANOVA) produces this result. The ANALYSIS OF COVARIANCE (ANCOVA) provides the same information when the explanatory variables are composed of a mixture of discrete and continuous variables.

This process can be implemented when there is more than one adjustor variable in the regression model. In that case, the procedure provides estimates of the strength of association between the variables of interest after making simultaneous adjustments for the influence of a battery of other variables.

—Lemuel A. Moyé

REFERENCES

Deming, W. E. (1997). *Statistical adjustment of data.* New York: Dover.

Kleinbaum, D. G., Kupper, L. L., & Morgensten, H. (1982). *Epidemiologic research: Principles and quantitative methods.* New York: Van Nostrand Reinhold.

Neter, J., Wasserman, W., & Kutner, M. (1990). *Applied linear statistical models* (3rd ed.). Homewood, IL: Irwin.

AGGREGATION

Aggregation is the process by which data are collected at a different level from the level of greatest interest. Data are aggregated because they are more available at the group level. For example, crime statistics are available from urban centers but, in general, not from every citizen of the city. Attempts are then made to identify exposure-response relationships among different aggregates. The ECOLOGICAL FALLACY is the false conclusion that a result about groups, or aggregates, of individuals applies to the individuals within these aggregates. Because each aggregate contains individuals who differ from each other, the conclusion about the aggregate does not apply to each of its member individuals.

—Lemuel A. Moyé

See also ECOLOGICAL FALLACY

AIC. *See* GOODNESS-OF-FIT MEASURES

ALGORITHM

Algorithm is a computer science term for a way of solving a problem and also refers to the instructions given to the computer to solve the problem. The study of algorithms is central to computer science and is of great practical importance to data analysis in the social sciences because algorithms are used in statistical programs.

An algorithm can be thought of as any step-by-step procedure for solving a task. Imagine five playing cards face down on a table and the task of sorting them. Picking them up one at a time with the right hand and placing them in the left hand in their proper place would be one way to solve this task. This is an algorithm, called *insertion sort* in computer science.

It is worth noting the subtle distinction between the concept of an algorithm and the concept of a method or technique. For example, a method would be LEAST SQUARES, matrix inversion would be a technique used therein, and *LU decomposition* and *Strassen's algorithm* would be alternative algorithms to accomplish matrix inversion. A single data analysis method may use more than one algorithm.

It is impossible to write statistical software without using algorithms, so the importance of algorithms to quantitative methodologists is ensured. However, user-friendly statistical software packages eliminate the need for end users to construct their own algorithms for most tasks. Nonetheless, at least a basic understanding of algorithms can be useful. For example, MAXIMUM LIKELIHOOD ESTIMATION methods can use an initial estimate as a starting point, and in some cases, failure to converge may be remediated by trivially altering the initial estimate. Without some familiarity of the underlying algorithm, a researcher may be stuck with a nonconverging function.

Another setting where some knowledge of algorithms is useful is presented in Figure 1, which shows two possible digraph depictions of the same network data. The left panel uses the *multidimensional scaling* algorithm, and the right uses *simulated annealing*. The data are identical, which may be verified by observing who is connected to whom, but the appearance of the graphs is different. Algorithms are important here because interpretation of the network data is affected by the appearance of the graph, which is affected in turn by the choice of algorithm. Although in many cases, different algorithms will produce the same result

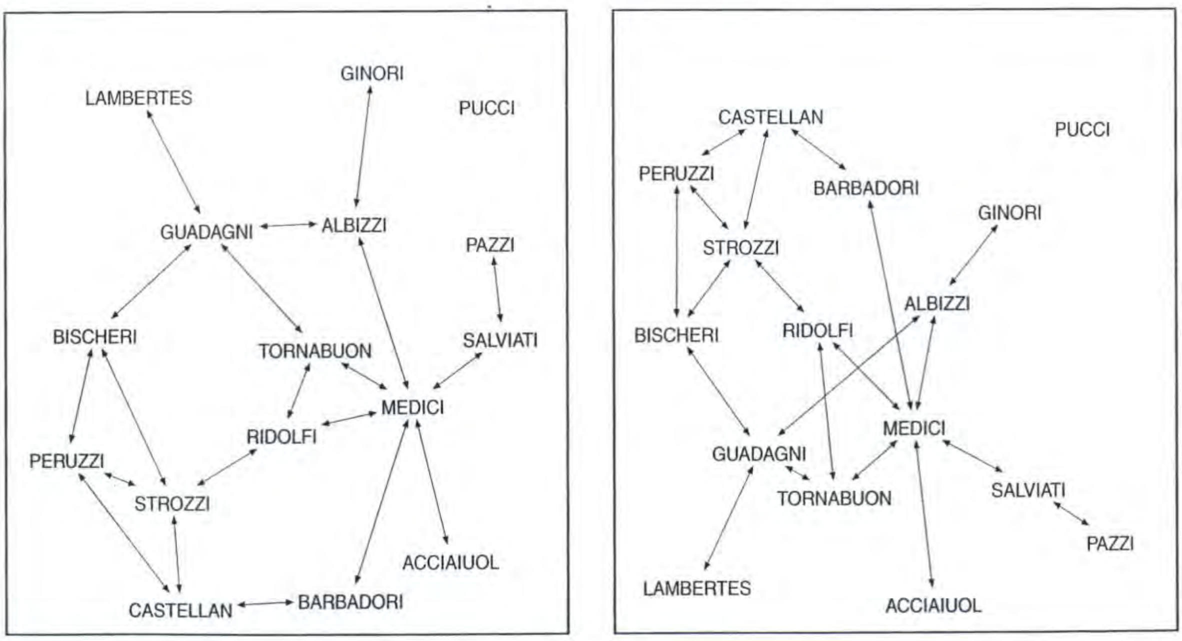

Figure 1 Two Possible Digraph Depictions of the Same Network Data

but differ in speed, in this case, different algorithms produce different results.

The term *algorithm* is sometimes used more broadly to mean any step-by-step procedure to solve a given task, whether or not a computer is involved. For instance, matching historical records from more than one archival source can be done by hand using an algorithm.

—Andrew Noymer

REFERENCES

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to algorithms* (2nd ed.). Cambridge, MA: MIT Press.

Knuth, D. E. (1997). *Fundamental algorithms: The art of computer programming* (Vol. 1, 3rd ed.). Reading, MA: Addison-Wesley.

ALPHA, SIGNIFICANCE LEVEL OF A TEST

The development of SIGNIFICANCE TESTING is attributed to Ronald Fisher. His work in agrarian statistics in the 1920s led him to choose the significance threshold of .05, a level that has been accepted by research communities in many fields. The underlying motivations for the acceptance of hypothesis testing and the *P* VALUE have been reviewed (Goodman, 1999). However, it must be emphasized that no mathematical theory points to .05 as the optimum TYPE I ERROR level—only tradition.

The alpha error level is that level prospectively chosen during the design phase of the research; the *p* value is that error level computed based on the research data. Alpha levels are selected because they indicate the possibility that random sampling error has produced the result of the study, so they may be chosen at levels other than .05. The selection of an alpha level greater than .05 suggests that the investigators are willing to accept a greater level of Type I error to ensure the likelihood of identifying a positive effect in their research. Alternatively, the consequences of finding a positive effect in the research effort may be so important or cause such upheaval that the investigators set stringent requirements for the alpha error rate (e.g., $\alpha = .01$). Each of these decisions is appropriate if it (a) is made prospectively and (b) is adequately explained to the scientific community.

One adaptation of significance testing has been the use of ONE-TAILED TESTING. However, one-sided testing has been a subject of contention in medical research and has been the subject of controversy in the social sciences as well. An adaptation of TWO-TAILED TESTING, which leads to nontraditional alpha error levels, is the implementation of asymmetric regions of significance. When testing for the effect of an intervention in improving education in elementary school age children, it is possible that the intervention may have a paradoxical effect, reducing rather than increasing testing scores. In this situation, the investigator can divide the Type I error rate, so that 80% of the available Type I error level (.04 if the total alpha error level is .05) is in the "harm" end of the tail, and put the remaining 1% in the benefit tail of the distribution. The prospective declaration of such a procedure is critical and, when in place, demonstrates the investigator's sensitivity to the possibility that the intervention may be unpredictably harmful. Guidelines for the development of such asymmetric testing have been produced (Moyé, 2000).

When investigators carry out several hypothesis tests within a clinical trial, the family-wise or overall Type I error level increases. A research effort that produces two statistical hypotheses resulting in *p* values at the .045 level may seem to suggest that the research sample has produced two results that are very likely to reflect findings in the larger population from which the sample was derived. However, an important issue is the likelihood that at least one of these conclusions is wrong. Assuming each of the two tests is independent, then the probability that at least one of these analyses produces a misleading result through sampling error alone is $1 - (.955)^2 = .088$. This can be controlled by making adjustments in the Type I error levels of the individual tests (e.g., BONFERRONI TECHNIQUES), which are useful in reducing the alpha error level for each statistical hypothesis test that was carried out. As an example, an investigator who is interested in examining the effect of a smoking cessation program on teenage smoking may place a Type I error rate of .035 on the effect of the intervention on self-reports of smoking and place the remaining .015 on the intervention's effect on attitudes about smoking. This approach maintains the overall error rate at .05 while permitting a positive result for the effect of therapy on either of these two endpoint measurements of smoking.

—Lemuel A. Moyé

REFERENCES

Goodman, S. N. (1999). Toward evidence-based medical statistics: 1. The *p* value fallacy. *Annals of Internal Medicine, 130*, 995–1004.

Moyé, L. A. (2000). *Statistical reasoning in medicine.* New York: Springer-Verlag.

ALTERNATIVE HYPOTHESIS

The HYPOTHESIS that opposes the NULL HYPOTHESIS is commonly called the alternative. It is usually the research, or dominant, hypothesis. Suppose, for example, the research hypothesis is that sugar consumption is related to hyperactivity in children. That hypothesis, H_A, is an alternative to the null, H_0, that sugar consumption is not related to hyperactivity.

—Michael S. Lewis-Beck

AMOS (ANALYSIS OF MOMENT STRUCTURES)

AMOS is a statistical software package for social sciences analyses applying STRUCTURAL EQUATION MODELS or COVARIANCE STRUCTURE MODELS. AMOS has good graphic interface. For more information, see the Web page of the software at www.spss.com/spssbi/amos/.

—Tim Futing Liao

ANALYSIS OF COVARIANCE (ANCOVA)

The methodology with this name grew out of a desire to combine ANALYSIS OF VARIANCE and REGRESSION analysis. It received considerable interest before the arrival of good computer packages for statistics, but the separate name for this methodology is now in decreasing use. More often, the analysis is simply referred to as multiple regression analysis, with both QUANTITATIVE VARIABLES and DUMMY VARIABLES as explanatory variables on a quantitative dependent variable.

DEFINITION

Analysis of covariance is a multivariate statistical method in which the dependent variable is a quantitative variable and the independent variables are a mixture of NOMINAL VARIABLES and quantitative variables.

EXPLANATION

Analysis of variance is used for studying the relationship between a quantitative dependent variable and one or more nominal independent variables. Regression analysis is used for studying the relationship between a quantitative dependent variable and one or more quantitative independent variables. This immediately raises the question of what to do when, in a common situation, the research hypothesis calls for some of the independent variables to be nominal and some to be quantitative variables.

With the arrival of statistical computer packages, it soon become clear that such a situation can easily be covered by introducing one or more dummy variables for the nominal variables and doing an ordinary multiple regression analysis. This can be done with or without the construction of INTERACTION variables, depending on the nature of the available data and the problem at hand. The use of a separate name for analysis of covariance has been decreasing because regression analysis can handle this analysis without any difficulties.

HISTORICAL ACCOUNT

A good deal of work on the development of analysis of covariance took place in the 1930s. As with so much of analysis of variance, Sir Ronald A. Fisher had a leading role in its development, which was highlighted in a special issue of the *Biometrics* journal, with an introductory paper by William G. Cochran (1957) and six other papers by leading statisticians of the time.

Analysis of covariance was a natural outgrowth of both analysis of variance and regression analysis. In analysis of variance, it may be reasonable to control for a quantitative variable and compare adjusted means instead of the raw, group means when comparing groups. Similarly, in regression analysis, it may be reasonable to control for a nominal variable. Both of these desires resulted in developments of analysis of covariance and shortcut methods for computations, using desk calculators.

APPLICATIONS

In its simplest form, analysis of covariance makes use of one quantitative variable and one nominal variable as explanatory variables. In regression notation, the model can be written as

$$Y = \alpha + \beta_1 X + \beta_2 D, \tag{1}$$

where Y and X are quantitative variables; D is a dummy variable with values 0 and 1, representing two groups; and α, β_1, and β_2 are the regression parameters.

It is now possible to do analysis of variance comparing the means for the two groups while controlling for the X variable. Instead of comparing the means of Y for the two groups directly, we compare *adjusted* means of Y. By substituting 0 and 1 for the dummy variable, we get the two parallel lines with the following equations:

$$Y_1 = \alpha + \beta_1 x_1, \tag{2}$$

$$Y_2 = (\alpha + \beta_2) + \beta_1 x_2. \tag{3}$$

For a common value of the X variable—say, the overall mean—the difference between the predicted Y values becomes the value of the coefficient β_2. Thus, when we control for the X variable by holding it constant at a fixed value, the adjusted difference in Y means becomes the vertical difference between the two lines above.

It is also possible to compare the two groups and control for the nominal group variable D. We could do that by dividing the data up into two groups and run one regression of Y on X for the observations where $D = 0$ and another for $D = 1$. This can be done when the two groups have identical slopes, and statisticians have developed elaborate computing formulas for this case using only desk calculators. But it would be more efficient to run one multivariate regression analysis with two independent variables instead of two separate, simple regressions.

The method generalizes to the use of more than one explanatory variable of each kind.

EXAMPLES

A comparison of incomes for females and males at a small college was done by comparing the mean salaries of the two groups. It was found that the mean for the men was larger than the mean for the women. However, it was pointed out that more women were in the lower ranks of the faculty, and perhaps that was the reason for the disparity in salaries. To control for age, we could keep age constant and divide the data into groups, one for each age, and compare the salary means for men and women within each age group. With only a small data set to begin with, this would involve several analyses with very few cases.

Similarly, it would be possible to control for gender by doing a separate analysis for each gender. That would only involve two groups, but the groups would still be smaller than the entire faculty used as one group.

Instead, one multivariate regression analysis with age as one variable and gender as the dummy variable would achieve the same purpose without running into the problem with small data sets. The coefficient b_1 (the estimate of β_1) would show the effect of age, controlling for gender, and the coefficient b_2 (the estimate of β_2) would show the effect of gender, controlling for age.

It is also possible to do a more sophisticated analysis, allowing for the regression lines for income on age not to be parallel, which would allow for an interaction effect between age and gender on salary. This is done by creating an interaction variable as the product of the age variable and the dummy variable. This product variable is then used in a multiple regression analysis, with salary as the dependent variable and age, dummy, and interaction as three explanatory variables.

—Gudmund R. Iversen

REFERENCES

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics, 13*(3), 261–281.

Hardy, M. A. (1993). *Regression with dummy variables* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–093). Newbury Park, CA: Sage.

Wildt, A. R., & Ahtola, O. T. (1978). *Analysis of covariance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–012). Beverly Hills, CA: Sage.

ANALYSIS OF VARIANCE (ANOVA)

Analysis of variance is the name given to a collection of statistical methods used to analyze the impact of one or more NOMINAL VARIABLES as independent variables on a QUANTITATIVE VARIABLE as the dependent variable. For one nominal variable, there is ONE-WAY ANALYSIS OF VARIANCE; for two nominal variables, there is TWO-WAY ANALYSIS OF VARIANCE and so on for higher models. Data on several dependent variables can be analyzed using MANOVA (MULTIVARIATE ANALYSIS OF VARIANCE).

EXPLANATION

A study of the impact of gender and race on income could employ analysis of variance to see whether mean incomes in the various gender/racial groups are different. It would perhaps be better to use a name such as *analysis of means* for these methods, but variances are actually used to determine whether group means are different—thus the name.

HISTORICAL ACCOUNT

Analysis of variance has its origins in the study of data from experiments. Many of the earliest applications were from agriculture in the study of how yield on a piece of land is affected by factors such as type of fertilizer and type of seed. The British statistician Sir Ronald A. Fisher did much of the early work developing analysis of variance.

Experiments can be set up in many different ways, and there are corresponding methods of analysis of variance for the analysis of the data from these designs. Thus, the name *analysis of variance* covers a large number of methods, and the ways in which an experiment is designed will determine which analyses are used. The ties between the design of EXPERIMENTS and analysis of variance are very strong.

There are also strong ties between analysis of variance and REGRESSION analysis. Historically, the two sets of statistical methods grew up separately, and there was little contact between the people who worked in the two areas. The old, elaborate methods of computations used in analysis of variance seemed very different from computations used in regression analysis. But now that statistical software has taken over the computations, it is not very difficult to demonstrate how the two sets of methods have many things in common. Most analyses of variance can be represented as regression analysis with DUMMY VARIABLES. Indeed, analysis of variance and regression analysis are special cases of the so-called general linear model. This became clearer when people started to use analysis of variance on data obtained not only from experiments but also from observational data.

APPLICATIONS AND EXAMPLES

One-Way Analysis of Variance

Many of the central ideas in analysis of variance can be illustrated using the simplest case in which there are data on a dependent variable Y from two groups. Thus, we have a nominal variable with two values as the independent variable. This becomes a one-way analysis of variance because there are data only on one independent variable. The data could have been obtained from a psychological experiment whereby one group consisted of the control group and the other group consisted of elements in an experimental group that received a certain stimulus. Then, is there any effect of the stimulus? Also, the data can come from an observational study in which there are incomes from people in different ethnic groups, and we want to see if the income means are different in the underlying ethnic populations.

Thus, we know what group an element (person) belongs to, and for each element, we have data on a quantitative variable Y (response, income). For the sake of numerical computing convenience from the days before electronic computing software, we use the notation commonly used in analysis of variance, which looks very different from the one used in regression analysis.

A typical observation of the variable Y is denoted y_{ij}. The first subscript i denotes the group the observation belongs to, and here $i = 1$ or $i = 2$. Within each group, the observations are numbered, and the number for a particular observation is denoted by j. For example, y_{25} is the observation for the fifth element in the second group. Similarly, the mean of the observations in group i is denoted $\bar{y}_i$, and the mean of all the observations is denoted by $\bar{y}$. With these symbols, we have the following simple identity:

$$y_{ij} - \bar{y} = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i). \qquad (1)$$

The deviation of an observation y_{ij} from the overall mean $\bar{y}$ is written as a sum of the deviation of the group mean $\bar{y}_i$ from the overall mean $\bar{y}$ plus the deviation of the observation from the group mean.

If people's incomes were not affected by the group a person belongs to or by any other variables, then all the observations would be equal to the overall mean. Everyone would have the same income. There is absolutely no way for the observations to differ because they are not affected by any variables. The observed variations in the observations around the overall mean

$\bar{y}$ are then possibly due to two factors: the effect of the independent, categorical variable and the effect of all other variables.

Similarly, the first term on the right side measures how large the effect is of the independent variable. If this variable had no effect, then the means in the two groups would be the same, and the difference between a group mean and the overall mean would equal zero. The second term on the right measures how different an observation is from the group mean. If all other variables had no effects, then the observations in each group would all be equal because group membership is the only thing that matters. The differences between the observations and the group means are therefore due to the effects of all other variables, also known as the residual variable. In a population, this is known as the ERROR variable, and the estimated values in a sample form the residual variable:

Combined effect of independent and residual variable	=	Effect of independent variable	+	Effect of residual variable

(2)

This identity holds for a single observation. Next, we want to summarize these effects across all the observations. We could simply add all the effects, but all three sums are equal to zero because the effects are all deviations from means.

Magnitudes of Effects

Instead, we square the left and right sides of the equation above to get positive terms. Squares are used for historical and mathematical convenience. There is nothing "correct" about using squares, and we have to realize that all the results we get from the analysis are heavily influenced by the choice of squares.

Adding the squares gives the sum of squared combined effects on the left side. On the right, we get the sum of squared effects for the independent variable, plus the sum of squared effects of the residual variable, plus two times the sum of the products of the effects of the independent and residual variables. It can then be shown that the sum of the cross-products always equals zero.

That gives the following equation for the squared effects across all the observations, written in symbols, words, and abbreviations:

$$\Sigma\Sigma(y_{ij} - \bar{y})^2 = \Sigma\Sigma(\bar{y}_i - \bar{y})^2 + \Sigma\Sigma(y_{ij} - \bar{y}_i)^2$$

Total sum of squares	=	Between-group sum of squares	+	Within-group squares

(3)

$$TSS = BSS + WSS \tag{4}$$

The two summations signs in front of the squares show that we add all the terms for the observations in each group and then add the terms for the two groups.

Thus, for each *individual*, we have broken down the combined effect into a sum of effects of the independent and residual variables. By using squares, we also have for the entire *group* of observations the combined effect broken down into the sum of effects of the independent variable (BETWEEN SUM OF SQUARES) and the residual variable. For more complicated analyses of variance, we get more extensive breakdowns of the total SUM OF SQUARES.

For the sake of simplicity, we can divide the three sums of squares by the total SUM OF SQUARES. That gives us the following:

$$\frac{TSS}{TSS} = \frac{BSS}{TSS} + \frac{WSS}{TSS}, \tag{6}$$

$$1.00 = R^2 + (1 - R^2). \tag{7}$$

This step defines the COEFFICIENT OF DETERMINATION R^2. It tells how large the effect of the independent variable is as a proportion of the total sum of squares. This quantity is often denoted eta^2 in an older notation. It plays the same role as R^2 in regression analyses. R itself can be thought of as a correlation coefficient measuring the strength of the relationship between the independent and dependent variables.

The quantities described above are often displayed in an analysis of variance table, as shown in Table 1.

Hypothesis Testing

Could the relationship have occurred by chance? This question translates into the null hypothesis that the population group means are equal:

$$H_0\colon \mu_1 = \mu_2. \tag{8}$$

Table 1 Sums of Squares and Proportions

Source of Effect	*Sum of Squares*	*Proportion of Effect*
Independent variable	$\sum\sum(\bar{y}_i - \bar{y})^2$	R^2
Residual variable (error)	$\sum\sum(y_{ij} - \bar{y}_i)^2$	$1-R^2$
Total	$\sum\sum(y_{ij} - \bar{y})^2$	1

With more than two groups and two population means, the test of the null hypothesis becomes more complicated. The null hypothesis states that all the population means are equal, but the logical opposite of several means being equal is only that at least *some* and not all of the means are different. After rejecting the null hypothesis, the question becomes *which* of the means are different.

A null hypothesis is studied by finding the corresponding p VALUE. Such a p value tells us how often (proportion of times) we get the data we got, or something more extreme, from a population in which the null hypothesis is true. This can also be seen as the probability of rejecting a true null hypothesis. If the p value is small—say, less than .05—we reject the null hypothesis. The p value is found from the data by computing the corresponding value of the statistical F variable. The actual computation is mostly of interest to statisticians only, but it is shown in Table 2, which corresponds to the table produced by most statistical software packages.

There are three more columns in Table 2 than in Table 1 that can be used to find the p value. The letter k stands for the number of groups, and n stands for the total number of observations in the study.

The computation of F is based on a comparison of the between and within sums of squares. But sums of squares cannot be compared directly because they depend on how many terms there are in the sums. Instead, each sum of squares is changed into a variance by dividing the sum by its appropriate DEGREES OF FREEDOM. That gives the quantities in the column labeled *Mean Square,* but the column could just as well have been denoted by the word *variance.* A mean square is an "average" square in the sense that a variance is an average squared distance from the mean. Finally, the ratio of the two mean squares is denoted by the letter F (in honor of Fisher).

By working through the mathematics of the derivations above, it is possible to show that when the null hypothesis is true, then the two mean squares (variances) are about equal. Thus, the ratio of the two mean squares (the F value) is approximately equal to 1.00. When the null hypothesis is false and the population means are truly different, then the value of F is a good deal larger than 1.00.

When the values of the residual variable follow a normal distribution, the F variable has the corresponding F-DISTRIBUTION (a variable like the NORMAL DISTRIBUTION t, and CHI-SQUARE DISTRIBUTION variables). The F-distribution can then be used to find the p value for the data. The critical values of F are dependent on how many groups and how many observations there are. This is dealt with by saying that we have F with $k - 1$ and $n - k$ degrees of freedom. Suppose we have data from two groups and a total of 100 observations, and $F = 5.67$. From the F-distribution with $2 - 1 = 1$ and $100 - 2 = 98$ degrees of freedom, we find that for this value of F, the p value = .02. Because it is small, we reject the null hypothesis and conclude that the difference between the two group means is statistically significant. Most statistical software provides an exact value of the p value. There are also tables for the F variable, but because there is a different F-distribution for each pair of degrees of freedom, the F tables tend to be large and not very detailed.

Data for two groups can also be analyzed using a t-test for the difference between two means or a regression analysis with a dummy variable. The (two-sided) p values would have been the same in all three cases. If we had more than two groups, there would not be any t-test, but we could use regression with the appropriate number of dummy variables instead of analysis of variance.

Two-Way Analysis of Variance

Here, two independent, nominal variables affect one dependent, quantitative variable. Similarly, there are three-way analysis or higher analyses, depending on how many independent variables there are.

Aside from having more than one independent variable, there is now a so-called interaction effect present in addition to the (main) effects of the independent variables. An interaction effect of the independent variables on the dependent variable is an effect of the two independent variables over and beyond their separate main effects.

Table 2 Analysis of Variance Table With Sums of Squares, Proportions, Mean Squares, and F Value

Source of Effect	*Sum of Squares*	*Proportion of Effect*	*Degree of Freedom*	*Mean Square*	*F Value*
Independent variable	$\sum\sum(\bar{y}_i - \bar{y})^2\,(BSS)$	R^2	$k-1$	$\frac{BSS}{k-1} = BMS$	$\frac{BMS}{WMS}$
Residual variable (error)	$\sum\sum(y_{ij} - \bar{y}_i)^2\,(WSS)$	$1 - R^2$	$n-k$	$\frac{WSS}{n-k} = WMS$	
Total	$\sum\sum(y_{ij} - \bar{y})^2\,(TSS)$	1	$n-1$		

The analysis is based on the following identity:

$$y_{ijk} - \bar{y} = (\bar{y}_{i.} - \bar{y}) + (\bar{y}_{.j} - \bar{y}) + (\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}) + (\bar{y}_{ijk} - \bar{y}_{ij}). \quad (9)$$

Each observation in this notation is identified by the three subscripts i, j, and k. The first subscript gives the value of the first nominal variable, the second subscript gives the value of the second nominal variable, and the last subscript gives the number of observations for a particular combination of the two nominal variables.

If nothing influenced the dependent variable, all the values of that variable would have to be equal to the overall mean $\bar{y}$. Everyone would have the same income if there were no effects of gender, race, and all other variables. Any deviation from this value represents effects of one or more independent variables, as shown on the left side above. On the right side, the first difference represents the effect of the first nominal variable, the second difference represents the effect of the second nominal variable, the third term represents the effect of the interaction variable, and the last term is the residual, which measures the effects of all other variables. Without the inclusion of the interaction term, there would not be an equality.

By squaring the equation above, the right side gives the square of each term plus two times the cross-products. Adding across all the observations, the sums of all the cross-products equal zero, and the resulting sums of squares give the effects of the independent variables and the residual variable. These sums can be expressed as proportions, and by dividing by the residual mean square, we get F values and thereby p values for the tests of the hypotheses of no effects.

Models I, II, and III: Fixed- or Random-Effect Models

When we have data on the dependent variable for all values of the independent variables, we have a MODEL I ANOVA (fixed effects). When we have data on the dependent variable for only a sample of values of the independent variables, we have a MODEL II ANOVA (random effects). MODEL III ANOVA makes use of a mix of fixed and random effects.

—Gudmund R. Iversen

REFERENCES

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–054). Beverly Hills, CA: Sage.

Cobb, G. W. (1998). *Introduction to design and analysis of experiments.* New York: Springer-Verlag.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–001). Beverly Hills, CA: Sage.

ANALYTIC INDUCTION

Analytic induction is sometimes treated as a model for qualitative RESEARCH DESIGN. It portrays inquiry as an iterative process that goes through the following stages: initial definition of the phenomenon to be explained; examination of a few cases; formulation of a HYPOTHESIS; study of further cases to test this hypothesis; and, if the evidence negates the hypothesis, either redefinition of the phenomenon to exclude the negative case or reformulation of the hypothesis so that it can explain it. This is followed by study of further cases, and if this further evidence supports the hypothesis, inquiry can be concluded. However, if any negative case arises in the future, inquiry will have to be resumed: Once again, either the hypothesis must be reformulated or the phenomenon redefined, and further cases must be studied, and so on.

Analytic induction shares with GROUNDED THEORY an opposition to the testing of hypotheses derived from armchair theorizing (see INDUCTION). It also shares with that approach an emphasis on the development of THEORY through investigating a relatively small number of cases and adopting a flexible mode of operation in which theory emerges out of—and guides future—data collection and analysis. At the same time, analytic induction has features that are at odds with the self-understanding of much QUALITATIVE RESEARCH today—notably, its explicit concern with testing hypotheses and discovering universal laws. Its most distinctive feature is recognition that the type of phenomenon to be explained may need to be reformulated in the course of research; this arises because the task is to discover theoretical categories all of whose instances are explained by a single type of cause. In the context of analytic induction, a theory specifies both necessary and sufficient conditions, and this is why a single exception is sufficient to refute any theory.

The term *analytic induction* seems to have been coined by the Polish American sociologist Florian Znaniecki in his book *The Method of Sociology* (Znaniecki, 1934). There, he contrasted it with *enumerative induction*, a term referring to the kind of inference that is characteristic of statistical method. Enumerative induction involves trying to discover the characteristics of phenomena belonging to a particular class by studying a large number of cases and describing what they have in common. Znaniecki points to a paradox here: If we do not know the essential features of a type of phenomenon, we cannot identify instances of it; however, if we *do* know those essential features, we already know what enumerative induction is designed to tell us. Following on from this, Znaniecki argued that, contrary to the claims of some contemporary social scientists, statistical method is not the method of natural science: Natural scientists do not study large numbers of cases and try to derive theories from their average characteristics. Rather, he suggested, they study a single case or a small number of cases in depth, and develop theories that identify the essential features of these cases as instances of general classes of phenomena. Thus, according to Znaniecki, the task of science is the search for universal laws—relations of causality or functional dependence—not statistical generalizations; and it is analytic, not enumerative, induction that is its method (see Gomm, Hammersley, & Foster, 2000; Hammersley, 1989).

Znaniecki does not provide any detailed examples of analytic induction, but two subsequent writers provided what are now taken to be key illustrations. Alfred Lindesmith (1968) investigated the causal process involved in people becoming addicted to opiates, arguing that the essential condition was their coming consciously to use a drug to alleviate withdrawal symptoms. Donald Cressey, one of his students, set out to identify the necessary and sufficient conditions of trust violation, the latter concept being a theoretical reformulation of the commonsense and legal concept of embezzlement, a reformulation that emerged during the course of his research (Cressey, 1950, 1953). He put forward quite a complicated theory whereby those in positions of financial trust become trust violators when they have a financial problem that they cannot disclose to others; when they realize that this can be secretly solved through embezzlement; and when they can appeal to some account, an excuse, which preserves their sense of themselves as trustworthy.

There has been much controversy about the character and value of analytic induction. In an early but still valuable critique, W. P. Robinson (1951) challenged the distinction between enumerative and analytic induction; and argued that the latter, as usually practiced, focuses almost entirely on the discovery of necessary, rather than sufficient, conditions. He saw this as a weakness that could be overcome by researchers studying situations in which a theory would predict a particular type of phenomenon to occur. At the same time, he argued that requiring the evidence to display a universal rather than a probabilistic distribution—all instances in line with the theory and none at odds with it—is too demanding a standard because it is rarely likely to be met if a large number of cases are studied. Ralph Turner (1951) reacted against this argument by emphasizing the difference between analytic and enumerative induction, but proposed that they are complementary in framing explanations. Although the concept of analytic induction is less frequently invoked today by qualitative researchers than previously, Howard Becker (1998) has recently reiterated and developed the idea that analytic induction best captures the logic of social scientific investigation.

—Martin Hammersley

REFERENCES

Becker, H. S. (1998). *Tricks of the trade.* Chicago: University of Chicago Press.

Cressey, D. (1950). The criminal violation of financial trust. *American Sociological Review, 15,* 738–743.

Cressey, D. (1953). *Other people's money.* Glencoe, IL: Free Press.

Gomm, R., Hammersley, M., & Foster, P. (Eds.). (2000). *Case study method.* London: Sage. [Includes the articles by Robinson and Turner.]

Hammersley, M. (1989). *The dilemma of qualitative method.* London: Routledge Kegan Paul.

Lindesmith, A. (1968). *Addiction and opiates.* Chicago: Aldine.

Robinson, W. (1951). The logical structure of analytic induction. *American Sociological Review, 16*(6), 812–818.

Turner, R. H. (1951). The quest for universals. *American Sociological Review, 18*(6), 604–611.

Znaniecki, F. (1934). *The method of sociology.* New York: Farrar and Rinehart.

ANONYMITY

Anonymity refers to the assurance that participants are not individually identified by disguising or withholding their personal characteristics. In conducting social science research, the protection of "human subjects" is a serious concern for institutional review boards and researchers alike. Guaranteeing and preserving individuals' anonymity is a key component of protecting their rights as research participants.

As researchers, we take several steps to ensure the anonymity of our participants. For example, we assign case numbers to each person instead of using respondents' real names to identify their data. When presenting our findings, we disguise any facts that would allow participants to be identified. For example, we use pseudonyms instead of individuals' real names. Researchers involved in the construction of data sets must remove all personal identifiers (such as name, address, and Social Security number) before permitting public use of the data.

We withhold direct identifiers as a matter of course, but indirect identifiers, such as state of residence, organizational membership, and occupation, might also be used to recognize an individual or group (Inter-University Consortium for Political and Social Research, 2002). Quantitative researchers routinely recode the variables into broader categories as the broader categories adequately capture the shared characteristics they believe may affect the outcome variable of interest. Data from a sample of physicians, for example, may include the *type* of medical school from which they received their medical degrees (e.g., private or public institutions) rather than the *name* of that institution. In qualitative research, indirect identifiers may pose a greater problem, in part because data are not aggregated. Qualitative social scientists strive to provide rich narrative accounts of the group they are studying. Part of the richness of the narrative is the extent to which the author can link the participants' social milieus to their experiences and perceptions. Thus, often we want the reader to know the occupation of a particular participant and how that may have affected other aspects of his or her life (or another variable of interest).

Protecting participants' anonymity may be especially important when dealing with groups who are distrustful of research and science and/or those involved in illegal or stigmatized activities, such as illicit drug use. Groups who have been victimized or exploited by past research projects are likely to be distrustful. The Tuskegee Syphilis Study, in which African American men were left untreated so that the natural course of syphilis could be studied, is an egregious example.

In my research of women with HIV/AIDS, I discovered many reasons women worried about preserving their anonymity, including illicit drug use, criminal behavior of their children, and nondisclosure of their HIV/AIDS diagnosis to family and friends. Interestingly, on the other hand, some participants may not want to remain anonymous as they are proud—proud of what they have overcome, proud of what they have learned in the process, and proud of who they are or have become—and they want to share their stories. In fact, Grinyer (2002) has reported that a cancer patient who participated in her study expressed that she wanted her real name used rather than a pseudonym. The desire of some participants notwithstanding, researchers are obliged to take every possible measure to ensure the confidentiality of data and provide anonymity to research participants.

—Desirée Ciambrone

REFERENCES

Grinyer, A. (2002). The anonymity of research participants: Assumptions, ethics, and practicalities. *Social Research Update, 36,* 1–4.

Inter-University Consortium for Political and Social Research. (2002). *Guide to social science data preparation and archiving* [online]. Supported by the Robert Wood Johnson Foundation. Available: www.ICPSR.umich.edu

APPLIED QUALITATIVE RESEARCH

Applied qualitative research usually denotes qualitative studies undertaken explicitly to inform policymaking, especially in government but also in commerce. The term was possibly first coined in 1985 in Britain as part of a venture to encourage the use of qualitative research methods in government (Walker, 1985). At the time, public officials were inclined to equate evidence with OFFICIAL STATISTICS (Weiss, 1977), although qualitative methodologies had already successfully penetrated marketing and advertising research. Such studies are often commissioned by companies or government agencies and undertaken to a closely specified contract.

Applied qualitative research is an activity rather than a movement or coherent philosophy (Ritchie & Lewis, 2003). The objective is frequently to develop, monitor, or evaluate a policy or practice, using qualitative techniques as an alternative to, or to complement, other approaches. The rationale for using qualitative methods is that they are better suited to the task, either in specific circumstances or inherently. This may be because the topic is ill-defined or not well theorized, thereby precluding the development of structured questionnaires. It might be sensitive, intangible, or ephemeral, or it could appertain to distressing or emotional events or deep feelings. It could be inherently complex or an institution. Earlier research may have produced conflicting results.

Practitioners of applied qualitative research are therefore apt to align themselves with pragmatism, matching method to specific research questions. Their ontological position is likely to correspond to "subtle realism" (Hammersley, 1992), accepting that a diverse and multifaceted world exists independently of subjective understanding but believing it to be accessible via respondents' interpretations. They will probably strive to be neutral and objective at each stage in the research process and thereby generate findings that are valid and reliable, ones that are true to the beliefs and understandings of their respondents and potentially generalizable to other settings. They will seek detailed descriptions of the realities, as understood by their respondents, but clearly delineate between these interpretations and their own in analysis, making sense of complexity by simplification and structuring.

Applied qualitative research is eclectic in its use of qualitative methods although heavily reliant on IN-DEPTH INTERVIEWS and FOCUS GROUPS. There is currently a trend toward repeat interviews with the same respondents to investigate the links between intention, behavior, and outcomes. PROJECTIVE TECHNIQUES are commonly used, typically to stimulate respondents' reflections on the process and outcome of projective exercises rather than to elicit scores on personality traits; occasionally, ETHNOGRAPHIC and ACTION RESEARCH techniques are also used. Different methods, including quantitative ones, are frequently used within the same project. When used with surveys, qualitative techniques are variously employed to refine concepts and develop questionnaires; to help interpret, illuminate, illustrate, and qualify survey findings; to triangulate between methods; and to provide complementary insight.

—Robert Walker

REFERENCES

Hammersley, M. (1992). *What's wrong with ethnography.* London: Routledge.

Ritchie, J., & Lewis, J. (2003). *Qualitative research practice.* London: Sage.

Walker, R. (Ed.). (1985). *Applied qualitative research.* Aldershot, UK: Gower.

Weiss, C. (1977). *Use of social research in public policy.* Lexington, MA: D. C. Heath.

APPLIED RESEARCH

Applied research is, by definition, research that is conducted for practical reasons and thus often has an immediate application. The results are usually actionable and recommend changes to increase effectiveness and efficiency in selected areas. Applied research rarely seeks to advance theoretical frameworks. THEORY is generally used instrumentally to help identify and define concepts and variables so that researchers can operationalize them for analysis.

Applied research is usually used in contrast to BASIC RESEARCH (sometimes called pure research) that does not yield immediate application upon completion. Basic research is designed to advance our understanding and knowledge of fundamental social processes and interactions, regardless of practical or immediate applications. This does not imply that applied research is more useful than basic research. Almost all basic research eventually results in some worthwhile application over the long run. The main distinction focuses on whether the research has *immediate use.* However, applied research often is easier to relate to by the general public as it yields actionable and practical results, whereas some basic research may have intuitive relevance only for researchers.

A common type of applied research is program evaluation. In some cases, researchers study the effectiveness of the current program to evaluate how effectively the program is in achieving its goal. In other cases, researchers might study the efficiency of the program's delivery system to examine the extent of resource wastage. In either case, the researchers' intent is to find ways to improve aspects of the program or its delivery system.

Another example of applied research is market research. Marketers may conduct research (focus groups, in-depth interviews, or surveys) to ferret consumers' product needs and preferences, enabling them to package the most appropriate products. Marketers may also test their communication materials with consumers to ensure that the intended marketing message is effectively delivered. The recommendations from this type of research could include word changes, packaging style, delivery approach, and even the color scheme of the materials.

In many cases, the audience of applied research is often limited to specific interested parties. As such, the results of applied research are often packaged as in-house reports that have limited circulation or are proprietary. However, some applied research is shared via publication in professional journals. For example, *The Gerontologist* has a section titled *Practice Concepts* that features applied research and short reports, in addition to publishing regular-length articles. An example from the *Practice Concepts* section is the report written by Si, Neufeld, and Dunbar (1999) on the relationship between having bedrails and residents' safety in a short-term nursing home rehabilitation unit. Si et al. found that in the selected unit, the dramatic decline in bedrail use did not lead to an increase in residents' injuries. They suggested that the necessity of bedrails depended on both the training and experience of the staff as well as the type and severity of the patients' injuries.

For further reading, please refer to the following books: *Applied Research Design: A Practical Guide* by Hendrick, Bickman, and Rog (1993) and *Focus Groups: A Practical Guide for Applied Research* by Krueger (2000).

—VoonChin Phua

REFERENCES

Hendrick, T., Bickman, L., & Rog, D. J. (1993). *Applied research design: A practical guide.* Newbury Park, CA: Sage.

Krueger, R. A. (2000). *Focus groups: A practical guide for applied research.* Thousand Oaks, CA: Sage.

Si, M., Neufeld, R. R., & Dunbar, J. (1999). Removal of bedrails on a short-term nursing home rehabilitation unit. *The Gerontologist, 39,* 611–614.

ARCHIVAL RESEARCH

For the social scientist, archival research can be defined as the locating, evaluating, and systematic interpretation and analysis of sources found in archives.

Original source materials may be consulted and analyzed for purposes other than those for which they were originally collected—to ask new questions of old data, provide a comparison over time or between geographic areas, verify or challenge existing findings, or draw together evidence from disparate sources to provide a bigger picture.

For the social scientist, using archives may seem relatively unexciting compared to undertaking fieldwork, which is seen as fresh, vibrant, and essential for many researchers. Historical research methods and archival techniques may seem unfamiliar; for example, social scientists are not used to working with handbreak written sources that historians routinely consult. However, these methods should be viewed as complementary, not competing, and their advantages should be recognized for addressing certain research questions. Indeed, consulting archival sources enables the social scientist to both enhance and challenge the established methods of defining and collecting data.

What, then, are the types of sources to be found of interest to the social scientist? Archives contain official sources (such as government papers), organizational records, medical records, personal collections, and other contextual materials. Primary research data are also found, such as those created by the investigator during the research process, and include transcripts or tapes of interviews, FIELDNOTES, personal DIARIES, observations, unpublished manuscripts, and associated correspondence. Many university repositories or special collections contain rich stocks of these materials. Archives also comprise the cultural and material residues of both institutional and theoretical or intellectual processes, for example, in the development of ideas within a key social science department.

However, archives are necessarily a product of sedimentation over the years—collections may be subject to erosion or fragmentation—by natural (e.g., accidental loss or damage) or manmade causes (e.g., purposive section or organizational disposal policies). Material is therefore subjectively judged to be worthy of preservation by either the depositor or archivist and therefore may not represent the original collection in its entirety. Storage space may also have had a big impact on what was initially acquired or kept from a collection that was offered to an archive.

Techniques for locating archival material are akin to excavation—what you can analyze depends on what you can find. Once located, a researcher is free to immerse himself or herself in the materials—to evaluate, review, and reclassify data; test out prior hypotheses; or discover emerging issues. Evaluating sources to establish their validity and reliability is an essential first step in preparing and analyzing archival data. A number of traditional social science analytic approaches can be used once the data sources are selected and assembled—for example, GROUNDED THEORY to uncover patterns and themes in the data, BIOGRAPHICAL METHODS to document lives and ideas by verifying personal documentary sources, or CONTENT ANALYSIS to determine the presence of certain words or concepts within text.

Access to archives should not be taken for granted, and the researcher must typically make a prior appointment to consult materials and show proof of status. For some collections, specific approval must be gained in advance via the archivist.

—Louise Corti

REFERENCES

Elder, G., Pavalko, E. K., & Clipp, E. C. (1993). *Working with archival lives: Studying lives* (Sage University Papers on Quantitative Applications in the Social Sciences, 07–088). Newbury Park, CA: Sage.

Foster, J., & Sheppard, J. (1995). *British archives: A guide to archival resources in the United Kingdom* (3rd ed.). London: Macmillan.

Hill, M. M. (1993). *Archival strategies and techniques* (Qualitative Research Methods Series No. 31). Newbury Park, CA: Sage.

Scott, J. (1990). *A matter of record: Documentary sources in social research.* Cambridge, UK: Polity.

ARCHIVING QUALITATIVE DATA

Qualitative data archiving is the long-term preservation of qualitative data in a format that can be accessed by researchers, now and in the future. The key to ensuring long-term accessibility lies in the strategies for data processing, the creation of informative documentation, and in the physical and technical procedures for storage, preservation, security, and access (see DATA ARCHIVES).

QUALITATIVE DATA are data collected using QUALITATIVE RESEARCH methodology and techniques across the range of social science disciplines. Qualitative research often encompasses a diversity of methods and tools rather than a single one, and the types of data collected depend on the aim of the study, the nature of the sample, and the discipline. As a result, data types extend to IN-DEPTH or UNSTRUCTURED INTERVIEWS, SEMI-STRUCTURED INTERVIEWS, FIELDNOTES, unstructured DIARIES, observations, personal documents, and photographs. Qualitative research often involves producing large amounts of raw data, although the methods typically employ small sample sizes. Finally, these data may be created in a number of different formats: digital, paper (typed and handwritten), and audiovisual.

Until 1994, no procedures existed for the systematic archiving and dissemination of qualitative data, although the ORAL HISTORY community had a professional interest in preserving tape recordings gathered from oral history interviewing projects. Many of these tape recordings may be found at the British Library National Sound Archive in London.

In 1994, the first qualitative data-archiving project on a national scale was established in the United

Kingdom, with support from the Economic and Social Research Council. The Qualidata Centre established procedures for sorting, describing, processing both raw data and accompanying documentation (meta-data), and establishing appropriate mechanisms for access.

The way data sets are processed can be split into three main activities: checking and anonymizing, converting the data set, and generating meta-data. Checking the data set consists of carrying out activities such as establishing the completeness and quality of data, as well as the relationships between data items (e.g., interviews, field notes, and recordings). Good-quality transcription is essential when archiving interviews without audio recordings, as is ensuring that the content of the data meets any prior consent agreements with participants. CONFIDENTIALITY and copyright are two key issues that arise in the archiving and dissemination of qualitative data. Converting data consists of transferring data to a format suitable for both preservation and dissemination. This may include digitization of paper resources.

The third main processing activity, generating meta-data (data about data), refers to the contextual information generated during processing, such as the creation of a "data list" of interview details, and ensuring that speaker and interviewer tags and questions or topic guide headers are added to raw text. A key function of meta-data is to enable users to locate transcripts or specific items in a data collection most relevant to their research. User guides are also created that bring together key documentation such as topic guides, personal research diaries, end of award reports, and publications. This type of information, collated by principal investigators at the fieldwork and analysis stage, is clearly of great benefit for future archiving.

Finally, the reuse or SECONDARY ANALYSIS OF QUALITATIVE DATA is a topic that has began to spark debate since the late 1990s. Qualitative data can be approached by a new analyst in much the same way as quantitative data can—to ask new questions of old data, provide comparisons, or verify or challenge existing findings. However, the shortcomings of using "old data" for secondary analysis should always be recognized.

First, the lack of the exact data required to answer the question at hand can be a frustrating experience. Second, as Hammersley (1997) notes, an assessment of the quality and validity of the data without access to full knowledge of the research process is more complicated than appraising one's own data collection. This is because the path of qualitative analysis is almost never linear and is almost certainly likely to involve a degree of trial and error in the pursuit of interesting lines of investigation.

Third, for qualitative data in particular, there is the difficulty of reconstructing the interview situation or fieldwork experience, which is often not fully captured in an archived interview transcript or set of field notes. Consequently, a secondary analyst does not have access to the tacit knowledge gained by the primary researcher, which, in some circumstances, may be important in helping interpret data. Listening to archived audio-recorded interviews, in addition to reading transcripts, can often help to reconstruct an interview setting.

—Louise Corti

REFERENCES

Corti, L., Foster, J., & Thompson, P. (1995). *Archiving qualitative research data* (Social Research Update No. 10). Surrey, UK: Department of Sociology, University of Surrey.

Hammersley, M. (1997). Qualitative data archiving: Some reflections on its prospects and problems. *Sociology, 31*(1), 131–142.

James, J. B., & Sørensen, A. (2000, December). Archiving longitudinal data for future research: Why qualitative data add to a study's usefulness. *Forum Qualitative Sozialforschung* [Forum: Qualitative Social Research] [online journal], *1*(3). Available: http://qualitative-research.net/fqs/fqs-eng.htm

Mruck, K., Corti, L., Kluge, S., & Opitz, D. (Eds.). (2000, December). Text archive: Re-analysis. *Forum Qualitative Sozialforschung* [Forum: Qualitative Social Research] [online journal], *1*(3). Available: http://qualitative-research.net/fqs/fqs-eng.htm

ARIMA

ARIMA is an acronym for autoregressive, integrated, moving average, denoting the three components of a general class of STOCHASTIC time-series models described by Box and Tiao (1965, 1975) and developed by Box and Jenkins (1976). Estimated univariate ARIMA models are often used successfully for forecasting and also form the basic foundation for multivariate analysis, such as INTERVENTION ANALYSIS assessment or TRANSFER FUNCTION models in the BOX-JENKINS MODELING approach to TIME SERIES. The use of ARIMA models in conjunction with transfer function models is often contrasted with the use of time-series REGRESSION models, but ARIMA models are increasingly used in tandem with regression models as well.

THE PIECES

ARIMA models have three components, with each one describing a key feature of a given time-series process. Briefly, the three components correspond to the effect of past values, called an autoregressive (AR) process; past shocks, called a moving average (MA) process; and the presence of a stochastic trend, called an integrated (I) process. Each of these is discussed below with reference to example time-series processes.

Autoregressive

Time-series processes are often influenced by their own past values. Consider the percentage of the public that says they approve of the way the president is doing his job. Public sentiment about the president's performance today depends in part on how they viewed his performance in the recent past. More generally, a time series is autoregressive if the current value of the process is a function of past values of the process. The effect of past values of the process is represented by the autoregressive portion of the ARIMA model, written generally as follows:

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p y_{t-p} + \mu_t. \quad (1)$$

Each observation consists of a random shock and a linear combination of prior observations. The number of lagged values determines the order of autoregression and is given by p, denoted AR(p).

Moving Average

ARIMA models assume that time-series processes are driven by a series of random shocks. Random shocks may be loosely thought of as a collection of inputs occurring randomly over time. More technically, random shocks are normally and identically distributed random variables with mean zero and constant variance. As an example, consider that reporting requirements for violent crime may change. This change "shocks" statistics about violent crime so that today's reports are a function of this (and other) shock(s). We can write this as follows:

$$y_t = a + \mu_t + c_1 \mu_{t-1} + c_2 \mu_{t-2} + \cdots + c_q \mu_{t-q}, \quad (2)$$

where q gives the number of past shocks, μ that affect the current value of the $\bar{y}$ process, here violent crime statistics. The number of lagged values of μ determines the moving average *order* and is given by q, denoted MA(q).

Integrated

ARIMA models assume that the effects of past shocks and past values decay over time: Current values of presidential approval reflect recent shocks and recent values of approval more so than distant values. For example, mistakes a president makes early in his tenure are forgotten over time. This implies that the PARAMETER values in the general AR and MA model components must be restricted to ensure that effects decay (i.e., that the time series is stationary). Loosely, a stationary time series is one with constant mean, variance, and covariances over time. A series that is not stationary is known as *integrated.* Some economic time series (e.g., economic growth) are thought to be integrated because they exhibit a historical tendency to grow.

Integrated time series must be transformed before analysis can take place. The most common TRANSFORMATION is called differencing, whereby we analyze the changes in y rather than the levels of y (we analyze $y_t - y_{t-1}$). The number of times a series must be differenced to be stationary gives the order of integration, denoted I(d), where d may be any real number (if d is a fraction, the process is said to be fractionally integrated, and the model is called an ARFIMA model). If the time series is integrated in levels but is not integrated in changes, the process is integrated of order $d = 1$.

The condition for stationarity for an AR(1) process is that b_1 must be between -1 and $+1$. If b_1 is not in this range, the effects of the past values of the process would accumulate over time, and the values of successive ys would move toward infinity; that is, the series would not be stationary and would need to be differenced before we (re)consider the order of autoregression that characterizes the process. In the MA(1) case, the restrictions require c_1 to be between -1 and $+1$. In the case of higher order processes, similar restrictions must be placed on the parameters to ensure stationarity.

THE GENERAL MODEL

The general ARIMA(p, d, q) model combines the model components given above and is written as follows:

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p y_{t-p} + \mu_t + c_1 \mu_{t-1} + c_2 \mu_{t-2} + \cdots + c_q \mu_{t-q}, \qquad (3)$$

where y has been differenced a suitable number (d, generally 0 or 1) of times to produce stationarity, p gives the number of AR lags, and q is the number of MA lags. Any time series may contain any number (including 0) of AR, I, and MA parameters, but the number of parameters will tend to be small. Generally, a simple structure underlies time-series processes; most processes are well explained as a function of just a few past shocks and past values. In fact, many social science processes behave as simple ARIMA(1,0,0) processes.

It is important to note that a given time-series process may be best characterized as an ARIMA(0,0,0) process, also called a WHITE-NOISE process. A white-noise process is a series of random shocks. It does not have autoregressive, integrated, or moving average components. When estimating regression models, inferences are only valid if the error is a white-noise process.

The ARIMA model that characterizes a particular series is identified through a specific methodology. In particular, an ARIMA model of the process that we care about is built from an observed time series in an empirically driven process of identification, estimation, and diagnosis (see BOX-JENKINS MODELING for an explanation of how this is done). For a more general treatment of ARIMA models and the Box-Jenkins approach, see McCleary and Hay (1980).

—Suzanna De Boef

REFERENCES

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

Box, G. E. P., & Tiao, G. C. (1965). A change in level of nonstationary time series. *Biometrika, 52,* 181–192.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association, 20,* 70–79.

McCleary, R., & Hay, R. A., Jr. (1980). *Applied time series analysis for the social sciences.* London: Sage.

ARROW'S IMPOSSIBILITY THEOREM

Kenneth J. Arrow (1921–) is one of the 20th century's leading economists. His most famous contribution to the social sciences is Arrow's impossibility theorem (Arrow, 1963). Arrow examined the logic of choice by a society based on the preferences of the individuals in that society and identified several reasonable conditions that a social choice rule used by a society should have. The impossibility theorem states that when there are three or more alternatives, it is impossible to find a social choice rule that always aggregates individual preferences into social preferences in a way that satisfies these desirable conditions. Thus, even if each individual's preferences are perfectly consistent and logical, the preferences of society as a whole might be incoherent.

Begin by assuming each individual in a society has preferences over a list of alternatives. A social choice rule is a rule that takes all individual preferences and combines them to determine the preferences of society as a whole. For example, individuals might have preferences over political candidates, and we seek a social choice rule (such as majority rule) to combine their preferences and determine who should take office. Arrow identified five reasonable conditions that a social choice rule should satisfy. The social choice rule should be the following:

- *Complete.* The social choice rule should provide a complete ranking of all alternatives. That is, for any alternatives A and B, the social choice rule should tell us if A is preferred to B, B is preferred to A, or if there is social indifference between A and B.
- *Paretian.* If every individual prefers A to B, then the social choice rule should rank A above B. Selecting alternative A over alternative B in this case would be a Pareto-optimal move (one that makes all individuals better off), and the social choice rule should rank A above B.
- *Transitive.* If the social choice rule ranks A above B and B above C, then A should be ranked higher than C.
- *Independent of irrelevant alternatives.* The ranking of A compared to B should not depend on preferences for other alternatives.

• *Nondictatorial.* The social choice rule should not depend on the preferences of only one individual (a dictator).

The proof of Arrow's impossibility theorem demonstrates that it is impossible to find a social choice rule that always satisfies all five of these conditions when there are three or more alternatives. That is, no matter what rule a society adopts to determine social preferences, it will either have some undesirable "irrational" features or be determined entirely by one person.

The result of the impossibility theorem is troubling to those who study democratic choice. However, relaxing some of the conditions imposed by the impossibility theorem or adding additional restrictions does allow for making social decisions through the political process. An important example is when alternatives are assumed to be arranged along some one-dimensional spectrum, and individual preferences over these alternatives are single-peaked, meaning that strength of preference peaks at the most preferred alternative and decreases across other alternatives as they become further from the most preferred alternative on the spectrum. In this case, with an odd number of individuals, majority voting will lead to adopting the preferences of the median voter (e.g., Black, 1958; Downs, 1957). Societal norms and institutions can also offer a means to escape the implications of the impossibility theorem (e.g., Riker, 1980).

—Garrett Glasgow

REFERENCES

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New Haven, CT: Yale University Press.

Black, D. (1958). *The theory of committees and elections.* Cambridge, UK: Cambridge University Press.

Downs, A. (1957). *An economic theory of democracy.* New York: Harper Collins.

Riker, W. H. (1980). Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review, 74,* 432–446.

ARTIFACTS IN RESEARCH PROCESS

The term *research artifacts* refers to the systematic biases, uncontrolled and unintentional, that can threaten the INTERNAL or EXTERNAL VALIDITY of one's research conclusions. The potential for research artifacts in social science research exists because human participants are sentient, active organisms rather than passive, inactive objects. As such, the interactions between the researcher and the agents of study can potentially alter the experimental situation in subtle ways unintended by the researcher.

The potential problems that research artifacts may cause in social science research were evident in the late 19th and early 20th centuries. For example, a controversial series of studies conducted at the Western Electric Company's Hawthorne Works in Illinois during the 1920s was interpreted as implying that simply observing someone's behavior is sufficient to alter that person's behavior. Earlier, the work of the eminent German psychologist Oskar Pfungst (1911/1965), involving a horse named "Clever Hans," at the turn of the century dramatically demonstrated how an individual's expectations could unconsciously influence another's behavior. In another classic work, Saul Rosenzweig (1933) argued that the experimental situation should be considered a psychological problem in its own right. He suggested that one must consider the possibility that the attitudes and motivations of the experimenter and research participant can subtly influence the research situation. Despite these warning signs of the potential problem of research artifacts, it was not until the late 1950s and 1960s that this problem and its ramifications for social science research were systematically investigated. This research, which was first pulled together in a volume edited by Rosenthal and Rosnow (1969), has commonly focused on two potential sources of research artifacts: the researcher and the research participant.

RESEARCHER-RELATED ARTIFACTS

Because researchers are not passive, disinterested participants in the scientific process, they can sometimes unknowingly or unwittingly engage in behaviors that introduce an artifact into the research design. Robert Rosenthal (1966), in a seminal book, defined two broad types of researcher-related artifacts as noninteractional and interactional artifacts. Noninteractional artifacts are biases that do not actually affect the research participants' behavior. Rather, they are systematic biases in how the researcher observes, interprets, or reports the research results. For example, the observations the researcher makes may be biased in a manner that increases the likelihood that the researcher's hypotheses will be supported. Similarly, how the researcher interprets the data may be biased in the direction of supporting the hypotheses of interest.

Although these two types of noninteractional biases may be unintentional, a third type is clearly intentional. The researcher may fabricate or fraudulently manipulate data to manufacture support for his or her hypotheses, which is clearly unethical.

The second category of researcher-related artifacts is associated with the interaction between the researcher and the research participants. For example, various biosocial characteristics of the researcher (such as ethnicity, sex, and age) have the potential to influence the participants' behavior in an unintended manner. Likewise, psychosocial attributes of the researcher (such as personality) may lead participants to respond in a manner ostensibly unrelated to the factors under investigation. Situational effects (such as changes in the researcher's behavior during the course of the study) can alter the manner in which individuals respond to the experimental stimuli. Finally, as Rosenthal (1966) demonstrated, even the simple knowledge of one's hypotheses can influence the researcher's behavior, leading to a self-fulfilling prophecy or EXPERIMENTER EXPECTANCY bias, which is by far the most thoroughly investigated source of interactional artifacts.

PARTICIPANT ARTIFACTS

The research participants are also active, motivated parties who are cognizant of their role as the object of study. Martin T. Orne (1962) argued that VOLUNTEER SUBJECTS, in particular, are motivated by a desire to help the cause of science and therefore will adopt the role of a "good subject." In this capacity, they seek out and respond to cues (DEMAND CHARACTERISTICS) that ostensibly enable them to behave in ways that are likely to support the researcher's hypotheses. Milton J. Rosenberg (1969), however, argued that EVALUATION APPREHENSION motivates research participants' behaviors. Their concern over being negatively evaluated, he argued, leads them to respond to experimental stimuli in a manner that presumably enables them to "look good." A third possible motivation underlying the participants' behavior in the research setting is the desire to obey the investigator who is perceived as an authority figure. Regardless of the underlying motivation, the implication is that individuals are responding to the stimuli under investigation in a systematically biased fashion.

MINIMIZING RESEARCH ARTIFACTS

Work on identifying potential sources of research artifacts has implied ways to either detect the presence of an artifact or minimize its introduction into the RESEARCH DESIGN. For example, potential noninteractional research-related artifacts may be identified through independent REPLICATION of research results and unimpeded access to research data by other scientists. Rosenthal (1966) described how an expectancy control design could be used to detect the presence of experimenter expectancy bias. Standardized research protocols, DOUBLE-BLIND PROCEDURE experimental designs, and the separation in time and place of the experimental treatment and measurement of the treatment effects represent other general strategies for minimizing researcher-related artifacts.

Strategies for avoiding certain participant artifacts include the introduction of alternative task-orienting cues and the creation of an experimental situation that reduces evaluation apprehension concerns or the motivation to respond to demand cues. Using participants who are unaware that they are participating in a research study can minimize the REACTIVITY problem, but this strategy can raise ethical issues regarding the invasion of PRIVACY.

Orne (1969) described how quasi-control subjects could be used to ferret out demand characteristics operating in a study. These participants essentially act as "consultants" to the researcher. Quasi-control subjects are asked to discuss the rationale behind their behavior after exposure to various aspects of the experimental design or speculate on how they might respond if they were actually exposed to the experimental stimuli. In the end, the most important mechanism that researchers have for identifying and controlling potential research artifacts is the independent replication of research results by different researchers.

—David B. Strohmetz and Ralph L. Rosnow

REFERENCES

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 776–783.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143–179). New York: Academic Press.

Pfungst, O. (1965). *Clever Hans: The horse of Mr. von Osten.* New York: Holt, Rinehart, & Winston. (Original work published 1911.)

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279–349). New York: Academic Press.

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts.

Rosenthal, R., & Rosnow, R. L. (Eds.). (1969). *Artifact in behavioral research.* New York: Academic Press.

Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review, 40,* 337–354.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research.* New York: W. H. Freeman.

Strohmetz, D. B., & Rosnow, R. L. (1994). A mediational model of research artifacts. In J. Brzezinski (Ed.), *Probability in theory-building: Experimental and nonexperimental approaches to scientific research in psychology* (pp. 177–196). Amsterdam: Editions Rodopi.

ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is the study of how to build intelligent systems. Specifically, AI is concerned with developing computer programs with intelligent behaviors, such as problem solving, reasoning, and learning.

Since the name *artificial intelligence* was coined in 1956, AI has been through cycles of ups and downs. From the beginning, game playing has been an important AI research subject for problem solving. In 1997, IBM's Deep Blue chess system defeated world-champion Garry Kasparov. Expert systems have been successfully developed to achieve expert-level performance in real-world applications by capturing domain-specific knowledge from human experts in knowledge bases for machine reasoning, and they have been used widely in science, medicine, business, and education. Machine learning arose as an important research subject in AI, enabling dynamic intelligent systems with learning ability to be built. Patterned after biological evolution, genetic algorithm spawns the population of competing candidate solutions and drives them to evolve ever better solutions. Agent-based models demonstrate that globally intelligent behavior can arise from the interaction of relatively simple structures. Based on the interaction between individual agents, intelligence is seen as emerging from society and not just as a property of an individual agent. Recently, AI has been moving toward building intelligent agents in real environments, such as Internet auctions. Multiagent systems, as the platform for the convergence of various AI technologies, benefit from the abstractions of human society in an environment that contains multiple agents with different capabilities, goals, and beliefs.

The AI approach has been applied to social, psychological, cognitive, and language phenomena. At the theory-building level, AI techniques have been used for theory formalization and simulation of societies and organizations. For example, a GAME THEORY problem, the "prisoner's dilemma," is a classic research problem that benefits from computer SIMULATION in exploring the conditions under which cooperation may arise between self-interested actors who are motivated to violate agreements to cooperate at least in the short term. More recently, the multiagent system emerged as a new approach of modeling social life. These models show how patterns of diffusion of information, emergence of norms, coordination of conventions, and participation in collective action can be seen as emerging from interactions among adaptive agents who influence each other in response to the influence they receive.

At a data analysis level, AI techniques have been used to intelligently analyze qualitative data (data collected using methods such as ethnography) using symbolic processors and quantitative data using AI-enhanced statistical procedures. For example, it has been shown that NEURAL NETWORKS can readily substitute for and even outperform multiple regression and other multivariate techniques. More recently, the process of knowledge discovery has shown great potential in social science research by taking the results from data mining (the process of extracting trends or patterns from data) and transforming them into useful and interpretable knowledge through the use of AI techniques.

—Xiaoling Shu

REFERENCES

Alonso, E. (2002). AI and agents: State of the art. *AI Magazine, 23*(3), 25–29.

Bainbridge, E. E., Carley, K. M., Heise, D. R., Macy, M. W., Markovsky, B., & Skvoretz, J. (1994). Artificial social intelligence. *Annual Review of Sociology, 20,* 407–436.

Conte, R., & Dellarocas, C. (2001). Social order in info societies: An old challenge for innovation. In R. Conte &

C. Dellarocas (Eds.), *Social order in multi-agent systems* (pp. 1–16). Boston: Kluwer Academic.

Luger, G. F. (2002). *Artificial intelligence: Structures and strategies for complex problem solving* (4th ed.). Reading, MA: Addison-Wesley.

Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology, 28,* 143–166.

ARTIFICIAL NEURAL NETWORK. *See* NEURAL NETWORK

ASSOCIATION

The analysis of the relation (association) between variables is a fundamental task of social research. The association between two (categorical) variables can be expressed in a two-way CONTINGENCY TABLE or in cross-classification by calculating row percentages, column percentages, or total percentages. This technique can also be implemented if variables are continuous. The variables must be grouped into categories. A summary measure of the association between two variables is given by association coefficients, which measure the strength of the relation between two variables X and Y and, for ordinal, interval, and ratio variables, the direction of the association. The direction can be positive (concordant; higher values in X correspond to higher values in Y) or negative (discordant; higher values in X correspond to lower values in Y).

Two broad classes of association coefficients can be distinguished:

- ASYMMETRIC (or directional) MEASURES or association coefficients assume that one of the two variables (e.g., X) can be identified as the independent variable and the other variable (e.g., Y) as the dependent variable. For example, the variable X might be gender, and variable Y might be occupational status.
- SYMMETRIC MEASURES or association coefficients do not require differentiation between dependent and independent variables. They can, therefore, be used if such a distinction is impossible; an example is a researcher who is interested in the relations between different leisure-time activities (doing sports and visiting the theater). However, symmetric coefficients can also be computed if identification as independent and dependent variables is possible.

Some examples of symmetric coefficients (e.g., Healey, 1995; SPSS, 2002) are the following: phi (ϕ); Cramér's V; the contingency coefficient C for nominal variables; Kendall's tau$_a$(τ_a), tau$_b$ (τ_b), and tau$_c$ (τ_c); Goodman and Kruskal's gamma (γ) for ordinal variables; Spearman's rho (ρ) for rank (ordinal) data; and Pearson's r for interval or ratio variables. Both the size of the table and the number of both variables' categories determine which measure is used within one measurement level. Directed symmetric coefficients are often synonymously labeled as CORRELATION coefficients; sometimes, the term *correlation* is reserved for directed symmetric associations between two continuous variables.

Examples of asymmetric coefficients (e.g., Healey, 1995; SPSS, 2002) are lambda (λ), Goodman and Kruskal's τ, the uncertainty coefficient for nominal variables, Somers's d for ordinal variables, and eta (η), if the dependent variable is interval or ratio scaled and the independent variable is nominal. Further symmetric coefficients, such as symmetric lambda, symmetric Somers, and so on, can be derived from asymmetric coefficients by means of averaging.

The quoted coefficients (except η) assume that both variables have the same measurement level. If the analyzed variables have different measurement levels, it is recommended to reduce the measurement level and then compute an appropriate association measure for the lower measurement level. Unfortunately, this is often necessary as coefficients for mixed measurement levels are not available in standard statistical software packages. Coefficients for mixed measurement levels can be derived from Daniels' generalized coefficient of correlation, for example (Daniels, 1944; Kendall, 1962, pp. 19–33).

The concept of association can also be generally applied to the multivariate case. Examples of multivariate association coefficients include the CANONICAL CORRELATION ANALYSIS between two sets of variables, $X = \{X_1, X_2, \ldots, X_p\}$ and $Y = \{Y_1, Y_2, \ldots, Y_p\}$; the MULTIPLE CORRELATION coefficient R (from multiple regression) for a set of variables, $X = \{X_1, X_2, \ldots, X_p\}$ and $Y = \{Y_1\}$; and the inertia between two nominal variables, X and Y, from BIVARIATE ANALYSIS.

An alternative strategy to analyze bivariate and multivariate contingency tables is ASSOCIATION MODELS based on LOG-LINEAR MODELING. Log-linear models allow one to specify and analyze different and complex patterns of associations, whereas association coefficients give a summary measure.

Association coefficients measure the strength of the relation between two variables. Users oftentimes demand guidelines to interpret the strength, but general guidelines are impossible. The strength of an association depends on the sample's homogeneity, the reliability of the variables, and the type of relationship between the variables. In psychological experiments (very homogeneous sample, variables with high reliability, direct casual link), a coefficient of 0.4 may signal weak correlation, whereas the same value could be suspiciously high in a sociological survey (heterogeneous sample, lower reliability, indirect causal links). Hence, when fixing threshold values, it is important to consider these factors along with the results of previous studies.

—Johann Bacher

REFERENCES

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika, 35,* 129–135.

Healey, J. F. (1995). *Statistics: A tool for social research* (3rd ed.). Belmont, CA: Wadsworth.

Kendall, M. G. (1962). *Rank correlation methods* (3rd ed.). London: Griffin.

SPSS. (2002, August). Crosstabs. In *SPSS 11.0 statistical algorithms* [online]. Available: http://www.spss.com/tech/stat/algorithms/11.0/crosstabs.pdf

ASSOCIATION MODEL

Although the term ASSOCIATION is used broadly, *association model* has a specific meaning in the literature on CATEGORICAL DATA ANALYSIS. By *association model,* we refer to a class of statistical models that fit observed frequencies in a cross-classified table with the objective of measuring the strength of association between two or more ordered categorical variables. For a two-way table, the strength of association being measured is between the two categorical variables that comprise the cross-classified table. For three-way or higher-way tables, the strength of association being measured can be between any pair of ordered categorical variables that comprise the cross-classified table. Although some association models make use of the a priori ordering of the categories, other models do not begin with such an assumption and indeed reveal the ordering of the categories through estimation. The association model is a special case of a LOG-LINEAR MODEL or log-bilinear model.

Leo Goodman should be given credit for having developed association models. His 1979 paper, published in the *Journal of the American Statistical Association,* set the foundation for the field. This seminal paper was included along with other relevant papers in his 1984 book, *The Analysis of Cross-Classified Data Having Ordered Categories.* Here I first present the canonical case for a two-way table before discussing extensions for three-way and higher-way tables. I will also give three examples in sociology and demography to illustrate the usefulness of association models.

GENERAL SETUP FOR A TWO-WAY CROSS-CLASSIFIED TABLE

For the cell of the ith row and the jth column ($i = 1, \ldots, I$, and $j = 1, \ldots, J$) in a two-way table of R and C, let f_{ij} denote the observed frequency and F_{ij} the expected frequency under some model. Without loss of generality, a log-linear model for the table can be written as follows:

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \mu_{ij}^{RC}, \qquad (1)$$

where μ is the main effect, μ^R is the row effect, μ^C is the column effect, and μ^{RC} is the interaction effect on the logarithm of the expected frequency. All the parameters in equation (1) are subject to ANOVA-type normalization constraints (see Powers & Xie, 2000, pp. 108–110). It is common to leave μ^R and μ^C unconstrained and estimated nonparametrically. This practice is also called the "saturation" of the marginal distributions of the row and column variables. What is of special interest is μ^{RC}: At one extreme, μ^{RC} may all be zero, resulting in an independence model. At another extreme, μ^{RC} may be "saturated," taking $(I-1)(J-1)$ degrees of freedom, yielding exact predictions ($F_{ij} = f_{ij}$ for all i and j).

Typically, the researcher is interested in fitting models between the two extreme cases by altering specifications for μ^{RC}. It is easy to show that all ODDS RATIOS in a two-way table are functions of the interaction parameters (μ^{RC}). Let θ_{ij} denote a local log-odds ratio for a 2×2 subtable formed from four adjacent cells obtained from two adjacent row categories and two adjacent column categories:

$$\theta_{ij} = \log\{[F_{(i+1)(j+1)}F_{ij}]/[F_{(i+1)j}F_{i(j+1)}]\},$$
$$i = 1, \ldots, I-1, j = 1, \ldots, J-1.$$

Let us assume that the row and column variables are ordinal on some scales x and y. The scales may be observed or latent. A linear-by-linear association model is as follows:

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \beta x_i y_j, \qquad (2)$$

where β is the parameter measuring the association between the two scales x and y representing, respectively, the row and column variables. If the two scales x and y are directly observed or imputed from external sources, estimation of equation (2) is straightforward via MAXIMUM LIKELIHOOD ESTIMATION for log-linear models.

ASSOCIATION MODELS FOR A TWO-WAY TABLE

If we do not have extra information about the two scales x and y, we can either impose assumptions about the scales or estimate the scales internally. Different approaches give rise to different association models. Below, I review the most important ones.

Uniform Association

If the categories of the variables are correctly ordered, the researcher may make a simplifying assumption that the ordering positions form the scales (i.e., $x_i = i$, $y_j = j$). Let the practice be called *integer scoring*. The integer-scoring simplification results in the following uniform association model:

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \beta_{ij}. \qquad (3)$$

The researcher can estimate the model with actual data to see whether this assumption holds true.

Row Effect and Column Effect Models

Although the uniform association model is based on integer scoring for both the row and column variables, the researcher may wish to invoke integer scoring only for the row *or* the column variable. When integer scoring is used only for the column variable, the resulting model is called the *row effect model*. Conversely, when integer scoring is used only for the row variable, the resulting model is called the *column effect model*. Taking the row effect model as an example, we can derive the following model from equation (2):

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + j\phi_i. \qquad (4)$$

This model is called the row effect model because the latent scores of the row variable ($\phi_i = \beta x_i$) are revealed by estimation after we apply integer scoring for the column variable. That is, ϕ_i is the "row effect" on the association between the row variable and the column variable. Note that the terms *row effect* and *column effect* here have different meanings than μ_i^R and μ_j^C, which are fitted to saturate the marginal distributions of the row and column variables.

Goodman's *RC* Model

The researcher can take a step further and treat both the row and column scores as unknown. Two of Goodman's (1979) association models are designed to estimate such latent scores. Goodman's Association Model I simplifies equation (1) to the following:

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + j\phi_i + i\varphi_j, \qquad (5)$$

where ϕ_i and φ_j are, respectively, unknown row and column scores as in the row effect and column effect models. However, it is necessary to add three normalization constraints to uniquely identify the $(I + J)$ unknown parameters of ϕ_i and φ_j.

Goodman's Association Model I requires that both the row and column variables be correctly ordered a priori because integer scoring is used for both, as shown in equation (5). This requirement means that the model is not invariant to positional changes in the categories of the row and column variables. If the researcher has no knowledge that the categories are correctly ordered or in fact needs to determine the correct ordering of the categories, Model I is not appropriate. For this reason, Goodman's Association Model II has received the most attention. It is of the following form:

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \beta\phi_i\varphi_j, \qquad (6)$$

where β is the association parameter, and ϕ_i and φ_j are unknown scores to be estimated. Also, ϕ_i and φ_j are

Table 1 Comparison of Association Models

Model	μ^{RC}	DF_m	θ_{ij}
Uniform association	βij	1	β
Row effect	$j\phi_i$	$(I-1)$	$\phi_{i+1}-\phi_i$
Column effect	$i\phi_j$	$(J-1)$	$\phi_{j+1}-\phi_j$
Association Model I	$j\phi_i+i\phi_j$	$I+J-3$	$(\phi_{i+1}-\phi_i)+(\phi_{j+1}-\phi_j)$
Association Model II (*RC*)	$\beta\phi_i\phi_j$	$I+J-3$	$(\phi_{i+1}-\phi_i)(\phi_{j+1}-\phi_j)$

subject to four normalization constraints because each requires the normalization of both location and scale.

As equation (6) shows, the interaction component (μ^{RC}) of Goodman's Association Model II is in the form of multiplication of unknown parameters—log-bilinear specification. The model is also known as the *log-multiplicative model*, or simply the RC model. The RC model is very attractive because it allows the researcher to estimate unknown parameters even when the categories of the row and the column variables may not be correctly ordered. All that needs to be assumed is the existence of the ordinal scales. The model can reveal the orders through estimation.

Table 1 presents a summary comparison of the aforementioned association models. The second column displays the model specification for the interaction parameters (μ^{RC}). The number of degrees of freedom for each μ^{RC} specification is given in the third column (DF_m). If there are no other model parameters to be estimated, the degrees of freedom for a model are equal to $(I-1)(J-1)-DF_m$. The formula for calculating the local log-odds ratio is shown in the last column.

Goodman's Association Model II (*RC* model) can be easily extended to have multiple latent dimensions so that μ^{RC} of equation (1) is specified as

$$\mu_{ij}^{RC}=\sum\beta_m\phi_{im}\varphi_{jm}, \tag{7}$$

where the summation sign is with respect to all possible m dimensions, and the parameters are subject to necessary normalization constraints. Such models are called *RC(M)* models. See Goodman (1986) for details.

ASSOCIATION MODELS FOR THREE-WAY AND HIGHER-WAY TABLES

Below, I mainly discuss the case of a three-way table. Generalizations to a higher-way table can be easily made. Let R denote row, C denote column, and L denote layer, with their categories indexed respectively by i ($i=1,\ldots,I$), j ($j=1,\ldots,J$), and k ($k=1,\ldots,K$). In a common research setup, the researcher is interested in understanding how the two-way association between R and C depends on levels of L. For example, in a trend analysis, L may represent different years or cohorts. In a comparative study, L may represent different nations or groups. Thus, research attention typically focuses on the association pattern between R and C and its variation across layers.

Let F_{ijk} denote the expected frequency in the ith row, the jth column, and the kth layer. The saturated log-linear model can be written as follows:

$$\log(F_{ijk})=\mu+\mu_i^R+\mu_j^C+\mu_k^L+\mu_{ij}^{RC}+\mu_{ik}^{RL}+\mu_{jk}^{CL}+\mu_{ijk}^{RCL}. \tag{8}$$

In a typical research setting, interest centers on the variation of the *RC* association across layers. Thus, the baseline (for the null hypothesis) is the following conditional independence model:

$$\log(F_{ijk})=\mu+\mu_i^R+\mu_j^C+\mu_k^L+\mu_{ik}^{RL}+\mu_{jk}^{CL}. \tag{9}$$

That is, the researcher needs to specify and estimate μ^{RC} and μ^{RCL} to understand the layer-specific RC association.

There are two broad approaches to extending association models to three-way or higher-way tables. The first is to specify an association model for the typical association pattern between R and C and then estimate parameters that are specific to layers or test whether they are invariant across layers (Clogg, 1982a). The general case of the approach is to specify μ^{RC} and μ^{RCL} in terms of the RC model so as to change equation (8) to the following:

$$\log(F_{ijk})=\mu+\mu_i^R+\mu_j^C+\mu_k^L+\mu_{ik}^{RL}+\mu_{jk}^{CL}+\beta_k\phi_{ik}\varphi_{jk}. \tag{10}$$

That is, the β, ϕ, and φ parameters can be layer specific or layer invariant, subject to model specification and statistical tests. The researcher may also wish to test special cases (i.e., the uniform association, column effect, and row effect models) where ϕ and/or φ parameters are inserted as integer scores rather than estimated.

The second approach, called the *log-multiplicative layer-effect model,* or the "unidiff model," is to allow a flexible specification for the typical association pattern between R and C and then to constrain its cross-layer variation to be log-multiplicative (Xie, 1992). That is, we give a flexible specification for μ^{RC} but constrain μ^{RCL} so that equation (8) becomes the following:

$$\log(F_{ijk}) = \mu + \mu_i^R + \mu_j^C + \mu_k^L + \mu_{ik}^{RL} + \mu_{jk}^{CL} + \phi_k \psi_{ij}. \qquad (11)$$

With the second approach, the RC association is not constrained to follow a particular model and indeed can be saturated with $(I-1)(J-1)$ dummy variables. In a special case in which the typical association pattern between R and C is the RC model, the two approaches coincide, resulting in the three-way RCL log-multiplicative model. Powers and Xie (2000, pp. 140–145) provide a more detailed discussion of the variations and the practical implications of the second approach. It should be noted that the two approaches are both special cases of a general framework proposed by Goodman (1986) and extended in Goodman and Hout (1998).

APPLICATIONS

Association models have been used widely in sociological research. Below, I give three concrete examples. The first example is one of scaling. See Clogg (1982b) for a detailed illustration of this example. Clogg aimed to scale an ordinal variable that measures attitude on abortion. The variable was constructed from a Guttman scale, and the cases that did not conform to the scale response patterns were grouped into a separate category, "error responses." As is usually the case, scaling required an "instrument." In this case, Clogg used a measure of attitude on premarital sex that was collected in the same survey. The underlying assumption was that the scale of the attitude on abortion could be revealed from its association with the attitude on premarital sex. Clogg used the log-multiplicative model to estimate the scores associated with the different categories of the two variables. Note that the log-multiplicative RC model assumes that the categories are ordinal but not necessarily correctly ordered. So, estimation reveals the scale as well as the ordering. Through estimation, Clogg showed that the distances between the adjacent categories were unequal and that those who gave "error responses" were in the middle in terms of their attitudes on abortion.

The second example is the application of the log-multiplicative layer-effect model to the cross-national study of intergenerational mobility (Xie, 1992). The basic idea is to force cross-national differences to be summarized by layer-specific parameters [i.e., ϕ_k of equation (11)] while allowing and testing different parameterizations of the two-way association between father's occupation and son's occupation (i.e., ψ_{ij}). The ϕ_k parameters are then taken to represent the social openness or closure of different societies.

The third example, which involves the study of human fertility, is nonconventional in the sense that the basic setup is not log-linear but log-rate. The data structure consists of a table of frequencies (births) cross-classified by age and country and a corresponding table of associated exposures (women-years). The ratio between the two yields the country-specific and age-specific fertility rates. The objective of statistical modeling is to parsimoniously characterize the age patterns of fertility in terms of fertility level and fertility control for each country. In conventional demography, this is handled using Coale and Trussell's *Mm* method. Xie and Pimentel (1992) show that this method is equivalent to the log-multiplicative layer-effect model, with births as the dependent variable and exposure as an "offset." Thus, the M and m parameters of Coale and Trussell's method can be estimated statistically along with other unknown parameters in the model.

ESTIMATION

Estimation is straightforward with the uniform, row effect, column effect, and Goodman's Association Model I models. The user can use any of the computer programs that estimate a log-linear model. What is complicated is when the RC interaction takes the form of the product of unknown parameters—the log-multiplicative or log-bilinear specification. In this case, a reiterative estimation procedure is required. The basic idea is to alternately treat one set of unknown parameters as known while estimating the other and to continue the iteration process until both are stabilized. Special computer programs, such as ASSOC and

CDAS, have been written to estimate many of the association models. User-written subroutines in GLIM and STATA are available from individual researchers. For any serious user of association models, I also recommend Lem, a program that can estimate different forms of the log-multiplicative model while retaining flexibility. See my Web site www.yuxie.com for updated information on computer subroutines and special programs.

—Yu Xie

REFERENCES

Clogg, C. C. (1982a). Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association, 77,* 803–815.

Clogg, C. C. (1982b). Using association models in sociological research: Some examples. *American Journal of Sociology, 88,* 114–134.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association, 74,* 537–552.

Goodman, L. A. (1984). *The analysis of cross-classified data having ordered categories.* Cambridge, MA: Harvard University Press.

Goodman, L. A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review, 54,* 243–309.

Goodman, L. A., & Hout, M. (1998). Understanding the Goodman-Hout approach to the analysis of differences in association and some related comments. In Adrian E. Raftery (Ed.), *Sociological methodology* (pp. 249–261). Washington, DC: American Sociological Association.

Powers, D. A., & Xie, Y. (2000). *Statistical methods for categorical data analysis.* New York: Academic Press.

Xie, Y. (1992). The log-multiplicative layer effect model for comparing mobility tables. *American Sociological Review, 57,* 380–395.

Xie, Y., & Pimentel, E. E. (1992). Age patterns of marital fertility: Revising the Coale-Trussell method. *Journal of the American Statistical Association, 87,* 977–984.

ASSUMPTIONS

Assumptions are ubiquitous in social science. In theoretical work, assumptions are the starting axioms and postulates that yield testable implications spanning broad domains. In empirical work, statistical procedures typically embed a variety of assumptions, for example, concerning measurement properties of the variables and the distributional form and operation of unobservables (such as HOMOSKEDASTICITY or NORMAL DISTRIBUTION of the ERROR). Assumptions in empirical work are discussed in the entries for particular procedures (e.g., ORDINARY LEAST SQUARES); here we focus on assumptions in theories.

The purpose of a scientific THEORY is to yield testable implications concerning the relationships between observable phenomena. The heart of the theory is its set of assumptions. The assumptions embody what Popper (1963) calls "guesses" about nature—guesses to be tested, following Newton's vision, by testing their logical implications. An essential feature of the assumption set is internal logical consistency. In addition, three desirable properties of a theory are as follows: (a) that its assumption set be as short as possible, (b) that its observable implications be as many and varied as possible, and (c) that its observable implications include phenomena or relationships not yet observed, that is, novel predictions.

Thus, a theory has a two-part structure: a small part containing the assumptions and a large and ever-growing part containing the implications. Figure 1 provides a visualization of the structure of a theory.

A theory can satisfy all three properties above and yet be false. That is, one can invent an imaginary world, set up a parsimonious set of postulates about its operation, deduce a wide variety of empirical consequences, and yet learn through empirical test that no known world operates in conformity with the implications derived from the postulated properties of the imaginary world. That is why empirical analysis is necessary, or, put differently, why theoretical analysis alone does not suffice for the accumulation of reliable knowledge.

A note on terminology: *Assumption* is a general term, used, as noted earlier, in both theoretical and empirical work. Sharper terms sometimes used in theoretical work include *axiom,* which carries the connotation of *self-evident,* and *postulate,* which does not, and is therefore more faithful to an enterprise marked by guesses and bound for discovery. Other terms include the serviceable *premise,* the colorful *starting principle,* and the dangerous *hypothesis,* which is used not only as a postulate (as in the HYPOTHETICO-DEDUCTIVE METHOD invented by Newton) but also as an observable proposition to be tested.

Where do assumptions come from? Typically, the frameworks for analyzing topical domains include a variety of relations and FUNCTIONS, some of which may prove to be fruitful assumptions. In general,

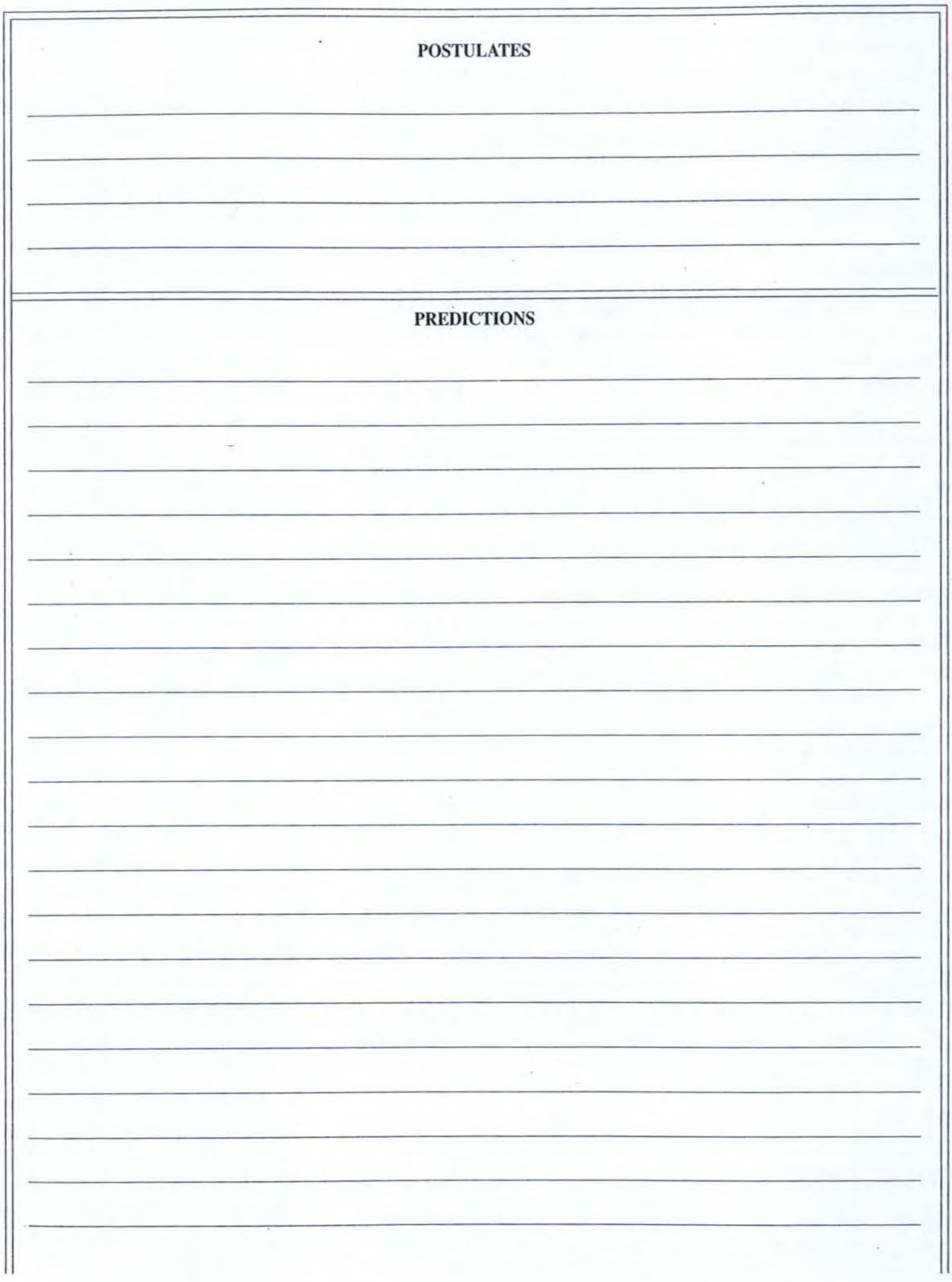

Figure 1 Structure of a Theory

Theory: ______________________________ *Date:* ________________

Page: ________________

PREDICTIONS	POSTULATES				

Notes:

Figure 2 Merton Chart of Theoretical Derivation mathematically expressed functions make promising assumptions, as (a) their very statement signals generality and breadth, and (b) they are amenable to manipulation via mathematical tools, which make it easy to derive a wealth of implications and, unlike verbal covering-law procedures, make it possible to derive novel predictions.

The tools for derivation of predictions include the full panoply of mathematical tools. Besides functions, these include DISTRIBUTIONS, MATRICES, inequalities, and partial differential equations. For example, the quantities related via functions can be characterized by their distributional properties, and the distributional form shapes the character of the predictions. The results can be surprising. Assumptions that look innocent turn out to touch vast domains in unexpected ways. To illustrate, the justice evaluation function, when used as an assumption, yields predictions for parental gift giving, differential mourning of mothers and fathers, differential risk of posttraumatic stress among combat veterans, differential social contributions of monastic and mendicant religious institutions, and inequality effects on migration (Jasso, 2001).

To assess the usefulness of assumptions, it is convenient to maintain a spreadsheet showing, for a particular theory, which of the assumptions in its assumption set are used in the derivation of each prediction. This layout tells at a glance which assumptions are doing most of the work and how many assumptions each prediction requires. Figure 2 provides a visualization of this type of layout, called a Merton Chart of Theoretical Derivation.

A theory and its assumptions may be more, or less, fundamental (BASIC RESEARCH). The Holy Grail is discovery of ultimate forces governing the sociobehavioral world. The theories and assumptions we work with currently, even the most fruitful and the ones with the longest reach, are like signposts along the road. But what signposts. Simultaneously, they assist our preparation and sharpening of tools for the deeper theories to come and illuminate patches of the road.

FURTHER READING

Cogent background in the philosophy of science is provided by Toulmin (1978). Elaboration of procedures for constructing sociobehavioral theories, including choice of assumptions and their mathematical manipulation, is provided in Jasso (1988, 2001, 2002). More fundamentally and more important, it is instructive, not to mention inspiring, to watch great theorists in action, choosing assumptions, refining them, deriving implications from them. A good starting point is Newton's (1686/1952) *Principia,* in which he worked out the method subsequently called the hypothetico-deductive method. Of course, the reader's taste will dictate the best starting point. Other possibilities include Mendel's (1866) paper, which laid the foundations for genetics, and the little Feynman and Weinberg (1987) volume, which also includes a foreword recalling Dirac's prediction of antimatter.

—Guillermina Jasso

REFERENCES

Feynman, R. P., & Weinberg, S. (1987). *Elementary particles and the laws of physics: The 1986 Dirac Memorial Lectures.* Cambridge, UK: Cambridge University Press.

Jasso, G. (1988). Principles of theoretical analysis. *Sociological Theory, 6,* 1–20.

Jasso, G. (2001). Formal theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 37–68). New York: Kluwer Academic/Plenum.

Jasso, G. (2002). Seven secrets for doing theory. In J. Berger & M. Zelditch (Eds.), *New directions in contemporary sociological theory* (pp. 317–342). Boulder, CO: Rowan & Littlefield.

Mendel, G. (1866). Versuche über Pflanzen-Hybriden [Experiments in plant hybridization]. In *Verhandlungen des naturforschenden Vereins* [Proceedings of the Natural History Society]. Available in both the original German and the English translation at www.mendelweb.org

Newton, I. (1952). *Mathematical principles of natural philosophy.* Chicago: Britannica. (Original work published 1686.)

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge.* New York: Basic Books.

Toulmin, S. E. (1978). Science, philosophy of. In *Encyclopaedia Britannica* (15th ed., Vol. 16, pp. 375-393). Chicago: Britannica.

ASYMMETRIC MEASURES

ASSOCIATION coefficients measure the strength of the relation between two variables X and Y and, for ordinal and quantitative (interval and ratio-scaled) variables, the direction of the association. Asymmetric (or directional) association coefficients also assume that one of the two variables (e.g., X) can be identified as the independent variable and the other variable (e.g., Y) as the dependent variable.

As an example, in a survey of university graduates, the variable X might be gender (nominal, dichotomous), ethnic group (nominal), the socioeconomic status of parents (ordinal), or the income of parents (quantitative). The variable Y might be the field of study or party affiliation (nominal), a grade (ordinal), the income of respondents (quantitative), or membership in a union (nominal, dichotomous, with a "yes" or "no" response category).

Available Coefficients

Most statistical software packages, such as SPSS, offer asymmetric coefficients only for special combinations of measurement levels. These coefficients are described first.

Measures for Nominal-Nominal Tables

Measures for nominal variables are lambda (λ), Goodman and Kruskal's tau (τ), and the uncertainty coefficient U. All measures are based on the definition of PROPORTIONAL REDUCTION OF ERROR (PRE):

$$\mathrm{PRE} = \frac{E_0 - E_1}{E_0}, \tag{1}$$

where E_0 is the error in the dependent variable Y without using the independent variable X to predict or explain Y, and E_1 is the error in Y if X is used to predict or explain Y.

Both τ and U use definitions of variance to compute E_0 and E_1. Similar to the well-known eta (η) from the ANALYSIS OF VARIANCE (see below), E_0 is the total variance of Y, and E_1 is the residual variance (variance within the categories of X). For τ, the so-called Gini concentration is used to measure the variation (Agresti, 1990, pp. 24–25), and U applies the concept of entropy. For λ, a different definition is used: It analyzes to what extent the dependent variable can be predicted both without knowing X and with knowledge of X.

All three coefficients vary between 0 (no reduction in error) and 1 (perfect reduction in error). Even if a clear association exists, λ may be zero or near zero. This is the case if one category of the dependent variable is dominant, so that the same category of Y has the highest frequencies and becomes the best predictor within each condition X.

Measure for Ordinal-Ordinal Tables

The most prominent measure for this pattern is Somers' d. Somers' d is defined as the least squares regression slope ($d_{y/x} = s_{xy}/s_x^2$) between the dependent variable Y and the independent variable X, if both variables are treated as ordinal (Agresti, 1984, pp. 161–163), and Daniels' formula for generalized correlation coefficient (see ASSOCIATION) is used to compute the variance s_x^2 of X and the covariance s_{xy} between X and Y. Somers' d is not a PRE coefficient. The corresponding PRE coefficient is the symmetric Kendall's τ_b^2, which can be interpreted as explained variance and is equal to the geometric mean if Somers' d is computed for X and Y as dependent variables ($= d_{y/x}$ and $d_{x/y}$). If both variables are dichotomous, Somers' d is equal to the difference of proportions (Agresti, 1984, p. 161).

Measures for Nominal-Interval (or Ratio) Tables

The well-known explained variance η^2 or, respectively, η (eta) from ANALYSIS OF VARIANCE (ANOVA) can be applied to situations in which the independent variable is nominal and the dependent variable is quantitative. η^2 is a PRE coefficient.

Formulas

Formulas for computing the described measures are given in Agresti (1990) and SPSS (2002), for example.

Measures for Other Combinations

Up until now, only pairings for nominal-nominal tables (= Constellations 1, 2, 5, and 6 in Table 1), ordinal-ordinal tables (= Constellation 11), and nominal-quantitative tables (= Constellations 4 and 8) have been analyzed.

Strategies to obtain measures for the other pairings, as well as those already analyzed, include the following:

- *Use of symmetric measures.* Association measures can be derived using Daniels's generalized correlation coefficient for all constellations (1, 3, 4, 9, 11 to 13, 15, 16), except those having one or two nominal-polytomous variables (Constellations 2, 5 to 8, 10, 14).

Table 1 Possible Pairings of Measurement Levels for Two Variables

	Dependent Variable Y			
Independent Variable X	*Nominal-Dichotomous*	*Nominal-Polytomous*	*Ordinal*	*Quantitative (Interval or Ratio)*
Nominal-dichotomous	1	2	3	4
Nominal-polytomous	5	6	7	8
Ordinal	9	10	11	12
Quantitative (interval or ratio)	13	14	15	16

• *Reduction of measurement level.* The measurement level of the variable with the higher measurement level is reduced. If, for example, X is nominal (polytomous) and Y is ordinal (Constellation 7), it is recommended to treat Y as nominal scaled (Constellation 6) and apply λ, τ, or U as an appropriate measure. This procedure results in a loss of information and can hide an association.

• *Upgrading of measurement level.* Ordinal variables are treated quantitatively. From a strict theoretical point of view, this strategy is incompatible. However, textbooks (see below) recommend this strategy, which is also frequently applied in survey research. In many surveys, I very often have found a nearly perfect linear relation between ordinal and quantitative measures. But convince yourself.

• *Use of logistic regression.* Appropriate LOGISTIC REGRESSION models have been developed for all constellations (Pairings 1 to 3, 5 to 7, 9 to 11, 13 to 15), except for those in which the dependent variable is quantitative (Pairings 4, 8, 12, and 16) (Agresti, 1984, pp. 104–131). Ordinal independent variables are treated as quantitative variables (Agresti, 1984, p. 114) or as nominal scaled.

• *Use of linear regression.* LINEAR REGRESSION can be used for Pairing 16 as well as for Pairings 4 and 8. For Constellation 8, dummies must be computed.

• *Use of special measures.* Special approaches have been developed: ridit analysis (Agresti, 1984, pp. 166–168) and the dominance statistic Δ (Agresti, 1984, pp. 166–168) can be applied to Constellations 3 and 7, as well as to 11 and 15, if the independent variable is treated as nominal scaled.

Dichotomous Variables. These take an exceptional position. They can be treated either as ordinal or quantitative. Therefore, all strategies for Constellation 11 can also be applied to Constellations 1, 3, and 9, and all strategies for Constellation 16 can be applied to Constellations 1, 4, and 13 and so on.

Generalizations. The great advantage to using logistic or linear regression is that generalizations exist when there is more than one independent variable. Hence, it is recommended to use these advanced techniques. However, logistic regression models do not solve the problem of interpreting the relation's strength. As with all association coefficients, general threshold values are not possible (see ASSOCIATION).

Significance of Coefficients. Test statistics are available for all measures and coefficients discussed. Computational formulas are given in SPSS (2002).

—Johann Bacher

REFERENCES

Agresti, A. (1984). *Analysis of ordinal categorical data.* New York: John Wiley.

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley.

SPSS. (2002, August). Crosstabs. In *SPSS 11.0 statistical algorithms* [Online]. Available: http://www.spss.com/tech/stat/algorithms/11.0/crosstabs.pdf

ASYMPTOTIC PROPERTIES

To make a STATISTICAL INFERENCE, we need to know the SAMPLING DISTRIBUTION of our ESTIMATOR. In general, the exact or finite sample distribution will not be known. The asymptotic properties provide us with the approximate distribution of our estimator by looking at its behavior as a data set gets arbitrarily large (i.e., infinite).

Asymptotic results are based on the sample size n. Let x_n be a RANDOM VARIABLE indexed by the size of the SAMPLE. To define consistency, we will need to define the notation of convergence in PROBABILITY. We will say that x_n converges in probability to a constant c if $\lim_{n\to\infty} \text{Prob}(|x_n - c| > \varepsilon) = 0$ for any positive ε. Convergence in probability implies that the values that the variable may take that are not close to c become increasingly unlikely as the sample size n increases. Convergence in probability is usually denoted by $\text{plim } x_n = c$. We can then say that an ESTIMATOR $\hat{\theta}_n$ of a parameter θ is a consistent estimator if and only if $\text{plim } \hat{\theta} = \theta$. Roughly speaking, this means that as the sample size gets large, the estimator's sampling DISTRIBUTION piles up on the true value.

Although consistency is an important characteristic of an estimator, to conduct statistical inference, we will need to know its sampling distribution. To derive an estimator's sampling distribution, we will need to define the notation of a limiting distribution. Suppose that $F(x)$ is the cumulative distribution function (cdf) of some random variable x. Furthermore, let x_n be a sequence of random variables indexed by sample size with cdf $F_n(x)$. We say that x_n converges in distribution to x if $\lim_{n\to\infty} |F_n(x) - F(x)| = 0$.

We will denote convergence in distribution by $x_n \xrightarrow{d} x$. In words, this says that the distribution of x_n gets close to the distribution of x as the sample size gets large. This convergence in distribution, along with the CENTRAL LIMIT THEOREM, will allow us to construct the asymptotic distribution of an estimator.

An asymptotic distribution is a distribution that is used to approximate the true finite sample distribution of a random variable. Suppose that $\hat{\theta}_n$ is a consistent estimate of a parameter θ. Then, using the central limit theorem, the asymptotic distribution of $\hat{\theta}_n$ is

$$\sqrt{n}(\hat{\theta}_n - \theta)^d \rightarrow N[0, V], \qquad (1)$$

where $N[\ ]$ denotes the cdf of the NORMAL DISTRIBUTION, and V is the asymptotic covariance matrix. This implies that $\hat{\theta}_n$ is distributed as $N[\theta, \frac{1}{n}V]$. This is a very striking result. It says that, regardless of the true underlying finite sample distribution, we can approximate the distribution of the estimator by the Normal distribution as the sample size gets large. We can thus use the Normal distribution to conduct HYPOTHESIS TESTING as if we knew the true sampling distribution.

However, it must be noted that the asymptotic distribution is only an approximation, although one that will improve with sample size; thus, the size, power, and other properties of the hypothesis test are only approximate. Unfortunately, we cannot, in general, say how large a sample needs to be to have confidence in using the asymptotic distribution for statistical inference. One must rely on previous experience, BOOTSTRAPPING simulations, or MONTE CARLO experiments to provide an indication of how useful the asymptotic approximations are in any given situation.

—Jonathan N. Katz

REFERENCE

White, H. (1984). *Asymptotic theory for econometricians.* Orlando, FL: Academic Press.

ATLAS.ti

ATLAS.ti is a software program designed for computer-assisted qualitative data analysis (see CAQDAS). As such, it allows qualitative data to be coded and retrieved. In addition, it supports the analysis of visual images as well as text. It also supports GROUNDED THEORY practices, such as the generation of MEMOS.

—Alan Bryman

ATTENUATION

Attenuation refers to the CORRELATION between two different measures being reduced due to measurement error. Of course, no test has scores that are perfectly reliable. Therefore, having unreliable scores results in poor predictability of the criterion. To obtain an estimate of the correlation, given predictor and criterion scores, with perfect RELIABILITY, Spearman (1904) proposed the following formula known as the correction for attenuation:

$$\bar{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{xx} \times r_{yy}}}. \qquad (1)$$

This formula has also been referred to as the double correction because both the predictor (x) and criterion

(y) scores are being corrected. Suppose that we are examining the correlation between depression and job satisfaction. The correlation between the two measures (r_{xy}) is .20. If the reliability (e.g., CRONBACH'S ALPHA) of the scores from the depression inventory (r_{xx}) is .80 and the reliability of the scores from the job satisfaction scale (r_{yy}) is .70, then the correlation corrected for attenuation would be equal to the following:

$$\bar{r}_{xy} = \frac{.20}{\sqrt{.80 \times .70}} = .26. \tag{2}$$

In this case, there is little correction. However, suppose the reliabilities of the scores for depression and job satisfaction are .20 and .30, respectively. The correlation corrected for attenuation would now be equal to .81. The smaller the reliabilities or sample sizes ($N < 300$), the greater the correlation corrected for attenuation (Nunnally, 1978). It is even possible for the correlation corrected for attenuation to be greater than 1.00 (Nunnally, 1978).

There are cases, however, when one might want to correct for unreliability for either the predictor or the criterion scores. These would incorporate the single correction. For example, suppose the correlation between scores on a mechanical skills test (x) and job performance ratings (y) is equal to .40, and the reliability of the job performance ratings is equal to .60. If one wanted to correct for only the unreliability of the criterion, the following equation would be used:

$$\bar{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{yy}}}. \tag{3}$$

In this case, the correlation corrected for attenuation would be equal to .51. It is also feasible, albeit less likely for researchers, to correct for the unreliability of the test scores only. This correction would occur when either the criterion reliability is measured perfectly or if the criterion reliability is unknown (Muchinsky, 1996). Hence, the following formula would be applied:

$$\bar{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{xx}}}. \tag{4}$$

Muchinsky (1996) summarized the major tenets of the correction for attenuation. First, the correction for attenuation does not increase the predictability of the test scores (Nunnally, 1978). Furthermore, the corrected correlations should neither be tested for STATISTICAL SIGNIFICANCE (Magnusson, 1967), nor should corrected and uncorrected VALIDITY coefficients be compared to each other (American Educational Research Association, 1985). Finally, the correction for attenuation should not be used for averaging different types of validity coefficients in META-ANALYSIS that use different estimates of reliabilities (e.g., internal consistencies, test-retest) (Muchinsky, 1996).

—N. Clayton Silver

REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological tests.* Washington, DC: American Psychological Association.

Magnusson, D. (1967). *Test theory.* Reading, MA: Addison-Wesley.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56,* 63–75.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15,* 72–101.

ATTITUDE MEASUREMENT

An attitude is an evaluation of an object (person, place, or thing) along a dimension ranging from favorable to unfavorable (Weiss, 2002). As such, an attitude has both affective (emotional) and cognitive aspects. Because they are internal states of people, attitudes can best be assessed via self-reports with quantitative psychological scales that can have several different formats (Judd, Smith, & Kidder, 1991).

THURSTONE SCALE

A THURSTONE SCALE (Thurstone & Chave, 1929) consists of a list of items varying in favorability toward the attitude object. The underlying idea is that an individual will agree with only those items that are close in favorability to where his or her attitude lies. For example, Spector, Van Katwyk, Brannick, and Chen (1997) created a scale of job attitudes that had items that ranged from extreme negative (e.g., *I hate my job*) to extreme positive (e.g., *My*

Table 1 Example of a Semantic Differential

The Supreme Court of the United States								
Good	___	___	___	___	___	___	___	Bad
Fair	___	___	___	___	___	___	___	Unfair
Weak	___	___	___	___	___	___	___	Strong
Passive	___	___	___	___	___	___	___	Active
Happy	___	___	___	___	___	___	___	Unhappy
Unimportant	___	___	___	___	___	___	___	Important

job is the most enjoyable part of my life), with more moderate positive (e.g., *I enjoy my job*) and negative items (e.g., *I don't really like my job very much*) in between.

A Thurstone scale is developed by first writing a number of items that are given to a panel of judges to sort into categories that vary from extremely unfavorable to extremely favorable. The categories are numbered, and the mean score across the judges is computed. To use the scale, respondents are asked to read the items presented in a random order and indicate for each whether they agree or not. The attitude score is the mean scale value of the items the individual endorsed.

LIKERT (SUMMATED RATING) SCALE

The LIKERT or SUMMATED RATING SCALE (Likert, 1932) consists of a series of statements that are either favorable or unfavorable about the attitude object. Respondents are given multiple response choices that vary along a continuum, usually agree to disagree, and they are asked to indicate their extent of agreement with each one. The scale is designed on the assumption that the favorable and unfavorable items are mirror images of one another, so that an individual who tends to agree with one set of items will disagree with the other. This is in contrast to the assumption of the Thurstone scale that people agree only with items that are close in favorability or unfavorability to their own attitude position. Spector et al. (1997) demonstrated that the Likert scale assumption is not always correct, and this can produce artifactual results when such items are factor analyzed.

The Likert scale has become the most popular for the assessment of attitudes, undoubtedly for two major reasons. First, it is relatively easy to develop (see LIKERT SCALE for the procedure), requiring no initial scaling study to get judges' ratings of the items. Second, this approach has been shown to yield reliable scales that are useful for assessing attitudes.

SEMANTIC DIFFERENTIAL

The SEMANTIC DIFFERENTIAL SCALE was developed by Osgood, Suci, and Tannenbaum (1957) to assess different aspects of attitude toward an object. The stem for a semantic differential is the name or description of the attitude object (e.g., Supreme Court of the United States). A series of bipolar adjectives is provided with a series of response choices in between. The respondent is asked to check the choice that best reflects his or her feeling, where the nearer the choice is to the adjective, the stronger the person endorses it (see Table 1).

The semantic differential items tend to group into three underlying dimensions of attitude: evaluation, potency, and activity. Evaluation indicates the extent to which the person has a favorable or unfavorable attitude, that is, whether he or she likes or dislikes the object. Potency is the strength of the object, which indicates the extent to which a person sees the attitude object as being powerful or not, regardless of favorability. Activity is the extent to which the attitude object is seen as being active or passive.

Semantic differentials tend to be used to assess attitudes about several objects at the same time rather than to develop an attitude scale for one object. It is possible to include only evaluation bipolar adjectives to yield a single score per attitude object as the sum of the scores across items. This can be an easy and efficient procedure for respondents, as there is less reading with this sort of scale than the Thurstone or Likert scale, and the same items can be used for each attitude object, making comparisons easier to make.

GUTTMAN SCALE

A GUTTMAN SCALE (Guttman, 1944) consists of a hierarchy of items that vary from unfavorable to favorable. Respondents are asked to indicate for each item whether he or she agrees with it. It is assumed that individuals with favorable attitudes will agree with all favorable items that are at their level or lower. Thus, a slightly favorable person will agree with a slightly favorable item but not an extremely favorable item, whereas an extremely favorable person will agree with both slightly and extremely favorable items. The Guttman scale has not been used as frequently as the other scales, probably for two reasons. First, the scale can be more difficult to develop than the more popular Likert and semantic differential scales. Second, the expected pattern of responses does not always occur because often those with extreme attitudes might disagree with slightly favorable items, or the items might reflect more than a single attitude dimension.

—Paul E. Spector

REFERENCES

Guttman, L. (1944). A basis for scaling quantitative data. *American Sociological Review, 9,* 139–150.

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations* (6th ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140,* 55.

Osgood, C. E., Suci, C. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management, 23,* 659–677.

Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude.* Chicago: University of Chicago Press.

Weiss, H. M. (2002). Deconstructing job satisfaction: Separating evaluations, beliefs and affective experiences. *Human Resources Management Review, 12,* 173–194.

ATTRIBUTE

One definition of *attribute* is a characteristic that a respondent, subject, or case does or does not have. A binary attribute variable has two values: the presence of the attribute and its absence. For example, adults in a survey may have the attribute of being married (score = 1) or not (score = 0). Another definition, from psychological or marketing research, refers to attributes as scalable items. In decision research, for example, attributes may refer to goals or criteria relevant to the decision maker. To illustrate, in deciding on a new car, the buyer may be faced with three alternatives and have to weigh attributes such as fuel efficiency, cost, and safety.

—Michael S. Lewis-Beck

See also BINARY, DICHOTOMOUS VARIABLES

ATTRITION

Attrition is a term used to describe the process by which a SAMPLE reduces in size over the course of a survey data collection process due to NONRESPONSE and/or due to units ceasing to be eligible. The term is also used to refer to the numerical outcome of the process.

In PANEL surveys, the number of sample units having responded at every wave decreases over waves due to the cumulative effects of nonresponse and changes in eligibility status. Such attrition may introduce unquantifiable bias into survey estimates due to the nonrandom nature of the dropout. For this reason, considerable efforts are often made to minimize the extent of sample attrition. The extent of attrition is affected also by factors such as the number of waves of data collection, the intervals between waves, the burden of the data collection exercise, and so on. WEIGHTING and other adjustment methods may be used at the analysis stage to limit the detrimental impacts of attrition.

Although perhaps most pronounced in the context of panel surveys, attrition can affect any survey with multiple stages in the data collection process. Many CROSS-SECTIONAL surveys have multiple data collection instruments—for example, a personal interview followed by a SELF-COMPLETION ("drop-off") QUESTIONNAIRE or an interview followed by medical examinations. In such cases, there will be attrition between the stages: Not all persons interviewed will also complete and return the self-completion questionnaire, for example. Even with a single data collection instrument, there may be multiple stages in the survey process. For example, in a survey of schoolchildren, it may be necessary first to gain the cooperation of a sample of schools to provide lists of pupils from which a sample can be selected. It may then be necessary to gain parental consent before approaching the sample children. Thus, there are at least three potential stages of sample dropout: nonresponse by schools, parents, and pupils. This, too, can be thought of as a process of sample attrition.

In attempting to tackle attrition, either by reducing it at the source or by adjusting at the analysis stage, different component subprocesses should be recognized and considered explicitly. There is an important conceptual difference between attrition due to nonresponse by eligible units and attrition due to previously eligible units having become ineligible. The latter need not have a detrimental effect on analysis if identified appropriately. Each of these two main components of attrition splits into distinct subcomponents. Nonresponse could be due to a failure to contact the sample unit ("noncontact"), inability of the sample unit to respond (for reasons of language, illness, etc.), or unwillingness to respond ("refusal"). Change of eligibility status could be due to death, geographical relocation, change in status, and so forth. The different components have different implications for reduction and adjustment. Many panel surveys devote considerable resources to "tracking" sample members over time to minimize the noncontact rate. The experience of taking part in a wave of the survey (time taken, interest, sensitivity, cognitive demands, etc.) is believed to be an important determinant of the refusal rate at subsequent waves; consequently, efforts are made to maximize the perceived benefits to respondents of cooperation relative to perceived drawbacks.

—Peter Lynn

REFERENCES

Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics, 15*(2), 269–282.

Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 259–272). New York: John Wiley.

Murphy, M. (1990). Minimising attrition in longitudinal studies: Means or end? In D. Magnusson & L. R. Bergman (Eds.), *Data quality in longitudinal research* (pp. 148–156). Cambridge, UK: Cambridge University Press.

AUDITING

Auditing is a means of establishing the dependability of findings in qualitative research. Dependability is associated with TRUSTWORTHINESS CRITERIA. Auditing entails keeping complete records of all phases of a qualitative research project and submitting them to peers who would assess whether proper procedures were followed, including the robustness of theoretical inferences and the ethical appropriateness of research practices.

—Alan Bryman

AUTHENTICITY CRITERIA

Authenticity criteria are criteria appropriate for judging the quality of inquiry (research, evaluation, and policy analysis) conducted within the protocols of emergent, nonpositivist paradigms such as CRITICAL THEORY, participant inquiry, CONSTRUCTIVISM, and others. Positivist criteria of rigor are well established: INTERNAL VALIDITY, EXTERNAL VALIDITY, RELIABILITY, and OBJECTIVITY. An early effort to devise similar but nonpositivist ways of assessing inquiry quality resulted in so-called TRUSTWORTHINESS CRITERIA (Guba, 1981), constructed as "parallels" to the four standard criteria and termed, respectively, *credibility, transferability, dependability,* and *confirmability.* But their very parallelism to the standard criteria of rigor makes their applicability to alternative inquiry approaches suspect. Moreover, the standard criteria are primarily *methodological,* overlooking such issues as power, ethics, voice, ACCESS, representation, and others raised by POSTMODERNISM.

The authenticity criteria were developed rooted in the propositions of alternative inquiry paradigms (Guba & Lincoln, 1989). This set is incomplete and primitive but may serve reasonably well as guidelines.

Certain *initial conditions* are prerequisite to all five authenticity criteria. These include the following: Respondents are drawn from all at-risk groups, fully INFORMED CONSENT procedures are in place, caring and trusting relationships are nurtured, inquiry procedures are rendered transparent to all respondents and audiences, and respondent-inquirer collaboration is built into every step, with full agreement on the rules to govern the inquiry and with information fully shared. The inquiry report is guaranteed to be available to all respondents and audiences. Finally, an appellate mechanism is established to be used in cases of conflict or disagreement.

Definitions of the criteria, together with recommended procedures to establish them, follow:

- *Fairness* is defined as the extent to which all competing constructions of reality, as well as their underlying value structures, have been accessed, exposed, deconstructed, and taken into account in shaping the inquiry product, that is, the emergent reconstruction. Certain *procedures* should be strictly followed. All prior constructions of respondents and inquirer are obtained, compared, and contrasted, with each enjoying similar privilege; prior and later constructions are compared and contrasted; respondents and inquirer negotiate data to be collected, methods to be employed, interpretations to be made, modes of reporting to be employed, recommendations to be made, and actions to be proposed (taken); introspective statements (testimony) about changes experienced by respondents and inquirer are collected; prolonged engagement and persistent observation are used; individual and group member checks/validations are employed; thick contextual description is provided; peer debriefers and auditors are used (implying the maintenance of an audit trail); and, finally, the degree of empowerment felt by respondents is assessed.
- *Ontological authenticity* is defined as the extent to which individual respondents' (and the inquirer's) early constructions are improved, matured, expanded, and elaborated, so that all parties possess more information and become more sophisticated in its use. Useful *procedures* include the following: explication of the respondents' and the inquirer's a priori positions; comparison of respondents' earlier and later personal constructions; solicitation of respondents' and the inquirer's introspective statements about their own growth, as well as the testimony of selected respondents regarding their own changing constructions; and the establishment of an audit trail demonstrating changes.
- *Educative authenticity* is defined as the extent to which individual respondents (and the inquirer) possess enhanced understanding of, appreciation for, and tolerance of the constructions of others outside their own stakeholding group. Useful *procedures* include the following: employment of a peer debriefer and an auditor by the inquirer, comparison of respondents' and the inquirer's assessments of the constructions held by others, respondents' and the inquirer's introspective statements about their understandings of others' constructions, respondent testimony, and maintenance of an audit trail.
- *Catalytic authenticity* is defined as the extent to which action (clarifying the focus at issue, moving to eliminate or ameliorate the problem, and/or sharpening values) is stimulated and facilitated by the inquiry process. Knowledge in and of itself is insufficient to deal with the constructions, problems, concerns, and issues that respondents bring to the inquiry process for elucidation; purposeful action must also be delineated. Useful *procedures* include the following: development of a joint construction (aiming at consensus when possible or explication of conflicting values when consensus is not possible), including the assignment of responsibility and authority for action; plans for respondent-inquirer collaboration; accessibility of the final report; and evidence of practical applications. Testimony of participants, the actual resolution of at least some of the inquiry concerns, and systematic follow-up over time to assess the continuing value of outcomes are also helpful techniques.
- *Tactical authenticity* is defined as the degree to which participants are empowered to take the action(s) that the inquiry implies or proposes. Useful *procedures* include the following: negotiation of data to be collected, as well as their interpretation and reporting; maintenance of CONFIDENTIALITY; use of consent forms; member checking/validation; and prior agreements about power. The best indicator that this criterion is satisfied is participant testimony. The presence of follow-up activities is highly indicative. Finally, the degree of empowerment felt by stakeholders is crucial.

The present state of developing authenticity criteria leaves much to be desired. Perhaps the most significant accomplishment to date is simply their existence, a demonstration of the fact that it is possible to "think outside the box" of conventional quality assessments. It is imperative for the inquirer to have thought through just what paradigm he or she elects to follow and, having made that decision, to select quality criteria that are appropriate. Applying positivist criteria to alternative paradigm inquiries (or conversely) can lead only to confusion and conflict.

—Egon G. Guba

REFERENCES

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal, 29,* 75–92.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

AUTOBIOGRAPHY

Autobiography refers to the telling and documenting of one's own life. Together with biography (researching and documenting the lives of others), autobiography has increasingly been drawn upon as a resource and method for investigating social life. Autobiographical research is part of a more general biographical turn within the social sciences, characterized by an emphasis on personal narratives and the LIFE HISTORY METHOD.

HISTORY OF AUTOBIOGRAPHICAL WORK

Autobiographical and life history work has a long genealogy, both generally and within the social sciences. Ken Plummer (2001) traces the rise of the "autobiographical society" and the various shifts that have taken place. He documents the ways in which the telling of lives has become inscribed within texts, with a move from oral to written traditions. He explores the possible origins of the autobiographical form and the rise of the individual "voice." He also considers the ways in which individual life stories can become part of collective explorations of shared lives and experiences. Liz Stanley (1992) has traced the role of autobiography within social science (particularly sociology) and FEMINIST RESEARCH. She examines the methodological aspects of autobiographical work, highlighting the relationships between feminist praxis and autobiographical practice. Feminism has had a particular influence on the role of autobiographies and personal narratives within contemporary social science, both as a way of "giving voice" and as an approach that places an emphasis on self-reflexivity and personal knowledge. Autobiography, and life history work more generally, can create textual and discursive spaces for otherwise hidden or muted voices, such as marginal or oppressed social groups.

AUTOBIOGRAPHICAL RESOURCES AND NARRATIVES

Autobiographical data can be treated as resources by social scientists. General features of lives and experiences can be revealed through individual accounts and life stories. Written autobiographies can be considered as a rich data set of "lives" to be explored and analyzed in their own right, in terms of what they can reveal about a life, setting, organization, culture, event, or moment in time. Hence, stories of lives can be treated as realist accounts of social life. Autobiographical work is also composed and then articulated through written texts. Hence, autobiographies can be considered as research events—the processes through which lives are remembered, reconstructed, and written. There are collective conventions of memory and "ways" of (re)telling a life, marked by key events, places, people, and times. Autobiographical texts thus have particular genres and narrative forms that can be systematically explored. They are stories that draw on a standard set of literary and textual conventions. These can be analyzed for structure and form through, for example, NARRATIVE ANALYSIS.

AUTOBIOGRAPHY AND METHOD(OLOGY)

The "methods" of autobiographical work can also be used to gather social science data. Although autobiography is not a social science method in its own right, it does draw on a range of empirical materials or "documents of life" (Plummer, 2001). These include letters, DIARIES and other documentary sources, artifacts and biographical objects, and visual media such as film, video, and photography. Many of these sources of data are relatively underused within social science research, despite their analytical potential. Hence, a consideration of how "lives" are remembered, invoked, constructed, and reproduced can encourage a more eclectic approach to what might be considered as data about the social world.

In a consideration of feminist social science, Stanley (1992) suggests that work on autobiography (and biography) can be conceptualized as part of a distinctive methodological approach. Included within this is a focus on REFLEXIVITY; the rejection of conventional dichotomies such as objectivity-subjectivity, self-other, and public-private; and an emphasis on researching and theorizing experience. Thus, (auto)biographical approaches to social science research can imply a greater self-awareness of the research process, research relationships, and the researcher-self, as well as a clearer appreciation of the value of LIVED EXPERIENCE and personal knowledge as part of social science scholarship.

WRITING AND RESEARCHING THE SELF

Personal narratives of the research process are a kind of autobiographical practice in their own right. The social researcher, as well as serving as a biographer of other lives, is simultaneously involved in autobiographical work of his or her own. Those working within QUALITATIVE RESEARCH, in particular, have reflected on and written about the self in a variety of autobiographical texts. FIELDNOTES and research journals have long been used to record the emotions, experiences, and identity work associated with social research. The genre of the confessional tale—revealing some of the personal aspects of fieldwork—allows for self-revelation and indiscretion, within a recognizable autobiographical format. Although such accounts are usually situated as separate from the research "data," it is also possible to blend autobiography with analyses. For example, the "story" of the Japanese workplace told in *Crafting Selves* (Kondo, 1990) is presented as a multilayered text that challenges the dichotomies between researcher and researched, self and other. Kondo's life stories form part of the representation of the research. She reveals herself and her biographical work as integral to the research process and product.

The self can also be a topic for explicit investigation. Thus, the methods of social research can be used to explicitly undertake autobiographical work. For example, ethnographic methods have been used to explore the personal experiences and life events of researchers themselves. These have included autobiographical analyses of illnesses, relationships, and bereavements. These personal autobiographical narratives can be located within a broader genre of AUTOETHNOGRAPHY (Reed-Danahay, 2001).

EVALUATION

It is difficult to evaluate autobiography as a particular approach to social research, given the variety of ways in which it may be used. Like all research that deals with life stories, autobiographical research should be reflexive of the voices that are being heard and the lives that are being told. Storytelling voices are differentiated and stratified, and although we should not privilege the powerful, we should also not valorize the powerless in our treatment of their life stories.

A criticism that has been leveled at particular forms of autobiographical social research is the potential for romanticizing the self or of engaging in self-indulgence. Although distinctions can be made between social research and autobiographical practice, their intertextuality should also be recognized. The self is inexplicably linked to the processes of social research. This should be recognized and understood. Personal experiences and autobiographical stories can be sources for insightful analysis and innovative social science.

—Amanda Coffey

REFERENCES

Kondo, D. K. (1990). *Crafting selves: Power, gender and discourses of identity in a Japanese workplace.* Chicago: University of Chicago Press.

Plummer, K. (2001). *Documents of life 2.* London: Sage.

Reed-Danahay, D. (2001). Autobiography, intimacy and ethnography. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography* (pp. 405–425). London: Sage.

Stanley, L. (1992). *The auto/biographical I: Theory and practice of feminist auto/biography.* Manchester, UK: Manchester University Press.

AUTOCORRELATION. *See* SERIAL CORRELATION

AUTOETHNOGRAPHY

Autoethnographic inquirers use their own experiences to garner insights into the larger culture or subculture of which they are a part. Autoethnography combines an autobiographical genre of writing with ethnographic research methods, integrated through both personal and cultural lenses and written creatively.

ETHNOGRAPHY first emerged as a method for studying and understanding the OTHER.

Fascination with exotic otherness attracted Europeans to study the peoples of Africa, Asia, the South Sea Islands, and the Americas. In the United States, for educated, White, university-based Americans, the others were Blacks, American Indians, recent immigrants, and the inner-city poor. To the extent that ethnographers reported on their own experiences as PARTICIPANT OBSERVERS, it was primarily through methodological reporting related to how they collected data and how, or the extent to which, they maintained detachment.

In contrast, an autoethnographic inquiry asks the following: How does my own experience of my own culture offer insights about my culture, situation, event, and/or way of life? Autoethnography raises the question of how an ethnographer might study her or his own culture without the burden or pretense of detachment (Patton, 2002). What if there is no *other* as the focus of study, but I want to study the culture of my own group, my own community, my own organization, and the way of life of people like me, the people I regularly encounter, or my own cultural experiences?

David Hayano (1979) is credited with originating the term *autoethnography* to describe studies by anthropologists of their own cultures. Autoethnographies are typically written in first-person voice but can also take the forms of short stories, poetry, fiction, photographic portrayals, personal essays, journal entries, and social science prose. Although autoethnographers use their own experiences to garner insights into the larger culture or subculture of which they are a part, great variability exists in the extent to which autoethnographers make themselves the focus of the analysis; how much they keep their role as social scientist in the foreground; the extent to which they use the sensitizing notion of culture, at least explicitly, to guide their analysis; and how personal the writing is. At the center, however, what distinguishes autoethnography from ethnography is self-awareness about and reporting of one's own experiences and introspections as a primary data source. Carolyn Ellis (Ellis & Bochner, 2000) has described this process as follows:

> I start with my personal life. I pay attention to my physical feelings, thoughts, and emotions. I use what I call systematic sociological introspection and emotional recall to try to understand an experience I've lived through. Then I write my experience as a story. By exploring a particular life, I hope to understand a way of life. (p. 737)

BOUNDARY-BLURRING INQUIRY

Many social science academics object to the way autoethnography blurs the lines between social science and literary writing. Laurel Richardson (2000), in contrast, sees the integration of art, literature, and social science as precisely the point, bringing together creative and critical aspects of inquiry. She suggests that what these various new approaches and emphases share is that "they are produced through *creative analytic practices,*" which leads her to call "this class of ethnographies *creative analytic practice ethnography*" (p. 929). Other terms are also being used to describe these boundary-blurring, evocative forms of qualitative inquiry—for example, *anthropological poetics, ethnographic poetics, ethnographic memoir, ethnobiography, evocative narratives, experimental ethnography, indigenous ethnography, literary ethnography, native ethnography, personal ethnography, postmodern ethnography, reflexive ethnography, self-ethnography, socioautobiography,* and *sociopoetics* (Patton, 2002, p. 85).

But how is one is to judge the quality of such nontraditional social scientific approaches that encourage personal and creative ethnographic writing? In addition to the usual criteria of substantive contribution and methodological rigor, autoethnography and other evocative forms of inquiry add aesthetic qualities, REFLEXIVITY (consciousness about one's point of view and what has influenced it), impact on the reader, and authenticity of voice as criteria. In this vein, Elliot Eisner (1997) has suggested that in "the new frontier in qualitative research methodology," an artistic qualitative social science contribution can be assessed by the "number and quality of the questions that the work raises" as much as by any conclusions offered (p. 268).

Autoethnography can take the form of creative nonfiction. Sociologist Michael Quinn Patton (1999), for example, wrote about a 10-day Grand Canyon hike with his son, during which they explored what it means to *come of age,* or be initiated into adulthood, in modern society. To make the study work as a story and make scattered interactions coherent, he rewrote conversations that took place over several days into a single evening's dialogue, he reordered the sequence of some conversations to enhance the plot line, and he revealed his emotions, foibles, doubts, weaknesses, and uncertainties as part of the data of inquiry.

Autoethnography integrates ethnography with personal story, a specifically autobiographical manifestation of a more general turn to BIOGRAPHICAL METHODS in social science that strives to "link macro and micro levels of analysis . . . [and] provide a sophisticated stock of interpretive procedures for relating the personal and the social" (Chamberlayne, Bornat, & Wengraf, 2000, pp. 2–3). Although controversial, the blurring of boundaries between social science and art is opening up new forms of inquiry and creative, evocative approaches to reporting findings.

—Michael Quinn Patton

REFERENCES

Chamberlayne, P., Bornat, J., & Wengraf, T. (2000). *The turn to biographical methods in social science.* London: Routledge.

Eisner, E. W. (1997). The new frontier in qualitative research methodology. *Qualitative Inquiry, 3,* 259–273.

Ellis, C., & Bochner, A. P. (2000). Autoethnography, personal narrative, reflexivity: Researcher as subject. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 733–768). Thousand Oaks, CA: Sage.

Hayano, D. M. (1979). Autoethnography: Paradigms, problems, and prospects. *Human Organization, 38,* 113–120.

Patton, M. Q. (1999). *Grand Canyon celebration: A father-son journey of discovery.* Amherst, NY: Prometheus.

Patton, M. Q. (2002). *Qualitative research and evaluation methods.* Thousand Oaks, CA: Sage.

Richardson, L. (2000). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 923–948). Thousand Oaks, CA: Sage.

AUTOREGRESSION. *See* SERIAL CORRELATION

AVERAGE

The *average* is a term used to represent typical values of variables, particularly those values that indicate the midpoint of a distribution, the MEAN or the MEDIAN. In general use, it is taken to refer to the arithmetic mean. The arithmetic mean is the sum of the values of the given variable, divided by the number of cases. The median is the midpoint of the distribution of values, such that exactly half the values are higher and half are lower than the median. The choice of which average to use depends on the particular advantages and disadvantages of the two measures for different situations and is discussed further under MEASURES OF CENTRAL TENDENCY.

Example

Most students, however innumerate, are able to calculate their grade point averages because it is typically the arithmetic mean of all their work over a course that provides their final grade. Table 1 shows a hypothetical set of results for students for three pieces of coursework. The table gives the mean and median for each student and also for the distribution of marks across each piece of work. It shows the different relationships the mean and median have with the different distributions, as well as the different ways that students can achieve a mark above or below the average. Dudley achieved a good result by consistent good marks, whereas Jenny achieved the same result with a much greater variation in her performance. On the other hand, Jenny achieved an exceptional result for Project 1, achieving 17 points above the mean and 19 above the median, whereas all students did similarly well on Project 2. If the apparently greater difficulty of Project 1 resulted in differential WEIGHTING of the scores, giving a higher weight to Project 1, then Jenny might have fared better than Dudley. The table provides a simple example of how the median can vary in relation to the mean. Of course, what would be the *fairest* outcome is a continuous matter of debate in educational institutions.

HISTORICAL DEVELOPMENT

The use of combining repeat observations to provide a check on their accuracy has been dated to Tycho Brahe's astronomical observations of the 16th century. This application of the arithmetic mean as a form of verification continued in the field of astronomy, but the development of the average for aggregation and estimation in the 18th and 19th centuries was intimately connected with philosophical debates around the meaning of nature and the place of mankind. Quetelet, the pioneer of analysis of social scientific data, developed the notion of the "average man" based on his measurements and used this as an ideal reference point for comparing other men and women. Galton, who owed much to Quetelet, was concerned rather with *deviation* from the average, with his stress

Table 1 Student Coursework Marks and Averages for Nine Students (Constructed Data)

	Project 1	Project 2	Project 3	Coursework Average: Mean	Coursework Average: Median
Adam	52	62	63	59	62
Akiko	51	63	55	56.3	55
Chris	55	65	57	59	57
Dudley	63	62	64	63	62
Jenny	71	66	52	63	66
Maria	48	64	65	59	64
Robin	41	61	47	49.7	47
Surinder	47	67	60	58	60
Tim	58	66	42	55.3	58
Mean	54	64	56	58	59
Median	52	64	57	59	60

on "hereditary genius" and eugenics, from which he derived his understanding of regression. Since then, "average" qualities have tended to be associated more with mediocrity than with perfection.

AVERAGES AND LOW-INCOME MEASUREMENT

A common use of averages is to summarize income and income distributions. Fractions of average resources are commonly used as de facto poverty lines, especially in countries that lack an official poverty line. Which fraction, which average, and which measure of resources (e.g., income or expenditure) varies with the source and the country and with those carrying out the measurement. Thus, in Britain, for example, the annual *Households Below Average Income* series operationalizes what has become a semi-official definition of *low income* as 50% of mean income. The series does also, however, quote results for other fractions of the mean and for fractions of the median. By comparison, low-income information for the European Union uses 60% of the median as the threshold, and it is the median of expenditure that is being considered. In both cases, incomes have been weighted according to household size by equivalence scales. As has been pointed out, the fact that income distributions are skewed to the right means that the mean will come substantially higher up the income distribution than the median. It is therefore often considered that the median better summarizes the center of an income distribution. However, when fractions of the average are being employed, the issue changes somewhat, both conceptually and practically. Conceptually, a low-income measure is attempting to grasp a point at which people are divorced from participation in expected or widespread standards and practices of living. Whether this is best expressed in relation to the median or the mean will depend in part on how the relationship of participation and exclusion is constructed. Similarly, at a practical level, a low-income measure is attempting to quantify the bottom end of the income distribution. Whether this is best achieved with an upper limit related to the mean or the median is a different question from whether the mean and the median themselves are the better summaries of the income distribution. In practice, in income distributions, 50% of the mean and 60% of the median typically fall at a very similar point. These two measures are consequently those that are most commonly used as they become effectively interchangeable.

An example of an income distribution is given in Figure 1. The figure shows gross U.S. household incomes from a survey of the late 1980s. Incomes have been plotted by household frequency in a form of histogram, although the histogram has been truncated so that the most extreme values (the highest incomes) have been excluded. The double line furthest to the right marks the mean of the distribution (at $2,447 per month), with the solid line to the left of it indicating where the median falls ($2,102). The mode, at the highest peak of the distribution, is also marked with

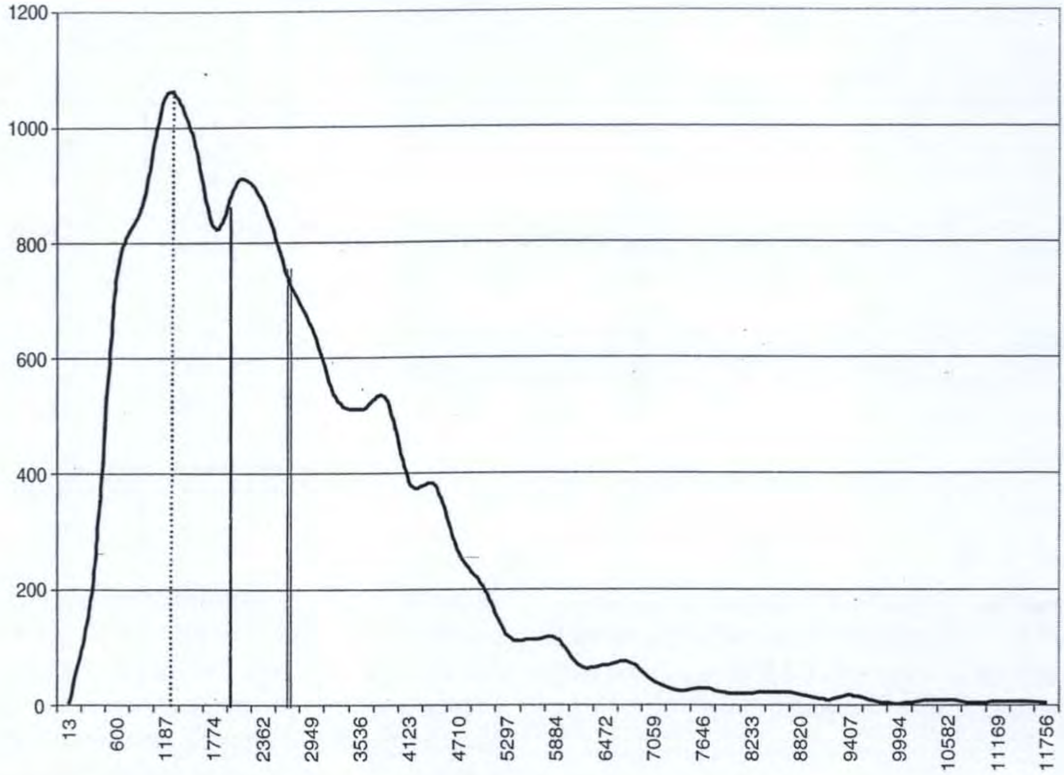

Figure 1 A Typical Income Distribution Showing Mode, Median, and Mean Income Points, U.S. Survey Data of Gross Household Incomes (Truncated), 1987

a dotted line. It is clear that the bulk of households fall to the left of the mean. The 60% of median and 50% of mean points will fall very close (at $1,261 and $1,223, respectively). They would thus cut off the bulk of the first hump of the distribution, which could then be deemed, on an inequality measure, to constitute low-income households.

—Lucinda Platt

See also MEAN, MEASURES OF CENTRAL TENDENCY, MEDIAN, MEDIAN TEST, *T*-TEST.

REFERENCES

Goodman, A., Johnson, P., & Webb, S. (1997). *Inequality in the UK.* Oxford, UK: Oxford University Press.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics: Vol. 1. Distribution theory* (5th ed.). London: Griffin.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Griffin.

B

BAR GRAPH

Bar graphs are one of the simplest ways of illustrating the relative frequencies or proportions of variables with discrete values. Typically, the bars representing the different values (e.g., age ranges, level of education, political affiliation, first language) rise from the x-axis, with the height of the bar representing the frequency, which can be read from the y-axis. It is also possible to produce bar graphs so that the bars are set along the y-axis and the frequencies are read from the x-axis, in which case the graph is called a horizontal bar graph. In this simple form, bar graphs display information that could also be displayed graphically using a two-dimensional PIE CHART. Figure 1 gives an example of a bar graph containing information on number of children in a group of 100 families. It gives an easy way of seeing the modal number of children (2) as well as the low frequency of larger families.

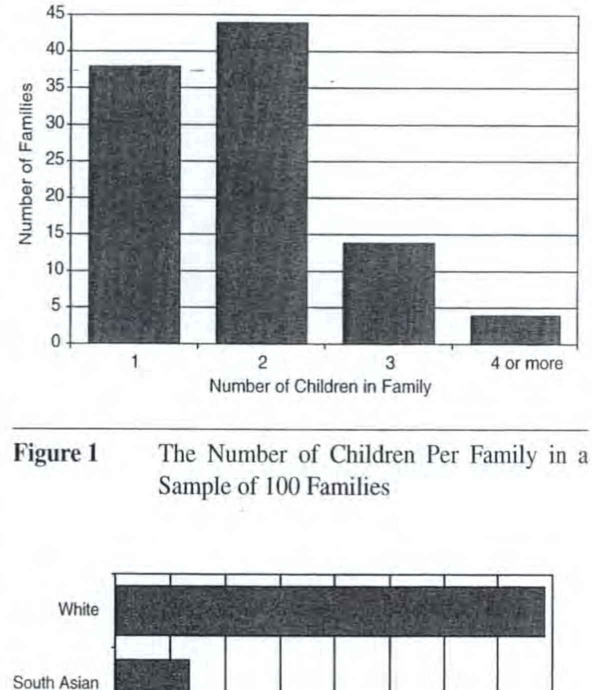

Figure 1 The Number of Children Per Family in a Sample of 100 Families

Figure 2 is an example of a horizontal bar graph that shows the proportions of different ethnic groups who made up the population of a British city according to the 1991 census.

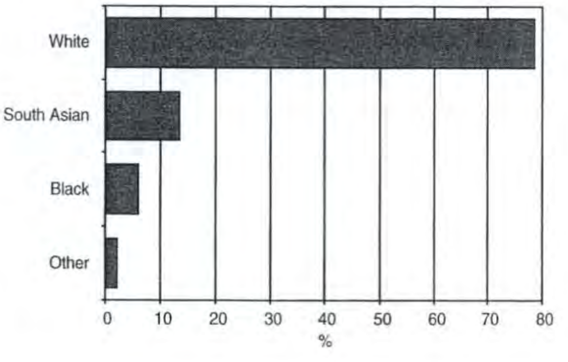

Figure 2 Bar Graph Showing the Ethnic Group Population Proportions in a British City, 1991

Bar graphs can be adapted to contain information on more than one variable through the use of clustering or stacking. Figures 3 and 4 illustrate the information on ethnicity contained in Figure 2 when the age structure of the population is taken into account. Figure 3 is an example of a clustered bar chart where each column sums to 100%. This shows us that the various ethnic groups have different population structures, but does not tell us what that means in terms of absolute numbers. Figure 4, on the other hand, stacks the populations next to each other, comparing white with all the minority groups, and retains the actual frequencies.

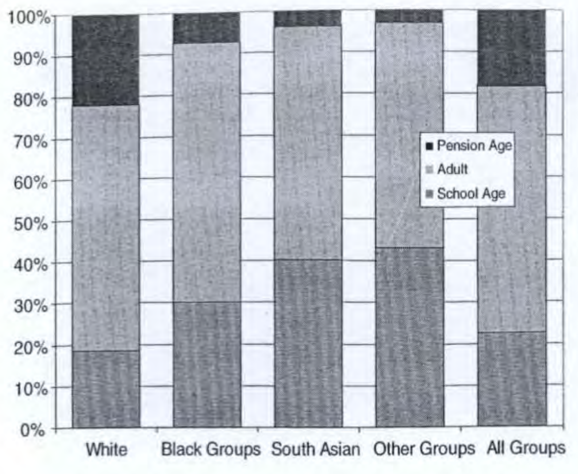

Figure 3 Stacked Bar Graph Showing the Population Structure of Different Ethnic Groups in a British City, 1991

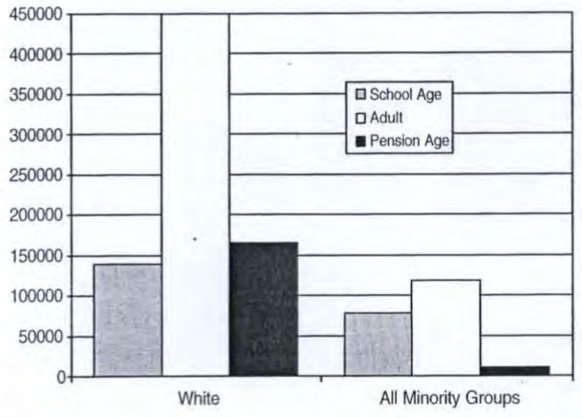

Figure 4 Clustered Bar Graph Comparing the Population Age Structure of White and Minority Ethnic Groups in a British City, 1991

The use of bar graphs to display data not only is a convenient way of summarizing information in an accessible form, but also, as John Tukey was at pains to argue, can lead us to insights into the data themselves. Therefore, it is perhaps surprising that the origins of the bar graph cannot be found before Charles Playfair's development of charts for statistical and social data at the end of the 18th century. Now, however, the development of software simplified the creation of graphs, including the more complex, three-dimensional stacked and clustered graphs, as well as the adjustment of their appearance, gridlines, and axes. The use of bar graphs in the social sciences to describe data and their characteristics swiftly and accessibly is not only widespread but also expected. Nevertheless, the choices (e.g., type of bar graph, which data should be contained within the bar and which should form the axis, frequency or proportion) can still require considerable thought and may have implications for how the data are understood.

—Lucinda Platt

REFERENCES

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Wallgren, A. et al. (1996). *Graphing statistics and data: Creating better data.* Thousand Oaks, CA: Sage.

BASELINE

A baseline is a point of reference for comparison of research findings. For example, before implementation of a drug prevention program, researchers might establish the prevailing average drug use in the target population to serve as a baseline in evaluating the effects of the program. In a REGRESSION context, another example might be the expected value of DEPENDENT VARIABLE Y when INDEPENDENT VARIABLE $X = 0$; that value could stand as a baseline for comparing the expected value of Y when $X = 1$, or some other nonzero value. Sometimes, the results of an entire model may serve as the baseline, to be compared to another model.

—Michael S. Lewis-Beck

BASIC RESEARCH

Curiosity is the hallmark of human mental activity. Since the Greeks, it has been thought that a fundamental human impulse is to know the causes of things. This passion for knowledge has produced not only knowledge but also the foundations for vast improvements in our daily lives (Kornberg, 1997).

Basic research is driven by curiosity about the way the world works. We know that sooner or later, the new knowledge will have practical beneficial consequences, sometimes astounding ones. But these cannot be foreseen, and, indeed, may coyly hide if searched for directly. The historical record makes abundantly clear that curiosity is the heartbeat of discovery and invention. Scientists engaged in basic research do not think about these things, except when called on to justify themselves; they are simply in thrall to their curiosities. But for those engaged in building the infrastructure of science or stocking the research portfolio, who often may be overpowered by concern for eradicating ills, the lessons about basic research are imperative (Kornberg, 1997).

Basic research can be characterized in several ways. Some note its tendency to be innovative and risk-taking. But at the bottom line, basic research is simply research whose purpose is to know the way some phenomenon or process works. Of course, phenomena and processes differ in the number and variety of domains they touch. For example, a high point in human understanding of physical nature was the realization that the same basic processes govern both earthly and celestial phenomena.

Thus, although curiosity drives all basic research, and all basic research seeks to understand the way something works, the more basic the process, the more basic the research. The Holy Grail is always discovery of the ultimate forces. Along the road, we lavish the utmost care and devotion on processes that, however ubiquitous or however long their reach, are not really ultimate forces. In so doing, we learn something about the way the world works, and we sharpen scientific tools in preparation for the more basic processes yet to be discerned.

Following Newton's (1686/1952) ideas for understanding physical nature, observed behavioral and social phenomena are usefully viewed as the product of the joint operation of several basic forces. Put differently, the sociobehavioral world is a *multifactor* world, a view that lies at the heart of both theoretical and empirical work (e.g., Parsons, 1968). The scientific challenges are two—one theoretical, the other empirical. The theoretical challenge is to identify the basic forces governing human behavior, describe their operation, and derive their implications. The empirical challenge is to test the derived implications. The theoretical and empirical work jointly lead to the accumulation of reliable knowledge about human behavioral and social phenomena.

Even the briefest consideration of basic research would not be complete without at least some speculation on the identity of the basic forces governing behavioral and social phenomena. Fermi used to spend an hour a day on pure speculation about the nature of the physical world (Segrè, 1970). In that spirit, an initial list of candidates for basic forces might include the four discussed in Jasso (2001, pp. 350–352):

1. To know the causes of things
2. To judge the goodness of things
3. To be perfect
4. To be free

—Guillermina Jasso

Further Reading

Kornberg's (1997) little essay should be required reading for all social scientists, all students, all deans, all review panels, and all R&D units.

REFERENCES

Jasso, G. (2001). Rule-finding about rule-making: Comparison processes and the making of norms. In M. Hechter & K.-D. Opp (Eds.), *Social norms* (pp. 348-393). New York: Russell Sage.

Kornberg, A. (1997). *Basic research, the lifeline of medicine.* Retrieved from http://www.nobel.se/medicine/articles/research/

Newton, I. (1952). *Mathematical principles of natural philosophy.* Chicago: Britannica. (Originally published in 1686)

Parsons, T. (1968). Émile Durkheim. In D. L. Sills (Ed.), *International encyclopedia of the social sciences, Vol. 4.* New York: Macmillan.

Segrè, E. (1970). *Enrico Fermi: Physicist.* Chicago: University of Chicago Press.

BAYES FACTOR

Given data y and two models M_0 and M_1, the Bayes factor

$$B_{10} = \frac{p(y\,|M_1)}{p(y\,|M_0)} = \left\{\frac{p(M_1\,|y)}{p(M_0\,|y)}\right\} \Big/ \left\{\frac{p(M_1)}{p(M_0)}\right\}$$

is a summary of the evidence for M_1 against M_0 provided by the data, and is also the ratio of posterior to prior odds (hence the name "*Bayes* factor"). Twice the logarithm of B_{10} is on the same scale as the deviance and likelihood ratio test statistics for model comparisons. Jeffreys (1961) suggests the following scale for interpreting the Bayes factor:

B_{10}	$2 \log B_{10}$	*Evidence for* M_1
< 1	< 0	Negative (support M_0)
1 to 3	0 to 2	Barely worth mentioning
3 to 12	2 to 5	Positive
12 to 150	5 to 10	Strong
> 150	> 10	Very strong

Note that B is a ratio of *marginal likelihoods*, that is, $p(y|M_k) = \int p(y|\theta_k, M_k)p(\theta_k)\, d\theta_k$, where θ_k is a vector of parameters of model M_k, $p(\theta_k)$ is the prior density of θ_k, and $p(y|\theta_k, M_k)$ is the likelihood of y under model $M_k, k \in \{0, 1\}$. Raftery (1996) provides a summary of methods of computing Bayes factors, and Good (1988) summarizes the history of the Bayes factor.

—Simon Jackman

See also BAYES' THEOREM

REFERENCES

Good, I. J. (1988). The interface between statistics and philosophy of science. *Statistical Science, 3,* 386–397.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Clarendon.

Raftery, A. E. (1996). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 163–187). London: Chapman and Hall.

BAYES' THEOREM, BAYES' RULE

At the heart of Bayesian statistics and decision theory is Bayes' Theorem, also frequently referred to as Bayes' Rule. In its simplest form, if H is a hypothesis and E is evidence, then the theorem is

$$\Pr(H|E) = \frac{\Pr(E \cap H)}{\Pr(E)} = \frac{\Pr(E|H)\Pr(H)}{\Pr(E)}$$

provided $\Pr(E) > 0$, so that $\Pr(H|E)$ is the PROBABILITY of belief in H *after* obtaining E, and $\Pr(H)$ is the *prior* probability of H before considering E. The left-hand side of the theorem, $\Pr(H|E)$ is usually referred to as the *posterior* probability of H. Thus, the theorem supplies a solution to the general problem of inference or induction (e.g., Hacking, 2001), giving us a mechanism for learning about a hypothesis H from data E.

Bayes' Theorem itself is uncontroversial: It is merely an accounting identity that follows from the axiomatic foundations of probability linking joint, conditional, and marginal probabilities, that is, $\Pr(A|B)\Pr(B) = \Pr(A \cap B)$, where $\Pr(B) \neq 0$. Thus, Bayes' Theorem is sometimes referred to as the *rule of inverse probability,* because it shows how a conditional probability B given A can be "inverted" to yield the conditional probability A given B (Leamer, 1978, p. 39). In addition, the theorem provides a way for considering two hypotheses, H_i and H_j, in terms of ODDS RATIOS; that is, it follows from Bayes' Theorem that

$$\frac{\Pr(H_i|E)}{\Pr(H_j|E)} = \frac{\Pr(E|H_i)\Pr(H_i)}{\Pr(E|H_j)\Pr(H_j)}$$

giving the posterior odds ratio for H_i over H_j (given E) equal to the *likelihood ratio* of E under H_i and H_j times the prior odds ratio. In the continuous case (e.g., learning about a real-valued parameter θ given data y), Bayes' Theorem is often expressed as

$$\pi(\theta|y) \propto f(y;\theta)\pi(\theta),$$

where $f(y;\theta)$ is the *likelihood function,* and the constant of proportionality $[f(y;\theta)\pi(\theta)\, d\theta]^{-1}$ does not depend on θ and ensures that the posterior density integrates to 1. This version of Bayes' Theorem is fundamental to Bayesian statistics, showing how the likelihood function (the probability of the data, given θ) can be turned into a probability statement about θ, given data y. Note also that the Bayesian approach treats the parameter θ as a random variable, and conditions on the data, whereas frequentist approaches consider θ a fixed (but unknown) property of a population from which we randomly sample data y.

The Reverend Thomas Bayes died in 1761. The result now known as Bayes' Theorem appeared in an essay attributed to Bayes, and it was communicated to the Royal Society after Bayes' death by Richard Price in 1763 (Bayes, 1763) and republished many times (e.g., Bayes, 1958). Bayes himself stated the result only for a uniform prior. According to Stigler (1986b), in 1774, Laplace (apparently unaware of Bayes' work) stated the theorem in its more general form (for discrete events). Additional historical detail can be found in Bernardo and Smith (1994, chap. 1), Lindley (2001), and Stigler (1986a, chap. 3).

—Simon Jackman

REFERENCES

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society, 53,* 370–418.

Bayes, T. (1958). An essay towards solving a problem in the doctrine of chances. *Biometrika, 45,* 293–315.

Bernardo, J., & Smith, A. F. M. (1994). *Bayesian theory.* Chichester, UK: Wiley.

Hacking, I. (2001). *An introduction to probability and inductive logic.* Cambridge, UK: Cambridge University Press.

Leamer, E. (1978). *Specification searches: Ad hoc inference with nonexperimental data.* New York: Wiley.

Lindley, D. V. (2001). Thomas Bayes. In C. C. Heyde & E. Seneta (Eds.), *Statisticians of the centuries* (pp. 68–71). New York: Springer-Verlag.

Stigler, S. (1986a). *The history of statistics: The measurement of uncertainty before 1900.* Cambridge, MA: Belknap.

Stigler, S. (1986b). Laplace's 1774 memoir on inverse probability. *Statistical Science, 1,* 359–378.

BAYESIAN INFERENCE

The Bayesian statistical approach is based on updating information using a PROBABILITY theorem from the famous 1763 essay by the Reverend Thomas Bayes. He was an amateur mathematician whose work was found and published 2 years after his death by his friend Richard Price. The enduring association of an important branch of statistics with his name is due to the fundamental importance of this probability statement, now called Bayes' Law or BAYES' THEOREM, which relates conditional probabilities.

Start with two events of interest, X and Y, which are not independent. We know from the basic axioms of probability that the conditional probability of X given that Y has occurred is obtained by

$$p(X|Y) = \frac{p(X, Y)}{p(Y)},$$

where $p(X, Y)$ is "the probability that both X and Y occur," and $p(Y)$ is the unconditional probability that Y occurs. The conditional expression thus gives the probability of X after some event Y occurs.

We can also define a different conditional probability in which X occurs first,

$$p(Y|X) = \frac{p(Y, X)}{p(X)}.$$

Because the joint probability that X and Y occur is the same as the joint probability that Y and X occur $[p(X, Y) = p(Y, X)]$, then it is possible to make the following simple rearrangement:

$$\begin{aligned} p(X, Y) &= p(X|Y)p(Y), \\ p(Y, X) &= p(Y|X)p(X), \\ p(X|Y)p(Y) &= p(Y|X)p(X), \\ \therefore\ p(X|Y) &= \frac{p(X)}{p(Y)} p(Y|X) \qquad (1) \end{aligned}$$

(for more details, see Bernardo and Smith, 1994, chap. 3). The last line is the famous Bayes' Law, and this is really a device for "inverting" conditional probabilities. It is clear that one could just as easily produce $p(Y|X)$ in the last line above by moving the unconditional probabilities to the left-hand side in the last equality. Bayes' Law is useful because we often know $p(X|Y)$ and would like to know $p(Y|X)$, or vice versa. The fact that $p(X|Y)$ is never equal to $p(Y|X)$ (that is, the probability that

$$\frac{p(X)}{p(Y)} = 1$$

is zero) is often called the *inverse probability problem.*

It turns out that this simple result is the foundation for the general paradigm of Bayesian statistics.

DESCRIPTION OF BAYESIAN INFERENCE

Bayesian inference is based on fundamentally different assumptions about data and parameters than are classical methods (Box & Tiao, 1973, chap. 1). In the Bayesian world, all quantities are divided into two groups: observed and unobserved. Observed quantities are typically the data and any known relevant constants. Unobserved quantities include PARAMETERS of interest to be estimated, MISSING DATA, and parameters of lesser interest that simply need to be accounted for. All observed quantities are fixed and are conditioned on. All unobserved quantities are assumed to possess distributional qualities and are therefore treated as random variables. Thus, parameters are now no longer treated as fixed unmoving in the total population, and all statements are made in probabilistic terms.

The inference process starts with assigning PRIOR DISTRIBUTIONS for the unknown parameters. These range from very informative descriptions of previous research in the field to deliberately vague and diffuse forms that reflect relatively high levels of ignorance. The prior distribution is not an inconvenience imposed by the treatment of unknown quantities; instead, it is an opportunity to systematically include qualitative, narrative, and intuitive knowledge into the statistical model. The next step is to stipulate a likelihood function in the conventional manner by assigning a parametric form for the data and inserting the observed quantities. The final step is to produce a POSTERIOR DISTRIBUTION by multiplying the prior distribution and the likelihood function. Thus, the likelihood function uses the data to *update* the prior knowledge conditionally.

This process, as described, can be summarized by the simple mnemonic:

$$\text{posterior probability} \propto \text{prior probability} \times \text{likelihood function.} \qquad (2)$$

This is just a form of Bayes' Law where the denominator on the right-hand side has been ignored by using proportionality. The symbol "$\propto$" stands for "proportional to," which means that constants have been left out that make the posterior sum or integrate to 1, as is required of probability mass functions and PROBABILITY DENSITY FUNCTIONS. Renormalizing to a proper form can always be done later, plus, using proportionality is more intuitive and usually reduces the calculation burden. What this "formula" above shows is that the posterior distribution is a compromise between the prior distribution, reflecting research beliefs, and the likelihood function, which is the contribution of the data at hand. To summarize, the Bayesian inference process is given in three general steps:

1. Specify a probability model that includes some prior knowledge about the parameters if available for unknown parameter values.
2. Update knowledge about the unknown parameters by conditioning this probability model on observed data.
3. Evaluate the fit of the model to the data and the sensitivity of the conclusions to the assumptions.

Nowhere in this process is there an artificial decision based on the assumption that some NULL HYPOTHESIS of no effect is true. Therefore, unlike the seriously flawed null hypothesis SIGNIFICANCE TEST, evidence is presented by simply summarizing this posterior distribution. This is usually done with quantiles and probability statements, such as the probability that the parameter of interest is less than/greater than some interesting constant, or the probability that this parameter occupies some region. Also note that if the posterior distribution is now treated as a new prior distribution, it too can be updated if new data are observed. Thus, knowledge about the parameters of interest is updated and accumulated over time (Robert, 2001).

THE LIKELIHOOD FUNCTION

Suppose collected data are treated as a fixed quantity and we know the appropriate probability mass function or probability density function for describing the data generation process. Standard likelihood and Bayesian methods are similar in that they both start with these two suppositions and then develop estimates of the unknown parameters in the parametric model. MAXIMUM LIKELIHOOD ESTIMATION substitutes the unbounded notion of likelihood for the bounded definition of probability by starting with Bayes' Law:

$$p(\theta|\mathbf{X}) = \frac{p(\theta)}{p(\mathbf{X})} p(\mathbf{X}|\theta), \qquad (3)$$

where θ is the unknown parameter of interest and $\mathbf{X}$ is the collected data. The key is to treat

$$\frac{p(\theta)}{p(\mathbf{X})}$$

as an unknown function of the data independent of $p(\mathbf{X}|\theta)$. This allows us to use $L(\theta|\mathbf{X}) \propto p(\mathbf{X}|\theta)$. Because the data are fixed, then different values of the likelihood function are obtained merely by inserting different values of the unknown parameter, θ.

The likelihood function, $L(\theta|\mathbf{X})$, is similar to the desired but unavailable inverse probability, $p(\mathbf{X}|\theta)$, in that it facilitates testing alternate values of θ to find a most probable value, $\hat{\theta}$. However, because the likelihood function is no longer bounded by 0 and 1, it is now important only *relative* to other likelihood functions based on differing values of θ. Note that the prior, $p(\theta)$, is essentially ignored here rather than overtly addressed. This is equivalent to assigning a uniform prior in a Bayesian context, an observation that has led some to consider classical inference to be a special case of Bayesianism: "Everybody is a Bayesian; some know it."

Interest is generally in obtaining the *maximum likelihood* estimate of θ. The value of the unconstrained and unknown parameter, θ, provides the maximum value of the likelihood function, $L(\theta|\mathbf{X})$. This value of θ, denoted $\hat{\theta}$, is the most likely to have generated the data given H_0 expressed through a specific parametric form relative to other possible values in the sample space of θ.

BAYESIAN THEORY

The Bayesian approach addresses the inverse probability problem by making distributional assumptions about the unconditional distribution of the parameter, θ, *prior* to observing the data, $\mathbf{X} : p(\theta)$. The prior and likelihood are joined with Bayes' Law:

$$\pi(\theta|\mathbf{X}) = \frac{p(\theta)L(\theta|\mathbf{X})}{\int_\Theta p(\theta)L(\theta|\mathbf{X})d\theta}, \tag{4}$$

where $\int_\Theta p(\theta)L(\theta|\mathbf{X})d\theta = p(\mathbf{X})$. Here, the $\pi(\,)$ notation is used to distinguish the posterior distribution for θ from the prior. The term in the denominator is generally not important in making inferences and can be recovered later by integration. This term is typically called the *normalizing constant*, the *normalizing factor*, or the *prior predictive distribution*, although it is actually just the marginal distribution of the data and ensures that $\pi(\theta|\mathbf{X})$ integrates to 1.

A more compact and useful form of (4) is developed by dropping this denominator and using proportional notation, because $p(\mathbf{X})$ does not depend on θ and therefore provides no relative inferential information about more likely values of θ:

$$\pi(\theta|\mathbf{X}) \propto p(\theta)L(\theta|\mathbf{X}), \tag{5}$$

meaning that the unnormalized *posterior* (sampling) distribution of the parameter of interest is proportional to the prior distribution times the likelihood function.

The maximum likelihood estimate is equal to the Bayesian posterior mode with the appropriate uniform prior, and they are asymptotically equal given *any* prior: Both are NORMALLY DISTRIBUTED in the limit. In many cases, the choice of a prior is not especially important because as the sample size increases, the likelihood progressively dominates the prior. Although the Bayesian assignment of a prior distribution for the unknown parameters can be seen as subjective (although *all* statistical models are actually subjective), there are often strong arguments for particular forms of the prior: Little or vague knowledge often justifies a diffuse or even uniform prior, certain probability models logically lead to particular forms of the prior (conjugacy), and the prior allows researchers to include additional information collected outside the current study.

SUMMARIZING BAYESIAN RESULTS

Bayesian researchers are generally not concerned with just getting a specific point estimate of the parameter of interest, θ, as a way of providing empirical evidence in probability distributions. Rather, the focus is on describing the shape and characteristics of the posterior distribution of θ. Such descriptions are typically in the form of direct probability intervals (credible sets and highest posterior density regions), quantiles of the posterior, and probabilities of interest such as $p(\theta_i < 0)$.

Hypothesis testing can also be performed in the Bayesian setup. Suppose Θ_1 and Θ_2 represent two competing hypotheses about the state of some unknown parameter, θ, and form a partition of the sample space: $\Theta = \Theta_1 \cup \Theta_2$, $\Theta_1 \cap \Theta_2 = \phi$. Prior probabilities are assigned to each of the two outcomes: $\pi_1 = p(\theta \in \Theta_1)$ and $\pi_2 = p(\theta \in \Theta_2)$. This leads to competing posterior distributions from the two priors and the likelihood function: $p_1 = p(\theta \in \Theta_1|\mathbf{X})$ and $p_2 = p(\theta \in \Theta_2|\mathbf{X})$. It is common to define the prior odds, π_1/π_2, and the posterior odds, p_1/p_2, as evidence for H_1 versus H_2. A much more useful quantity, however, is $(\pi_1/\pi_2)/(p_1/p_2)$, which is called the BAYES FACTOR. The Bayes Factor is usually interpreted as odds favoring H_1 versus H_2 given the observed data. For this reason, it leads naturally to the Bayesian analog of hypothesis testing.

AN EXAMPLE

Suppose that $x_1, x_2, \ldots, x_n$ are observed independent random variables, all produced by the same Bernoulli probability mass function with parameter θ so that their sum,

$$y = \sum_{i=1}^{n} x_i,$$

is distributed binomial (n, θ) with known n and unknown θ. Assign a beta (A, B) prior distribution for θ,

$$p(\theta|A, B) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}\theta^{A-1}(1-\theta)^{B-1},$$
$$0 < \theta; 0 < A \text{ or } B,$$

with researcher-specified values of A and B.

The posterior distribution for θ is obtained by the use of Bayes' Law:

$$\pi(\theta|y) = \frac{p(\theta|A, B)p(y|\theta)}{p(y)}, \tag{6}$$

where $p(y|\theta)$ is the probability mass function for y. The numerator of the right-hand side is easy to calculate now that there are assumed parametric forms:

$$\begin{aligned} p(y,\theta) &= p(y|\theta)p(\theta) \\ &= \left[\binom{n}{y}\theta^{y}(1-\theta)^{n-y}\right] \\ &\quad \times \left[\frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}\theta^{A-1}(1-\theta)^{B-1}\right] \\ &= \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} \\ &\quad \times \theta^{y+A-1}(1-\theta)^{n-y+B-1} \end{aligned} \tag{7}$$

and the denominator can be obtained by integrating θ out of this joint distribution:

$$\begin{aligned} p(y) &= \int_0^1 \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} \\ &\quad \times \theta^{y+A-1}(1-\theta)^{n-y+B-1}dp \\ &= \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} \\ &\quad \frac{\Gamma(y+A)\Gamma(n-y+B)}{\Gamma(n+A+B)}. \end{aligned} \tag{8}$$

For more technical details, see Gill (2002, p. 67). Performing arithmetic, the operations in (6) provide

$$\begin{aligned} \pi(\theta|y) &= \frac{p(\theta)p(y|\theta)}{p(y)} \\ &= \frac{\Gamma(n+A+B)}{\Gamma(y+A)\Gamma(n-y+B)} \\ &\quad \times \theta^{(y+A)-1}(1-\theta)^{(n-y+B)-1}. \end{aligned} \tag{9}$$

This is a new beta distribution for θ with parameters $A' = y + A$ and $B' = n - y + B$ based on updating the prior distribution with the data y. Because the prior and the posterior have the same distributional family form, this is an interesting special case, and it is termed a *conjugate model* (the beta distribution is conjugate for the binomial probability mass function). Conjugate models are simple and elegant, but it is typically the case that realistic Bayesian specifications are much more complex analytically and may even be impossible.

MARKOV CHAIN MONTE CARLO

A persistent and troubling problem for those developing Bayesian models in the 20th century was that it was often possible to get a sufficiently complicated posterior from multiplying the prior and the likelihood, that the mathematical form existed, but quantities of interest such as means and quantiles could not be calculated analytically. This problem pushed Bayesian methods to the side of mainstream statistics for quite some time. What changed this unfortunate state of affairs was the publication of a review essay by Gelfand and Smith (1990) that described how similar problems had been solved in statistical physics with MARKOV CHAIN simulation techniques. Essentially, these are iterative techniques where the generated values are (eventually) from the posterior of interest (Gill, 2002). Thus, the difficult posterior form can be described empirically using a large number of simulated values, thus performing difficult integral calculations through simulation. The result of this development was a flood of papers that solved a large number of unresolved problems in Bayesian statistics, and the resulting effect on Bayesian statistics can be easily described as revolutionary.

Markov chains are composed of successive quantities that depend probabilistically only on the value of their immediate predecessor. In general, it is possible to set up a chain to estimate multidimensional probability structures, the desired posterior distributions, by starting a Markov chain in the appropriate sample space and letting it run until it settles into the desired target distribution. Then, when it runs for some time confined to this particular distribution, it is possible to collect summary statistics such as means, variances, and quantiles from the simulated values. The two most common procedures are the Metropolis-Hastings algorithm and the GIBBS SAMPLING, which have been shown to possess desirable theoretical properties that lead to empirical samples from the target distribution. These methods are reasonably straightforward to implement and are becoming increasingly popular in the social sciences.

—Jeff Gill

REFERENCES

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* New York: Wiley.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis.* New York: Wiley.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*, 389–409.
Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach.* New York: Chapman and Hall.
Robert, C. P. (2001). *The Bayesian choice: A decision theoretic motivation* (2nd ed.). New York: Springer-Verlag.

BAYESIAN SIMULATION. *See* MARKOV CHAIN MONTE CARLO METHODS

BEHAVIOR CODING

Behavior coding is a systematic, flexible, and cost-efficient method of examining the interaction between a survey interviewer and respondent. Coders listen to live or taped interviews and assign CODES for behaviors such as the accuracy with which interviewers read questions and the difficulties respondents exhibit in answering. Results of this coding can be used to monitor interviewer performance, identify problems with survey questions, or analyze the effects of methodological experimentation. Although each of these objectives requires somewhat different procedures, the general process is the same. A researcher must identify behavior to code; develop an appropriate coding scheme; select and train coders; assess reliability; supervise the coding task; and, finally, analyze the results.

Beginning in the late 1960s, survey methodologists conducted a variety of studies of interviews based on behavior coding of taped interviews (e.g., Cannell, Fowler, & Marquis, 1968). These studies provided valuable information about various components of the survey interview, including differences in behaviors in OPEN-ENDED and CLOSED-ENDED QUESTIONS, in characteristics of respondents, and in experienced and inexperienced interviewers. The most significant finding in these studies was that the surveys did not proceed as the researcher had expected. Questions were not administered as specified, nor were other techniques followed according to instructions. The first reaction was to blame the interviewer, attributing deficiencies in performance to poor training, inadequate supervision, or simply a lack of interviewer effort. More careful analysis, however, suggested that much of the difficulty could be attributed to the techniques in which they were trained or inadequacies in the survey questions.

Not surprisingly, the first practical routine use of behavior coding in survey research was to monitor interviewer performance. Behavior coding is now used as a technique in the developmental stages of survey questions and is used routinely by many organizations in evaluating questions during PRETESTS (Oksenberg, Cannell, & Kalton, 1991). Most coding schemes developed for the purpose of evaluating questions examine the initial reading of the question by the interviewer, as well as all of the subsequent respondent behaviors. Interviewer reading codes evaluate the degree to which the question was read exactly as worded. Respondent behavior codes, at a minimum, attempt to capture behaviors related to comprehension difficulties (e.g., request for clarification); potential problems related to question wording or length (e.g., interruptions); and the adequacy of the response (e.g., inadequate response, qualified response, adequate response) as well as inability to respond (e.g., "don't know" responses and refusals). High rates of "poor" behaviors are indicative of potential problems with a question wording or response options.

—Nancy Mathiowetz

See also INTERVIEWER EFFECTS, INTERVIEWER TRAINING, PRETESTING

REFERENCES

Cannell, C., Fowler, J., & Marquis, K. (1968). *The influence of interviewer and respondent psychological and behavioral variables on the reporting of household interviews* (Vital and Health Statistics Series 2, No. 6). Washington, DC: National Center for Health Statistics.
Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics, 7*(3), 349–365.

BEHAVIORAL SCIENCES

The behavioral sciences are focused on the explanation of human behavior. Psychology is the primary example, although other social sciences have behavioral subfields. For example, in political science, the study of voting behavior is well-developed, as is the study of consumer behavior in economics.

—Michael S. Lewis-Beck

BEHAVIORISM

Behaviorism, a term invented by John B. Watson (Watson, 1959), is the name of both a method and a metaphysics. On the grounds that first-person reports of one's private consciousness are suspect because they cannot be independently verified, the *methodological behaviorist* seeks to replace personal introspection of one's own states of mind with controlled observations of one's behavior by other people. Instead of taking an angry child's description of her emotion as definitive, the behaviorist would note whether she tries to damage the sorts of things that reportedly annoy her. On the grounds that we should regard as unreal what can be observed by only one person in only one way, the *metaphysical behaviorist* maintains that emotions and other so-called states of mind are not private possessions knowable only by the people and animals who have them, but publicly observable dispositions to behave in determinate ways. Thus, metaphysical behaviorists contend that, because anger normally *consists in* a disposition to damage things, we are observing a child's emotion by observing how it is manifest in her behavior. As thus defined, behaviorism is well caricatured by a joke its critics used to tell about two behaviorists who meet in the street. One says to the other, "You feel fine; how do I feel?"

The first-person method of introspecting what were called the phenomena of (human) consciousness dominated psychology during the 18th and 19th centuries. During most of the 20th century, the third-person method of observing behavior, both animal and human, under carefully controlled conditions dominated what was called behavioral science. However, by the end of the 20th century, many cognitive scientists became convinced that behaviorist strictures against introspection and consciousness had been *too* restrictive. In the new view, sensations, emotions, thoughts, and other states of mind are really states of the brain. These states will normally give rise to distinctive sorts of behavior, but they are related to the behavior as causes are related to their effects. So, it is argued, their true nature can no more be discovered merely by observing the behavior to which they give rise than the physical structure of magnets can be discovered merely by watching them attract iron filings. On the theory that the human brain has a special capacity to monitor its own states, some scientists and philosophers have therefore undertaken once again to study consciousness, the favorite topic of 19th-century philosophers, by combining introspection with research in brain physiology. Behaviorists have responded to this renewal of an old program not by denying the relevance of brain physiology, but by insisting that the so-called phenomena of consciousness are illusions, not hard data, and that—as Howard Rachlin tries to show—behavior is still the key to the mind's secrets (Rachlin, 2000).

—Max Hocutt

REFERENCES

Rachlin, H. (2000). *The science of self-control.* Cambridge, MA: Harvard University Press.

Watson, J. B. (1959). *Behaviorism.* Chicago: University of Chicago Press.

BELL-SHAPED CURVE

The bell-shaped curve is the term used to describe the shape of a NORMAL DISTRIBUTION when it is plotted with the x-axis showing the different values in the distribution and the y-axis showing the frequency of their occurrence. The bell-shaped curve is a symmetric distribution such that the highest frequencies cluster around the midpoint of the distribution with a gradual tailing off toward 0 at an equal rate on either side in the frequency of values as they move away from the center of the distribution. In effect, it resembles a church bell, hence the name. Figure 1 illustrates such a bell-shaped curve. As can be seen from the symmetry and shape of the curve, all three MEASURES OF CENTRAL TENDENCY—the mean, the mode, and the median—coincide at the highest point of the curve.

The bell-shaped curve is described by its mean, μ, and its STANDARD DEVIATION, σ. Each bell-shaped curve with a particular μ and σ will represent a unique distribution. As the frequency of distributions is greater toward the middle of the curve and around the mean, the probability that any single observation from a bell-shaped distribution will fall near to the mean is much greater than that it will fall in one of the tails. As a result, we know, from normal probabilities, that in the bell-shaped curve, 68% of values will fall within 1 standard deviation of the mean, 95% will fall within roughly 2 standard deviations of

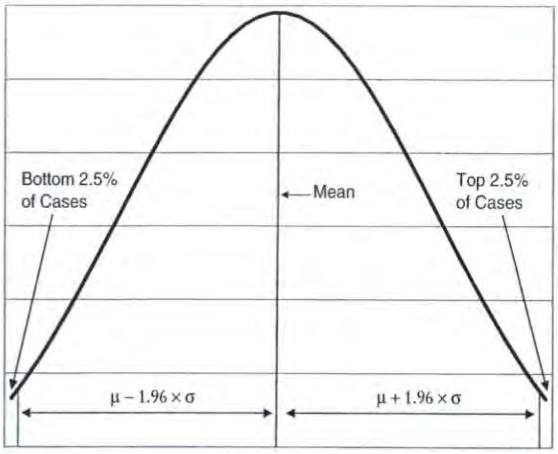

Figure 1 Example of a Bell-Shaped Curve

the mean, and nearly all will fall within 3 standard deviations of the mean. The remaining observations will be shared between the two tails of the distribution. This is illustrated in Figure 1, where we can see that 2.5% of cases fall into the tails beyond the range represented by $\mu \pm 1.96 \times \sigma$. This gives us the probability that in an approximately bell-shaped sampling distribution, any case will fall with 95% probability within this range, and thus, by statistical extrapolation, the population parameter can be predicted as falling within such a range with 95% confidence. (See also CENTRAL LIMIT THEOREM, NORMAL DISTRIBUTION.)

A particular form of the bell-shaped curve has its mean at 0 and a standard deviation of 1. This is known as the standard normal distribution, and for such a distribution the distribution probabilities represented by the equation $\mu \pm z^{*}\sigma$ simplify to the value of the multiple of σ (or Z-SCORE) itself. Thus, for a standard normal distribution, the 95% probabilities lie within the range ± 1.96.

Bell-shaped distributions, then, clearly have particular qualities deriving from their symmetry, such that it is possible to make statistical inferences for any distributions that approximate this shape. By extrapolation, they also form the basis of statistical inference even when distributions are not bell-shaped. In addition, distributions that are not bell-shaped can often be transformed to create an approximately bell-shaped curve. Thus, for example, income distributions, which show a skew to the right, can be transformed into an approximately symmetrical distribution by taking the log of the values. Conversely, distributions with a skew to the left, such as examination scores or life expectancy, can be transformed to approximate a bell shape by squaring or cubing the values.

HISTORICAL DEVELOPMENT AND USE

The normal-distribution, or bell-shaped, curve has held a particular attraction for social statisticians, especially since Pierre Laplace proved the central limit theorem in 1810. This showed how a bell-shaped curve could arise from a non-bell-shaped distribution, and how a set of samples, all with their own errors, could, through their means, take on a symmetric and predictable form. Laplace was building on the work of Carl Friedrich Gauss, who gave his name to the normal distribution (it is alternatively known as the Gaussian distribution). Astronomical observation, which employed the arithmetic MEAN as a way of checking repeat observations and providing the "true" value, also showed empirically that the other observations followed the shape of a bell-shaped curve, clustering around the mean. The bell curve therefore represented a set of errors clustering around a true value. The idea of the mean as representing some uniquely authentic value was taken up by Adolphe Quetelet in his work on the "average man" (see AVERAGE). Quetelet also developed the (erroneous) idea that normal distributions are the form to which aspects of social and biological life naturally tend and, conversely, assumed that where bell-shaped curves were found, the population creating the curve could be considered homogeneous. The tendency to regard bell-shaped distributions as having some innate validity and the temptation to reject those that do not continued to be a feature of statistical work into the 20th century, as Stigler (1999) has pointed out. However, although some distributions in nature (e.g., adult height) approximate a bell-shaped curve distribution in that they show a clustering around the mean and a tailing off in either direction, most do not; and perfect symmetry is rarely found. Nevertheless, the theoretical developments based on the bell-shaped curve and its properties, and the Gaussian distribution in particular, remain critical to statistical inference in the social sciences.

The visual and conceptual attraction of the bell-shaped curve has meant its continuing use as a marker for grasping truths or expressing beliefs about society beyond, although drawing on, its statistical inferential properties. Thus, for example, Richard J. Herrnstein and Charles Murray's (1994) somewhat notorious discussion of the distribution of intelligence among whites and African Americans in the United States used the term *The Bell Curve* as the title for their book.

—Lucinda Platt

REFERENCES

Aggresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Marsh, C. (1988). *Exploring data: An introduction to data analysis for social scientists.* Cambridge, UK: Polity.

Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods.* Cambridge, MA: Harvard University Press.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Charles Griffin.

BERNOULLI

Bernoulli is the name given to a PROBABILITY feature known as the Bernoulli process, used for describing the probabilities of an event with two outcomes, such as flipping a coin. The BINOMIAL probability of k successes in n trials of a Bernoulli process is given by

$$P(E) = \binom{n}{k} p^k (1-p)^{n-k},$$

where the event E has outcomes of exactly k heads during n flips of a fair coin.

—Tim Futing Liao

BEST LINEAR UNBIASED ESTIMATOR

The term *best linear unbiased estimator* (BLUE) comes from application of the general notion of unbiased and efficient estimation in the context of linear estimation. In statistical and econometric research, we rarely have populations with which to work. We typically have one or a few SAMPLEs drawn from a population. Our task is to make meaningful statistical inferences about parameter estimates based upon sample statistics used as estimators. Here, the statistical properties of estimators have a crucial importance because without estimators of good statistical properties, we could not make credible statistical inferences about population parameters. We desire estimators to be unbiased and efficient. In other words, we require the expected value of estimates produced by an estimator to be equal to the true value of population parameters. We also require the variance of the estimates produced by the estimator to be the smallest among all unbiased estimators. We call an estimator the *best unbiased estimator* (BUE) if it satisfies both conditions. The class of BUE estimators may be either linear or nonlinear. If a BUE estimator comes from a linear function, then we call it a best linear unbiased estimator (BLUE).

For a more sophisticated understanding of the term BLUE, we need to delve into the meaning of the terms "estimator," "unbiased," "best," and "linear" in more detail.

ESTIMATOR

An estimator is a decision rule for finding the value of a population parameter using a sample statistic. For any population parameter, whether it is a population average, dispersion, or MEASURE OF ASSOCIATION between variables, there are usually several possible estimators. Good estimators are those that are, on average, close to the population parameter with a high degree of certainty.

UNBIASEDNESS

With unbiasedness, we require that, under repeated sampling, the expected value of the sample estimates produced by an estimator will equal the true value of the population parameter. For example, suppose we estimate from a sample a bivariate LINEAR REGRESSION $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i$ in order to find the parameters of a population regression function $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. Unbiasedness requires that the expected values of $\hat{\beta}_0$ and $\hat{\beta}_1$ equal the actual values of β_0 and β_1

in the population. Mathematically, this is expressed $E[\hat{\beta}_0] = \beta$ and $E[\hat{\beta}_1] = \beta_1$. Intuitively, unbiasedness means that the sample estimates will, on average, produce the true population parameters. However, this does not mean that any particular estimate from a sample will be correct. Rather, it means that the estimation rule used to produce the estimates will, on average, produce the true population parameter.

EFFICIENCY

The term *best* refers to the efficiency criterion. Efficiency means that an unbiased estimator has minimum variance compared to all other unbiased estimators. For example, suppose we have a sample drawn from a normal population. We can estimate the midpoint of the population using either the mean or the median. Both are unbiased estimators of the midpoint. However, the mean is the more efficient estimator because, under repeated sampling, the variance of the sample medians is approximately 1.57 times larger than the variance of the sample means. The rationale behind the efficiency criterion is to increase the probability that sample estimates fall close to the true value of the population parameter. Under repeated sampling, the estimator with the smaller variance has a higher probability of yielding estimates that are closer to the true value.

LINEARITY

For an estimator to be linear, we require that predicted values of the dependent variable be a linear function of the estimated parameters. Notice that we are interested here in linearity in parameters, not in variables. For example, suppose we estimate a sample REGRESSION such as $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1\sqrt{x_i} + \varepsilon_i$. In this regression, the relation between $\hat{y}_i$ and x_i is not linear, but the relation between $\hat{y}_i$ and $\sqrt{x_i}$ is linear. If we think of $\sqrt{x_i}$, instead of x_i, as our explanatory variable, then the linearity assumption is met. However, this cannot be applied equally for all estimators. For example, suppose we estimate a function like

$$\hat{y}_i = \hat{\beta}_0 + \frac{\sqrt{\hat{\beta}_1}}{\hat{\beta}_0}x_i + \varepsilon_i.$$

This is not a linear estimator, but requires nonlinear estimation techniques to find $\hat{\beta}_0$ and $\hat{\beta}_1$. An estimator of this sort might be BUE, but would not be BLUE.

THE GAUSS-MARKOV THEOREM

BLUE implies meeting all three criteria discussed above. BLUE estimators are particularly convenient from a mathematical standpoint. A famous result in statistics is the Gauss-Markov theorem, which shows that the LEAST SQUARES estimator is the best linear unbiased estimator under certain assumptions. This means that when an estimation problem is linear in parameters, it is always possible to use a least squares estimator to estimate population parameters without bias and with minimum variance. This result is convenient mathematically because it is unnecessary to prove the statistical properties of least squares estimators. They always have good statistical properties as long as the Gauss-Markov assumptions are met.

LIMITS OF BLUE

Although the BLUE property is attractive, it does not imply that a particular estimator will always be the best estimator. For example, if the disturbances are not normal and have been generated by a "fat-tailed" distribution so that large disturbances occur often, then linear unbiased estimators will generally have larger variances than other estimators that take this into account. As a result, nonlinear or robust estimators may be more efficient. Additionally, linear estimators may not be appropriate in all estimation situations. For example, in estimating the variance of the population disturbance (σ_e^2), quadratic estimators are more appropriate.

—B. Dan Wood and Sung Ho Park

REFERENCES

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics.* New York: Oxford University Press.

Greene, W. (2003). *Econometric analysis* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H., & Lee, T. (1988). *Introduction to the theory and practice of econometrics.* New York: Wiley.

BETA

This Greek letter, β, refers to a REGRESSION slope in two different ways. In the first, it is the symbol for the slope value in the regression equation for the POPULATION (as opposed to the SAMPLE). In the second, it is the symbol for a standardized slope (expressing change in STANDARD DEVIATION units) and is sometimes called a "beta weight." Usually, the context will make clear which of the two meanings of beta is intended.

—Michael S. Lewis-Beck

See also STANDARDIZED REGRESSION COEFFICIENT

BETWEEN-GROUP DIFFERENCES

A relationship exists between two variables if there is covariation between two variables; that is, if certain values of one variable tend to go together with certain values of the other variable. If people with more education tend to have higher incomes than people with less education, a relationship exists between education and income.

To study a possible relationship, we create groups of observations based on the independent variable in such a way that all elements that have the same value on the independent variable belong to the same group. The independent variable can be a NOMINAL VARIABLE, such as gender, where all the females are in one group and all the males in another group. Next, the two groups are compared on the dependent variable, say, income. If the two groups are found to be different, a relationship exists.

The independent variable can also be a QUANTITATIVE VARIABLE, such as years of completed school. Here, everyone with 8 years of education is in one group, everyone with 9 years of education is in the next group, and so on. Again, with income as the dependent variable, we compare the incomes in the various groups.

One way to compare the groups is to compute the value of a summary statistic, such as the mean, and then compare these means across the groups. We could prefer to compare the medians from the different groups, or we can create a graph such as BOXPLOTS and compare them. Using the means, the question becomes, How do we compare them? Also, how large are the differences in the means in comparison with the spread of the values of the dependent variable?

With a nominal independent variable, the method most commonly used to compare the means of the groups is ANALYSIS OF VARIANCE; with a quantitative independent variable, REGRESSION analysis is used the most. In analysis of variance, we compare the mean incomes of the two genders to the overall mean. In regression, we find the predicted income values of each year of education and compare these values to the overall mean. Because these points lie on a line, we need not ask if the points are different; rather, does the regression line through these points have a nonzero slope, thereby showing that the predicted incomes are different? Had we run an analysis of variance on the education-income data, we would have found a smaller RESIDUAL SUM OF SQUARES and a better fit than for the regression analysis. But an analysis of variance model uses more DEGREES OF FREEDOM than the corresponding regression model, and a regression analysis would therefore typically have a smaller *P* VALUE.

In analysis of variance, we compare the group means to the overall mean, taking the sizes of the groups into account, by computing the BETWEEN-GROUP SUM OF SQUARES. The more the groups vary, the larger the sum of squares becomes. This sum is also referred to as the *between sum of squares*. Presumably, the means are different because the dependent variable has been affected by one or more independent variables. If gender is the one independent variable, the sum might be labeled *gender sum of squares* or *model sum of squares*.

The similar quantity in regression analysis is most often simply called the *regression sum of squares*. It takes into account how steep the regression line is and how spread out the observations are on the independent variable.

—Gudmund R. Iversen

REFERENCES

Cobb, G. W. (1997). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07–001, 2nd ed.). Newbury Park, CA: Sage.

Lewis-Beck, M. S. (1980). *Applied regression: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07–022). Beverly Hills, CA: Sage.

BETWEEN-GROUP SUM OF SQUARES

In principle, in the study of relationships between variables, the data are first divided into groups according to one or more independent variables. Within each group, we often compute a summary measure of the dependent variable, such as a mean, and then we compare these means.

If the means are different, then we have established that a relationship exists between the independent and the dependent variable. With two groups, the comparison is easy: We simply subtract one mean from the other to see if they are different. With more than two groups, the comparisons of the means become more extensive. We want to take into account how different the means are from each other as well as how large the groups are.

With an independent NOMINAL VARIABLE having k values, we do an ANALYSIS OF VARIANCE to determine how different the groups are. In such an analysis, for mathematical and historical reasons, we compute the following sum:

$$n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2 + \cdots + n_k(\bar{y}_k - \bar{y})^2,$$

where the ns are the number of observations in the groups, the y-bars with subscripts are the k group means, and the one mean without a subscript is the overall mean of the dependent variable. This is the between-group sum of squares. Sometimes, this is abbreviated to simply the *between sum of squares* (BSS) or even the *model sum of squares*. The magnitude of the sum clearly depends upon how different the group means are from each other and how large the groups are. With several independent variables, there is a between-group sum of squares for each variable, usually labeled with the name of the variable.

In and of itself, the magnitude of the between-group sum of squares does not tell much. But when it is divided by the total sum of squares, the fraction gives the proportion of the variation in the dependent variable accounted for by the independent variable. Also, when divided by its degrees of freedom $(k - 1)$, the fraction becomes the numerator of the F value used to find the p VALUE used to determine whether to reject the null hypothesis of no group differences.

When the independent variable is a QUANTITATIVE VARIABLE, REGRESSION analysis is usually preferred over analysis of variance. A linear model assumes the population means of the dependent variable on a straight line. Rather than a NULL HYPOTHESIS asking whether these means are equal, we need to look only at the slope of the regression line. A horizontal line with slope equal to zero implies equal means.

The regression sum of squares takes the place of the between-group sum of squares. Its major advantage is that its number of DEGREES OF FREEDOM is based only on the number of independent variables (1 in simple regression), not on the number of groups. However, if we also can do an analysis of variance on the same data, the BSS is typically somewhat smaller than the regression sum of squares, meaning that analysis of variance fits the data better than regression.

—Gudmund R. Iversen

See also SUM OF SQUARES

REFERENCES

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07–001, 2nd ed.). Newbury Park, CA: Sage.

Lewis-Beck, M. S. (1980). *Applied regression: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07–022). Beverly Hills, CA: Sage.

BIAS

In common usage, the term *bias* means predisposition, prejudice, or distortion of the truth. Although the general idea is the same, in statistics, we use the term bias in connection with the relationship between a (sample) STATISTIC (also called an ESTIMATOR) and its true (PARAMETER) value in the POPULATION.

The difference between the EXPECTED VALUE of a statistic and the parameter value that the statistic is trying to estimate is known as bias. To see what all this means, one can proceed as follows.

Let X be a RANDOM VARIABLE from some PROBABILITY DENSITY FUNCTION (PDF) with the following parameters: a (population) mean value of μ_x and (population) variance of σ_x^2. Suppose we draw a RANDOM SAMPLE of n observations from this population. Let

$$\bar{X} = \frac{\sum_{i=1}^{i=n} X_i}{n}$$

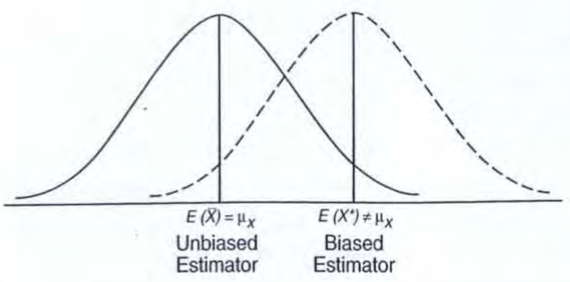

Figure 1 A Biased (X^*) and Unbiased ($\bar{X}$) Estimator of Population Mean Value, μ_x.

be the sample mean value. Suppose we use $\bar{X}$, the (sample) statistic, as an estimator of μ_x.

If the expected value of $\bar{X}$ is equal to μ_x, that is, $E(\bar{X}) = \mu_x$, then it can be said that $\bar{X}$ is an UNBIASED estimator of μ_x. If this equality does not hold, then $\bar{X}$ is a biased estimator of μ_x. That is,

$$\text{bias} = E(\bar{X}) - \mu_x,$$

which is, of course, zero if the expected value of the sample mean is equal to the true mean value. The bias can be positive or negative, depending on whether $E(\bar{X})$ is greater than μ_x or less than μ_x.

In other words, if one draws repeated samples of n observations of X, the average value of the means of those repeated samples should equal the true mean value μ_x. But note that there is no guarantee that any single value of $\bar{X}$ obtained from one sample will necessarily equal the true population mean value.

To find out if the expected value of the sample means equals the true mean value, one can proceed as follows:

$$\begin{aligned} E(\bar{X}) &= E\left[\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right] \\ &= \frac{1}{n}[Ex_1 + Ex_2 + \cdots + Ex_n] \\ &= \frac{1}{n}[n.\mu_x] = \mu_x, \end{aligned}$$

because each X has the same mean value μ_x. Because $E(\bar{X}) = \mu_x$, it can be said that the sample mean is an unbiased estimator of the population mean.

Geometrically, the bias can be illustrated as in Figure 1.

To further illustrate bias, consider two estimators of the population variance σ_x^2, namely,

$$S_x^{*2} = \sum \frac{(X_i - \bar{X})^2}{n} \quad \text{and} \quad S_x^2 = \sum \frac{(X_i - \bar{X})^2}{n-1},$$

which are the sample variances. The difference between the two is that the former has n DEGREES OF FREEDOM (df) and the latter has $(n-1)\,df$. It can be shown that

$$E(S_x^{*2}) = \sigma_x^2\left(1 - \frac{1}{n}\right) \quad \text{but} \quad E\left(S_x^2\right) = \sigma_x^2.$$

Because the expected value of the former estimator is not equal to true σ_x^2, it is a biased estimator of the true variance.

Unbiasedness is one of the desirable properties that an estimator should satisfy. The underlying idea is intuitively simple. If an estimator is used repeatedly, then, on average, it should be equal to its true value. If this does not happen, it will systematically distort the true value.

Unbiasedness is a small, or finite, sample property. An estimator may not be unbiased in small samples, but it may be unbiased in large samples, that is, as the sample size n increases indefinitely, the expected, or mean, value of the estimator approaches its true value. In that case, the estimator is called asymptotically unbiased. This is clearly the case with the estimator S_x^{*2}, because

$$\lim_{n\to\infty} E\left(S_x^{*2}\right) = \sigma_x^2.$$

Hence, asymptotically, S_x^{*2} is unbiased. In statistics, one comes across situations in which an estimator is biased in small samples, but the bias disappears as the sample size increases indefinitely.

The examples given thus far relate to a single random variable. However, the concept of bias can be easily extended to the multivariate case. For instance, consider the following MULTIPLE REGRESSION:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i,$$

where Y is the dependent variable, the Xs are the explanatory variables, and u_i is the STOCHASTIC error term. In matrix notation, the preceding multiple regression can be written as

$$\mathbf{y} = \beta\mathbf{X} + \mathbf{u},$$

where $\mathbf{y}$ is an $(n \times 1)$ vector of observations on the dependent variable and $\mathbf{X}$ is an $(n \times k)$ data matrix on the k explanatory variables (including the intercept) and where $\mathbf{u}$ is an $(n \times 1)$ vector of the stochastic error terms.

Using the method of ORDINARY LEAST SQUARES (OLS), the estimators of the beta coefficients are obtained as follows:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

where $\hat{\beta}$ is a $(k \times 1)$ vector of the estimators of the true β coefficients.

Under the assumptions of the classical LINEAR REGRESSION model (the classical linear regression model assumes that values of the explanatory variables are fixed in repeated sampling, that the errors terms are not correlated, and that the error variance is constant, or HOMOSKEDASTIC), it can be shown that $\hat{\beta}$ is an unbiased estimator of the true β. The proof is as follows:

$$\begin{aligned} E(\hat{\beta}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u})] \\ &= \beta + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = \beta, \end{aligned}$$

where use is made of the assumptions of the classical linear regression model (note that the explanatory variables are nonstochastic).

As this expression shows, each element of the *estimated* β vector is equal to the corresponding value of the true β vector. In other words, each $\hat{\beta}_i$ is an unbiased estimator of the true β_i.

—Damodar N. Gujarati

REFERENCES

Gujarati, D. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

BIC. *See* GOODNESS-OF-FIT MEASURES

BIMODAL

It is a characteristic of a statistical DISTRIBUTION having two separate peaks or MODES. In contrast, the NORMAL DISTRIBUTION is a UNIMODAL DISTRIBUTION.

—Tim Futing Liao

BINARY

Binary means two, and it may have at least three kinds of connotation to social scientists. The *binary system,* unlike the decimal system, has only two digits (0 and 1) and serves as the foundation of the operation of the modern computer. A *binary variable* measures a quantity that has only two categories. Social scientists often engage in *binary thinking* in categorizing social phenomena, such as the public and the private, Gemeinschaft and Gesellschaft, the micro and the macro, and so on.

—Tim Futing Liao

See also ATTRIBUTE, DICHOTOMOUS VARIABLES

BINOMIAL DISTRIBUTION

In a binomial distribution, there are a finite number of independently sampled observations, each of which may assume one of two outcomes. Thus, the binomial distribution is the DISTRIBUTION of a binary variable, such as males versus females, heads versus tails of a coin, or live versus dead seedlings. The SAMPLE proportion or PROBABILITY, p, that a given observation has a given outcome is calculated as the count, X, of observations sampled having that outcome divided by sample size, n. The sum of probabilities of each of a pair of binary states is always 1.0.

The binomial distribution is approximately what one would get if marbles were dropped at the apex of a Galton board, which is a triangular array of pegs such that there is a 50% chance of a dropped marble going either left or right. It is also the distribution resulting from the binomial expansion of the formula $(p+q)^n$, where p and q are the probabilities of getting or not getting a particular dichotomous outcome in n trials. For instance, (see Figure 1) for four trials selecting for Democrats and Republicans from a population with equal numbers of each, the binomial expansion is

$$\begin{aligned} (p+q)^4 &= (1/2 + 1/2)^4 \\ &= 1p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + 1q^4 \\ &= 1/16 + 4/16 + 6/16 + 4/16 + 1/16. \end{aligned}$$

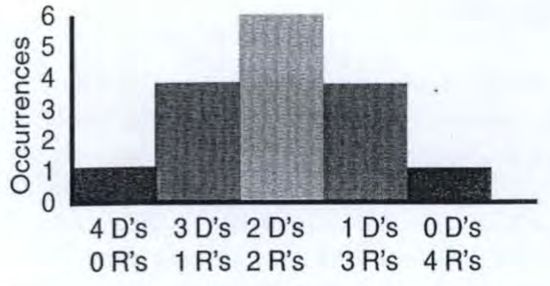

Figure 1 Binomial Distribution

The equation for calculating probabilities in a binomial distribution is

$$p(r) = {}_nC_r p^r q^{n-r},$$

where $p(r)$ refers to the probability of getting r observations in one category of a DICHOTOMOUS VARIABLE and $(n - r)$ observations in the other category, given sample size n; p is the probability of getting the first state of the binary variable, and q is the probability of getting the other state; and ${}_nC_r$ is the binomial coefficient, which is the number of combinations of n things taken r at a time. When $p = q = .5$, the binomial distribution will be symmetrical, otherwise not.

For example, when tossing coins, the probability of getting either a head or a tail is .5. For three coin tosses, the probability of getting three heads is $(.5)(.5)(.5) = .125$. We could also get a head, a head, and a tail; or we could get a head, a tail, and a head. The number of combinations where we could get two heads in three tries is ${}_3C_2$, which is $3!/[2!(3-2)!]$, or 3. The probability of getting two heads is then

$$p(r) = {}_nC_r p^r q^{n-r} = (3)(.5^2)(.5^{3-2})$$
$$= (3)(.25)(.5) = .375.$$

A table of the binomial distribution, based on similar calculations, is printed in many statistics books or calculated by statistical software. For $n = 3$, the binomial table is

	p		
r	*0.1*	*0.5*	*0.9*
0	0.729	0.125	0.001
1	0.243	0.375	0.027
2	0.027	0.375	0.243
3	0.001	0.125	0.729

Thus for $p = .1$, .5, or .9, the table above shows the probability of getting $r = 0$, 1, 2, or 3 heads or other binomial states in three trials. Actual binomial tables typically show additional columns in p increments of .1, and also provide tables for $n = 2$ through $n = 20$. Some provide both one-sided and double-sided probabilities, where the former are appropriate to "as large/small" or "larger/smaller than" questions and the latter are appropriate to "as or more different from" questions.

When sample size n is large and the probability of occurrence is moderate, the sample proportion p and the count X will have a distribution that approximates the NORMAL DISTRIBUTION. The mean of this distribution will be approximately np, and its variance will be approximately $np(1 - p)$, which are the mean and variance of a binomial distribution. These normal approximations may be inaccurate for small values of n, so a common rule of thumb is that they may be used only if the product np is more than or equal to 10 and the product $np(1 - p)$ is also more than or equal to 10. For less than that, exact binomial probabilities are calculated. Statistical software usually will not use normal approximations to the binomial probability but will calculate exact binomial probabilities even for larger samples.

When sample size is large and the probability of occurrence is very small or large, the binomial distribution approaches the POISSON DISTRIBUTION. When samples are taken from a small POPULATION, the hypergeometric distribution may be used instead of the binomial distribution.

—G. David Garson

REFERENCES

Blalock, H. M. (1960). *Social statistics.* New York: McGraw-Hill.

Von Collani, E., & Drager, K. (2001). *Binomial distribution handbook.* Basel, Switzerland: Birkhauser.

BINOMIAL TEST

The binomial test is an exact probability test, based on the rules of PROBABILITY. It is used to examine the DISTRIBUTION of a single dichotomous variable when the researcher has a small SAMPLE. It is a nonparametric statistic that tests the difference between a sampled proportion and a given proportion, for one-sample tests.

The binomial test makes four basic assumptions. It assumes the variable of interest is a DICHOTOMOUS VARIABLE whose two values are mutually exclusive and exhaustive for all cases. Like all SIGNIFICANCE TESTS, the binomial test also assumes that the data have been sampled at random (see RANDOM SAMPLING). It assumes that observations are independently sampled. Finally, the binomial test assumes that the probability of any observation, p, has a given value that is constant across all observations. If an error-free value of p cannot be assumed, then alternative tests (Fisher's text, CHI-SQUARE TEST, t-TEST) are recommended. The binomial test is nonparametric and does not assume the data have a NORMAL DISTRIBUTION or any other particular distribution.

In a binomial test, we determine the probability of getting r observations in one category of a dichotomy and $(n-r)$ observations in the other category, given sample size n. Let $p =$ the probability of getting the first category and let $q = (1-p) =$ the probability of getting the other category. Let ${}_nC_r$ be the number of combinations of n things taken r at a time. The formula for the binomial probability is

$$p(r) = {}_nC_r p^r q^{n-r} = (n!p^r q^{n-r})/[r!(n-r)!].$$

For example, assume we know that a particular city is 60% Democratic ($p = .60$) and 40% non-Democratic ($q = .40$), and we sample a particular fraternal organization in that city and find 15 Democrats ($r = 15$) in a sample of 20 people ($n = 20$). We may then ask if the 15 : 5 split found in our sample is significantly greater than the 60:40 split known to exist in the POPULATION. This question can be reformulated as asking, "What is the probability of getting a sample distribution as strong or stronger than the observed distribution?" The answer is the sum of the binomial probabilities $p(15), p(16), \ldots, p(20)$. For instance, $p(15) = [(20!)(60^{15})(40^5)]/[(15!)(5!)] = .0746$. It can be seen that when n is at all large, computation requires a computer. Some statistics books also print a table of individual or cumulative binomial probabilities for various levels of n, r, and p.

The binomial mean is np and is the expected number of successes in n observations. The binomial standard deviation is the square root of npq, decreasing as p approaches 0 or 1.

When n is greater than 25 and p is near .50, and the product of npq is at least 9, then the binomial distribution approximates the normal distribution. In this situation, a normal curve Z-TEST may be used as an approximation of the binomial test. Note that the normal approximation is useful only when manual methods are necessary. SPSS, for instance, always computes the exact binomial test.

—G. David Garson

REFERENCES

Conover, W. J. (1998). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical inference.* New York: Wiley.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

BIOGRAPHIC NARRATIVE INTERPRETIVE METHOD (BNIM)

This methodology for conducting and analyzing biographic narrative interviews has been used in a variety of European research projects, either directly (e.g., Chamberlayne, Rustin, & Wengraf, 2002; Rosenthal, 1998) or in a modified version (e.g., FREE ASSOCIATION INTERVIEWING). Assuming that "narrative expression" is expressive both of conscious concerns and also of unconscious cultural, societal, and individual presuppositions and processes, it is psychoanalytic *and* sociobiographic in approach.

In each BNIM interview, there are three subsessions (Wengraf, 2001, chap. 6). In the first, the interviewer offers only a carefully constructed single narrative question (e.g., "Please tell me the story of your life, all the events and experiences that have been important to you personally, from wherever you want to begin up to now").

In the second, sticking strictly to the sequence of topics raised and the words used, the interviewer asks for more narratives about them. A third subsession can follow in which nonnarrative questions can be posed.

The transcript thus obtained is then processed twice (Wengraf, 2001, Chap. 12). The strategy reconstructs the experiencing of the "interpreting and acting" subject as he or she interpreted events; lived his or her life; and, in the interview, told his or her story. This strategy requires the analysts to go forward through the events as did the subject: *future-blind,* moment by moment, not knowing what comes next or later.

First, a chronology of objective life events is identified. "Objective life events" are those that could be checked using official documents, such as records of school and employment and other organizations. Each item (e.g., "failed exams at age 16"), stripped of the subject's current interpretation, is then presented separately to a research panel, which is asked to consider how this event might have been experienced *at the time,* and, if that experiential hypothesis were true, what might be expected to occur next or later in the life (following hypotheses). After these are collected and recorded, the next life-event item is presented: Its implications for the previous experiential and following hypotheses are considered, and a new round of hypothesizing commences. A process of imaginative identification and critical understanding is sought, constantly to be corrected and refined by the emergence of future events as they are presented one-by-one.

The transcript is then processed into "segments." A new segment is said to start when there is a change of speaker, of topic, or of the manner in which a topic is addressed. Again, each segment is presented in turn to a research panel that attempts to imagine how such interview events and actions might have been experienced at that moment of the interview, with subsequent correction and refinement by further segments.

In both series, separate structural hypotheses are sought, and only later are structural hypotheses developed relating the two. A similar future-blind procedure is also carried out for puzzling segments of the verbatim text (microanalysis). The question about the dynamics of the case can then be addressed: "Why did the people who lived their lives like *this,* tell their stories like *that?*"

—Tom Wengraf

REFERENCES

Chamberlayne, P., Rustin, M., & Wengraf, T. (Eds.). (2002). *Biography and social exclusion in Europe: Experiences and life journeys.* Bristol, UK: Policy Press.

Rosenthal, G. (Ed.). (1998). *The Holocaust in three generations: Families of victims and persecutors of the Nazi regime.* London: Cassell.

Wengraf, T. (2001). *Qualitative research interviewing: Biographic-narrative and semi-structured method.* London: Sage.

BIOGRAPHICAL METHOD.

See LIFE HISTORY METHOD

BIPLOTS

The biplot is an essential tool for any analysis of a data matrix of subjects by variables in which not only the structure of the variables is of interest, but also differences between individuals. Surprisingly, in many large statistical software packages, it is not a standard tool. Several variants of the biplot exist, but only the basic is mentioned here. The biplot was probably first conceived by Tucker (1960), but Gabriel's (1971) seminal work first showed the full potential of the technique. The most extensive reference at present is Gower and Hand (1996).

A standard biplot is the display of a rectangular matrix $\mathbf{X}$ decomposed into a product $\mathbf{AB}'$ of an $I \times S$ matrix $\mathbf{A} = (a_{is})$ with row coordinates and a $J \times S$ matrix $\mathbf{B} = (b_{js})$ with column coordinates, where S indicates the number of coordinate axes, which, preferably, is two or three. Using this decomposition for $\hat{\mathbf{X}}$, a two-dimensional approximation of $\mathbf{X}$, each element $\hat{x}_{ij}$ of this matrix can be written as

$$\hat{x}_{ij} = a_{i1}b_{j1} + a_{i2}b_{j2},$$

which is the *inner* (or *scalar*) *product* of the row vectors (a_{i1}, a_{i2}) and (b_{j1}, b_{j2}). A biplot is obtained by representing each row as a point a_i with coordinates (a_{i1}, a_{i2}), and each column as point b_j with coordinates (b_{j1}, b_{j2}), in a two-dimensional graph (with origin O). These points are generally referred to as *row markers* and *column markers,* respectively.

For example, suppose new mothers have been scored on variables that assess how well they are coping with their infants, such as parenting stress, depression, and cognitive stimulation of her child. The biplot of these data will display which mothers are coping in similar or in different ways, as expressed through their scores on the variables, and it shows which variables have similar patterns (say, depression and parenting stress) or different patterns of scores (depression and cognitive stimulation of the child).

In situations such as this, the variables typically will have been standardized and the coordinates of the

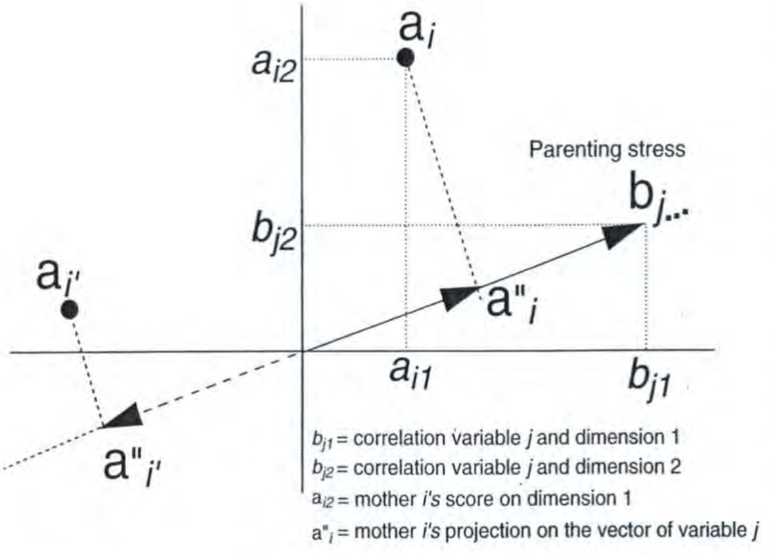

Figure 1 Main Elements of a Biplot: Column Vector, Row Points, Row Coordinates, Column Coordinates, and Projections

NOTE: The inner product between a_i and b_j is the product of the lengths of a''_i and b_j.

variables will have been scaled in such a way that these coordinates are the correlation between the variables and the coordinates (see also PRINCIPAL COMPONENTS ANALYSIS). In this case, the variables have *principal coordinates,* and the subjects have *standard coordinates,* that is, their coordinate axes have zero means and unit lengths. Because of this, the variables are usually displayed either as vectors originating from the origin, with the length of the vector indicating its variability, or as (biplot) axes on which the size of the original scale is marked, so-called calibrated axes. The subjects are indicated as points, and a subject's estimated score on a variable is deduced from the projection of its point on the variables, as illustrated in Figure 1. Because it is not easy to evaluate inner products in a three-dimensional space, the most commonly used biplots have only two dimensions and display the best two-dimensional approximation $\hat{\mathbf{X}}$ to the data matrix $\mathbf{X}$.

Given that a two-dimensional biplot is a reasonable approximation to the data matrix, the relative scores of two mothers on the same variable, say, parenting stress, can be judged by comparing the lengths of their projections onto the vector of parenting stress. In Figure 1 this entails comparing the length of a''_i with that of $a''_{i'}$. If a mother scores above the average, her projected score will be on the positive side of the variables; if she scores below the average, her projected score will be on the negative side of the vector; when her projected score equals the average, it coincides with the origin of the biplot. Two variables having an acute (obtuse) angle with each other have a positive (negative) correlation. Absence of correlation results in perpendicular vectors. All of these statements are conditional on the biplot being a sufficiently good approximation to the data in $\mathbf{X}$.

An important point in constructing graphs for biplots is that the physical vertical and horizontal coordinate axes should have the same physical scale. This will ensure that when row points are projected onto a column vector, they will end up in the correct place. Failing to adhere to this scaling (as is the case in some software packages) will make it impossible to evaluate inner products in the graph.

Besides being a valuable technique in its own right, the biplot also plays an important role in CORRESPONDENCE ANALYSIS of CONTINGENCY TABLES and in three-mode data analysis.

—Pieter M. Kroonenberg

REFERENCES

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika, 58,* 453–467.

Gower, J. C., & Hand, D. J. (1996). *Biplots.* London: Chapman & Hall.

Tucker, L. R. (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychometric scaling: Theory and applications* (pp. 155–167). New York: Wiley.

BIPOLAR SCALE

Bipolar scales are scales with the opposite ends represented by contrasting concepts such as *cold* and *hot,* *slow* and *fast, antisocial* and *social,* or *introvert* and *extrovert.* There can be 2 to 13 points requested for such a scale.

—Tim Futing Liao

BISERIAL CORRELATION

Biserial correlation (r_{bis}) is a correlational index that estimates the strength of a relationship between an artificially dichotomous variable (X) and a true continuous variable (Y). Both variables are assumed to be normally distributed in their underlying populations. It is used extensively in item analysis to correlate a dichotomously scored item with the total score. Assuming that an artificially dichotomous item is scored 1 for a correct answer and 0 for an incorrect answer, the biserial correlation is computed from

$$r_{\text{bis}} = \frac{\bar{Y}_1 - \bar{Y}_0}{S_Y}\left(\frac{n_1 n_0}{u n_.^2}\right),$$

where $\bar{Y}_1$ is the mean of Y for those who scored 1 on X, $\bar{Y}_0$ is the mean of Y for those who scored 0 on X; S_Y is the standard deviation of all Y scores; n_1 is the number of observations scored 1 on X; n_0 is the number of observations scored 0 on X; $n_. = n_1 + n_0$, or the total number of observations; and u is the ordinate (i.e., vertical height or the probability density) of the standard normal distribution at the cumulative probability of $p_1 = n_1/n_.$ (i.e., the proportion of observations scored 1 on X).

Assuming that the following data set is obtained from 30 examinees who took the Science Aptitude Test, their total test scores and item scores on Item #13 are as shown in the following table.

Treating the total score as the Y variable and the Item #13 score as the X variable, one calculates $\bar{Y}_0 = 12.93$, $\bar{Y}_1 = 29$, $S_Y = 11.92$, $n_0 = 15$, $n_1 = 15$, $n_. = 15 + 15 = 30$, and $p_1 = 15/30 = 0.50$. Taking p_1 to the standard normal distribution, one determines the ordinate, u, to be 0.3989. Substituting these values into the formula above, one obtains the biserial correlation as follows:

$$\begin{aligned} r_{\text{bis}} &= \frac{29.00 - 12.93}{11.92}\left[\frac{(15)(15)}{(0.3989)(30)^2}\right] \\ &= (1.3482)(0.6267) = 0.8449 \approx 0.85. \end{aligned}$$

The above value for biserial correlation can be interpreted as a measure of the degree to which the total score (continuous variable) differentiates between correct and incorrect answers on Item #13 (artificially dichotomous variable). The value of 0.85 suggests that the total score correlates strongly with the performance on this item. If the 15 students who failed this item had scored the lowest on the total test, the biserial correlation would have been 1.00 (the maximum). Conversely, if "pass" and "fail" on Item #13 were matched randomly with total scores, the biserial correlation would be 0.00 (the absolute minimum).

The biserial correlation is an estimate of the product-moment correlation, namely, Pearson's *r,* if the normality assumption holds for X and Y population distributions. Its theoretical range is from -1 to $+1$. In social sciences research, r_{bis} computed from empirical data may be outside the theoretical range if the normality assumption is violated or when $n_.$ is smaller than 15. Thus, r_{bis} may be interpreted as an estimate of Pearson's r only if (a) the normality assumption is not

Examinee	*Item #13 score (X)*	*Total score (Y)*	*Examinee*	*Item #13 score (X)*	*Total score (Y)*	*Examinee*	*Item #13 score (X)*	*Total score (Y)*
1	0	8	11	1	14	21	0	28
2	0	12	12	1	13	22	1	33
3	0	6	13	0	10	23	1	32
4	0	12	14	0	9	24	1	32
5	0	8	15	0	8	25	1	33
6	0	8	16	1	33	26	0	34
7	0	8	17	0	28	27	1	35
8	0	11	18	1	29	28	1	34
9	1	13	19	1	30	29	1	38
10	0	4	20	1	29	30	1	37

violated, (b) the sample size is large (at least 100), and (c) the value of p_1 is not markedly different from 0.5. Biserial correlation should not be used in secondary analyses such as regression.

—Chao-Ying Joanne Peng

See also CORRELATION

REFERENCE

Glass, G. V, & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

BIVARIATE ANALYSIS

Bivariate analysis is a statistical analysis of a pair of variables. Such analysis can take the form of, for example, CROSS-TABULATION, SCATTERPLOT, CORRELATION, ONE-WAY ANOVA, or simple REGRESSION. It is a step often taken after a UNIVARIATE ANALYSIS of the DATA but prior to a multivariate analysis.

—Tim Futing Liao

BIVARIATE REGRESSION. *See* SIMPLE CORRELATION (REGRESSION)

BLOCK DESIGN

Block design is an experimental or QUASI-EXPERIMENTAL design in which units of study are grouped into categories or "blocks" in order to control undesired sources of variation in observed measurements or scores. An undesired source of variation is caused by nuisance variables that affect observed measurements or scores of observations but are not the main interest or focus of a researcher. Nuisance variables are also called concomitant variables.

For example, in studying the effect of different treatments on depression, a researcher may wish to control the severity of a patient's depression. Thus, in this study, the severity of depression is the nuisance variable. Levels of severity of depression are "blocks." Let's further assume that there are three treatments for depression and five levels of the severity of depression. The design layout looks as in the table below.

The statistical model assumed for a block design is

$$Y_{ij} = \mu + \alpha_i + \pi_i + \varepsilon_{ij},$$
$$(i = 1, \ldots, n;\ \text{and}\ j = 1, \ldots, p),\ \text{where}$$

- Y_{ij} is the score of the ith block and the jth treatment condition
- μ is the grand mean, therefore, a constant, of the population of observations
- α_j is the treatment effect for the jth treatment condition; algebraically, it equals the deviation of the population mean ($\mu_{.j}$) from the grand mean (μ or $\mu_{..}$). It is a constant for all observations' dependent score in the jth condition, subject to the restriction that all α_j sum to zero across all treatment conditions as in a fixed-effects design, or the sum of all α_js equals zero in the populations of treatments, as in the random effects model
- π_i is the block effect for population i and is equal to the deviation of the ith population mean (μ_i) from the grand mean($\mu_{..}$). It is a constant for all observations' dependent score in the ith block, normally distributed in the underlying population
- ε_{ij} is the random error effect associated with the observation in the ith block and jth treatment condition. It is a random variable that is normally distributed in the underlying population of the ith block and jth treatment condition and is independent of π_i.

	Treatment 1	*Treatment 2*	*Treatment 3*	*Row mean*
Block 1 = most severe	Y_{11} (score)	Y_{12}	Y_{13}	$\bar{Y}_{1.}$
Block 2	Y_{21}	Y_{22}	Y_{23}	$\bar{Y}_{2.}$
Block 3	Y_{31}	Y_{32}	Y_{33}	$\bar{Y}_{3.}$
Block 4	Y_{41}	Y_{42}	Y_{43}	$\bar{Y}_{4.}$
Block 5 = least severe	Y_{51}	Y_{52}	Y_{53}	$\bar{Y}_{5.}$
Column mean	$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	$\bar{Y}_{.3}$	Grand mean = $\bar{Y}_{..}$

A block design is also called *randomized block design* if observations are first randomly assigned to treatment conditions within each block. The procedure by which observations are assigned to treatment conditions is called "blocking." There are various methods by which blocks are formed or observations are blocked on the nuisance variable. These methods attempt to achieve the goal of (a) including equal segments of the range of the nuisance variable in each block, or (b) including equal proportion of the observation population in each block. Most methods try to accomplish the second goal. Using the example above, a researcher may first rank order patients on their level of depression (such as the score on a depression scale). Second, he or she randomly assigns the top three scorers (most severely depressed patients) to the three treatment conditions within the first block. The second block is formed by the second set of three depressed patients who were judged to be less depressed than the top three but more depressed than the rest of sample. The process continues until all blocks are formed.

If more than one observation is assigned to the combination of block and treatment condition, the design is referred to as the *generalized block design.* Alternatively, the same patient (i.e., an observation) can serve as a block and is subject to all treatments in a random sequence. Some authors prefer to call this specific variation of block designs REPEATED-MEASURES DESIGN.

A randomized block design is appropriate for experiments that meet the usual ANOVA assumptions (a correct model, normality, equal variance, and independent random error) and the following additional assumptions:

a. One treatment factor with at least two levels and a nuisance variable, also with two or more levels.

b. The variability among units of analysis within each block should be less than that among units in different blocks.

c. The observations in each block should be randomly assigned to all treatment levels. If each block consists of only one experimental unit that is observed under all treatment levels, the order of presentation of treatment levels should be randomized independently for each experimental unit, if the nature of treatment permits.

In general, there are three approaches to controlling the undesired source of variation in observed measurements:

• First, hold the nuisance variable constant; for example, use only patients who are mildly depressed.

• Second, assign observations randomly to treatment conditions with the assumption that known and unsuspected sources of variation among the units are distributed randomly over the entire experiment and thus do not affect just one or a limited number of treatment conditions in a specific direction. In other words, the principle of randomization will "take care of" the undesired source of variation in replications of the same study in the long run.

• Third, include the nuisance variable as one of the factors in the experiment, such as in block designs or analysis of covariance (ANCOVA). Thus, the main objective of block designs is to better investigate the main effect of the treatment variable (i.e., α_j), not the interaction between the treatment variable and a nuisance (also blocking) variable.

—Chao-Ying Joanne Peng

See also BLOCK DESIGN, EXPERIMENT, NUISANCE PARAMETERS

REFERENCES

Huck, S. (2000). *Reading statistics and research* (3rd ed.). New York: Addison, Wesley, Longman.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Kirk, R. E. (1999). *Statistics: An introduction* (4th ed.). Orlando, FL: Harcourt Brace.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

BLUE. *See* BEST LINEAR UNBIASED ESTIMATOR

BMDP

BMDP is a software package containing 40+ powerful programs for statistical analysis, ranging from simple data description to complex multivariate analysis and time-series analysis. The output produced by BMDP is very descriptive and largely self-explanatory.

—Tim Futing Liao

BONFERRONI TECHNIQUE

The Bonferroni technique is used to hold a familywise TYPE I ERROR rate to a preselected value when a family of related SIGNIFICANCE TESTS (or confidence intervals) is conducted. This situation, sometimes called MULTIPLE COMPARISONS, or, more generally, "simultaneous inference," may arise when statistical tests are carried out on data from the same units (e.g., when people respond to several questionnaire items or when correlations among several variables are calculated on the same people), or when statistical estimates such as means or proportions are used more than once in a set of hypotheses (e.g., when a set of planned or post hoc comparisons are not statistically independent). Although the Bonferroni technique could be applied when the tests are independent, its use in such cases is extremely rare.

The Bonferroni technique has several advantages that make it an attractive alternative to other simultaneous inference procedures. Among these are that it may be applied to any statistical test resulting in a P VALUE, that it is very easy to use, and that it does not result in much wasted power when compared with other simultaneous significance testing procedures.

DEVELOPMENT

The basis of the Bonferroni technique is Bonferroni's inequality: $\alpha_f \leq \alpha_1 + \alpha_2 + \cdots + \alpha_m$. Because it is α_f that the researcher wishes to control, then choosing values for the several α_j terms that make $\alpha_1 + \alpha_2 + \cdots + \alpha_m = \alpha_j$ will result in control over α_f, the familywise Type I error rate. Usually, the several α_j values are set equal to each other, and each test is run at $\alpha_t = \alpha_f/m$. The value of α_f is therefore, at most, $m\alpha_t$.

APPLICATION

In order to apply the Bonferroni technique, we must have a family (i.e., set) of m (at least two) significance tests and will define two significance (α) levels: the per-test Type I error rate and the familywise Type I error rate. The per-test Type I error rate, α_t, is the significance level of each statistical test (i.e., the null hypothesis is rejected if the p value of the test falls below α_t). The familywise Type I error rate, α_f, is the probability that one or more of the m tests in the family is declared significant as a Type I error (i.e., when its null hypothesis is true).

Use of the Bonferroni technique involves adjusting the per-test Type I error rate such that $\alpha_t = \alpha_f/m$. In practice, the familywise Type I error rate is usually held to .05. Thus, α_t is usually set at $.05/m$.

Although tables of these unusual percentiles of various distributions (e.g., t, chi-square) exist, in practice, the Bonferroni technique is most easily applied using the p values that are standard output in statistical packages. The p value of each statistical test is compared with the calculated α_t and if $p \leq \alpha_t$, the result is declared statistically significant; otherwise, it is not.

EXAMPLE

A questionnaire with four items has been distributed to a group of participants who have also been identified by gender. The intent is to test, for each item, whether there is a difference between the distributions of males and females. CHI-SQUARE TESTS based on the four two-way tables (gender by item response) will be used. However, if each test were conducted at the $\alpha = .05$ level, the probability of at least one of them reaching statistical significance by chance (assuming all null hypotheses true) is greater than .05 because there are four opportunities for a Type I error, not just one. In this example, there are $m = 4$ significance tests, so instead of comparing the p value for each test with .05, it is compared with $.05/4 = .0125$.

Similarly, if each of four experimental group means is to be compared with that of a control group, the Bonferroni technique could be used. Each of the four tests (t or F) would be accomplished by comparing the p value of the resulting statistic with .0125. Although the tests are correlated because they share a mean (that of the control groups), the familywise Type I error rate would be controlled to at most .05.

—William D. Schafer

BOOTSTRAPPING

Bootstrapping is a computer intensive, nonparametric approach to STATISTICAL INFERENCE. Rather than making assumptions about the SAMPLING DISTRIBUTION of a statistic, bootstrapping uses the variability within a sample to estimate that sampling distribution empirically. This is done by randomly resampling with replacement from the sample many, many times in a way that mimics the original sampling scheme. There are various approaches to constructing CONFIDENCE INTERVALS with this estimated sampling distribution to make statistical inferences.

The goal of inferential statistical analysis is to make PROBABILITY statements about a POPULATION parameter, θ, from a STATISTIC, $\hat{\theta}$, calculated from sample data drawn from a population. At the heart of such an inference is the statistic's SAMPLING DISTRIBUTION, which is the range of values it could take on in a random sample from a given population and the probabilities associated with those values. In the standard parametric inferential statistics that social scientists learn in graduate school (with the ubiquitous *T*-TESTS and *P* VALUES), we get information about a statistic's sampling distribution from mathematical analysis. For example, the CENTRAL LIMIT THEOREM gives us good reason to believe that the sampling distribution of a sample mean is normal, with an expected value of the population mean and a STANDARD DEVIATION of approximately the standard deviation of the variable in the population divided by the square root of the sample size. However, there are situations where either no such parametric statistical theory exists or the assumptions needed to apply it do not hold. In these cases, one may be able to use bootstrapping to make a probability-based inference to the population parameter.

Bootstrapping is a general approach to statistical inference that can be applied to virtually any statistic. The basic procedure has two steps: (a) estimating the statistic's sampling distribution through resampling, and (b) using this estimated sampling distribution to construct confidence intervals to make inferences to population parameters.

First, a statistic's sampling distribution is estimated by treating the sample as the population and conducting a form of MONTE CARLO SIMULATION on it. This is done by randomly resampling with replacement a large number of samples of size n from the original sample of size n. Replacement sampling causes the resamples to be similar to, but slightly different from, the original sample because an individual case in the original sample may appear once, more than once, or not at all in any given resample.

Resampling should be conducted to mimic the sampling process that generated the original sample. Any stratification, WEIGHTING, clustering, stages, and so forth used to draw the original sample need to be used to draw each resample. In this way, the random variation that was infused into the original sample will be infused into the resamples in a similar fashion. The ability to make inferences from complex RANDOM SAMPLES is one of the central advantages of bootstrapping over parametric inference. In addition to mimicking the original sampling procedure, resampling ought to be conducted only on the random component of a statistical model. For example, this may be the raw data for simple descriptive statistics and survey samples or the ERROR term for a REGRESSION model.

For each resample, one calculates the sample statistic to be used in inference, $\hat{\theta}^*$. Because each resample is slightly and randomly different from each other resample, these $\hat{\theta}^*$s will also be slightly and randomly different from one another. The central assertion of bootstrapping is that a relative FREQUENCY DISTRIBUTION of these $\hat{\theta}^*$s is an estimate of the sampling distribution of $\hat{\theta}$, given the sampling procedure used to derive the original sample being mimicked in the resampling procedure.

To illustrate the effect of resampling, consider the simple example in Table 1. The original sample was drawn as a simple random sample from a standard normal distribution. The estimated mean and standard deviation vary somewhat from the population parameters (0 and 1, respectively) because this is a random sample. Note several things about the three resamples. First, there are no values in these resamples that do not appear in the original sample, because these resamples were generated from the original sample. Second, due to resampling with replacement, not every value in the original sample is found in each resample, and some of the original sample values are found in a resample more than once. Third, the sample statistics estimated from the resamples (in this case, the means and standard deviations) are close to, but slightly different from, those of the original sample. The relative frequency distribution of these means (or standard deviations or any other statistic calculated from these resamples) is the bootstrap estimate of the sampling distribution of the population parameter.

How many of these resamples and $\hat{\theta}^*$s are needed to generate a workable estimate of the sampling distribution of $\hat{\theta}$? The estimate is asymptotically unbiased, but

Table 1 Example of Bootstrap Resamples

Case Number	*Original Sample* [$N(0,1)$]	*Resample #1*	*Resample #2*	*Resample #3*
1	0.697	−0.270	−1.768	−0.270
2	−1.395	0.697	−0.152	−0.152
3	1.408	−1.768	−0.270	−1.779
4	0.875	0.697	−0.133	2.204
5	−2.039	−0.133	−1.395	0.875
6	−0.727	0.587	0.587	−0.914
7	−0.366	−0.016	−1.234	−1.779
8	2.204	0.179	−0.152	−2.039
9	0.179	0.714	−1.395	2.204
10	0.261	0.714	1.099	−0.366
11	1.099	−0.097	−1.121	0.875
12	−0.787	−2.039	−0.787	−0.457
13	−0.097	−1.768	−0.016	−1.121
14	−1.779	−0.101	0.739	−0.016
15	−0.152	1.099	−1.395	−0.27
16	−1.768	−0.727	−1.415	−0.914
17	−0.016	−1.121	−0.097	−0.860
18	0.587	−0.097	−0.101	−0.914
19	−0.270	2.204	−1.779	−0.457
20	−0.101	0.875	−1.121	0.697
21	−1.415	−0.016	−0.101	0.179
22	−0.860	−0.727	−0.914	−0.366
23	−1.234	1.408	−2.039	0.875
24	−0.457	2.204	−0.366	−1.395
25	−0.133	−1.779	2.204	−1.234
26	−1.583	−1.415	−0.016	−1.121
27	−0.914	−0.860	−0.457	1.408
28	−1.121	−0.860	2.204	0.261
29	0.739	−1.121	−0.133	−1.583
30	0.714	−0.101	0.697	−2.039
Mean	−0.282	−0.121	−0.361	−0.349
Standard deviation	1.039	1.120	1.062	1.147

under what conditions is the variance small enough to yield inferences that are precise enough to be practical? There are two components to this answer. First, the asymptotics of the proof of unbiasedness for the bootstrap estimate of the sampling distribution appears to require an original sample size of 30 to 50, in terms of degrees of freedom. Second, the number of resamples needed to flesh out the estimated sampling distribution appears to be only about 1,000. With high-powered personal computers, such resampling and calculation requires a trivial amount of time and effort, given the ability to write an appropriate looping ALGORITHM.

After one estimates the sampling distribution of $\hat{\theta}$ with this resampling technique, the next step in bootstrap statistical inference is to construct confidence intervals using this estimate. There are several ways to do this, and there has been some controversy as to which confidence interval approach is the most practical and statistically correct. Indeed, much of the discussion of the bootstrap in the 1980s and 1990s was devoted to developing and testing these approaches, which are too complicated to discuss in this article. The references listed below give details and instructions on these confidence interval approaches.

DEVELOPMENT

Bootstrapping is of a fairly recent vintage, with its genesis in a 1979 article by Bradley Efron in the *Annals of Statistics*, titled, "Bootstrap Methods: Another Look at the Jackknife." But as this title suggests, although Efron is widely regarded as the father of the bootstrap, the idea of assessing intrasample variability to make nonparametric statistical inferences is not new. Statisticians have long understood the limitations of parametric inference and worked to overcome them. Although extremely powerful when appropriate, parametric statistical inference can be applied confidently to only a fairly narrow range of statistics in very specific circumstances. Ronald Fisher's exact test is one old way to make inferences without parametric assumptions, but it is limited in that it requires a complete delineation of all permutations of a given population. This is clearly not practical in most modern social science situations. Several techniques were discussed in the 1950s and 1960s that used the variability of a statistic in various subsamples to assess overall sampling variability. The most general of these approaches was jackknifing, which systematically eliminated cases from a sample one at a time and assessed the impact of this casewise elimination on a statistic's variability. But it was not until the advent of high-powered personal computers that Efron could take the leap to using unlimited numbers of randomly developed subsamples to estimate a statistic's sampling distribution, and thus the bootstrap was born.

APPLICATION

The two situations where bootstrapping is most likely to be useful to social scientists are (a) when making inferences using a statistic that has no strong parametric theory associated with it; and (b) when making inferences using a statistic that has strong parametric theory under certain conditions, but these conditions do not hold. For example, the bootstrap might be used in the former case to make inferences from indirect effects of path models, EIGENVALUES, the switch point in a switching regression, or the difference between two medians. This use of the bootstrap may become especially important as greater computer power and statistical sophistication among social scientists lead to the development of estimators designed to test specific hypotheses very precisely, but without any clear understanding of the properties of their sampling distributions. The second use of the bootstrap may be important as a check on the robustness of parametric statistical tests in the face of assumption violations. For example, because the asymptotic parametric theory for MAXIMUM LIKELIHOOD ESTIMATION (MLEs) appears to hold reasonably well only as sample sizes approach 100, the standard t-tests on MLEs may be inaccurate with a sample of, say, 75 cases. Bootstrap statistical inference can be used in such a situation to see if this potential inaccuracy is a problem.

—Christopher Z. Mooney

REFERENCES

Chernick, M. R. (1999). *Bootstrap methods: A practioner's guide.* New York: Wiley-Interscience.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge, UK: Cambridge University Press.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7,* 1–26.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference.* Thousand Oaks, CA: Sage.

BOUNDED RATIONALITY

Bounded rationality is a term used to describe a decision-making process. It is less a unified theory, and more a recognition that decision making in practice often does not conform to the concept of rationality that is the foundation of many formal models of behavior. The term was introduced by Herbert Simon in the 1950s (e.g., Simon, 1957) and is an active area of research today (e.g., Rubenstein, 1998). It is important to note that bounded rationality is not irrationality. Under bounded rationality, actors are still trying to make good decisions, but their decision-making process does not meet all of the criteria of full rationality.

Bounded rationality is perhaps best understood by comparison to the fully rational actors assumed in most formal models of behavior. Fully rational actors always make the best decision possible given the information available. Fully rational actors have clearly defined preferences, use all relevant information in an unbiased way, and are able to solve difficult optimization problems without error. Although most researchers who assume full rationality are aware that this is not a completely accurate description of decision making in reality, they often assume actors behave "as if" they were fully rational.

However, those who study bounded rationality have found many examples where actors do not behave "as if" they were fully rational. In reality, actors may not be capable of making the complex computations required for fully rational behavior. For example, in an infinitely repeated game, full rationality requires that actors look an infinite number of periods into the future to plan their actions today, yet this most likely exceeds the computational abilities of even the most capable organizations or individuals. Under bounded rationality, actors still try to plan for the best actions today but are not capable of the complex computations undertaken by fully rational actors. Perhaps the most common distinction between bounded and unbounded rationality is due to Simon, who holds that under full rationality, actors optimize, finding the best solution to a problem, whereas under bounded rationality, actors often satisfice, adopting a solution that is "close enough."

Other examples of boundedly rational behavior have been documented, such as a tendency to simplify problems and framing effects. Fully rational actors are indifferent to logically equivalent descriptions of alternatives and choice sets. However, in a well-known EXPERIMENT, Tversky and Kahneman (1986) demonstrated that the framing of a decision problem can have an impact on the choices individuals make. Subjects in the experiment were told that an outbreak of a disease will cause 600 people to die in the United States. The subjects were divided into two groups. The first group of subjects was asked to choose between two mutually exclusive treatment programs (A and B) that would yield the following results:

A. Two hundred people will be saved.
B. Six hundred people will be saved with probability 1/3; nobody will be saved with probability 2/3.

The second group of subjects was asked to choose between two mutually exclusive programs (C and D) with the following results:

C. Four hundred people will die.
D. Nobody will die with probability 1/3; all six hundred will die with probability 2/3.

It is clear that A is equivalent to C, and B is equivalent to D. However, 72% of individuals in the first group chose A, while 78% of individuals in the second group chose D. This result is more evidence that there are differences between actual behavior and the assumptions of full rationality, and thus, many actors in decision-making problems are best described as boundedly rational.

—Garrett Glasgow

REFERENCES

Rubenstein, A. (1998). *Modeling bounded rationality.* Cambridge: MIT Press.
Simon, H. A. (1957). *Models of man.* New York: Wiley.
Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* 251–278.

BOX-AND-WHISKER PLOT.

See BOXPLOT

BOX-JENKINS MODELING

The methodology associated with the names of Box and Jenkins applies to TIME-SERIES data, where observations occur at equally spaced intervals (Box & Jenkins, 1976). Unlike DETERMINISTIC MODELS, Box-Jenkins models treat a time series as the realization of a STOCHASTIC process, specifically as an ARIMA (autoregressive, integrated, moving average) process. The main applications of Box-Jenkins models are forecasting future observations of a time series, determining the effect of an intervention in an ongoing time series (INTERVENTION ANALYSIS), and estimating dynamic input-output relationship (transfer function analysis).

ARIMA MODELS

The simplest version of an ARIMA model assumes that observations of a time series (z_t) are generated by random shocks (u_t) that carry over only to the next time point, according to the moving-average parameter (θ). This is a first-order moving-average model, MA(1), which has the form (ignore the negative sign of the parameter)

$$z_t = u_t - \theta u_{t-1}. \quad (1)$$

This process is only one short step away from pure randomness, called "white noise." The memory today (t) extends only to yesterday's news ($t-1$), not to any before then. Even though a moving-average model can be easily extended, if necessary, to accommodate a larger set of prior random shocks, say u_{t-2} or u_{t-3}, there soon comes the point where the MA framework proves too cumbersome. A more practical way of handling the accumulated, though discounted, shocks of the past is by way of the AUTOREGRESSIVE model. That takes us to the "AR" part of ARIMA:

$$z_t = \phi z_{t-1} + u_t. \tag{2}$$

In this process, AR(1), the accumulated past (z_{t-1}), not just the most recent random shock (u_{t-1}), carries over from one period to the next, according to the autoregressive parameter (ϕ). But not all of it does. The autoregressive process requires some leakage in the transition from time $t-1$ to time t. As long as ϕ stays below 1.0, the AR(1) process meets the requirements of stationarity (constant mean level, constant variance, and constant covariance between observations).

Many time series in real life, of course, are not stationary. They exhibit trends (long-term growth or decline) or wander freely, like a RANDOM WALK. The observations of such time series must first be transformed into stationary ones before autoregressive and/or moving-average components can be estimated. The standard procedure is to difference the time series z_t,

$$\nabla z_t = z_t - z_{t-1}, \tag{3}$$

and to determine whether the resulting series (∇z_t) meets the requirements of stationarity. If so, the original series (z_t) is considered integrated at order 1. That is what the "I" in ARIMA refers to. I simply counts the number of times a time series must be differenced to achieve stationarity. One difference $(I = 1)$ is sufficient for most nonstationary series.

The analysis of a stationary time series then proceeds from model identification, through parameter estimation, to diagnostic checking. To identify the dynamic of a stationary series as either autoregressive (AR) or moving average (MA), or as a combination of both—plus the order of those components—one examines the AUTOCORRELATIONs and partial autocorrelations of the time series up to a sufficient number of lags. Any ARMA process leaves its characteristic signature in those correlation functions (ACF and PACF). The parameters of identified models are then estimated through iterative MAXIMUM LIKELIHOOD ESTIMATION procedures. Whether or not this model is adequate for capturing the ARMA dynamic of the time series depends on the error diagnostic. The Ljung-Box Q-statistic is a widely used summary test of white-noise residuals.

FORECASTING

Box-Jenkins models were designed to avoid the pitfalls of forecasting with deterministic functions (extrapolating time trends). An ARIMA forecast is derived from the stochastic dynamic moving a time series. Once the parameters of the model have been obtained, no other data are required to issue the forecast. Forecasts come automatically with the model estimates, and they arrive before the event to be predicted takes place. Hence, ARIMA forecasts are true ex-ante, unconditional forecasts that incorporate all the information about the phenomenon's dynamic behavior. For example, an AR(2) model of the U.S. presidential vote, estimated with aggregate results from 1860 to 1992, was able to forecast quite early the Democratic victory in 1996 (Norpoth, 1995).

INTERVENTION EFFECTS

There are many forms of behavior and performance that government intends to change by adopting policies. Wherever those outcomes are measured frequently, Box-Jenkins models are able to deliver statistical estimates of the effect of such interventions (Box & Tiao, 1975; Clarke, Norpoth, & Whiteley, 1998). The basic form of an intervention model, say for a policy aimed at reducing traffic fatalities, is as follows:

$$y_t = [\omega/(1-\delta B)]\, I_t + N_t, \tag{4}$$

where

y_t = fatalities in the tth month,
I_t = the intervention variable (1 for month of policy adoption, and 0 elsewhere),
ω = the step change associated with the intervention variable,
δ = rate of decay of the step change,
B = backward-shift operator (such that, for example, $BI_t = I_{t-1}$),
N_t = "noise" modeled as an ARIMA process.

This model does not require the impact of the intervention to be durable, nor that the impact is gone the moment the intervention is over. By modeling the time series as an ARIMA process, one avoids confounding the alleged intervention impact with many other possible effects. Such a randomization precludes mistaking a trend for an intervention effect.

TRANSFER FUNCTIONS

Instead of a discrete intervention changing the course of a time series, consider another time series continuously affecting the ups and downs of the time series of interest. This influence need not be instant, but may take time to materialize fully. This can be expressed with a transfer function model (Box & Jenkins, 1976):

$$y_t = [\omega/(1 - \delta B)]x_t + e_t. \qquad (5)$$

In this model, ω and δ describe an influence constantly at work between X and Y. The moment X moves, Y does, too, with a magnitude set by the ω parameter. Then, a moment later, that movement of X triggers yet a further reaction in Y, according to the δ parameter, and so on. To identify a dynamic model of this sort for variables that are each highly autocorrelated, Box and Jenkins (1976) introduced the following strategy: "Prewhiten" the X series by removing its ARIMA dynamic; then, filter the Y series accordingly; and then obtain cross-correlations between the transformed time series. This is a drastic strategy that may end up "throwing the baby out with the bath water." Caution is advised.

Although Box-Jenkins models cover a wide array of time-series problems, this is not a methodology where something always shows up. The user faces the risk of coming up empty-handed. Time series is a domain where near-perfect correlations can dissolve into randomness in a split second. Difference a random walk, where successive observations are perfectly correlated, and you get white noise, where the correlation is zero.

—Helmut Norpoth

REFERENCES

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association, 70,* 70–79.

Clarke, H. D., Norpoth, H., & Whiteley, P. F. (1998). It's about time: Modeling political and social dynamics. In E. Scarborough & E. Tanenbaum (Eds.), *Research strategies in the social sciences* (pp. 127–155). Oxford, UK: Oxford University Press.

Granger, C. W. J., & Newbold, P. (1986). *Forecasting economic time series* (2nd ed.). New York: Academic Press.

Mills, T. C. (1990). *Time series techniques for economists.* Cambridge, UK: Cambridge University Press.

Norpoth, H. (1995). Is Clinton doomed? An early forecast for 1996. *Political Science & Politics, 27,* 201–206.

BOXPLOT

Boxplots were developed by Tukey (1977) as a way of summarizing the midpoint and range of a distribution and comparing it with other distributions, or comparing the distributions of a given variable grouped by another, categorical, variable. The size of the box indicates the range of the middle 50% of values of the chosen variable, that is, the range from the 25th to the 75th quartile, or, as it is also known, the INTERQUARTILE RANGE. A line across the box represents the MEDIAN value. *Whiskers* extend from the box to reach the furthest values at either end, excluding outliers. That is, they reach to the furthest values that fall from either end of the box within 1.5 times the size of the box. These limits to the length of the whiskers are known as *fences.* The furthest values reached by the whiskers will not necessarily reach to the fences. Whether they do or not will depend on the shape of the distribution and the amount of dispersion. OUTLIERS are considered to be values that fall beyond the fences and may be noted separately and labeled individually. *Extremes* are an extreme version of outliers that fall further than three times the size of the box from either end of the box. Again, they may be noted and labeled individually. Thus, boxplots provide, in condensed, graphical form, a large amount of key information about a distribution and are particularly useful for identifying its extremes and outliers.

Figure 1 provides an example of a boxplot comparing the earnings of men and women in Britain in 1999. The boxplot indicates the shape of the earnings distribution. We can see that it is skewed to the right (it clusters closer to 0) for both men and women. However, the median for women falls lower in the box, indicating a greater clustering of lower incomes. In addition, the whole distribution is situated lower for women, indicating that their earnings are systematically lower. There are a number of outliers (labeled by their case number), showing that a few people have

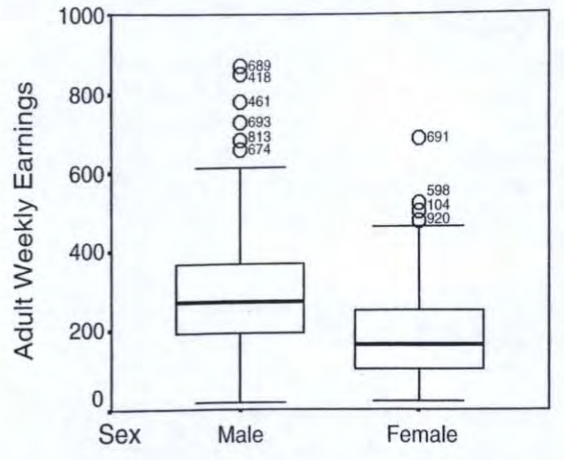

Figure 1 Boxplot of Male and Female Weekly Earnings

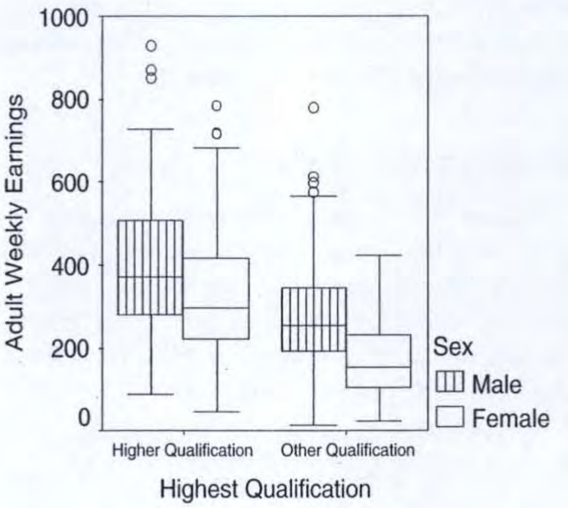

Figure 2 Boxplot of Adult Weekly Earnings by Sex and Highest Qualification, Britain 1999

very high earnings. It is interesting to note, though, that most of the female outliers fall within the fence of the male distribution.

Figure 2 is an example of a clustered bar chart, which breaks down the information in Figure 1 by education level, but supplying all the same information as in Figure 1 for each subcategory. It explores whether the patterns illustrated in Figure 1 are replicated when observations are clustered by sex *within* education categories. It shows how compressed the earnings distribution is for women without a higher qualification and the absence of outliers for this group. It also shows how the interquartile range narrows steadily across the four groups, indicating most heterogeneity in earnings among highly qualified men and most homogeneity among less well-qualified women. It demonstrates the power of graphical representation not only to convey essential information about our data, but also to encourage us to think about it in an exploratory way.

—Lucinda Platt

REFERENCES

Marsh, C. (1988). *Exploring data: An introduction to data analysis for social scientists.* Cambridge, UK: Polity.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

C

CAIC. *See* GOODNESS-OF-FIT MEASURES

CANONICAL CORRELATION ANALYSIS

The basic idea of canonical correlation analysis is clearly illustrated by Russett (1969), who analyzed the ASSOCIATION between some economic and political characteristics of 47 countries. Five indicators were used for measuring economic inequality: the division of farmland, the GINI COEFFICIENT, the percentage of tenant farmers, the gross national product (GNP) per capita, and the percentage of farmers. Russett measured political instability with four indicators: the instability of leadership, the level of internal group violence, the occurrence of internal war, and the stability of democracy. The research hypothesis was that Alexis de Tocqueville was right: There is not one nation that is capable of maintaining a democratic form of government for an extensive period of time if economic resources are unevenly distributed among its citizens. In other words, there is a significant association between the economic *in*equality and the political *in*stability of nations.

The two theoretical concepts in this research problem are "economic inequality" (X^*) and "political instability" (Y^*). They are called canonical variables with an expected high CORRELATION, known as *canonical correlation.* The first canonical variable, X^*, is measured by $p = 5$ indicators, X_1 to X_5, and we will

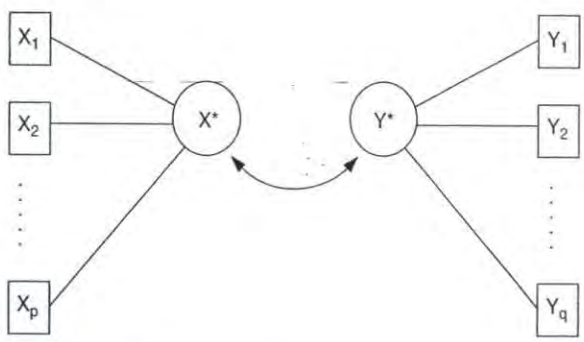

Figure 1 Canonical Correlation

consider X^* as a linear combination (a weighted sum) of these X variables. In an analogous fashion, Y^*, the second canonical variable, is a linear combination of the $q = 4$ indicators, Y_1 to Y_4. In the most general case, in which the X set contains p variables and the Y set q variables, the diagram can be represented as in Figure 1.

In Russett's (1969) research problem, the canonical correlation between the canonical variables X^* and Y^* is in fact a causal relationship because he asserts that nations with greater economic inequality display greater political instability. The arrow in the other direction, according to which the political characteristics would influence the economic characteristics, is not implied in de Tocqeville's or Russett's view. However, considering the fact that a canonical correlation is a "correlation," the statistical analysis is not asymmetrical. For this reason, we do not draw a causal arrow from X^* to Y^* in the figure but a double-sided curved arrow, indicating that the question of CAUSALITY remains open.

THE MODEL OF CANONICAL CORRELATION ANALYSIS

In canonical correlation analysis, we want to see if there is a significant association between a set of X variables and a set of Y variables. For this reason, we look for a linear combination of the X set, $X^* = a_1X_1 + a_2X_2 + \cdots + a_pX_p$ and a linear combination of the Y set, $Y^* = b_1Y_1 + b_2Y_2 + \cdots + b_qY_q$ in such a way that X^* and Y^* are maximally correlated. The two linear combinations, X^* and Y^*, are not observed. In our example, they have been given a name a priori: economic inequality and political instability. Sometimes such an a priori theory is absent at the start of research, and a name must be devised afterwards.

To find these two linear combinations (canonical variables X^* and Y^*), the a and b weights must be calculated. Canonical correlation analysis aims at determining these weights in such a way that the canonical correlation ρ is as high as possible. The square of this canonical correlation is the proportion of variance in one set (e.g., political characteristics) that is explained by the variance in the other set (e.g., economic characteristics).

Just as in PRINCIPAL COMPONENTS ANALYSIS, in which not one but several uncorrelated components can be determined, here various pairs of canonical variables can be found. Each pair is uncorrelated with the preceding pair and is calculated each time so that the canonical correlation is maximal. The number of pairs of canonical variables or, which comes down to the same, the number of canonical correlations is equal to the number of variables in the smallest set, that is, $\min[p, q]$.

The calculation of each of the canonical correlations and of the a and b weights will be done by examining the eigenstructure of a typical matrix. In DISCRIMINANT ANALYSIS and principal components analysis, this typical matrix is $\mathbf{W}^{-1}\,\mathbf{B}$ and $\mathbf{R}$, respectively. In canonical correlation analysis this matrix is $\mathbf{R}_{yy}^{-1}\mathbf{R}'_{xy}\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}$. In this matrix notation, a distinction is made between different types of within- and between-correlation matrices. Between correlations are calculated ($\mathbf{R}'_{xy}$ and $\mathbf{R}_{xy}$), and a control for within correlations is built in ($\mathbf{R}_{yy}^{-1}$ and $\mathbf{R}_{xx}^{-1}$). We will now explain this in more detail.

THREE TYPES OF CORRELATIONS

A legitimate method for examining the correlations between the economic and political characteristics is an analysis of the separate correlations between, for example, the Gini index and the (in)stability of leadership (X_1, Y_1) or between the percentage of farmers and the level of violence (X_2, Y_2). These are the correlations between X and Y variables and are therefore called *between correlations.*

Canonical correlation analysis, however, goes a step further than this examination of separate correlations.

First of all, in canonical correlation analysis, one aims to examine to what degree two sets—X variables on one hand and Y variables on the other—show association "as a whole." To this end, each set is replaced by a linear combination. The correlation between these two linear combinations is the objective of the research.

Second, in canonical correlation analysis, one controls for the associations within each of the sets, as it is possible that the variables within the X set of economic characteristics are also correlated in pairs. These types of correlations are called *within correlations.* They form a problem that is comparable to multicollinearity in multiple regression analysis. There, the regression coefficients that are calculated are partial coefficients (i.e., we control each variable for association with the other variables in the X set). Something similar is done in canonical correlation analysis, only here we do not just control on the side of the X set but on the side of the Y set of political characteristics as well because we now have various Y variables that could also be correlated in pairs.

We can, therefore, distinguish three types of correlations:

1. Correlations between X variables; the correlation matrix is $\mathbf{R}_{xx}$.
2. Correlations between Y variables; the correlation matrix is $\mathbf{R}_{yy}$.
3. Correlations between X and Y variables; the correlation matrix is $\mathbf{R}_{xy} = \mathbf{R}'_{yx}$.

The fact that canonical correlation analysis aims to examine the between correlations and not the within correlations can be illustrated more clearly by comparing this technique to double principal components analysis (PCA twice). PCA looks for the correlations within one set of variables (i.e., within correlations). If we were to perform two PCAs, one on the X set and one on the Y set, we would obtain a first principal component in each PCA. The correlation between these two principal components is absolutely not the same as the (first) canonical correlation because the correlation between the two components measures to what degree the within associations in the set of economic characteristics are comparable to the within associations in the set of political characteristics. Canonical correlation, on the other hand, measures the degree to which the economic and political characteristics are correlated, controlling for (i.e., assuming the absence of) within associations within the economic set, on one hand, and within the political set, on the other.

GEOMETRIC APPROACH

To explain canonical correlation from a geometric viewpoint, we create two coordinate systems: one with the x variables as axes and one with the y variables as axes. Canonical correlation analysis aims at finding a first pair of canonical variables, one X^* in the space of X variables and one Y^* in the space of Y variables, in such a way that the correlation between X^* and Y^* is maximal. Afterwards, one looks for a second pair of canonical variables, uncorrelated with the first pair, in such a way that the (canonical) correlation between this second pair is as high as possible. The high canonical correlations between the first pair, X_1^* and Y_1^*, as well as the second pair, X_2^* and Y_2^*, can be seen by looking at the coordinate systems in which the canonical variables are axes. The fact that canonical variables of different pairs—for example, X_1^* and X_2^* (and also Y_1^* and Y_2^*)—are uncorrelated does not necessarily mean that they are orthogonal in the space of x variables. It does mean that the regression line in (X_1^*, X_2^*) space has a zero slope.

OBJECTIVES OF THE TECHNIQUE

Three goals are pursued in canonical correlation analysis. With Russett's (1969) study in mind, these goals can be described as the following:

1. We look for a first pair of linear combinations, one based on the set of economic variables and the other based on the set of political variables, in such a way that the correlation between both of the linear combinations is maximal. Next we look for a second pair of linear combinations, also maximally correlated and uncorrelated with the first pair. We do this as many times as there are variables in the smallest set. We call the linear combinations *canonical variables*. The weights are known as *canonical weights*. The maximized correlations are called *canonical correlations*. On the basis of the canonical weights, we attempt to interpret the associations "between" the sets. One can also attempt to give a name to each canonical variable per pair.

2. Making use of the canonical weights, the scores of the canonical variables are calculated. The correlations between the original variables and the canonical variables often offer additional possibilities for interpretation. These are called *structure correlations*.

3. We check to see whether the associations between the two sets can be generalized to the population of the countries. This is done by testing the canonical correlations for significance. We can perform this test for all of the canonical correlations together or for each of the separate canonical correlations or their subgroups. Because the first canonical correlation is the largest and the following correlations become smaller and smaller, it seems obvious to test these separately and only to interpret the ones that give us the most significant result.

A CAUSAL APPROACH TO CANONICAL CORRELATION ANALYSIS

In 1968, Stewart and Love developed an asymmetrical version of canonical correlation analysis. The canonical correlations calculated above are, after all, symmetrical coefficients. For example, the square of the first canonical correlation (i.e., the first eigenvalue λ_1), which represents the squared correlation between the first pair of canonical variables, can be interpreted as the proportion of variance in the Y set (political characteristics) that is explained by the X set (economic characteristics). One could, however, just as easily reason the other way round: λ_1 is also the proportion of variance in economic inequality that is explained by political instability. In other words, no distinction is made between independent and dependent variables, between cause and effect.

For this reason, Stewart and Love proposed an *index of redundancy*. This index measures the degree to which the Y set can be reconstructed on the basis of the knowledge of the X set. It appears that this index is equal to the arithmetic mean of all multiple determination coefficients that are obtained by repeated multiple regression analyses with one Y variable on the function of the entire X set.

THE SIGNIFICANCE TESTS

To test whether the canonical correlations are significant, we make use of Wilk's lambda, Λ, a measure—just like Student's t, Hotelling's T, and Fisher's F—that confronts the ratio of the between dispersion and the within dispersion from the sample with this ratio under the null hypothesis. Wilk's Λ is the multivariate analog of this ratio in its most general form, so that t, T, and F are special cases of Λ. This measure is therefore used for more complex multivariate analysis techniques, such as multiple discriminant analysis, multivariate analysis of variance and covariance, and canonical correlation analysis.

In canonical correlation analysis, Wilk's Λ can be calculated as the product of error variances, in which the error variances are determined by the eigenvalues (i.e., the squared canonical correlations and, therefore, the proportions of explained variance) subtracted from 1. Bartlett (1947) constructed a V measure that is distributed as chi-square with pq degrees of freedom, where p and q represent the number of variables of the X set and the Y set:

$$V = -[n - 1 - (p + q + 1)/2] \ln \lambda.$$

When, for pq degrees of freedom, this value is greater than the critical chi-square value that is found under the null hypothesis for a significance level of $\alpha = 0.05$, then there is a significant association between the economic inequality and political instability of countries. The hypothesis of de Tocqueville and Russett is then empirically supported.

The V measure can also be split additively, so that a separate test can be performed as to whether the significant correlation between the economic and political characteristics also holds for the second and following canonical correlations.

EXTENSIONS

Cases can arise in applied multivariate research in which the research problem is composed of more than two sets of characteristics for the same units; for example, next to an economic X set and a political Y set, there is also a Z set of sociocultural characteristics. Generalized canonical correlation is the name given to procedures used to study correlation between three or more sets.

In the foregoing discussion, it was taken for granted that the variables were all measured at the interval or ratio level. It was also assumed that the relationships were all linear. These two requirements can, however, be relaxed. The computer program CANALS performs a canonical analysis, which is nonmetric as well as nonlinear. This promising new development deserves our attention. The interested reader is referred to Gifi (1980).

—Jacques Tacq

REFERENCES

Bartlett, M. S. (1947). Multivariate analysis. *Journal of the Royal Statistical Society, Series B, 9,* 176–197.

Cooley, W., & Lohnes, P. (1971). *Multivariate data analysis.* New York: John Wiley.

Gifi, A. (1980). *Nonlinear multivariate analysis.* Leiden, The Netherlands: Leiden University Press.

Green, P. (1978). *Analyzing multivariate data.* Hinsdale, IL: Dryden.

Russett, B. (1969). Inequality and instability: The relation of land tenure to politics. In D. Rowney & J. Graham (Eds.), *Quantitative history: Selected readings in the quantitative analysis of historical data* (pp. 356–367). Homewood, IL: Dorsey.

Stewart, D. K., & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin, 70,* 160–163.

Van de Geer, J. (1971). *Introduction to multivariate analysis for the social sciences.* San Francisco: Freeman.

CAPI. *See* COMPUTER-ASSISTED DATA COLLECTION

CAPTURE-RECAPTURE

Capture-recapture, also called *mark-recapture,* is a method for estimating the size of a POPULATION as follows: Capture and mark some members of the population. Take a random SAMPLE from the population (recapture), and count the members of the sample with marks. The fraction marked in the sample estimates the fraction marked in the population. The estimated size of the population is the number marked in the population divided by the fraction marked in the sample. For example, suppose that 1,000 members of a population were marked and that in a random sample of 500 members of the population, 200 had marks. Then the 1,000 marked members of the population comprise about $200/500 = 0.4$ of the population, so the population size is about $1{,}000/0.4 = 2{,}500$. Laplace was the first to use capture-recapture to estimate the size of a population—that of France in 1786 (Cormack, 2001). Capture-recapture has several crucial assumptions:

1. The sample (recapture) is a simple random sample from the population. So every population member must have the same chance of being in the sample, and the chance that a given member is in the sample cannot depend on whether the member was marked. This is called the independence assumption; violations cause correlation bias.

2. The population of marked and unmarked individuals is constant between capture and recapture (no births, deaths, immigration, or emigration).

3. The marks are unambiguous.

Refinements of the method try to account for population changes, unequal PROBABILITIES of recapture, time-varying probabilities of recapture, "trap effects," and so on (Cormack, 2001; Otis, Burnham, White, & Anderson, 1978; Pollock, Nichols, Brownie, & Hines, 1990). Capture-recapture is sometimes used to estimate the size of animal populations and has been adapted to estimate other population parameters, such as survival rates (Lebreton, Burnham, Clobert, & Anderson, 1992). Capture-recapture is the basis for a method that has been proposed to adjust the U.S. census for undercount (see CENSUS ADJUSTMENT).

—Philip B. Stark

REFERENCES

Cormack, R. (2001). Population size estimation and capture-recapture methods. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 17). Amsterdam: Elsevier.

Lebreton, J.-D., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs, 62,* 67–118.

Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs, 62,* 1–135.

Pollock, K. H., Nichols, J. D., Brownie, C., & Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs, 107,* 1–97.

CAQDAS (COMPUTER-ASSISTED QUALITATIVE DATA ANALYSIS SOFTWARE)

Computer-assisted qualitative data analysis software (CAQDAS) is a term introduced by Fielding and Lee (1998) to refer to the wide range of software now available that supports a variety of analytic styles in QUALITATIVE RESEARCH. The software does not "do" the analysis; it merely assists. It needs people to read, understand, and interpret the data. An essential part of qualitative data analysis is the effective handling of data, and this requires careful and complex management of large amounts of texts, video, codes, memos, notes, and so on. Ordinary word processors and database programs as well as dedicated text management programs have been used for this to good effect. However, a core activity in much qualitative analysis is CODING: the linking of one or more passages of text (or sequences of video) in one or more different documents or recordings with named analytic ideas or concepts. A second generation of CAQDAS introduced facilities for coding and for manipulating, searching and reporting on the text, images, or video thus coded. Such code-and-retrieve functions are now at the heart of the most commonly used programs.

PROGRAM FUNCTIONS

Most programs work with textual data. Although it is not necessary to transcribe ethnographic notes, narratives, and so forth, it is usually advantageous because most software supports the rapid recontextualization of extracted and/or coded passages by displaying the original text. Most programs support the coding of text, and such codes can be arranged hierarchically, modified, and rearranged as the analysis proceeds. Some software (e.g., MaxQDA, Qualrus, ATLAS.TI, and NVIVO) can work with rich text that includes different fonts, styles, and colors. Several programs support the analysis of digitized audio, images, and video, in some cases allowing the direct coding of digitized material (e.g., HyperRESEARCH, Atlas.ti, CI-SAID, AnnoTape, and the Qualitative Media Analyser). So far, there is no software that can automatically transcribe from audio recordings. Even the best speech recognition software still relies on high-quality recordings and familiarization with one speaker and cannot cope with multiple respondents' voices.

Most software allows the writing of MEMOS, code definitions, FIELDNOTES, and other analytic texts as part of the process of analysis, and some (e.g., NVivo, NUD*IST/N6, Atlas.ti, and Qualrus) allow this text to be linked with data and codes in the project database. A key feature of many programs is the ability to search online text for words and phrases. Typically, such lexical searching allows the marking of all the matching text wherever it is found in the documents. Some programs support the searching of the database using the coded text. The most sophisticated of these allow complex searches involving various Boolean and proximity combinations of words, phrases, coded text, and VARIABLES (e.g., NVivo, MaxQDA, WINMAX, Atlas.ti, NUD*IST/N6, HyperRESEARCH, and Qualrus).

A third generation of software has extended these functions in two ways. Some assist analytic procedures with a variety of facilities to examine features and relationships in the texts. Such programs (Qualrus, HyperRESEARCH, Ethno, and AQUAD Five) are often referred to as theory builders or model builders because they contain various tools that assist researchers to develop theoretical ideas and test a HYPOTHESIS (Kelle, 1995). Another example in some programs is a model-building facility that uses diagrams linked to the main database of codes, text, or images (e.g. Atlas.ti, Qualrus, and NVivo). Second, some programs enable the exchange of data and analyses between collaborating researchers. Some do this by the export and import of data and analyses in a common format such as XML (e.g., Tatoe and Atlas.ti). Others are able to merge project databases from different researchers (e.g., NVivo, Atlas.ti, and NUD*IST/N6).

CHOOSING SOFTWARE

As Weitzman and Miles (1995) have shown, CAQDAS programs differ in facilities and operation. Which program is appropriate will depend partly on practical issues such as funding and available expertise, but the type of analysis needed is also relevant. Some software programs are better for large data sets with structured content and for linking with quantitative data (NUD*IST /N6), but others are better at dealing with digitized video and audio (Qualitative Media Analyser, AnnoTape). Some allow complex, hierarchically and multiply coded text; others allow only simple coding of passages. Analytic style matters too. Styles involving thematic analysis such as GROUNDED THEORY and various kinds of PHENOMENOLOGY are well supported. However, case-based approaches and NARRATIVE, DISCOURSE, and CONVERSATION ANALYSIS are less well supported.

ADVANTAGES AND DISADVANTAGES

Qualitative data tend to be voluminous, and it is all too easy to produce partial and biased analyses. The use of CAQDAS helps, not least because it removes much of the sheer tedium of analysis. It is easier to be exhaustive, to check for NEGATIVE CASES, and to ensure that text has been coded in consistent and well-defined ways. An audit trail of analysis can be kept so that emerging analytic ideas can be recalled and justified. Most programs allow the production of simple statistics so that the researcher can check (and give evidence) whether examples are common or idiosyncratic. In general, the advantage of CAQDAS is not that it can do things, such as working in teams, with large data sets or with audio and video, that could not be done any other way but rather that, without the software, such things are impossibly time-consuming.

Nevertheless, the use of CAQDAS remains contested. Early users felt that they were closer to the words of their respondents or to their field notes when using paper-based analysis rather than computers. It is likely that this was because users were familiar with traditional approaches but new to computers. Now, most software has good facilities for recontextualization, and there is little reason for those using CAQDAS to feel distant from the data. Other critics suggest that CAQDAS reinforces certain biases such as the dominance of grounded theory and the code-and-retrieve approach. Although it is true that much software has been inspired by grounded theory and other approaches that give a primacy to coding text, the programs are now more sophisticated and flexible and have become less dedicated to any one analytic approach. Despite this, there is evidence that most published research describing the use of CAQDAS remains at a very simple level of pattern analysis and shows little evidence of good reliability or validity. As users become more experienced and more researchers use the software, this is likely to change, and innovative use of the software will increase.

—Graham R. Gibbs

REFERENCES

Fielding, N. G., & Lee, R. M. (1998). *Computer analysis and qualitative research.* London: Sage.

Kelle, U. (Ed.). (1995). *Computer-aided qualitative data analysis: Theory, methods and practice.* London: Sage.

Weitzman, E. A., & Miles, M. B. (1995). *Computer programs for qualitative data analysis: A software source book.* Thousand Oaks, CA: Sage.

CARRYOVER EFFECTS

Carryover effects occur when an experimental treatment continues to affect a participant long after the treatment is administered. Carryover effects are ubiquitous because virtually everything that happens to an organism continues to affect the organism in some fashion for some time, like ripples on the surface of a pond radiating from a thrown stone. Some treatments may produce carryover effects that are fairly modest in size and/or fairly limited in duration, but others may produce carryover effects that are substantial and/or sustained.

The extended impact of a prescription drug provides the most accessible example of carryover effects because people are well aware of the possibility of a drug continuing to exert its influence for a substantial period of time after it is administered. In fact, many drug manufacturers advertise claims about the extended effect of their drugs. People are also aware of the potential complications from the interaction of different drugs, leading them to avoid taking one drug while another is still present in their systems.

Researchers can eliminate or minimize the impact of carryover effects in their studies by means of different strategies. The most obvious approach would be to eliminate the impact of carryover effects by using an independent groups design. In such a design, only one treatment is administered to each participant, data are collected, and the participant's role in the study is complete. Thus, any carryover effects that occur would have no impact on the data collected in other conditions of the study (although the treatment may continue to affect the participant).

Carryover effects are of great concern in experiments using REPEATED-MEASURES DESIGNS because participants are exposed to a series of treatments, after each of which data are collected. Thus, carryover effects from an early treatment could continue to affect the participant's responses collected after a subsequent treatment. In such a case, of course, the response to the later treatment cannot be interpreted as unambiguously due to that treatment but is instead driven by the combination of both treatments.

One strategy to minimize carryover effects, and thereby enhance the likelihood that a response is due solely to the immediately preceding treatment, is to allow sufficient time to elapse between treatments. With sufficient posttreatment delay, the participant should return to a pretreatment state before the next treatment is administered. However, it is often difficult to know how long to wait for the carryover effects to subside. Furthermore, if the delay requires the participant to return on subsequent days, attrition may become a problem.

Another essential strategy, which might be used in combination with some delay between treatments, is to counterbalance the order in which participants are exposed to the treatments. COUNTERBALANCING ensures that any carryover effects, as well as any ORDER EFFECTS (e.g., practice or fatigue effects), will fall equally on all conditions, thereby eliminating the confound between treatment and position.

—Hugh J. Foley

REFERENCES

Girden, E. R. (1992). *ANOVA: Repeated measures* (Sage University Paper Series on Quantitative Applications in the Social Science, 07–084). Newbury Park, CA: Sage.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Maxwell, S. E., & Delaney, H. D. (2000). *Designing experiments and analyzing data: A model comparison perspective.* Mahwah, NJ: Lawrence Erlbaum.

CART (CLASSIFICATION AND REGRESSION TREES)

The acronym CART is used in the data-mining industry both as (a) a generic for all methods of growing classification (unsupervised learning) and regression (supervised learning) trees and (b) a specific set of algorithms for growing binary trees introduced in the book by Breiman, Friedman, Olshen, and Stone (1984). The latter rely on sophisticated methods of cross-validation to adjust for problems introduced by the limitation of binary splits, which contrast with multisplit tree-growing algorithms such as CHAID (see description of CHAID).

—Jay Magidson

REFERENCE

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees.* Monterey, CA: Wadsworth.

CARTESIAN COORDINATES

The Cartesian coordinates are composed of either two or three axes (number lines) formed at 90 degrees to one another. In the two-dimensional case, x and y are often used to refer to the two axes, whereas in the three-dimensional case, a third axis, z, is added. CASES representing DATA points are located in the coordinates so that we can know the patterns of distribution of the data and the relationships between and among the variables represented by the axes.

—Tim Futing Liao

CASE

A case is the unit of study. In an aggregate analysis—say, of welfare expenditures in the American states—the case is the state, and there are 50 cases. In a SURVEY, the respondent is the case. For an EXPERIMENT, the subjects are the cases. In general, the SAMPLE size (N) indicates the number of cases, and each unit of that N represents one case.

—Michael S. Lewis-Beck

See also CASE STUDY

CASE CONTROL STUDY

The case control study is a method for estimating how various factors affect the risks of specified outcomes for large and heterogeneous populations. The method was first developed in epidemiology to identify the factors associated with heightened risks of contracting various diseases (Rothman, 1986). However, the method appears well suited for problems in the social sciences, in which researchers often aim to establish how various factors influence the risks of certain events, such as war, for a large number of countries.

Case control methods have not been exploited much in the social sciences, perhaps because their origins in epidemiology have placed them outside most social scientists' training. This is unfortunate because case control studies have several advantages that make them particularly well suited for studying large cross-national data sets.

The method itself is quite simple. When studying rare events such as war or revolution, if one is interested in the *onset* of such conditions from prior periods of peace, then the number of such onsets is relatively few. However, years of peace will be extremely numerous. Typically, a data set of country-years (consisting of X countries observed for Y years) that includes both years of peace, as well as years in which wars or revolutions began, will have hundreds or even thousands of country-years of peace and only a few dozen country-years of event onsets. The case control method seeks to economize on data analysis by treating the thousands of peace country-years as having relatively little new information in each year. Thus, instead of pooling all country-years into the analysis—as is routinely done with most commonly used statistical methods—the case control method selects a RANDOM SAMPLE of the peace years and pools that with the event onset years. The event onset country-years are the "cases," and the peace or nonevent years are the "controls." A profile of independent variables is developed for each of the cases and controls, generally from the year just prior to event onset for cases and for the year selected for the controls. Conventional statistical methods such as LOGISTIC or ORDINARY LEAST SQUARES (OLS) REGRESSION can then be used to seek which variables are more strongly associated with the "cases" relative to the "controls."

The case control method generally reduces a data set of country-year data from thousands of observations to a few hundred. However, if the controls are indeed selected randomly, the efficiency of estimating independent variable effects is virtually the same as from analysis of the entire data set (King & Zeng, 2001).

Researchers working with large cross-national data sets to study wars, revolutions, or other rare events face several major obstacles to valid inferences regarding the effects of specific variables. Two of these are practical difficulties; two others involve more substantive problems. Case control methods help with all of them.

The first practical difficulty is that large cross-national data sets are often only available for certain kinds of data, often demographic and economic factors. Although some data (e.g., on political regime type) are available for many countries and many years, other types of data (e.g., characteristics of the regime leadership or of particular leaders) would have to be laboriously coded by hand to obtain full coverage. For other kinds of political units—such as cities, regions, or provinces—relevant data may be even more limited. Collecting or coding data for thousands of observations might be prohibitive, but doing so for a few hundred cases is often feasible. Thus, use of the case control method allows analysis of more varied and interesting variables than merely those already available with complete coverage of all countries in all years.

The second difficulty also concerns data availability. Even when data sets exist that claim to cover all countries and all years, these in fact often have extensive gaps of missing data or provide data only at 5- or 10-year intervals. Conducting conventional analyses requires filling in values for all countries and all years, thus leading to reliance on varied imputational and interpolation techniques. However, the years of interest to investigators are precisely those just prior to the onset of major disruptive events; it is in such country-years that data are most likely to be incomplete, yet routine interpolation or imputation is most likely to fail. Thus, the population of Nigeria is probably known only to within 20% and that of Canada to less than 1% accuracy. Yet to test the effects of various demographic variables on ethnic conflict, the former case is especially important. By using case control methods to reduce the number of country-years, one does not have to fill in every gap in missing data or to rely on interpolation or imputation for long stretches of data. Researchers can focus on the specific country-years drawn in the case control sample and seek to refine their accuracy on those specific observations.

There are two more substantive issues for which case control methods offer significant advantages. First, when using all country-years in conventional regression analyses, most countries will have long series of consecutive peace years. These create problems of SERIAL CORRELATION. Although econometric methods routinely counter these problems to bring out the impact of time-varying factors for individual countries, there is no good method of doing so for many different countries at once. Because case control methods use only a sample of the peace country-years, they have no such long strings of consecutive peace years in the analysis. Thus, the serial correlation problem does not arise.

Second, when using country-year data sets, using all country-years often involves comparing "apples" to "oranges." That is, most country-years of peace will come from rich, democratic countries, whereas most country-years of war or violence onset will come from poorer, nondemocratic ones. This might simply be because wealth and democracy lead to peace. But it might also be because peaceful but poor and nondemocratic countries happen to be underrepresented in the world as we find it or because rich, democratic countries have some other characteristic (such as being clustered in Europe and North America, far from poorer areas) that conduces to peace. In any event, in using case control methods, one can help resolve such issues by matching controls with individual cases.

For example, one can stratify the random draw of controls to match the controls with the cases by year, region, colonial history, or other factors. This would increase the frequency of underrepresented groups in the data being analyzed and prevent the results from being driven by the chance distribution of countries on certain variables. Matching is particularly appropriate for variables that cannot be changed by policy but that may shift the baseline risks for groups of countries, such as region, colonial history, or time period.

Whether using simple random selection of controls or matching on certain criteria, the number of controls can be the same as the number of cases, or multiple controls can be selected for each case. What matters is having sufficient observations for significant statistical analysis, not a certain ratio of cases to controls. Thus, if one has 500 cases, a 1:1 ratio of cases to controls will provide a large enough N for most purposes; however, if one has only 100 cases, one may wish to select three or four controls for each case to achieve a larger N. Matched case control studies may require even larger numbers of controls to gain sufficient numbers of cases and controls for each stratum on the matched variable.

Epidemiology texts (e.g., Rothman 1986) provide easy guides to the estimation techniques appropriate for use with case control samples. Social scientists should make greater use of these techniques. If they can help medical analysts determine which factors cause cancer or other illnesses in a large population of individuals, they may help social scientists determine which factors cause war or other forms of violent events in a large population of varied societies.

—Jack A. Goldstone

REFERENCES

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis, 9,* 137–163.

Rothman, K. J. (1986). *Modern epidemiology.* Boston: Little, Brown.

CASE STUDY

The term *case study* is not used in a standard way; at face value, its usage can be misleading because there is a sense in which all research investigates cases. Nevertheless, we can identify a core meaning of the term, as referring to research that studies a small number of cases, possibly even just one, in considerable depth; although various other features are also often implied.

Case study is usually contrasted with two other influential kinds of research design: the social SURVEY and the EXPERIMENT. The contrast with the survey relates to dimensions already mentioned: the number of cases investigated and the amount of detailed information the researcher collects about each case. Other things being equal, the less cases studied, the more information can be collected about each of them. Social surveys study a large number of cases but usually gather only a relatively small amount of DATA about each one, focusing on specific features of it (cases here are usually, though not always, individual respondents). By contrast, in case study, large amounts of information are collected about one or a few cases, across a wide range of features. Here the case may be an individual (as in LIFE HISTORY work), an event, an institution, or even a whole national society or geographical region. A complication here is that when each case studied is large, social survey techniques may be used to collect data within it.

Case study can also be contrasted with experimental research. Although the latter also usually involves investigation of a small number of cases compared to survey work, what distinguishes it from case study is the fact that it involves direct control of variables. In experiments, the researcher creates the case(s) studied, whereas case study researchers identify cases out of naturally occurring social phenomena. This, too, is a dimension, not a dichotomy: QUASI-EXPERIMENTS and FIELD EXPERIMENTS occupy mid-positions.

The term *case study* is also often taken to carry implications for the *kind* of data that are collected, and perhaps also for *how these are analyzed.* Frequently, but not always, it implies the collection of unstructured data, as well as QUALITATIVE analysis of those data. Moreover, this relates to a more fundamental issue about the purpose of the research. It is sometimes argued that the aim of case study research should be to capture cases in their uniqueness, rather than to use them as a basis for wider empirical or theoretical conclusions. This, again, is a matter of emphasis: It does not necessarily rule out an interest in coming to general conclusions, but it does imply that these are to be reached by means of inferences from what is found in particular cases, rather than through the cases being selected to test a hypothesis. In line with this, it is frequently argued that case studies should adopt an INDUCTIVE orientation, but not all case study researchers accept this position.

Another question that arises in relation to case study concerns OBJECTIVITY, in at least one sense of that term. Is the aim to produce an account of each case from an external or research point of view, one that may contradict the views of the people involved? Or is it solely to portray the character of each case "in its own terms"? This contrast is most obvious when the cases are people, so that the aim may be to "give voice" to them rather than to use them as RESPONDENTS or even as informants. However, although this distinction may seem clear-cut, in practice it is more complicated. Multiple participants involved in a case may have diverse perspectives, and even the same person may present different views on different occasions. Furthermore, there are complexities involved in determining whether what is presented can ever "capture" participant views rather than presenting an "external" gloss on them.

Some commentators point out that a case is always a "case of" something, so that an interest in some general category is built in from the beginning, even though definition of that category may change over the course of the research. In line with this, there are approaches to case study inquiry that draw on the COMPARATIVE METHOD to develop and refine theoretical categories. Examples include GROUNDED THEORIZING, historical approaches employing John Stuart Mill's methods of agreement and difference (see Lieberson in Gomm, Hammersley, & Foster, 2000), and ANALYTIC INDUCTION.

Another area of disagreement concerns whether case study is a method—with advantages and disadvantages, which is to be used as and when appropriate, depending on the problem under investigation—or a paradigmatic approach that one simply chooses or rejects on philosophical or political grounds. Even when viewed simply as a method, there can be variation in the specific form that case studies take:

- in the number of cases studied;
- in whether there is comparison and, if there is, in the role it plays;
- in how detailed the case studies are;
- in the size of the case(s) dealt with;
- in what researchers treat as the context of the case, how they identify it, and how much they seek to document it;
- in the extent to which case study researchers restrict themselves to description, explanation, and/or theory or engage in evaluation and/or prescription.

Variation in these respects depends to some extent on the purpose that the case study is intended to serve. When it is designed to test or illustrate a theoretical point, it will deal with the case as an instance of a type, describing it in terms of a particular theoretical framework (implicit or explicit). When it is exploratory or concerned with developing theoretical ideas, it is likely to be more detailed and open-ended in character. The same is true when the concern is with describing and/or explaining what is going on in a particular situation for its own sake. When the interest is in some problem in the situation investigated, the discussion will be geared to diagnosing that problem, identifying its sources, and perhaps outlining what can be done about it.

Variation in purpose may also inform the selection of cases for investigation. Sometimes the focus will be on cases that are judged to be typical of some category or representative of some population. Alternatively, there may be a search for deviant or extreme cases, on the grounds that these are most likely to stimulate theoretical development and/or provide a test for some hypothesis.

Many commentators, however, regard case study as more than just a method: It involves quite different assumptions about how the social world can and should be studied from those underlying other approaches (see, e.g., Hamilton, 1980). Sometimes, this is formulated in terms of a contrast between POSITIVISM, on one hand, and NATURALISM, INTERPRETIVISM, or CONSTRUCTIONISM, on the other. At the extreme, case study is viewed as more akin to the kind of portrayal of the social world that is characteristic of novelists, short story writers, and even poets. Those who see case study in this way may regard any comparison of it with other methods in terms of advantages and disadvantages as fundamentally misconceived.

A series of methodological issues arises from these differences in view about the purpose and nature of case study and have been subjected to considerable debate:

1. *Generalizability.* In some case study work, the aim is to draw, or to provide a basis for drawing, conclusions about some general type of phenomenon or about members of a wider population of cases. A question arises here, though, as to how this is possible. Some argue that what is involved is a kind of inference or generalization that is quite different from statistical analysis, being "logical," "theoretical," or "analytical" in character (see Mitchell in Gomm et al., 2000; Yin, 1994). Others suggest that there are ways in which case studies can be used to make what are, in effect, the same kind of generalizations as those that survey researchers produce (see Schofield in Gomm et al., 2000). Still others argue that case studies need not make any claims about the generalizability of their findings; what is crucial is the use readers make of them—that they feed into processes of "naturalistic generalization" (see Stake in Gomm et al., 2000) or facilitate the "transfer" of findings from one setting to another on the basis of "fit" (see Lincoln & Guba in Gomm et al., 2000). There are questions, however, about whether this latter approach addresses the issue of generalizability effectively (Gomm et al., 2000, chap. 5).

2. *Causal or narrative analysis.* Case study researchers sometimes claim that by examining one or two cases, it is possible to identify CAUSAL processes in a way that is not feasible in survey research. This is because the case(s) are studied in depth, and over time rather than at a single point. It is also often argued that, by contrast with experiments, case study research can investigate causal processes "in the real world" rather than in artificially created settings. Other formulations of this argument emphasize that outcomes can always be reached by multiple pathways, so that narrative accounts of events in particular cases are essential if we are to understand those outcomes (see Becker in Gomm et al., 2000). Here, parallels may be drawn with the work of historians. In relation to all these arguments, however, there are questions about how to distinguish contingent from necessary relationships among events if only one or a small number of cases are

being studied, as well as about what role THEORY plays in causal/narrative analysis (see Gomm et al., 2000, chap. 12).

As already noted, some case study researchers argue that they can identify causal relations through comparative analysis, for example, by means of Mill's methods of agreement and difference or via analytic induction. Sometimes, the comparative method is seen as analogous to statistical analysis, but often a sharp distinction is drawn between the "logics" involved in "statistical" and "case study" work. Nevertheless, questions have been raised about whether there is any such difference in logic (Robinson in Gomm et al., 2000), as well as about the adequacy of Mill's canons and of analytic induction as means of producing theory via case study (Lieberson in Gomm et al., 2000).

3. *The nature of theory.* Although many case study researchers emphasize the role of theory, they differ in their views about the character of the theoretical perspective required. For some, it must be a theory that makes sense of the case as a bounded system. Here, the emphasis is on cases as unique configurations that can only be understood as wholes (see, e.g., Smith, 1978). For others, the task of theory is more to locate and explain what goes on within a case in terms of its wider societal context (see Burawoy, 1998). Without this, it is argued, intracase processes will be misunderstood. Indeed, it is often argued that analysis of a case always presumes some wider context; so the issue is not whether or not a macro theory is involved but rather how explicit this is and whether it is sound.

4. *Authenticity and authority.* Sometimes, case study research is advocated on the basis that it can capture the unique character of a person, situation, group, and so forth. Here there may be no concern with typicality in relation to a category or with generalizability to a population. The aim is to represent the case authentically "in its own terms." In some versions, this is seen as a basis for discovering symbolic truths of the kind that literature and art provide (see Simons, 1996). There are questions here, though, about what this involves. After all, different aesthetic theories point in divergent directions.

The commitment to authenticity may also be based on rejection of any claim to authority on the part of the case study researcher and/or on the idea that case study can be used to amplify the unique voices of those whose experience in, as well as perspective on, the world often go unheard. However, questions have been raised about this position, not just by those committed to a scientific approach or by those who emphasize the role of macro theory, but also by some constructionists and POSTMODERNISTS. The latter's arguments undermine the notion of authenticity by denying the existence of any real phenomenon that is independent of investigations of it, by questioning the legitimacy of researchers speaking on behalf of (or even acting as mediators for) others, and/or by challenging the idea that people have unitary perspectives that are available for case study description.

In summary, although case study has grown in popularity in recent years and is of considerable value, the term is used to refer to a variety of different approaches, and it raises some fundamental methodological issues.

—Martyn Hammersley

REFERENCES

Burawoy, M. (1998). The extended case method. *Sociological Theory, 16*(1), 4–33.

Gomm, R., Hammersley, M., & Foster, P. (Eds.). (2000). *Case study method.* London: Sage.

Hamilton, D. (1980). Some contrasting assumptions about case study research and survey analysis. In H. Simons (Ed.), *Towards a science of the singular: Essays about case study in educational research and evaluation* (pp. 78–92). Norwich, UK: Centre for Applied Research in Education University of East Anglia.

Simons, H. (1996). The paradox of case study. *Cambridge Journal of Education, 26*(2), 225–240.

Smith, L. M. (1978). An evolving logic of participant observation, educational ethnography and other case studies. *Review of Research in Education, 6,* 316–377.

Yin, R. (1994). *Case study research* (2nd ed.). Thousand Oaks, CA: Sage.

CASI. *See* COMPUTER-ASSISTED DATA COLLECTION

CATASTROPHE THEORY

Catastrophe theory refers to a type of behavior among some nonlinear dynamic mathematical models that experience NONLINEAR DYNAMICS such that sudden or rapid large-magnitude changes in the value of one variable are a consequence of a small change that occurs in the value of a parameter (called a *control parameter*). In this sense, catastrophe theory can model phenomena that loosely follow a "straw that broke the camel's back" scenario, although catastrophe theory can be very general in its application. The modern understanding of catastrophe theory has its genesis in work by Thom (1975).

Nearly all early work with catastrophe theory employed polynomial functions in the specification of differential equation mathematical models. In part, this was an important consequence of the generality of Thom's (1975) findings. Because all sufficiently smooth functions can be expanded using a Taylor series approximation (which leads us to a polynomial representation of the original model), it is possible to analyze the polynomial representation directly (see Saunders, 1980, p. 20). However, scientists can avoid using one of Thom's canonical polynomial forms by working with their own original theory-rich specifications as long as it is clear that the original specification has catastrophe potential (Brown, 1995a).

Fundamental to catastrophe theory is the idea of a bifurcation. A bifurcation is an event that occurs in the evolution of a dynamic system in which the characteristic behavior of the system is transformed. This occurs when an attractor in the system changes in response to change in the value of a parameter (called a *control parameter* because its value controls the manifestation of the catastrophe). A catastrophe is one type of bifurcation (as compared with, say, subtle bifurcations or explosive bifurcations). The characteristic behavior of a dynamic system is determined by the behavior of trajectories, which are the values of the variables in a system as they change over time. When trajectories intersect with a bifurcation, they typically assume a radically different type of behavior as compared with what occurred prior to the impact with the bifurcation. Thus, if a trajectory is "hugging" close to an attractor or equilibrium point in a system and then intersects with a bifurcation point, the trajectory may suddenly abandon the previous attractor and "fly" into the neighborhood of a different attractor. The fundamental characteristic of a catastrophe is the sudden disappearance of one attractor and its basin, combined with the dominant emergence of another attractor. Because multidimensional surfaces can also attract (together with attracting points on these surfaces), these gravity centers within dynamical systems are referenced more generally as attracting hypersurfaces, limit sets, or simply attractors.

The following model, which is due to Zeeman (1972), illustrates a simple cusp catastrophe and is used to model the change in muscle fiber length (variable x) in a beating heart. The control parameter A (which in this instance refers to the electrochemical activity that ultimately instructs the heart when to beat) can change in its value continuously, and it is used to move trajectories across an equilibrium hypersurface that has catastrophe potential. The parameter q identifies the overall tension in the system, and f is a scaling parameter. The two differential equations in this system are $dx/dt = -f(x^3 - qx + A)$ and $dA/dt = x - x_1$. Here, x_1 represents the muscle fiber length at systole (the contracted heart equilibrium). Setting the derivative $dx/dt = 0$, we will find between one and three values for x, depending on the other values of the system. When there are three equilibria for x for a given value of the control parameter A, one of the equilibria is unstable and does not attract any trajectory. The other two equilibria compete for the attention of the surrounding trajectories, and when a trajectory passes a bifurcation point in the system, the trajectory abandons one of these equilibria and quickly repositions itself into the neighborhood (i.e., the basin) of the other equilibrium. This rapid repositioning of the trajectory is the catastrophe.

Two social scientific examples of nonlinear differential equation dynamic catastrophe specifications have been developed and explored by Brown (1995b, chaps. 3 and 5). One example involves the interaction between candidate preferences, feelings for a political party, and the quality of an individual's political context or milieu during the 1980 presidential election in the United States. The other example addresses the interaction between the partisan fragmentation of the Weimar Republic's electorate, electoral deinstitutionalization, and support for the Nazi party. Both examples are fully estimated; the first uses both individual and aggregate data, whereas the second employs aggregate data only.

—Courtney Brown

REFERENCES

Brown, C. (1995a). *Chaos and catastrophe theories.* Thousand Oaks, CA: Sage.

Brown, C. (1995b). *Serpents in the sand: Essays on the nonlinear nature of politics and human destiny.* Ann Arbor: University of Michigan Press.

Saunders, P. T. (1980). *An introduction to catastrophe theory.* New York: Cambridge University Press.

Thom, R. (1975). *Structural stability and morphogenesis.* Reading, MA: Benjamin.

Zeeman, E. C. (1972). Differential equations for the heartbeat and nerve impulse. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 8–67). Edinburgh, UK: Edinburgh University Press.

CATEGORICAL

Categorical is a term that includes such terms as NOMINAL VARIABLE and ordinal variable. It derives its name from the fact that, in essence, a categorical variable is simply a categorization of observations into discrete groupings, such as the religious affiliations of a sample.

—Alan Bryman

See also ATTRIBUTE, DISCRETE, NOMINAL VARIABLE

CATEGORICAL DATA ANALYSIS

Social science data, when quantified, can be subject to a variety of statistical analyses, the most common of which is regression analysis. However, the requirements of a continuous dependent variable and of other regression assumptions make linear regression sometimes a less desirable analytic tool because a lot of social science data are categorical.

Social science data can be categorical in two common ways. First, a variable is categorical when it records nominal or discrete groups. In political science, a political candidate in the United States can be a Democrat, Republican, or Independent. In sociology, one's occupation is studied as a discrete outcome. In demography, one's contraceptive choice such as the pill or the condom is categorical. Education researchers may study discrete higher education objectives of high school seniors: university, community college, or vocational school.

Furthermore, a variable can take on an ordinal scale such as the LIKERT SCALE or a scale that resembles it. Such a scale is widely used in psychology in particular and in the social sciences in general for measuring personality traits, attitudes, and opinions and typically has five ordered categories ranging from *most important* to *least important* or from *strongly agree* to *strongly disagree* with a neutral middle category. An ordinal scale has two major features: There exists a natural order among the categories, and the distance between a lower positioned category and the next category in the scale is not necessarily evenly distributed throughout the scale. Variables not measuring personality or attitudes can also be represented with an ordinal scale. For example, one's educational attainment can be simply measured with the three categories of "primary," "secondary," and "higher education." These categories follow a natural ordering, but the distance between "primary" and "secondary" and that between "secondary" and "higher education" is not necessarily equal. It is apparent that an ordinal variable has at least three ordered categories. Although there is no upper limit for the total number of categories, researchers seldom use a scale of more than seven ordered categories.

The examples of categorical data above illustrate their prevalence in the social sciences. Categorical data analysis, in practice, is the analysis of categorical *response* variables.

HISTORICAL DEVELOPMENT

The work by Karl Pearson and G. Udny Yule on the association between categorical variables at the turn of the 20th century paved the way for later development in models for discrete responses. Pearson's contribution is well known through his namesake chi-square statistic, whereas Yule was a strong proponent of the ODDS RATIO in analyzing association. However, despite important contributions by noted statisticians such as R. A. Fisher and William Cochran, categorical data analysis as we know it today did not develop until the 1960s.

The postwar decades saw a rising interest in explaining social issues and a burgeoning need for skilled social science researchers. As North American universities increased in size to accommodate postwar baby boomers, so did their faculties and the bodies of university-based social science researchers. Categorical scales come naturally for measuring attitudes, social class, and many other attributes and concepts in the social sciences. Increasingly in the 1960s, social surveys were conducted, and otherwise quantifiable data were obtained. The increasing methodological sophistication in the social sciences satisfied the increasing need for analytic methods for handling the increasingly available categorical data. It is not surprising that many major statisticians who developed regression-type models for discrete responses were all academicians with social sciences affiliations or ties, such as Leo Goodman, Shelby Haberman, Frederick Mosteller, Stephen Fienberg, and Clifford Clogg. These methodologists focused on log-linear models, whereas their counterparts in the biomedical sciences concentrated their research on regression-type models for categorical data. Together, they and their biomedical colleagues have taken categorical data analysis to a new level.

The two decades before the 21st century witnessed a greater concern with models studying the association between ordinal variables, especially when their values are assumed to be latent or not directly measurable. The same period also saw a resurgence of interest in latent variable models studied by Lazarsfeld in the 1950s (e.g., latent class analysis) and their extension to regression-type settings. At the forefront of these two areas were Leo Goodman, Clifford Clogg, and Jacque Hagenaars, who represented the best of both North American and European social science methodologists.

Another tradition in categorical data analysis stems from the concerns in econometrics with a regression-type model with a dependent variable that is limited in one way or another (in observed values or in distribution), a notion popularized by G. S. Maddala. When only two values (binary) or a limited number of ordered (ordinal) or unordered (nominal) categories can be observed for the dependent variable, LOGIT and PROBIT models are in order. During the past quarter century, computer software has mushroomed to implement these binary, ordinal, and multinomial logit or probit models. In statistics, many models for categorical data such as the logit, the probit, and Poisson regression as well as linear regression can be more conveniently represented and studied as the family of GENERALIZED LINEAR MODELS.

TYPES OF CATEGORICAL DATA ANALYSIS

There are several ways to classify categorical data analysis. Depending on the parametric nature of a method, two types of categorical data analysis arise.

1. *Nonparametric methods.* These methods make minimal assumptions and are useful for hypothesis testing. Examples include the Pearson CHI-SQUARE TEST, Fisher's exact test (for small expected frequencies), and the Mantel-Haenszel test (for ordered categories in linear association).

2. *Parametric methods.* This model-based approach assumes random (possibly stratified) sampling and is useful for estimation purposes and flexible for fitting many specialized models (e.g., symmetry, quasi-symmetry). It also allows the estimation of the STANDARD ERROR, COVARIANCE, and CONFIDENCE INTERVAL of model parameters as well as predicted response probability.

The second type is the most popular in the social sciences, evidenced by the wide application of parametric models for categorical data. Variables in social science research can occupy dependent or independent positions in a statistical model. They are also known as response and explanatory variables, respectively. The type (categorical or not) and the positioning of a variable (DEPENDENT or INDEPENDENT) give rise to the statistical model(s) for (categorical) data, providing another way to classify categorical data analysis. For simplicity, we view all variables as either categorical or continuous (i.e., with interval or ratio scales). In some models, all variables can be considered dependent because only the association among them is of interest to the researcher. For the models that distinguish between dependent and independent variables, we limit our attention to those with single dependent variables. That is, in such models, there is only one dependent variable per model regardless of the number of independent variables, which can be all continuous, all categorical, or a mixture of both.

Models of association do not make a distinction between response and explanatory variables in the sense that the variables under investigation depend on each other. When both variables in a pair are continuous, we use correlation analysis; when they are categorical, we use contingency table and log-linear analysis. The situation can be generalized to multiple variables as in the case of partial correlation analysis and multiway contingency tables.

When the two variables in a pair are of different types (i.e., one categorical and the other continuous), a causal order is assumed with the separation of the independent and the dependent variables. For analyzing the distribution of a continuous-response variable in explanatory categories, we apply ANOVA or linear regression, which can be extended into including additional explanatory variables that are continuous. In that case, we use ANCOVA and linear regression. The flexibility of logit and probit models is evidenced by their applicability to all three cells of Table 1 for a categorical response variable, regardless of the type of explanatory variables—continuous, categorical, or mixed. Often

Table 1 Correspondence Between Variable Types and Statistical Models

		Dependent	
		Continuous	*Categorical*
Dependent	Continuous	Correlation	
	Categorical		Contingency table analysis, log-linear models
Independent	Continuous	Linear regression	Logit and probit models
	Categorical	Continuous	Logit and probit models
	Mixed	ANCOVA, linear regression	Logit and probit models

Table 2 Types of Latent Variable Models According to the Type of Variables

	Latent Continuous	*Latent Categorical*
Observed continuous	Factor-analytic model; SEM	Latent profile model
Observed categorical	Latent trait model; SEM	Latent class model

the same data may be analyzed using different methods, depending on the assumptions the researcher is willing to make and the purposes of the research.

Thus, in Table 1, the cells in the column under the heading of categorical dependent variables comprise the methods of categorical data analysis. A further consideration of these methods is whether the response variable has ordered categories as opposed to pure nominal ones. Excluded from the table, for instance, is Poisson regression that models an integer-dependent variable following a Poisson distribution, which is also a limited-dependent variable model.

Depending on whether a variable is discrete or ordinal, a model for categorical data can be further classified. For example, a logit (or probit) model is binary when the dependent variable is dichotomous, is ordinal (or ordered) when the dependent variable has ordered categories, and is multinomial when it has more than two discrete categories. Log-linear models for the association between ordinal variables can be extended into linear-by-linear association and log-multiplicative models, to name just two examples (see LOG-LINEAR MODEL and ASSOCIATION MODEL for further details).

We have so far examined categorical data analysis with observed variables only. Social scientists also employ in categorical data analysis the LATENT VARIABLE to represent concepts that cannot be measured directly. Here we have yet another way of classifying categorical data analysis, depending on whether the latent variables, the observed variables, or both in a model are categorical.

Specifically, categorical data analysis is concerned with the second row in Table 2, in which the observed variables are categorical. LATENT CLASS ANALYSIS, which estimates categorical latent classes using categorical observed variables, has received most attention in the social sciences.

Several major computer programs have made the latent class model much more accessible by many social scientists. These include C. C. Clogg's mainframe program MLLSA, which is currently available in S. Eliason's Categorical Data Analysis System at the University of Minnesota; S. Haberman's NEWTON and DNEWTON; J. Hagenaars's LCAG; and J. Vermunt's LEM, among others.

MODEL FITTING AND TEST STATISTICS

Yet one more way to view and classify categorical data analysis looks at how model fitting is defined. Much of current social science research using categorical data derives from three general yet related traditions: one that fits EXPECTED FREQUENCIES to observed frequencies, another that fits expected response values to observed response values, and a last that fits expected measures of association through correlation to the observed counterparts. The third tradition describes how categorical data are handled in factor-analytic and structural equation modeling approaches, and it receives separate attention elsewhere in the encyclopedia. We focus here on the first two traditions.

The first tradition fits a particular statistical model to a table of frequencies. Let f_i indicate observed frequency in cell i and F_i indicate expected frequency in cell i in the table. Then our purpose in model fitting is to reduce some form of the difference between the observed and the expected frequencies in the table. Thus, we obtain the Pearson chi-square statistic and the LIKELIHOOD RATIO STATISTIC L^2 (or G^2):

$$\chi^2 = \sum_i \frac{(f_i - F_i)^2}{F_i} \quad \text{and} \quad L^2 = 2\sum_i f_i \log\left(\frac{f_i}{F_i}\right).$$

The summation is over all cell i except the ones with structural zeros. These statistics can be used alone in a nonparametric method for a table of frequencies or in a parametric method such as a log-linear model, which typically expresses $\log(F_i)$ as a linear combination of the effects of the rows, columns, layers, and so on and their interactions in a table. The two test statistics given above are actually special cases of a family of test statistics known as power divergence statistics, which includes a third member, the CRESSIE-READ STATISTIC, that represents a compromise between χ^2 and L^2.

The second tradition is that of the generalized linear model, which attempts to model the expected value μ of the response variable y as a function of a linear combination of the explanatory variables and their parameters. The data are expected to follow a form of exponential distribution, and the function linking $g(\mu)$ to the linear combination of the independent variables and parameters may take on various forms, such as the widely used function of the logit.

The two traditions are closely related. For example, when all variables—response and explanatory—are categorical, $F_i = n_i\mu_i$, where n_i is the number of observations in cell i. The test statistics described above also apply to the assessment of model fitting of generalized linear models.

MODEL COMPARISON AND SELECTION

This is a prominent issue in categorical data analysis. Researchers often feel the need to compare statistical models and to choose a better fitting model from a set of models. The latter situation describes hierarchical log-linear model fitting when a number of log-linear models embedded in one another are considered. These models have different degrees of freedom because they include a different number of parameters. Differences in L^2 give comparative likelihood ratio statistics for drawing conclusions.

Sometimes researchers need to compare models that are nonhierarchical. This can be a comparison of the model fitting of two (or more) subgroups of observations or two (or more) models that include different sets of variables. Although sometimes treating these subgroups or submodels as members of a single combined supermodel is a useful exercise, this approach is not always realistic because the submodels under comparison can be just incompatible. Therefore, a statistical criterion of a model's goodness of fit that applies across different models with different degrees of freedom is necessary. Two widely applied criteria belonging to this category are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

Relying on information-theoretic considerations, the AIC adjusts the likelihood ratio statistic with twice the number of degrees of freedom in a given model:

$$\text{AIC} = L^2 - 2(df),$$

where df represents the number of degrees of freedom in the model. Between two models, hierarchical or not, the model with the smaller AIC value is the better one.

Based on Bayesian decision theory, the BIC further adjusts the likelihood ratio statistic with the total sample size (N):

$$\text{BIC} = L^2 - \log(N)(df).$$

The BIC gives an approximation of the logarithm of the Bayes factor for model comparison. It has been shown that the BIC penalizes the model with a larger degree of freedom more so than the AIC if sample sizes are not very small. When comparing hierarchical models in which L^2 differences are also applicable, the BIC usually favors a parsimonious model more so than the other goodness-of-fit statistics.

MISSING DATA IN CATEGORICAL DATA ANALYSIS

MISSING DATA can be a serious concern in any multivariate analysis and deserve particular attention in models of frequency tables that may have zero cells. Of special interest are the so-called partially observed data. If an observation has missing information on all variables, the observation cannot be used at all. But if an observation has missing information on some of the variables in the model, the conventional approach of listwise deletion would waste useful information. Added to this problem is the mechanism with which missing information occurs. Of the three mechanisms of missing data—missing completely at random, missing at random, and missing not at random—the last may create bias in drawing inference about relationships between variables in the model. Developments in recent years have sought to deal with estimating statistical models of categorical data under various missing data mechanisms. The LEM software by J. Vermunt is the most flexible for such purposes.

—Tim Futing Liao

REFERENCES

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley.

Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical models for ordinal variables.* Thousand Oaks, CA: Sage.

Everitt, B. S. (1992). *The analysis of contingency tables.* London: Chapman & Hall.

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: MIT Press.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models.* Thousand Oaks, CA: Sage.

Long, S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

Powers, D. A., & Xie, Y. (2000). *Statistical models for categorical data analysis.* New York: Academic Press.

CATI. *See* COMPUTER-ASSISTED DATA COLLECTION

CAUSAL MECHANISMS

Causal mechanisms involve the processes or pathways through which an outcome is brought into being. We explain an outcome by offering a HYPOTHESIS about the cause(s) that typically bring it about. So a central ambition of virtually all social research is to discover causes. Consider an example: A rise in prices causes a reduction in consumption. The causal mechanism linking cause to effect involves the choices of the rational consumers who observe the price rise, adjust their consumption to maximize overall utility, and reduce their individual consumption of this good. In the aggregate, this rational behavior at the individual level produces the effect of lower aggregate consumption.

There are two broad types of theories of causation: the Humean theory ("causation as regularities") and the causal realist theory ("causation as causal mechanism"). The Humean theory holds that causation is entirely constituted by facts about empirical regularities among observable variables; there is no underlying causal nature, causal power, or causal necessity. The causal realist takes notions of causal mechanisms and causal powers as fundamental and holds that the task of scientific research is to arrive at empirically justified THEORIES and hypotheses about those causal mechanisms. Consider these various assertions about the statement, "X caused Y":

- X is a necessary and/or sufficient condition of Y.
- If X had not occurred, Y would not have occurred.
- The conditional probability of Y given X is different from the absolute probability of $Y[(P(Y|X) \neq P(Y)]$.
- X appears with a nonzero coefficient in a REGRESSION equation PREDICTING the value of Y.
- There is a causal mechanism leading from the occurrence of X to the occurrence of Y.

The central insight of causal realism is that the final criterion is in fact the most fundamental. According to causal realism, the fact of the existence of underlying causal mechanisms linking X to Y accounts for each of the other criteria; the other criteria are symptoms of the fact that there is a causal pathway linking X to Y.

Causal reasoning thus presupposes the presence of a causal mechanism; the researcher ought to attempt to identify the unseen causal mechanism joining the variables of interest. The following list of causal criteria suggests a variety of ways of using available evidence to test or confirm a causal hypothesis: apply Mill's methods of similarity and difference as a test for necessary and sufficient conditions, examine conditional probabilities, examine CORRELATIONS and regressions among the variables of interest, and use appropriate parts of social theory to hypothesize about underlying causal mechanisms. Causal realism insists, finally, that EMPIRICAL evidence must be advanced to assess the credibility of the causal mechanism that is postulated between cause and effect.

What is a causal mechanism? A causal mechanism is a sequence of events or conditions, governed by lawlike regularities, leading from the explanans to the explanandum. Wesley Salmon (1984) puts the point this way: "Causal processes, causal interactions, and causal laws provide the mechanisms by which the world works; to understand *why* certain things happen, we need to see *how* they are produced by these mechanisms" (p. 132). Nancy Cartwright likewise places real causal mechanisms at the center of her account of scientific knowledge. As she and John Dupré put the point, "Things and events have causal capacities: in virtue of the properties they possess, they have the power to bring about other events or states" (Dupré & Cartwright, 1988). Most fundamentally, Cartwright argues that identifying causal relations requires substantive theories of the causal powers or capacities that govern the entities in question. Causal relations cannot be directly inferred from facts about association among variables.

The general nature of the mechanisms that underlie social causation has been the subject of debate. Several broad approaches may be identified: agent-based models, structural models, and social influence models. Agent-based models follow the strategy of aggregating the results of individual-level choices into macro-level outcomes, structural models attempt to demonstrate the causal effects of given social structures or institutions (e.g., the tax collection system) on social outcomes (levels of compliance), and social influence models attempt to identify the factors that work behind the backs of agents to influence their choices. Thomas Schelling's (1978) apt title, *Micromotives and Macrobehavior,* captures the logic of the former approach, and his work profoundly illustrates the sometimes highly unpredictable results of the interactions of locally rational-intentional behavior. Jon Elster has also shed light on the ways in which the tools of rational choice theory support the construction of large-scale sociological explanations (Elster, 1989). Emirbayer and Mische (1998) provide an extensive review of the current state of debate on the concept of agency. Structuralist and social influence approaches attempt to identify socially salient influences such as institution, state, race, gender, and educational status and provide detailed accounts of how these factors influence or constrain individual trajectories, thereby affecting social outcomes.

—Daniel Little

REFERENCES

Cartwright, N. (1989). *Nature's capacities and their measurement.* Oxford, UK: Oxford University Press.

Dupré, J., & Cartwright, N. (1988). Probability and causality: Why Hume and indeterminism don't mix. *Nous, 22,* 521–536.

Elster, J. (1989). *The cement of society: A study of social order.* Cambridge, UK: Cambridge University Press.

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology, 103*(4), 962–1023.

Little, D. (1998). *Microfoundations, method and causation: On the philosophy of the social sciences.* New Brunswick, NJ: Transaction Publishers.

Mackie, J. L. (1974). *The cement of the universe: A study of causation.* Oxford, UK: Clarendon.

Miller, R. W. (1987). *Fact and method: Explanation, confirmation and reality in the natural and the social sciences.* Princeton, NJ: Princeton University Press.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

Schelling, T. C. (1978). *Micromotives and macrobehavior.* New York: Norton.

CAUSAL MODELING

In the analysis of social science data, researchers to a great extent rely on methods to infer causal relationships between CONCEPTS represented by measured VARIABLES—notably, those between DEPENDENT VARIABLES and INDEPENDENT VARIABLES. Although the issue of CAUSALITY can be rather philosophical, a number of analytical approaches exist to assist the researcher in understanding and estimating causal relationships. Causal thinking underlies the basic design of a range of methods as diverse as EXPERIMENTS, and models based on the MARKOV CHAIN. One thing that distinguishes experimental from nonexperimental design is the fact that the independent variables are actually set at fixed values, or manipulated, rather than merely observed. Thus, with nonexperimental or observational design, causal inference is much weaker. Given an observational design, certain conditions can strengthen the argument that the finding of a relationship between, say, X and Y is causal rather than SPURIOUS. Of particular importance is temporality (i.e., a demonstration that X occurred before Y in time). Also important is meeting the ASSUMPTIONS necessary for causal inference. In classical REGRESSION analysis, these are sometimes referred to as the GAUSS-MARKOV assumptions.

Methods for analyzing observational data also include all kinds of regression-type models subsumed under GENERAL LINEAR MODELS or GENERALIZED LINEAR MODELS, especially GRAPHIC MODELING, PATH ANALYSIS, or STRUCTURAL EQUATION MODELING (SEM), which allow for multiequation or SIMULTANEOUS EQUATION systems. SEM, which can include LATENT VARIABLES to better represent CONCEPTS, can be estimated with software such as AMOS, EQS, LISREL, or M-PLUS.

A typical simultaneous equation system may contain RECIPROCAL RELATIONSHIPS (e.g., X causes Y and Y causes X), which renders them NONRECURSIVE. Here is an example of such a system, which may be called a causal model:

$$Y_1 = b_{10} + b_{11}X_1 + b_{12}Y_2 + e_1,$$
$$Y_2 = b_{20} + b_{21}Y_1 + b_{22}X_2 + e_2.$$

The phrase *causal model* is taken to be a synonym for a system of simultaneous, or structural, equations. Sometimes, too, a sketched arrow diagram, commonly called a *path diagram,* accompanies the presentation of the system of equations itself. Before estimation proceeds, the IDENTIFICATION PROBLEM must be solved through examination of the relative number of EXOGENOUS and ENDOGENOUS variables to see if the order and rank conditions are met. In this causal model, the X variables are exogenous, the Y variables endogenous. As it turns out, when each equation in this system is exactly identified, the entire system is identified, and its parameters can be estimated. By convention, exogenous variables are labeled X and endogenous variables labeled Y, but it must be emphasized that simply labeling them such does not automatically make them so. Instead, the assumptions of exogeneity and endogeneity must be met.

Once the model is identified, estimation usually goes forward with some INSTRUMENTAL VARIABLES method, such as two-stage least squares. A three-stage least squares procedure also exists, but it has not generally been shown to offer any real improvement over the two-stage technique. The method known as indirect least squares can be employed if each equation in the system is exactly identified. ORDINARY LEAST SQUARES cannot be used for it will generally produce biased and inconsistent estimates unless the system is RECURSIVE, whereby causality is one-way and the error terms are uncorrelated across equations. There are MAXIMUM LIKELIHOOD estimation analogs to two-stage and three-stage least squares: limited-information maximum likelihood and full-information maximum likelihood, respectively.

Causal modeling, as the discussion so far suggests, relies on statistical analysis that has PROBABILITY theory as its foundation. Causality describes a process imbued in uncertainty, and it is the job of the analyst to take into account the uncertainty while capturing the causal patterns.

—Michael S. Lewis-Beck and Tim Futing Liao

REFERENCES

Berry, W. (1984). *Nonrecursive causal models.* Beverly Hills, CA: Sage.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley.

Davis, J. (1985). *The logic of causal order.* Beverly Hills, CA: Sage.

Kmenta, J. (1997). *Elements of econometrics* (2nd ed.). Ann Arbor: University of Michigan Press.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

CAUSALITY

Often a hesitation is expressed when causality stands central in social science discussions and debates. "We are afraid of causality," so it sounds. But the discussants all admit that they cannot do without causality: "In the end, we need causal explanations." This hesitation also holds for the notion of causality in the history of philosophy. The historical train of causality is like a local train; it resembles a procession of *Echternach,* going now two steps forward and then one step backward. Aristotle, the schoolmen, Leibniz, Kant, and Schopenhauer were causalists. But Hume, Mach, Pearson, Reichenbach, the Copenhagen school of quantum physicists, and Russell were a-causalists. It was Bertrand Russell (1912) who left us with the following bold expression: "The law of causality, I believe, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm" (p. 171).

ARISTOTLE: THE FOUNDING FATHER

We may consider Aristotle to be the founding father of causal thinking. He was a philosopher who distinguished many types of causes (e.g., *causa materialis, causa formalis, causa efficiens,* and *causa finalis*). In modern science of the 17th century, his *causa efficiens* or labor cause became the guiding if not the only principle. Causation was then associated with "production." The labor cause was the active agent, the external power was like the knock with a hammer, and the effect was that which undergoes in a passive way, like the chestnut that bursts into pieces. This notion of production is present in Newton's external power, which brings a resting body into movement. It is also present in social policymaking. For example, policymakers in France in the 19th century started a policy of raising the birth rate and of encouraging immigration to establish an excess of births over deaths, so that the aging population structure could be reversed. It is also present in the stimulus-response scheme of the school of BEHAVIORISM. And it is even present in our language because many verbs such as *poison, cure, calm, humidify,* and *illuminate* contain the idea of production.

GALILEI: NECESSARY AND SUFFICIENT CONDITION

It was also in modern science that cause was clarified in terms of a necessary and sufficient condition by Galilei, where *necessary* means "*conditio sine qua non*" or "If not X, then not Y," and *sufficient* means "If X, then always Y." (One should notice that "X necessitates Y" means "X is a sufficient condition of Y." Confusion is possible here!)

DAVID HUME: A TURNING POINT IN HISTORY

David Hume in the 18th century is really a turning point in the history of causality. In fact, no author today can still write about this subject without first discussing Hume. An outstanding example is Karl Popper (1972), who, in his book *Objective Knowledge,* has a first chapter titled "The Hume Problem." Hume was not against causality (see Hume, 1786). He believed that the world was full of it. But he was skeptical about the possibility of science getting insight into the question of why a cause is followed by an effect. His starting point is purely *empirical.* He argued that knowledge of causal relations was never brought about in an a priori fashion by means of pure deduction but that it was totally based on experience. Adam could not deduce from seeing water that it could suffocate him. Hume also believed that causes were external to their effects. If a billiard ball collides with a second ball and the latter starts moving, then there is nothing present in the first ball that gives us the slightest idea about what is going to happen to the second one. As for the middle term in between cause and effect (i.e., the causal arrow), Hume stated that such concepts as production, energy, power, and so forth belonged to an obscure philosophy that served as a shelter for superstition and as a cloak for covering foolishness and errors. We see the fire and feel the heat, but we cannot even guess or imagine the connection between the two. We are not even directly conscious of the energy with which our will influences the organs of our bodies, such that it will always escape our diligent investigations. The question of why the will influences the tongue and the fingers, but not the heart and the liver, will always bring us embarrassment. And the idea that the will of a Supreme Being is responsible here brings us far beyond the limits of our capabilities. Our perpendicular line is too short to plumb these yawning chasms. He concluded that when we saw both the cause and the effect, then there would be *constant conjunction.* After some time of getting used to this, *custom* arises. And then, via some mechanism of *psychological association,* we gradually start to develop a *belief.* And on the basis of this belief, we use causal terminology and make *predictions.* In short, the only thing that really exists is constant conjunction (regularity theory); the rest is psychology. Loosely speaking, correlation is the matter; the rest is chatter.

IMMANUEL KANT: CAUSALITY IS AN A PRIORI CATEGORY

With the reactions on David Hume, one can fill a whole library. Among others, Immanuel Kant, who first admitted that Hume had awakened him out of his dogmatic half-sleep, refused to accept that only experience is the basis of causality (see Kant, 1781). He believed in prior categories of understanding, which, together with experience, brought us the synthesis called objective knowledge. Causality was, in his view, one of these categories of understanding. This view of causality as an a priori category is not undisputed. John Mackie (1974) gave the example of a piece of wood that is cut into two parts and argued that from a long enough distance, we would first see the parts fall apart and then hear the sound of the axe due to the difference in the speed of the light and the sound, but up close, we would first hear it and then see the parts fall apart due to the inertia of the bodies. And if a little train on the table bumps into a second train, whereby the second train starts moving, then this has nothing to do with our prior faculty of the mind because there might be a hidden magnet under the table that brings about the second movement. So, there is not much of a priori causal knowledge here but rather the act of hypothesizing knowledge in a tentative way, with experience and observation needed to come to a conclusion.

MODERN SCIENCE: HYPOTHESIS TESTING

The tradition of hypothesis testing, often used today in our scientific research, was initiated by positivist philosophers of the 19th century such as August Comte and John Stuart Mill and became standard procedure within positivist research in the 20th century. Karl Popper (1959) and Carl Gustav Hempel (1965) also contributed to this general attitude by means of their deductive-nomological approach, which is well known as the covering law theory (i.e., every concrete causal statement is covered by general laws that operate in the background and serve as general theories and hypotheses).

THE PROBABILISTIC THEORY OF CAUSALITY

In the meantime, the debate on causality goes on. Most proposals contain a policy of encirclement because causal production is not approached in a direct way but rather indirectly via some other criteria. One such criterion is "probability," used by a school of scientists who may be called the adherents of *A Probabilistic Theory of Causality*. This is the title of a book by Patrick Suppes, who initiated this school of thought in 1970. His opinion corresponds with the opinion of the majority of scientific researchers in the world—that is, X is a cause of Y if and only if X exists (the probability of X is greater than zero), X is temporally prior to Y (the moment in time t_X comes before t_Y), there is a statistical relationship between X and Y (the probability of Y given X is greater than the probability of Y by itself), and there is no spuriousness (the statistical relationship between X and Y does not disappear when controlling other potentially CONFOUNDING factors). These criteria for causality are in use in social research, especially in statistical analysis. The method of path coefficients of Sewell Wright in genetics in the 1920s, the simultaneous equation models of Herman Wold in econometrics in the 1950s, the causal models of Simon and Blalock in sociology and other social sciences in the 1960s and 1970s, and the linear structural relations system (LISREL) of Karl Jöreskog in the 1970s and after are but a few of the many examples. Three main problems will always haunt this school of thought. The first is the notion of probability, which explains the statistical relationship (i.e., correlation) but not causation. The second is the dependence on theory, which is expressed by the relations between the variables in the causal model, especially by the factors to control for spuriousness. The third problem is that temporal priority is used instead of causal priority, which means that research practice is building on the sophism "Post hoc, ergo propter hoc" (i.e., Thereafter, so therefore).

MARIO BUNGE: PARADISE OF SIMPLISSIMUS

Another policy of encirclement is the use of Galilei's criterion of a "necessary" and "sufficient" condition, which is done by many modern authors but in the most intelligent way by John Mackie (1974). He reacts against the extreme standpoint of Mario Bunge (1959), who tells us that we commit the sin of meaning-inflation if we liberalize the concept of causality so that it covers nearly everything. In his view, there are many forms of determination in reality—probabilistic, dialectical, functional, structural, mechanical, and other determinations—of which the causal determination is only one. The core of this causal determination is the criterion given by Aristotle's causa efficiens (i.e., necessary production). Next to productive, the causal relation is also conditional (if X, then Y), unique (one cause, one effect), asymmetrical (when X causes Y, then Y does not cause X), and invariable (no exceptions, no probabilistic causality). Bunge also gives other strict criteria, such as linearity, additivity, continuity, and the like. His notion of causality is very rigid. He reacts against functionalists, interactionists, and dialecticians, who see everything in terms of interdependence and are, therefore, romanticists. In their view, causality is a key to every door, a panacea. In Bunge's view, causal relations, if strictly defined, are only a small part of reality, but they exist. They are neither myth nor panacea.

JOHN MACKIE: INUS CONDITION IN A CAUSAL FIELD

It goes without saying that Bunge's (1959) rigid definition of causality is a paradise of simplicity. John Mackie (1974) reacts against this rigidity in developing a very liberal notion of causality that is close to everyday language and accepts probabilism, multicausality, teleology, functionalism, and many other forms of determination. His approach is in terms of necessary and sufficient condition. A causal factor is, in his view, a necessary condition "in the circumstances," which means that silent premises and particulars should, as much as possible, be made explicit. For example, a fire breaks out in a house. Experts claim that this is due to a short circuit. They do not give a necessary condition because other things than a short circuit, such as the falling of an oil heater, could have caused the fire. They do not give a sufficient condition because even if there was a short circuit, other conditions such as inflammable materials or the absence of an efficient fire extinguisher are necessary for a fire to start. Therefore, there is a complex set of conditions, positive and negative, that together with a short circuit are sufficient but not necessary for the outbreak of a fire because other factors could have caused the fire. A short circuit is a necessary part of the set of conditions, for without it, the inflammable materials and absence of an efficient fire extinguisher could not cause a fire. A short circuit is then an insufficient but necessary part of a complex set of conditions, of which the total is unnecessary but sufficient for the result. In short, a short circuit is an INUS condition for fire—that is, an **I**nsufficient but **N**ecessary part of a set, which is **U**nnecessary but **S**ufficient for the result. Mackie adds to this that there will always be factors that cannot vary but are fixed in the causal field. For example, being born is, in the strict logical sense, an INUS condition of dying, but it cannot be a candidate cause of dying because a statement on the causes of death refers to people who have lived. A less evident example is the syphilis example of Scriven (1959). *Treponema pallidum,* a bacteria, is the unique cause of syphilis. However, only a small percentage of people contaminated by the syphilis bacteria come into the third and last phase of *paralysis generalis,* a brain paralysis accompanied by motoric disorder and mental disturbances and causing death. Now, the first statement about the unique cause refers to a different causal field than the second statement about paralysis generalis. The first is the causal field of all persons who are susceptible to the bacteria, for example, all persons who have sexual intercourse. The second is the causal field of all persons who are already contaminated by the bacteria.

In research practice, this notion of causal field is of crucial importance because it also contains, next to self-evident fixed factors related to the research problem, factors that are fixed due to pragmatic considerations related to time and space (living in a country and doing research in that country) in addition to factors that are actually INUS conditions but based on causal a priori considerations and, due to the danger of heterogeneity, have been fixed (doing research in the causal field of the unemployed and leaving young workers out because youth unemployment is a different problem entirely). Capital sin number one in the social sciences is really the neglect of the causal field.

—Jacques Tacq

REFERENCES

Aristotle. (1977). *Metaphysica,* Baarn: Het Wereldvenster.

Bunge, M. (1959). *Causality: The place of the causal principle in modern science.* Cambridge, MA: Harvard University Press.

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science.* London: Collier-Macmillan.

Hume, D. (1786). *Treatise on human nature.* Oxford, UK: Claredon.

Kant, I. (1781). *Kritik der reinen Vernunft* [Critique of pure reason]. Hamburg: Felix Meiner.

Mackie, J. (1974). *The cement of the universe.* Oxford, UK: Oxford University Press.

Popper, K. (1959). *The logic of scientific discovery.* London: Hutchinson.

Popper, K. (1972). *Objective knowledge: An evolutionary approach.* Oxford, UK: Clarendon.

Russell, B. (1912). *Mysticism and logic.* London: Allen & Unwin.

Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science, 130,* 477–482.

Suppes, P. (1970). *A probabilistic theory of causality.* Amsterdam: North Holland.

Tacq, J. J. A. (1984). *Causaliteit in Sociologisch Onderzoek. Een Beoordeling van Causale Analysetechnieken in het Licht van Wijsgerige Opvattingen over Causaliteit* [Causality in sociological research: An evaluation of causal techniques of analysis in the light of philosophical theories of causality]. Deventer, The Netherlands: Van Loghum Slaterus.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5,* 161–215.

CCA. *See* CANONICAL CORRELATION ANALYSIS

CEILING EFFECT

A ceiling effect occurs when a measure possesses a distinct upper limit for potential responses and a large concentration of participants score at or near this limit (the opposite of a FLOOR EFFECT). Scale attenuation is a methodological problem that occurs whenever variance is restricted in this manner. The problem is commonly discussed in the context of experimental research, but it could threaten the validity of any type of research in which an outcome measure demonstrates almost no variation at the upper end of its potential range.

Ceiling effects can be caused by a variety of factors. If the outcome measure involves a performance task with an upper limit (such as number of correct responses), participants may find the task too easy and score nearly perfect on the measure. Rating scales also possess upper limits, and participants' responses may fall outside the upper range of the response scale.

Due to the lack of variance, ceiling effects pose a serious threat to the validity of both experimental and nonexperimental studies, and any study containing a ceiling effect should be interpreted with caution. If a dependent variable suffers from a ceiling effect, the experimental or quasi-experimental manipulation will appear to have no effect. For example, a ceiling effect may occur with a measure of attitudes in which a high score indicates a favorable attitude and the highest response fails to capture the most positive evaluation possible. Because the measure cannot assess more favorable attitudes, a persuasion study using this measure may yield null findings if the prior attitudes of the participants were already at the high end of the attitude distribution.

Ceiling effects are especially troublesome when interpreting INTERACTION EFFECTS. Many interaction effects involve failure to find a significant difference at one level of an independent variable and a significant difference at another. Failure to find a significant difference may be due to a ceiling effect, and such interactions should be interpreted with caution.

Ceiling effects are also a threat for nonexperimental or correlational research. Most inferential statistics are based on the assumption of a normal distribution of the variables involved. It may not be possible to compute a significance test, or the restricted range of variance may increase the likelihood of rejecting the null hypotheses when it is actually true (TYPE II ERROR).

The best solution to the problem of ceiling effects is pilot testing, which allows the problem to be identified early. If a ceiling effect is found, the problem may be addressed several ways. If the outcome measure is task performance, the task can be made more difficult to increase the range of potential responses. For rating scales, the number of potential responses can be expanded. An additional solution involves the use of extreme anchor stimuli in rating scales. Prior to assessing the variable of interest, participants can be asked to rate a stimulus that will require them to use the extreme ends of the distribution. This will encourage participants to use the middle range of the rating scale, decreasing the likelihood of a ceiling effect.

—Robert M. Hessling,
Nicole M. Traxel, and Tara J. Schmidt

CELL

The term *cell* is used to refer to the point at which, in any CONTINGENCY TABLE, a row and a column intersect. Thus, a contingency table relating to two binary VARIABLES will have four cells.

—Alan Bryman

CENSORED DATA

Statistical data files are described as censored when four conditions are met: (a) data are available for a PROBABILITY SAMPLE of a POPULATION, (b) the DEPENDENT VARIABLE has one or more limit values, (c) the values of the INDEPENDENT VARIABLES are available for all cases regardless of the value of the dependent variable for any case, and (d) the limit value is not due to the selection of this value by the cases. A limit value is a value of a dependent variable either below or above which no other values can be measured or observed. As an example of a lower limit, neighborhood crime rates cannot be less than zero. Among neighborhoods with crime rates of zero, there could be two types of areas that have substantial differences in "crime proneness." Neighborhoods of the first type could be so safe that they would virtually never have a crime occur; those of the second type could have several "crime-facilitating" characteristics but, just by chance alone, did not have a crime in a particular year. Despite the differences between these areas, both types would have a crime rate of zero as their lower limit, and there would not be any value of the dependent variable that distinguishes these neighborhoods from each other. Sellouts of various types of events are examples of upper limits on sales as a measure of total demand. Some events may have more unmet demand than others, but there is no way of telling.

Breen (1996) describes dependent variables with limit values as truncated. If cases have been not been excluded from the sample being studied because of their values on the dependent or independent variables, then the data as a whole meet the third condition for being described as censored in the sense that measurement of the dependent variable either below a lower limit or above an upper limit is not possible and is, therefore, censored. For neighborhood crime, the data are censored if all neighborhoods in a city are included in a sample regardless of their crime rate and regardless of the values of the independent variables. If a study excludes neighborhoods based on the value of a variable, then the data are truncated, not censored. The data for a study of only high-crime or low-crime neighborhoods, or only minority or nonminority areas, would be described as truncated because data collection was stopped or truncated at a certain value of a variable or characteristic of the neighborhoods.

To meet the fourth condition for being censored, the cases with limit values must not be the result of any self-selection by the cases themselves, as can happen with data on individuals. As Breen (1996) noted, households who do not buy luxury items have chosen or selected themselves not to be buyers, whereas neighborhoods do not decide whether they will have crime in them, regardless of how many security precautions residents may take. For the buying example, the analyses of the purchasing data can require sample selection rather than censored data methods.

—Dennis W. Roncek

See also CENSORING AND TRUNCATION

REFERENCE

Breen, R. (1996). *Regression models: Censored, sample selected, or truncated data.* Thousand Oaks, CA: Sage.

CENSORING AND TRUNCATION

Both *censoring* and *truncation* refer to situations in which empirical data on random variables are incomplete or partial. These terms have not been consistently used throughout the existing literature, and what one author calls *censoring,* another may term *truncation.* Consequently, an author's definition of one of these terms always requires careful attention.

Usually, censoring refers to a situation in which some values of a random variable are recorded as lying within a certain range and are not measured exactly, for at least some members of a sample. In contrast, truncation customarily refers to a situation in which information is not recorded for some members of the population when a random variable's values are within a certain range. Thus, censoring is associated with having incomplete or inexact measurements, whereas truncation is associated with having an incomplete sample, that is, a sample chosen conditional on values of the random variable. Hald (1952) is one of the first authors to distinguish between censoring and truncation. Since then, statistical methods for analyzing censored variables and truncated samples have been developed extensively (Schneider, 1986).

A censored *variable* is one in which some observations in a sample are measured inexactly because some values of the variable lie in a certain range or ranges. A censored *observation or case* refers to a sample member for which a particular variable is censored (i.e., known only to lie in a certain range).

For example, everyone whose actual age was reported as 90 years or older in the 1990 U.S. Census of Population has been censored at age 90 in the census data made publicly available. The U.S. Census Bureau decided to censor age at 90 to ensure confidentiality and to avoid inaccurate information (because very old people sometimes exaggerate their age). Any person whose actual age was reported to be 90 to 120 on the household census form appears in the public micro-level data as a case with age censored at 90.

A truncated *variable* is one in which observations are measured only for sample members if the variable's value lies in a certain range or ranges; equivalently, it is a variable for which there are no measurements at all if the variable's value does not fall within selected ranges. Conceptually, a truncated sample is closely related to a sample that has been trimmed to eliminate extreme values that are suspected of being outliers.

One difference between truncation and trimming is that samples are customarily truncated on the basis of a variable's numerical values but trimmed on the basis of a variable's order statistics (e.g., the top and bottom five percentiles are excluded, even though they exist in the data). For example, very short people are typically excluded from military service. As a result, the height of a sample of military personnel is a truncated variable because people in the entire population are systematically excluded from the military if their height is below a certain level.

TYPES OF CENSORING

There are different ways of classifying censoring and censored variables. One basic distinction is between random censoring and censoring on the basis of some predetermined criteria. The two common predetermined criteria used to censor variables are called Type I and Type II censoring. (A few authors also refer to Types III and IV censoring, but the definitions of other types are less consistent from author to author.)

Censoring that results because a variable's value falls above or below some exogenously determined value is called length-based or Type I censoring. When the U.S. Census Bureau censored people's ages at 90 years old, as described above, it performed length-based censoring.

Type II censoring occurs when sample values of the variable are ordered (usually from the lowest to the highest value), and exact measurements cease after the ordered sample values have been recorded for a predetermined fraction of the sample. Type II censoring mainly occurs when the variable refers to time (e.g., time to an event, such as death), and measurements terminate after the time of the event has been recorded for a preselected fraction of the sample. The remaining sample members are known only to have times exceeding the time at which the measurement process stopped (i.e., greater than the last observed time). For example, social scientists might record the duration of joblessness for 1,000 people who have just lost their jobs because a factory closed. If the researchers decide a priori to stop collecting data on the duration of joblessness after 60% of the sample members have found new jobs, they are implementing Type II censoring based on the duration of joblessness. The fraction of censored cases is fixed in advance (40% in this example); however, the maximum observed duration of joblessness depends on the length of time that it takes 600 of the 1,000 people in the sample to find new jobs.

Often, censoring is the result of some other random process. To avoid biased conclusions, the random variable of interest must be statistically independent of the random process causing censoring. If the two are statistically independent, censoring is said to be noninformative; if they are statistically dependent, censoring is said to be informative.

Suppose, for example, that the variable of interest is a woman's age at the birth of her first child. Some women die childless; for these women, age at first birth is censored at their age at death. Death is clearly a random process. However, a key issue is whether a woman's age at death is statistically independent of her age at the birth of her first child. There is no statistical basis for determining if the process generating the outcome of interest and the random censoring process are statistically independent. Consequently, decisions about their independence are made in practice on the basis of substantive arguments. In any given application, some scholars may argue that the two processes are statistically independent, whereas others may argue that they are dependent. Still others may argue that these two processes are statistically independent (for all practical purposes) under some conditions but are dependent under other conditions. For example, the process governing the birth of a woman's first child and the process causing women to die may be (almost) independent in a developed country such as Sweden but not in a less developed country in sub-Saharan Africa.

Right, left, double, and multiple censoring are also ways of distinguishing censoring patterns. Right censoring means that values of the random variable are not observed exactly if they exceed a certain level τ_R; left censoring means that they are not observed exactly if they are less than a certain level τ_L. Double censoring refers to the combination of left and right censoring; it means that values are observed when they are above a certain level τ_L and below some other level τ_R, that is, between τ_L and τ_R. Finally, multiple censoring occurs when there are multiple censoring levels, and the data record only the intervals in which the values of the variable fall.

For example, a person's annual earnings may be right censored at \$100,000 (i.e., higher earnings are recorded as censored at \$100,000) or left censored at \$5,000 (i.e., lower earnings are recorded as censored at \$5,000). In addition, earnings may be recorded exactly for values between \$5,000 and \$100,000 but censored at these two extremes; then there is double censoring. Finally, surveys often ask people to report their annual earnings in terms of several ordered ranges (e.g., less than \$5,000, \$5,000–\$10,000, \$10,000–\$20,000, ..., \$90,000–\$100,000, more than \$100,000). Then the responses are multiply censored; an ordinal variable has been recorded instead of the original continuous variable.

ANALYSES OF CENSORED OUTCOMES

Tobin's (1958) suggestion of a censored normal regression model of consumer expenditures was among the first to incorporate the idea of censoring into empirical social scientific research. In the censored normal regression model, the true dependent variable Y^* is fully observed above some level τ_L (and/or below some other level τ_R); otherwise, it is known only to be below τ_L (or above τ_R, or outside the range from τ_L to τ_R). That is, the data on the continuous variable Y may be left censored, right censored, or censored on the left and the right. Formally, these three cases mean the following.

1. Censored on the left:

$$\begin{aligned} Y &= Y^*, \quad \text{if } Y^* \geq \tau_L \\ &= \tau_L, \quad \text{if } Y^* < \tau_L. \end{aligned}$$

2. Censored on the right:

$$\begin{aligned} Y &= Y^*, \quad \text{if } Y^* \leq \tau_R \\ &= \tau_R, \quad \text{if } Y^* > \tau_R. \end{aligned}$$

3. Censored on the left and the right:

$$\begin{aligned} Y &= Y^*, \quad \text{if } \tau_L \leq Y^* \leq \tau_R \\ &= \tau_L, \quad \text{if } Y^* < \tau_L \\ &= \tau_R, \quad \text{if } Y^* > \tau_R. \end{aligned}$$

It is assumed that Y^* has a cumulative probability distribution function, $F(Y^*)$. If Y^* has a normal distribution and the mean of $Y^* = X'\beta$, then the model is called a TOBIT model. This model can easily be estimated by the method of maximum likelihood. Other censored regression models are also readily estimated by maximum likelihood as long as the cumulative probability distribution of Y^* can be specified.

If the true dependent variable Y^* is subject to multiple, interval censoring at n points, $\tau_1 < \tau_2 < \tau_3 < \cdots < \tau_n$, and once again $F(Y^*)$ has some known or postulated distribution, then REGRESSION MODELS FOR ORDINAL VARIABLES can also be estimated by maximum likelihood. If $F(Y^*)$ is the cumulative normal distribution, the model is known as an ordinal probit model; if $F(Y^*)$ is the cumulative logistic distribution, the model is known as the ordinal logit model (Maddala, 1983).

In the same year that Tobin proposed the censored normal regression model, Kaplan and Meier (1958) proposed a product-limit estimator of $(1 - F(t))$ for right-censored data on the length of the random time T to an event (e.g., time until death) when the probability distribution of T, $F(t)$, was not specified a priori. Their proposed estimator, now known as the Kaplan-Meier estimator, which is unbiased and asymptotically normal, triggered a huge growth in treatments of censoring in biostatistical, engineering, and social scientific analyses of censored data on the time to an event, such as death or failure.

TRUNCATION

Similarly to censoring, samples may be truncated on the left, right, or doubly. A sample is truncated on the left when members of the population are excluded from a sample for values of the variable less than τ_L. A sample is truncated on the right when members of the population are excluded for values of the variable greater than τ_R. Double truncation occurs when members of the population are excluded if the values of the variable are less than τ_L or greater than τ_R (i.e., the values are not between τ_L and τ_R). In some instances, τ_L and τ_R are known a priori; in other instances, τ_L and τ_R are unknown and are estimated from the sample data, along with other sample statistics for the random variable.

Analyzing data from a truncated sample is almost always more problematic than analyzing censored data. When data are censored, one knows the fractions of the sample that have complete and incomplete measurements, as well as the value(s) at which censoring occurs. When analyzing data from a truncated sample, there are usually lingering doubts about the true distribution of the variable in the entire population. Furthermore, one rarely knows the fraction of the population that is excluded because of their values on the variable of interest. However, if the true probability distribution of the variable is known, one can estimate a model of the conditional distribution of the outcome, given that it is greater than τ_L, less than τ_R, or between τ_L and τ_R, depending on whether there is left, right, or double truncation.

Wachter and Trussell (1982) discussed estimating the distribution of heights of men in 18th-century England using historical data on recruits to the British navy, who were required to be taller than a certain level. This requirement meant that the naval recruits were a truncated sample of 18th-century Englishmen. In their application, they assumed that height has a normal distribution because a normal distribution fits data on people's heights in other, nontruncated samples very well. They dealt with other unusual problems, such as random fluctuations in the truncation level that was used, heaping at certain values, and rounding.

—Nancy Brandon Tuma

REFERENCES

Hald, A. (1952). *Statistical theory with engineering applications.* New York: John Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53,* 457–481.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics.* Cambridge, UK: Cambridge University Press.

Schneider, H. (1986). *Truncated and censored samples from normal populations.* New York: Marcel Dekker.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica, 26,* 24–36.

Wachter, K. W., & Trussell, J. (1982). Estimating historical heights. *Journal of the American Statistical Association, 77*(378), 279–293.

CENSUS

A census is an exercise in which an entire statistical POPULATION is counted. Although the term can be applied to many types of population (e.g., a traffic census counts all vehicles passing a particular point over a particular period), it is usually used, and is used here, to refer to a census of population—a demographic exercise that involves the counting of the entire population of a specified geographical area, usually a country, over a short period of time. Censuses are both CROSS-SECTIONAL (they provide results for a single point in time) and LONGITUDINAL (they provide a series of regular counts that allow the characteristics of places to be tracked over time).

Censuses are the most ubiquitous social research exercises carried out around the world. The United Nations coordinates census activities by disseminating good practice guidelines and definitions and publishing comparable population counts for most countries of the world.

The minimum information collected in a census relates to the age, sex, and household composition of the population. Further personal characteristics that are collected typically include ethnicity, employment status, and educational qualifications. In addition, many countries also collect information about housing characteristics. Some countries include other issues such as religious denomination or details of journeys to work. Financial information, particularly income, is not consistently collected around the world because of reluctance, in certain cultures, to reveal such information.

The organizational challenges of a census are formidable and involve organizing a field workforce large enough, sufficiently well trained, and sufficiently well coordinated to visit every location in a country where people may have stayed on census day or, more usually, over the night of the census. Such a workforce needs to be organized using a geographical framework that ensures that every dwelling space (which, in addition to houses, communal buildings, and hotels, must include areas where the homeless sleep on the streets or without shelter or where people work over census night) is included.

To collect census data effectively, one must have a complete list of the locations where information needs to be collected. Every dwelling and those resident at each dwelling must only be counted once or, if they are counted twice (e.g., students may be counted at both their home and their university locations), that must be noted and appropriate adjustments must be made in the total population count. This is achieved by having a geographical framework for a census.

The most frequently used framework is based on a map, preferably of large enough scale to show every potentially occupied building. In countries without adequate mapping, high-resolution satellite or aerial photography can be used. The map is divided into a number of nonoverlapping districts, each of which will be visited by a single census enumerator or a single enumeration team. These areas must completely fill the map, ensuring that every location that needs to be covered is included and that every place to be visited is allotted only once. These "enumeration districts" constitute the "input" geography of the census. Census results are also often published for enumeration districts. Modern censuses also record an accurate map reference (latitude and longitude, or local coordinates) for every house or building visited. These can either be read from the map or recorded using a Global Positioning System (GPS) receiver.

The advantage of collecting such map references, known as geo-codes, for every census location is that they allow census figures to be calculated simply (usually using a geographical information system) for any desired "output" geography. This allows the areas for which census data are published to be different from the "input" units, which need to be defined for the efficient management of the data collection teams rather than for meeting the needs of census data users.

Alternatives to systems of census geography that are based on maps are systems based on address lists and associated street maps. These require the production of a reliable "master address file" prior to the census. This needs to be complete and easy to relate to a map. A good example of this approach is the TIGER (Topologically Integrated Geographic Encoding and Referencing) system used in the United States, which is now being augmented with the Master Address file (MAF), which allows census forms to be mailed out rather than delivered by enumerators. A difficulty with address files is keeping them sufficiently up-to-date to provide a reliable base for the census. The enumeration process validates and completes such files.

Censuses need to be as accurate as possible because the results are used for an important set of legal, administrative, and research purposes. In the United States, the obligation to carry out a simple head count census is enshrined in the Constitution so that the boundaries of the areas that elect a member of Congress can be redrawn. This ensures fair elections based on equal populations electing each representative. Census results are used to allocate central government funds to local administrations based on the size of their population and the needs of that population. Censuses also provide the quantitative base for a very large number of other social research exercises by providing a "denominator." This is the figure for the base population of an area so that rates can be calculated. For example, a count of how many unemployed people are claiming benefits in an area is not very meaningful unless one can estimate what percentage of the total economically active population in that area the count represents. The census provides that base figure, allowing percentage unemployment rates to be calculated and compared reliably between areas.

Although censuses are an internationally comparable way of achieving population estimates, they are becoming more difficult to carry out, particularly in urban areas with a mobile and hard-to-count population. Censuses intrude on the PRIVACY of those filling in census forms; most citizens accept this intrusion as a necessary part of efficient government or because there may be penalties for noncompliance, but they are becoming more difficult to conduct. Census agencies apply very high standards of confidentiality to census data to satisfy the population that the invasion of their privacy will not be abused; nevertheless, it is becoming increasingly difficult to count certain parts of the population. For example, immigrants whose legal status is questionable are unlikely to willingly be counted; it also appears to be the case that in some countries, single young men are difficult to count because they are unlikely to be at home when enumerators call and have little incentive to fill in census forms.

Because of the increasing costs and difficulties in conducting censuses, many countries are investigating and implementing alternative register-based systems of tracking populations. Such registers may either be specific population registers that require citizens to register any change of address or family composition, or they may be based on other routinely collected information such as health registration, voter registration, property tax details, or drivers' licenses. It appears likely that register-based systems will become the norm in most developed countries in the near future, with censuses, or "sample censuses" (more accurately, "population surveys" based on drawing a statistical sample of the population), being used to ensure that register-based data are accurate. This will have the advantage of keeping better track of populations over time than 10 yearly, or even 5 yearly, censuses can achieve.

The census remains the "gold standard" of demographic research in most countries and is usually the single most important source of social data. However, the move to a "knowledge-based society," in which information technology is all-pervasive and information about individuals from many sources can be combined reliably and legally, will eventually make the census obsolete. Until then, it is the basis of demographic science and provides the base data on which much other social research is based.

—Robert Barr

REFERENCES

Barr, R. (1996). A comparison of aspects of the US and UK censuses of population. *Transactions in GIS, 1*, 49–60.

Dale, A., & Marsh, C. (Eds.). (1993). *The 1991 census user's guide*. London: HMSO.

National Research Council. (2001). *The 2000 census: Interim assessment*. Washington, DC: National Academy Press.

Rees, P., Martin, D., & Williamson, P. (Eds.). (2002). *The census data system*. Chichester, UK: Wiley.

United Nations Statistics Division. (2001). *Handbook on census management for population and housing censuses–Series F* (No. 83, Revision 1). New York: United Nations.

CENSUS ADJUSTMENT

The U.S. census tries to enumerate all residents of the United States, block by block, every 10 years. (A block is the smallest unit of census geography; the area of blocks varies with POPULATION density. There are about 7 million blocks in the United States.) State and substate counts matter for apportioning the House of Representatives, allocating federal funds, congressional redistricting, urban planning, and so forth. Counting the population is difficult, and two kinds of ERROR occur: *gross omissions* (GOs) and *erroneous enumerations* (EEs). A GO results from failing to count a person; an EE results from counting a person in error. Counting a person in the wrong block creates both a GO and an EE. Generally, GOs slightly exceed EEs, producing an undercount that is uneven demographically and geographically. In 1980, 1990, and 2000, the U.S. Census Bureau tried unsuccessfully to adjust census counts to reduce differential undercount using *dual-systems estimation* (DSE), a method based on CAPTURE-RECAPTURE. (Some other countries adjust their censuses using different methods.) For discussion, see the special issues of *Survey Methodology* (1992), *Journal of the American Statistical Association* (1993), *Statistical Science* (1994), and *Society* (2001). DSE involves the following:

- taking a random sample of blocks (in 2000, DSE sampled 25,000 blocks);
- trying to enumerate the residents of those blocks after census day, in the Post-Enumeration Survey or PES;
- trying to match PES records to census records, on the basis of data that often are incomplete or erroneous;
- estimating the undercount within demographic groups called *post strata* (in 2000, DSE used 448 post strata) by comparing capture-recapture estimates with census counts.

Being counted in the census is considered capture; being counted in the PES is considered recapture. The PES attempts to identify EEs and to account for movers. Ultimately, this yields an adjustment factor for each post stratum: the estimated population of the post stratum divided by the census count for the post stratum. The adjustment for each block in the country is found by separating the census count in that block into its components, post stratum by post stratum; applying the adjustment factor for each post stratum to the corresponding count; and then summing the adjusted counts to get the adjusted total for the block. This procedure is justified by the synthetic ASSUMPTION that the undercount rate is constant within each post stratum, regardless of geography. Failure of the synthetic assumption is called HETEROGENEITY.

There is another way to estimate the population, called *demographic analysis* (DA). DA estimates the population from administrative records and estimates of immigration and emigration using the identity

$$\text{population} = \text{births} - \text{deaths} + \text{immigration} - \text{emigration}.$$

Historically, DA estimates were the primary evidence of net census undercount and of differential undercount by race—motivating census adjustment.

In 1980, the U.S. Census Bureau decided not to adjust the census: Too many data were missing. In 1990, the Bureau sought to adjust, but the administration overruled the Bureau, finding that adjustment was unlikely to improve accuracy. The Bureau planned to adjust the 2000 census but in the end decided not to because DSE disagreed with DA. According to the 2000 census, there were about 281.4 million people residing in the United States. Adjustment would have added 1.2% to this number, but DA indicated that the census had found 0.7% too many people.

The U.S. census is remarkably accurate; proposed adjustments are relatively small. For example, in 2000, adjustment would have increased the population share of Texas by 0.043%, from 7.4094% to 7.4524%. That would have been the biggest state share change. Adjustment must be extremely accurate for such tiny changes to improve the census. Response errors in the census or the PES lead to problems in matching records and produce *processing errors* in DSE that are large on the relevant scale. Heterogeneity and CORRELATION bias also produce large errors.

DSE begins with capture-recapture but has layer upon layer of complexity in which details have big effects on the population estimate. For example, there are procedures for getting data by proxy interviews, searching neighboring blocks for missing records, detecting duplicate records, accounting for people who move between census day and the PES, and imputing missing data. The keys to DSE are matching PES records to census records accurately, independence, and the synthetic assumption—not counting better.

DSE would have added 3.3 million people net to the 2000 census. As of October 2001, the Census Bureau estimated net processing error in DSE to be 5 to 6 million and gross error in the DSE to be more than 12 million. In comparison, gross census error was estimated to be about 10 million (U.S. Census Bureau, 2001). Error in the adjustment is at least as large as the census error that DSE is intended to fix: Adjusting the U.S. census could easily make it worse.

—Philip B. Stark

REFERENCES

Freedman, D. A., & Wachter, K. W. (2001). *On the likelihood of improving the accuracy of the census through statistical adjustment* (Tech. Rep. 612). Berkeley: University of California, Department of Statistics.

The 1980 U.S. Census [Special issue]. (1992). *Survey Methodology, 18.*

The 1990 U.S. Census [Special issue]. (1993). *Journal of the American Statistical Association, 88.*

The 1990 U.S. Census [Special issue]. (1994). *Statistical Science, 9.*

The 2000 U.S. Census. (2001). *Society, 39,* 2–53.

U.S. Census Bureau. (2001). *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses* (With supporting documentation, Reps. 1–24). Washington, DC: Government Printing Office.

CENTRAL LIMIT THEOREM

This is a theorem that shows that as the size of a sample, n, increases, the SAMPLING DISTRIBUTION of a statistic approximates a NORMAL DISTRIBUTION even when the distribution of the values in the population are skewed or in other ways not normal. In addition, the spread of the sampling distribution becomes narrower as n increases; that is, the bell shape of the normal distribution becomes narrower. It is used in particular in relation to the sample mean: Thus, the sampling distribution of a sample mean approximates a normal distribution centered on the true population mean. With larger samples, that normal distribution will be more closely focused around the population mean and will have a smaller standard deviation. The theorem is critical for statistical inference as it means that population parameters can be estimated with varying degrees of confidence from sample statistics. For if the means of repeated samples from a population with population mean μ would form a normal distribution about μ, then for any given sample, the probability of the mean falling within a certain range of the population mean can be estimated. It does not require repeat sampling actually to be carried out. The knowledge that repeated sampling produces a normal and (for larger n) narrow distribution of sample averages concentrated around the population mean enables us to predict at varying levels of confidence just how close to the true mean our sample mean is. For large samples, the relationship will hold even if the population distribution is very skewed or discrete. This is important because in the social sciences, the distributions of many variables commonly investigated (e.g., income) are skewed. Large n can mean as little as 30 cases for estimation of a mean but might need to be more than 500 for other statistics.

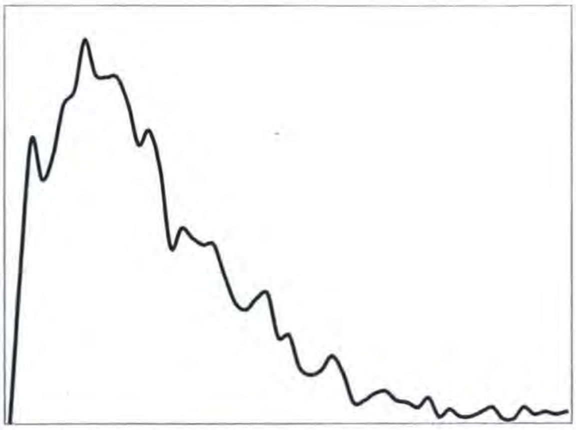

Figure 1 An Example of a Skewed Distribution

To illustrate, Figure 1 shows an example of a skewed distribution. It is a truncated British income distribution from a national survey of the late 1990s. The means of around 60 repeat random samples of sample size

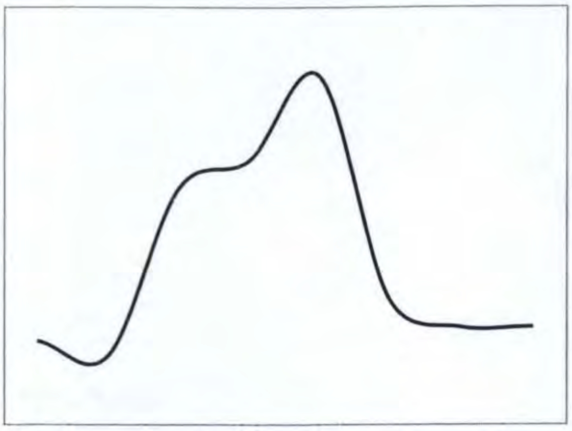

Figure 2 The Distribution of Sample Means From the Distribution Shown in Figure 1, in Which Sample Size = 10

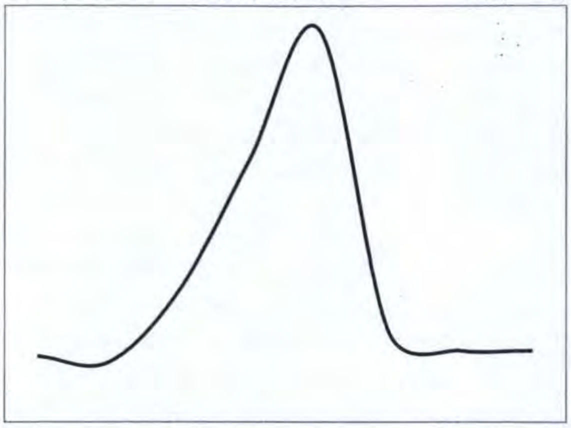

Figure 3 The Distribution of Sample Means From the Distribution Shown in Figure 1, in Which Sample Size = 100

$n = 10$ were taken from this distribution and result in the distribution shown in Figure 2. This distribution is less skewed than that in Figure 1, though not exactly normal. If a further 60 repeat random samples—this time, of $n = 100$—are taken from the distribution illustrated in Figure 1, the means of these samples form the distribution in Figure 3, which clearly begins to approximate a normal distribution.

EXAMPLE

As a result of this theorem, we can, for example, estimate the population mean of, say, the earnings of female part-time British employees. A random sample of 179 such women in the late 1980s provided a sample mean of earnings of £2.57 per hour with a sample standard deviation, *SD*, of 1.24, for the particular population under consideration. We can take the sample standard deviation to estimate the population standard deviation, which gives us a standard error of $1.24/\sqrt{179} = 0.093$. On the basis of the calculated probabilities of a normal distribution, we know that the population mean will fall within the range $2.57 \pm (1 \times 0.093)$, that is, between £2.48 and £2.66 with 68% probability; that it falls within the range $2.57 \pm (1.96 \times 0.093)$, that is, between £2.39 and £2.75 with 95% probability; and that it falls within the range $2.57 \pm (2.58 \times 0.093)$, that is, between £2.33 and £2.81 with 99% probability.

We can be virtually certain that the range covered by the sample mean, plus or minus 3 standard errors, will contain the value of the population mean. But as is clear, there is a trade-off between precision and confidence. Thus, we might be content being 95% confident that the true mean falls between £2.39 and £2.75 while acknowledging that there is a 1 in 20 chance that our sample mean does not bear this relationship to the true mean. That is, it may be one of the samples whose mean falls into one of the tails of the bell-shaped distribution.

HISTORICAL DEVELOPMENT

The central limit theorem can be traced to the year 1810 and to a proof offered by mathematician Pierre Laplace, although it was, itself, a development of Abraham De Moivre's limit theorem. The appeal of the theorem was immediate and constituted a significant step in statistical understanding. The binomial and the Poisson distributions are instances of the theorem in operation. However, a caveat was presented by the Cauchy density, named after Augustin Cauchy but also identified by Poisson. This is an instance of a distribution, with an undefined mean and standard deviation, for which the central limit theorem does not hold. The existence of this distribution indicates the existence of exceptions to the large sample rule and the fact that the theorem is based on certain assumptions about the underlying distribution, which can be shown not to be fulfilled in every case.

Nevertheless, the central limit theorem remains critical to statistical inference and estimation in that it enables us, with large samples, to make inferences based on assumptions of normality. The theorem can also be proved for multivariate variables under similar conditions as those that apply to the univariate case.

—Lucinda Platt

See also BELL-SHAPED CURVE, NORMAL DISTRIBUTION, SAMPLING DISTRIBUTION

REFERENCES

Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods.* Cambridge, MA: Harvard University Press.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics: Vol. 1. Distribution theory* (5th ed.). London: Griffin.

CENTRAL TENDENCY. *See* MEASURES OF CENTRAL TENDENCY

CENTROID METHOD

The *centroid method* is an agglomerative CLUSTERING method, in which the similarities (or dissimilarities) among clusters are defined in terms of the *centroids* (i.e., the multidimensional means) of the clusters on the variables being used in the clustering. Specifically, the centroid method is a sequential agglomerative hierarchical clustering or SAHN ALGORITHM (Sneath & Sokal, 1973), that is, a method in which all entities to be clustered are initially separate, then are combined in steps to form higher order clusters. The centroid method was first described by Sokal and Michener (1958).

The input for the centroid method is a set of N entities or "cases" to be clustered, in which each case (denoted O_i) is a multivariate observation on a set of relevant numeric descriptive variables, $X_1, \ldots, X_M$. In other words,

$$O_i = [X_{i,1}, X_{i,2}, \ldots, X_{i,M}].$$

The output of the method is an assignment of cases to a hierarchical (and nested) set of clusters. Usually, this hierarchical clustering solution is represented by a TREE DIAGRAM or "dendrogram."

DESCRIPTION OF THE ALGORITHM

As in any agglomerative clustering method, the centroid method algorithm proceeds in iterative steps. At the first step, there are N entities or cases to be clustered. However, at each step, two similar cases are combined, resulting in a cluster that is itself then treated as a new case (and the constituent cases are removed). Thus, the next step of the algorithm proceeds on the reduced set of $N-1$ cases.

At the first step, the algorithm begins by defining the dissimilarity between each pair of cases O_i and O_j as the Euclidean distance (or squared Euclidean distance) between the two cases:

$$d_{i,j} = \left[\sum_{m=1}^{M} (X_{i,m} - X_{j,m})^2 \right]^{\frac{1}{2}}.$$

Note that the values of case O_i on the variables $X_1, \ldots, X_M$ can be considered to be the centroid of a trivial cluster with one member. The result of these distance calculations is a matrix $\mathbf{D}$ of dissimilarities between each pair of cases. The first step of the sequential agglomerative algorithm continues as follows. The pair of cases O_i, O_j with the smallest dissimilarity is selected to be combined, and these two cases are combined to form a cluster that can be considered to be a new case $O_{N'}$. This new case $O_{N'}$ is added to the set of cases being clustered, and its constituent cases O_i and O_j are removed. In other words, the multivariate observations O_i and O_j are replaced by a single new case $O_{N'}$, whose data values are the centroid of the cluster $O_{N'}$:

$$O_{N'} = [\bar{X}_{N',1}, \bar{X}_{N',2}, \ldots, \bar{X}_{N',M}].$$

At this point, Step 1 of the iterative algorithm has been completed, and the next step commences. Note that the current number of cases is now $N-1$, rather than N (because two cases have been replaced by a single new cluster, $O_{N'}$). For this reduced proximity matrix, the dissimilarity of the new cluster $O_{N'}$ to each other case O_k is defined as the Euclidean distance from the centroid of the cluster to the vector of values of case O_k on the variables. That is,

$$d_{N',k} = \left[\sum_{m=1}^{M} (\bar{X}_{N',m} - X_{k,m})^2 \right]^{\frac{1}{2}}.$$

The second step proceeds with the selection of the new closest pair of cases (i.e., the pair of cases with the smallest dissimilarity) to be combined. At the second stage and later stages, either of the two cases being combined may be a cluster. The two closest cases are combined, the set of cases is reduced by one, and the proximity matrix is recomputed using the centroid of the new cluster.

These sequential steps are repeated until only two cases remain. These two cases may be combined into a single cluster containing all the cases.

WEIGHTED VERSUS UNWEIGHTED CENTROID METHOD

One complication has been omitted from the above description of the algorithm. This issue is how the centroid should be computed when the two cases being combined comprise different numbers of the original N cases. In the most common version of the algorithm, the centroid is computed using the data (equally weighted) from the original set of cases comprising that cluster. This version is referred to by Sneath and Sokal (1973) as the *unweighted pair-group centroid method*. On the other hand, if the centroid of a cluster is computed as the simple multidimensional average of the two cases being combined, without regard to the number of original cases comprising the two current clusters or cases, that has the effect of weighting individual cases in a larger cluster less than cases in a smaller cluster. This variant of the algorithm is termed the *weighted pair-group centroid method*.

An Example

A numerical example is presented in Table 1. In this hypothetical application, the data consist of math and verbal test scores for five students, and the Euclidean distance measure is used. At Step 1, the minimum distance (= 82.46) is between Cases $s03$ and $s04$. These two cases are averaged to form Cluster $c34$. At Step 2, the minimum distance between entities is between Case $s01$ and Cluster $c34$, so these are combined (using the unweighted method) to form Cluster $c134$ at Step 3. The algorithm proceeds in this manner until all cases are combined in a single cluster c^*.

Table 1

Step	*Subject*	*Math*	*Verbal*	*Minimum Distance*
1	$s01$	600	380	$d(s03, s04) = 82.46$
	$s02$	800	790	
	$s03$	760	330	
	$s04$	740	410	
	$s05$	520	780	
2	$s01$	600	380	$d(s01, c34) = 150.33$
	$s02$	800	790	
	$c34$	750	370	
	$s05$	520	780	
3	$c134$	700	373.33	$d(s02, s05) = 280.18$
	$s02$	800	790	
	$s05$	520	780	
4	$c134$	700	373.33	$d(c134, c25) = 413.61$
	$c25$	660	785	
5	c^*	684	538	

NONMONOTONICITY

Monotonicity is a property that is satisfied by many sequential agglomerative hierarchical clustering methods, such as single-link or complete-link clustering. To define monotonicity, let $O_{N'}$ be a cluster formed by combining cases O_i and O_j. A hierarchical clustering method satisfies *monotonicity* if $d_{N',k}$, the dissimilarity of the new cluster $O_{N'}$ to any other case O_k, is always equal to or greater than the two dissimilarities $d_{N',i}$ and $d_{N',j}$. The centroid method, in both its unweighted and weighted variants, does not satisfy monotonicity (Anderberg, 1973; Milligan, 1979). The method has been criticized for this failure because satisfying monotonicity ensures that the tree diagram used to represent a hierarchical cluster structure can be plotted with the height of each node representing the dissimilarity between clusters at a particular level or stage of the algorithm.

APPLICABILITY IN THE SOCIAL SCIENCES

Because the input for the centroid method is a set of multivariate observations in which the variables are assumed to be numeric, it might be applicable when the data consist of a sample of patient records or subjects measured on a test with multiple subscales. Use of the Euclidean distance or squared Euclidean distance to define the dissimilarities between clusters implies that the variables or subscales ought to be measured on comparable scales. In the social sciences, the centroid method has been used less frequently than other SAHN techniques such as Ward's method.

—James E. Corter

REFERENCES

Anderberg, M. R. (1973). *Cluster analysis for applications.* New York: Academic Press.

Milligan, G. W. (1979). Ultrametric hierarchical clustering algorithms. *Psychometrika, 44*(3), 343–346.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification.* San Francisco: Freeman.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin, 38,* 1409–1438.

CETERIS PARIBUS

Ceteris paribus is a Latin phrase meaning "other things being equal." It therefore refers to the process of comparing like with like when asserting a causal relationship or the effect of one variable on another. For example, we might be interested in ascertaining whether certain forms of training increase subsequent earnings. To identify the specific effect of the training, however, we would need to CONTROL for a number of individual characteristics such as gender, qualifications, employment history, age, and so on, as well as perhaps local labor market conditions. We would then be assessing the effect of training on individuals who were matched on these other factors, that is, who had similar qualifications, similar work history, and so forth and only differed in whether they received additional training. If there remained a significant effect of training net of these controls, we could then say that, ceteris paribus, training increased earnings. In fact, when we use statistical methods to compare like with like, we do not need to have exactly matched individuals within our sample data to be able to claim effects, other things being equal. Rather, multivariate methods estimate the effect of each different factor net of each other. Then the effect of training will be that at a fixed position on all other variables (such as men compared with men at one of a particular level of qualifications, etc.), even if some of those combinations of fixed positions do not occur within our data.

For example, we might be interested in whether women have lower earnings than men. Table 1 shows the results from LINEAR REGRESSIONS on a sample of British data from the late 1980s in terms of hourly earnings. The question is, Do women earn less per hour than men, on average? As Model 1 in Table 1 shows,

Table 1 Multiple Regression Showing Effects on Average Hourly Wage of Sex, Occupation, and Education, Britain, Late 1980s

	Model 1	*Model 2*
Constant	4.97	3.94
Sex		
(Comparison = male)		
Female	−1.96***	−1.79***
Occupation		
(Comparison = manual occupations)		
Professional		2.23***
Intermediate		0.78**
Educational qualifications		
(Comparison = no qualifications)		
Vocational and intermediate		0.16 (*ns*)
Higher qualifications		0.93***

$^{**}p < .01$; $^{***}p < .001$; *ns* = not statistically significant at the .05 level.

the answer would appear to be yes. They would appear to earn nearly £2 less per hour than men.

However, it might be argued that the reason for the lower earnings of women is due to the fact that they occupy different occupational niches than men and that their average levels of education also play a role, rather than the critical factor being sex per se. Therefore, to compare like with like, we control occupational level and educational level, which gives the results of Model 2 in Table 1. We now see that, ceteris paribus, the effect of sex is reduced to a difference of £1.79 per hour. On the other hand, occupying a professional occupation increases the earnings of men and women by £2.23 compared to being in a manual occupation. Some of the difference in hourly earnings between the sexes can therefore be attributed to different professional and educational patterns among the sexes. Only part, however. Ceteris paribus, we can still say that women are likely to earn substantially less than their male counterparts and that this difference is statistically significant.

—Lucinda Platt

See also CONTROL, MULTIVARIATE ANALYSIS

REFERENCE

Lewis-Beck, M. S. (Ed.). (1993). *Regression analysis.* London: Sage/Toppan.

CHAID

CHAID (CHi-squared Automatic Interaction Detection) is a tree-based segmentation technique useful in the exploratory analysis of a categorical dependent variable Y in terms of a large number of categorical explanatory variables. The basic algorithm works by successively splitting the sample (the initial "parent" group) into smaller and smaller subgroups that differ significantly with respect to the dependent variable, thereby growing a tree. The process continues until one of the stopping rules is triggered.

At any given point in the analysis, a given parent group is evaluated for possible splitting into two or more subgroups according to that predictor determined to be the most statistically significant (lowest p value) in its relationship to the dependent variable. In the original algorithm (Kass, 1980), the p value for a predictor is computed using the chi-squared test for independence after merging those categories that do not differ significantly with respect to the dependent variable and is adjusted using an appropriate Bonferroni multiplier.

Predictors are classified by the user as "free," "monotonic," or "floating" to establish those categories that are eligible to be merged; the merging algorithm combines categories, two at a time, in a stepwise fashion. A category of a free predictor is free to merge with any other categories of that predictor. For monotonic variables, only adjacent categories are eligible to be merged. Floating variables are treated as monotonic except for the final category, often representing missing data, which is eligible to merge with any other category.

Extensions include the following:

1. When Y is ordered categorical, the test for independence is replaced by the test for no Y association (Magidson, 1993). This test assesses the significance of differences between mean Y scores, analogous to ANOVA but without the normality assumption.

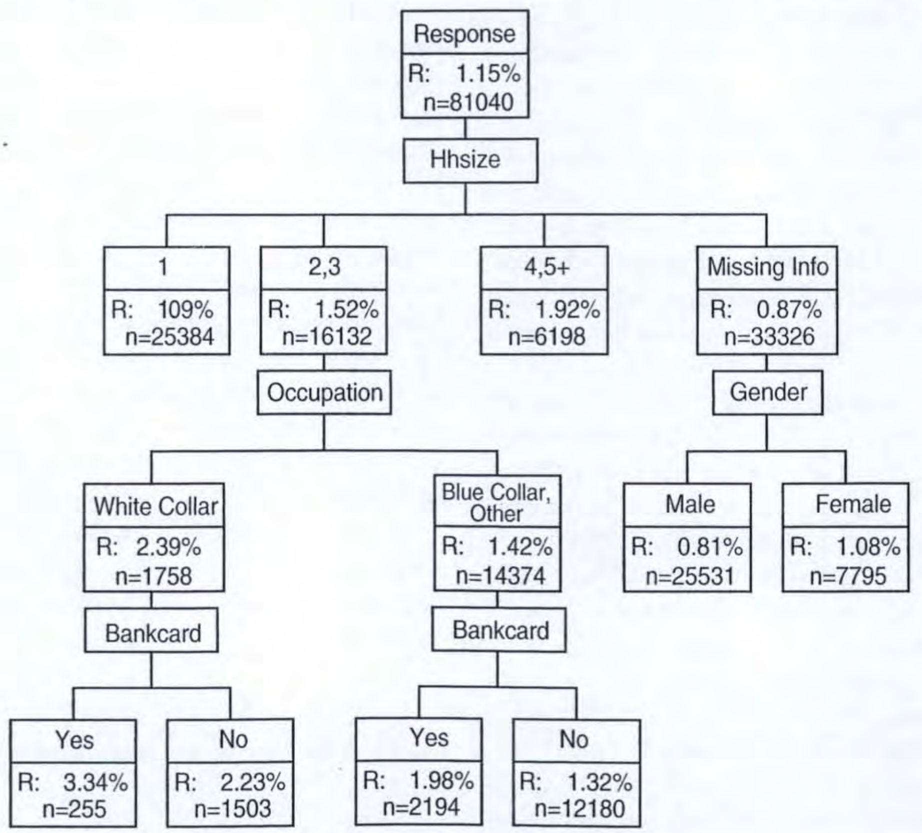

Figure 1 CHAID

2. For continuous dependent variables, the F statistic is used in place of the chi-squared statistic.
3. The merging algorithm is replaced by an exhaustive search through all possibilities (Biggs, De Ville, & Suen, 1991).
4. A hybrid CHAID adds the power of a model such as the LOGIT, or ordinal logit, to structure the tree, or latent class (LC) to extend the analysis to multiple dependent variables.

A latent class/CHAID hybrid analysis is used when there are multiple measures of response (profitability, purchase of different catalog items, quantity purchased), or in profiling the resulting segments obtained from a LATENT CLASS ANALYSIS. First, a LC analysis is performed to obtain an underlying latent variable that is passed on to CHAID as the dependent variable, which is then analyzed as a function of demographic or other exogenous variables.

Each node in the resulting tree predicts the latent class segments in addition to each of the original dependent variables, the latter obtained using LC model parameters. This type of analysis also results in a *common* splitting of the predictor categories appropriate for all the original dependent variables.

As an example of a CHAID/logit hybrid, data from Magidson (1993) was re-analyzed. The best segment using the original algorithm, HHSIZE = 2 or 3, OCCUP = "white collar," achieved a response rate of 2.39%. Figure 1 shows the result from a hybrid CHAID where the increased power of logit modeling allowed fitting a main-effect of CARD to both OCCUP nodes, resulting in an additional segment having a predicted response rate of 3.34%.

—Jay Magidson

REFERENCES

Biggs, D., De Ville, B., & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics, 18,* 49–62.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119–127.

Magidson, J. (1993). The CHAID approach to segmentation modeling: CHi-squared Automatic Interaction Detection. In R. Bagozzi (Ed.), *Handbook of marketing research* (pp. 118–159). London: Blackwell.

CHANGE SCORES

A change score is the difference between the value of a variable measured at one point in time (Y_t) from the value of the variable for the same unit at a previous time point (Y_{t-1}). Change scores are also referred to as DIFFERENCE SCORES, or with the symbol ΔY, so that

$$\Delta Y = Y_t - Y_{t-1}. \quad (1)$$

Change scores are central to the analysis of PANEL DATA, or data collected more than once on the same individuals or units over time, with the goal of most analyses being to model how changes in Y are related to levels or changes in INDEPENDENT VARIABLES over time. Change scores are also the primary DEPENDENT VARIABLE in PRETEST/posttest EXPERIMENTAL DESIGNS.

One common approach to the analysis of change scores is to model ΔY as a function of the independent variables X_1, X_2, and so forth, as well as the value of Y itself at the earlier time period $t-1$, as in

$$\Delta Y = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \varepsilon, \quad (2)$$

where the βs represent REGRESSION COEFFICIENTS and ε the equation's ERROR term. The effects of X_{t-1} and other lag values of X may also be included in the MODEL. Y_{t-1} could be added to both sides of change score equation (2) to produce an equivalent model predicting the *level* of Y at time t:

$$Y_t = \beta_0 + \beta_1 X_t + (\beta_2 + 1) Y_{t-1} + \varepsilon. \quad (3)$$

Y_{t-1}, or the "LAGGED endogenous variable," is usually included in change score models for a variety of reasons. Perhaps the most important is due to the phenomenon in LONGITUDINAL data known as REGRESSION TOWARD THE MEAN. Regression effects reflect the tendency for large values of a variable at one point in time to be associated with greater subsequent change in the *negative* direction; as the large initial values were produced in part by large RANDOM ERRORS, they are likely to be smaller in the next time period. The inclusion of Y_{t-1} may also be justified on substantive grounds if it can be argued that prior values of Y are a CAUSE of its subsequent values.

Estimation of change score models is complicated by the presence of random MEASUREMENT ERROR in Y, in which case the observed change in Y will not accurately reflect the "true change" over time. Correcting for measurement error typically requires more indicators of Y for two-wave data (i.e., additional variables that measure the same concept or LATENT VARIABLE as Y) or more waves of measurement for single-indicator models of Y (Kessler & Greenberg, 1981). Another potential problem is the presence of AUTOCORRELATED disturbances, in which the error terms for Y_t and Y_{t-1} are related, as occurs frequently in longitudinal data. In that case, Y_{t-1} will be related to the error term, and it will be necessary to estimate the model with an INSTRUMENTAL VARIABLE or related approach.

Change scores also figure prominently in models that attempt to control for FIXED EFFECTS, that is, characteristics of a given individual or unit that are unobserved by the researcher and that are stable over time. If Y_t is determined by a series of Xs, unobserved stable variables Z, and an error term, then the potential BIASING effects of Z can be eliminated by constructing a change score model, that is, by subtracting the equation for Y_{t-1} from the equation for Y_t (or, in multiwave data, subtracting the Y for each unit in each year from that unit's grand mean). If Y_{t-1} is included as an independent variable in a dynamic fixed-effect model, however, it will necessarily be related to the model's error term, with instrumental variable estimation and multiwave data being necessary to correct this potential bias (Hsiao, 1986).

—Steven E. Finkel

See also DIFFERENCE SCORES

REFERENCES

Hsiao, C. (1986). *Analysis of panel data.* Cambridge, UK: Cambridge University Press.

Kessler, R., & Greenberg, D. (1981). *Linear panel analysis.* New York: Academic Press.

CHAOS THEORY

Chaos theory refers to a type of behavior with NONLINEAR DYNAMICS that is both irregular and oscillatory. It is encountered mathematically with certain sets of nonlinear deterministic dynamic models in which patterns of overtime behavior are not repeated no matter how long the model continues to operate. Some discrete-time models using nonlinear difference equations are known to exhibit chaotic dynamics under certain conditions using only one equation. However, in continuous-time models, the possibility of chaos normally requires a minimum of three independent variables, which usually requires an interdependent system of three differential equations. This requirement for continuous-time models can be changed in special cases if the system also has a forced oscillator or time lags, in which case chaos can appear even in single-equation continuous-time models.

Chaos can occur only in nonlinear situations. In multidimensional settings, this means that at least one term in one equation must be nonlinear while also involving several of the variables. Because most nonlinear models (and nearly all of the substantively interesting ones) have no analytical solutions, they must be investigated using numerically intensive methods that require computers. Because the dimensionality and nonlinearity requirements of chaos do not guarantee its appearance, chaotic behavior is typically discovered in a model through computational experimentation that involves finding variable ranges and parameter values that cause a model to display chaotic properties.

Chaotic processes may be quite common in real physical and even social systems, even though our present ability to identify and model such processes is still developing. Discovering real chaotic processes in physical and social systems is often quite difficult because STOCHASTIC noise is nearly always present as well in such systems, and it is not easy to separate truly RANDOM behavior from chaotic behavior. Nonetheless, mathematical tools continue to be developed that are aimed at sorting out these processes and issues.

Chaos has three fundamental characteristics. They are (a) irregular periodicity, (b) sensitivity to initial conditions, and (c) a lack of predictability. These characteristics interact within any one chaotic setting to produce highly complex nonlinear variable trajectories. Irregular periodicity refers to the absence of a repeated pattern in the oscillatory movements of the chaotically driven variables. Because of the irregular periodicity, Fourier analysis, graphing techniques, and other methods are commonly used to build a case for identifying chaotic processes (see Brown, 1995a).

The nonlinear model that has been among the most well studied with regard to chaos in discrete settings is a general form of a logistic map, and its chaotic properties were initially investigated by May (1976). This general logistic map is $Y_{t+1} = aY_t(1-Y_t)$. Under the right conditions, this map can produce the standard S-shaped trajectory that is characteristic of all logistic structures. However, oscillations in the trajectory occur when the value of the parameter a is sufficiently large. For example, when the value of parameter a is set to 2.8, the trajectory of the model oscillates around the equilibrium value of $Y_{t+1} = Y_t = Y^*$ while it converges asymptotically toward this equilibrium limit. But when the value of the parameter a is set equal to, say, 4.0, the resulting longitudinal trajectory never settles down toward the equilibrium limit and instead continues to oscillate irregularly around the equilibrium in what seems to be a random manner that is caused by a DETERMINISTIC process.

The most famous continuous-time model that exhibits chaotic behavior is the so-called "Lorenz attractor" (Lorenz, 1963). This model is an interdependent nonlinear system involving three first-order differential equations, and it was originally used to analyze meteorological phenomena.

Models with forced oscillators are sometimes good candidates for exhibiting chaotic or near chaotic (i.e., seemingly random) longitudinal properties. In the social sciences, such a model has been developed and explored by Brown (1995b, chap. 6). This model is a nonlinear system of four interdependent differential equations that uses a forced oscillator with respect to a parameter specifying alternating partisan control of the White House (or other relevant governmental institution). The dynamics of the system are investigated with regard to longitudinal damage to the environment, public concern for environmental damage, and the cost of cleaning up the environment. Variations in certain parameter values yield a variety of both stable and unstable nonlinear dynamic behaviors, including behaviors that have apparent random-like properties typically associated with chaos.

—Courtney Brown

REFERENCES

Brown, C. (1995a). *Chaos and catastrophe theories.* Thousand Oaks, CA: Sage.

Brown, C. (1995b). *Serpents in the sand: Essays on the nonlinear nature of politics and human destiny.* Ann Arbor: University of Michigan Press.

Lorenz, E. N. (1963). Deterministic non-periodic flow. *Journal of Atmospheric Science, 20,* 130–141.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature, 26,* 459–467.

CHI-SQUARE DISTRIBUTION

The chi-square distribution is a distribution that results from the sums of squared standard normal variables. Even though the sums of normally distributed variables have a NORMAL DISTRIBUTION (with the shape of a BELL-SHAPED CURVE), the sums of their squares do not. The commonly used chi-square distribution refers to one type known as the central chi-square distribution. The following discussion focuses on this type. As both negative and positive values square to a positive number, the curve only covers nonnegative numbers, with its left-hand tail starting from near zero and with a skew to the right. The curve is specified by its DEGREES OF FREEDOM (*df*), which is the number of unconstrained variables whose squares are being summed. The curve has its MEAN as the *df*, and the square root of twice the *df* forms its STANDARD DEVIATION. The highest point of the curve occurs at *df*–1. As the number of *df* increases, the distribution takes on a more humpbacked shape with less skew but with a broader spread, given that its standard deviation will be increasing and that its furthest left point will remain at 0. Thus, a distribution with 5 *df* will have a mean of 5 and a standard deviation of 3.1, whereas a distribution with 10 *df* will have a mean of 10 and a standard deviation of 4.5. At high values of *df*, the curve tends toward a normal distribution, and in such (uncommon) cases, it is possible to revert to normal probabilities rather than calculating those for the chi-square distribution. The right tail probability of a chi-square distribution, which is the area in the right tail above a certain value, divided by the total area under the curve represents the likelihood that the chi-square value is randomly distributed. If the area in the tail beyond the given chi-square value represented, say, 5% of the area under the curve, then that chi-square value would be that at which we could be 95% confident that the association being tested held.

APPLICATION

These probabilities have been calculated and tabulated for certain ranges of chi-square values and *df*. Conversely, tables are commonly produced that provide the chi-square values that will produce certain probability levels at certain *df*. For example, such tables tell us that we can be 95% confident at a chi-square value of 9.49 for 4 *df*. Figure 1 shows the chi-square distribution for 4 *df* (with mean = 4 and standard

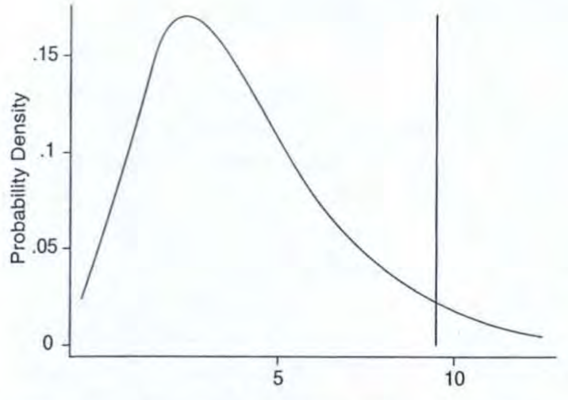

Figure 1 An Example of Chi-Square Distribution With 4 Degrees of Freedom

deviation = 2.83). The right tail probability is 5% at the point where the illustrated cutoff line takes the value of 9.49. This means that with statistics that approximate a chi-square distribution, we can compare the test statistic with such tables and ascertain its level of significance.

The most common such statistic is that resulting from Pearson's CHI-SQUARE TEST. However, this test is not the only use of the chi-square distribution. Other statistics that test the relationship between observed and "expected" or theoretical values (i.e., statistics that measure goodness of fit) approximate a chi-square distribution. For example, the likelihood ratio test statistic, also known as G^2 or L^2, which is derived from maximum likelihood method, approximates a chi-square distribution. It does this through a (natural) log transformation of the ratio between the maximum likelihood based on the null hypothesis (the –2 log likelihood) and the maximum likelihood based on the actual data values found. G^2 is a statistic commonly quoted for log-linear models. In practice, chi-square and G^2 statistics, though distinct, tend to give similar results and lead to similar conclusions about the goodness of fit of a given model, although chi-square is more robust for small samples.

EXAMPLE

For an example, Table 1 shows some data on voting behavior and class from 2 years in the British election of the late 1980s. The null hypothesis would suggest that the cell frequencies are randomly distributed. However, the chi-square for this is 84.7, and the G^2 (or likelihood ratio chi-square) is 88.4 for 4 *df*. These are clearly high values on a chi-square distribution based around

Table 1 Voting Data From Two British Elections in the Late 1980s, by Class

		Vote		
Year	*Class*	*Labor*	*Conservative*	*Total*
Year 1	Working class	50	20	70
	Middle class	10	40	50
Year 2	Working class	70	30	100
	Middle class	20	80	100
Total		150	170	320

4 *df* and, therefore, make it highly improbable that this model reflects reality. Constraining the model so that expected values of vote and class are expected to vary with each other, but so that the variation is constant across the 2 years, produces a model with Pearson and likelihood ratio chi-squares of 2.1 for 3 *df*. Clearly, such a value, when compared with the chi-square distribution having 3 *df*, is very probable (p = .54) and therefore represents a good fit with relative parsimony. The change in the chi-square and G^2 values (roughly 86 or 82 for 1 *df*) clearly represents a highly significant change.

HISTORY

The probabilities for a number of chi-square distributions were tabulated by Palin Elderton in 1901, and it was these tables that were used by Karl Pearson and his researchers in demonstrations of his chi-square test in the first decades of the 20th century. However, it is now widely acknowledged that Pearson's calculation of the numbers of degrees of freedom was misspecified, although Fisher's attempt to correct him was rejected and was only later acknowledged as being the correct solution. Pearson was therefore assessing his probability values from distributions that had too many degrees of freedom. Maximum likelihood ratios have been traced in origin to the 17th century, although it was the latter half of the 20th century that saw the real development and use of logit and log-linear models, as well as their accompanying GOODNESS-OF-FIT statistics, with the work of researchers such as Leo Goodman and Stephen Fienberg.

—Lucinda Platt

See also CHI-SQUARE TEST, DEGREES OF FREEDOM, LOG-LINEAR MODEL, LOGIT MODEL

REFERENCES

Agresti, A. (1996). *An introduction to categorical data analysis.* New York: John Wiley.

Fienberg, S. E. (1977). *The analysis of cross-classified categorical data.* Cambridge: MIT Press.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Griffin.

CHI-SQUARE TEST

The chi-square test is the most commonly used significance test for categorical variables in the social sciences. It was developed by Karl Pearson around 1900 as a means of assessing the relationship between two categorical variables as tabulated against each other in a CONTINGENCY TABLE. The test compares the actual values in the cells of the table with those that would be expected under conditions of independence (i.e., if there was no relationship between the variables being considered). Expected values are calculated for each cell by cross-multiplying the row and column proportions for that cell and taking as a share of the total number of cases considered. For example, Table 1 shows the expected counts in a hypothetical example of voters, which have been obtained on the basis of the row and column totals (known as the marginal distributions) and the total number of cases. It also shows the observed counts, which can be seen to differ from those determined by independence. In this situation of independence, voting behavior does not vary with sex.

Of course, even where there is independence in the population, the observed values in a sample are very unlikely exactly to mimic the expected values. What the chi-square test ascertains is whether any differences between observed and expected values in the cells of the table according to the sample imply real differences in the population (i.e., that the counts are not independent). It does this by comparing the actual cell counts with the cell counts expected if the proportions were consistent across each category of the explanatory variable. The value of the chi-square statistic is calculated as

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e},$$

Table 1 Expected Voting Behavior by Sex Under Conditions of Independence

	Party A	*Party B*	*Other Parties*	*Total*
Male	0.46 · 0.5 · 70,100	0.46 · 0.3 · 70,100	0.46 · 0.2 · 70,100	
Expected	16,123	9,674	6,449	32,246 (46%)
Observed	13,669	13,811	4,766	
Female	0.54 · 0.5 · 70,100	0.54 · 0.3 · 70,100	0.54 · 0.2 · 70,100	
Expected	18,927	11,356	7,571	37,854 (54%)
Observed	21,381	7,219	9,254	
Total	35,050 (50%)	21,030 (30%)	14,020 (20%)	70,100 (100%)

Table 2 Difficulty Making Ends Meet by Social Class, Britain 1987 (expected counts in parentheses)

	Easy Making Ends Meet	*Neither Easy Nor Difficult*	*Difficult Making Ends Meet*	*Total*
Professional/managerial	124 (84)	76 (86)	57 (87)	257 26%
Intermediate	73 (75)	84 (77)	73 (73)	230 23.3%
Working class	125 (163)	171 (167)	204 (169)	500 50.7%
Total	322 32.6%	331 33.5%	334 33.8%	987 100%

where f_o = the observed cell count and f_e = the expected cell count. The test then takes into account the size of the sample and the DEGREES OF FREEDOM in the table to determine whether the differences in the sample are likely to be due to chance or the probability that they are reflected in the population (i.e., that they are significant). The degrees of freedom are calculated as (number of rows − 1) × (number of columns − 1). The size of the chi-square test statistic can then be related to the CHI-SQUARE DISTRIBUTION for those degrees of freedom to ascertain the probability that the divergence of the observed from the expected values could occur within that distribution or whether it likely falls outside it.

As an example, Table 2 shows the difficulty of making ends meet tabulated by class for a sample of 987 people living in Britain in the late 1980s. The chi-square statistic resulting from comparing the observed counts in each cell with the expected counts (given in brackets) is 47.9 for 4 degrees of freedom. This figure is significant at the .000 level and thus shows that the probability that financial difficulty and social class are independent is less than 1 in a thousand.

Chi-square does not provide information about the strength of an association or its direction, only about the probability of dependence between the variables. The value of chi-square increases with sample size, and thus a low p value (high significance) in a large sample may come with a fairly weak relationship between the variables. The chi-square test also does not indicate which values of the response variable vary with the explanatory variable. It may be that only one or two cell counts deviate greatly from their expected counts. This can be ascertained by examining the table in more detail. It is not appropriate to use the chi-square test when an expected cell count for any cell in the table is less than 5. This may occur with small samples or with a large number of cells in the table. In these situations, an alternative test, such as Fisher's exact test, may be more useful.

—Lucinda Platt

REFERENCES

Agresti, A. (1996). *An introduction to categorical data analysis.* New York: John Wiley.

Everitt, B. S. (1977). *The analysis of contingency tables.* London: John Wiley.

Fienberg, S. E. (1977). *The analysis of cross-classified categorical data.* Cambridge: MIT Press.

CHOW TEST

In his article, "Tests of Equality Between Sets of Regression Coefficients in Linear Regression Models," econometrician Gregory Chow (1961) showed how the F test, or the VARIANCE RATIO TEST, which is widely used in ANALYSIS OF VARIANCE to test the homogeneity or equality of a set of means, can be used to find out the structural or parameter stability of REGRESSION models.

To explain the test, suppose we have data for the United States on savings (Y) and personal income (X), say, for the years 1970 to 1995, with Y and X being measured in billions of dollars.

Suppose we consider the following regression model, or the savings function:

$$Y_t = \beta_1 + \beta_2 X_t + u_{it}, \tag{1}$$

where Y is savings, X is income, u_{it} is the stochastic error term, and t is the time index.

One can use ORDINARY LEAST SQUARES (OLS) to estimate the parameters of the preceding savings regression. The question that the Chow test poses is this: Do the regression coefficients, β_1 and β_2, remain stable over the entire sample period? There was a severe economic recession in the United States in 1981–1982. Has the recession changed people's savings behavior? To answer the question, suppose one divides the sample data into two periods, 1970–1981 and 1982–1995, and estimates two separate regressions as follows:

$$\text{Period 1970–1981: } Y_t = \alpha_1 + \alpha_2 X_t + u_{2t}, \tag{2}$$

$$\text{Period 1981–1995: } Y_t = \lambda_1 + \lambda_2 X_t + u_{3t}, \tag{3}$$

where the us are the time-specific error terms.

If the relationship between savings and income is stable over time, one would expect that $\beta_1 = \alpha_1 = \lambda_1$ (i.e., the intercepts are statistically the same) and $\beta_2 = \alpha_2 = \lambda_2$ (i.e., the slope coefficients are the same). But how does one find that out? Chow suggests the following steps: Estimate the three regressions given above and obtain their respective RESIDUAL SUMS OF SQUARES (RSS), $S_1(df = 24)$, $S_2(df = 10)$, and $S_3(df = 12)$, respectively. Note that in all, there are 26 observations: 12 in 1970–1981 and 14 in 1982–1995.

If it is assumed that the error terms u_2 and u_3 are normally distributed, have the same variance (i.e., $\text{var}(u_2) = \text{var}(u_3) = \sigma^2$), and are independently distributed, the Chow test proceeds as follows: The assumption that the error terms are independently distributed implies that the two samples (i.e., sample periods) are independent. Hence, add S_2 and S_3 and obtain what may be called the *unrestricted residual sum of squares* (RSS_{ur}), which will have $(n_1 + n_2 - 2k)df$, where n_1 and n_2 are the number of observations in the first and second periods, and k is the number of parameters estimated in the model (two in the present instance).

Because the RSS, S_1, was obtained under the assumption that there is parameter stability (i.e., the regression coefficients do not change over the period), one can call S_1 the *restricted residual sum of squares* (RSS_{r}). This RSS will have $(n - k)df$, where $n = n_1 + n_2$. Now if there is parameter stability, RSS_{ur} and RSS_{r} should not be statistically different. But if there is no parameter stability, the two RSSs will differ. This can be seen, given the assumptions underlying the Chow test, by showing that

$$F = \frac{(\text{RSS}_{\text{r}} - \text{RSS}_{\text{ur}})/k}{(\text{RSS}_{\text{ur}})/(n_1 + n_2 - 2k)} \sim F_{[k,(n_1+n_2-2k)]}.$$

That is, the F value thus computed follows the F distribution with the stated degrees of freedom. Therefore, if the computed F in an application exceeds, say, the 1%, 5%, or 10% level of significance, one may reject the hypothesis that there is parameter stability. In the present case, that would mean that the savings function has changed over time. On other hand, if the computed F value is not significant, one can assume that there is parameter stability. In this case, there may be no harm in running just one regression pooling all the observations.

Some numerical results of the Chow test can be shown from the data on savings and income that were obtained for the United States for 1970–1995, with savings and income being measured in billions of dollars. The various sums of residuals squares discussed previously were as follows:

$$S_1 = 23{,}248.30\ (df = 24);$$
$$S_2 = 1785.032\ (df = 10);$$
$$S_3 = 10{,}005.22\ (df = 12).$$

Therefore, in the present instance, $\text{RSS}_{\text{r}} = S_1 = 23{,}248.30$ and $\text{RSS}_{\text{ur}} = S_2 + S_3 = 11{,}790.252$. Inserting these values in the previous F ratio gives the following:

$$F = \frac{(23{,}248.30 - 11{,}790.252)/2}{(11{,}790.252)/22} = 10.69.$$

Under the assumed conditions, this F value follows the F distribution with 2 and 22 *df* in the numerator and denominator, respectively. The 1% critical F value (i.e., the probability of committing a TYPE I ERROR) is 5.72. Because the computed F value exceeds this critical value, we can say that the savings-income relation in the United States over the sample period has undergone a structural change: The savings function in the two subperiods is not the same.

In using the Chow test, it must be remembered that the assumptions underlying the test must be fulfilled and that the researcher must know when a change in the relationship might have occurred. In the preceding example, if one is not sure that the savings-income relationship changed in 1982, the Chow test cannot be applied. Of course, one can choose different break points and obtain the relevant F values, thereby choosing that break point that gives the largest F value. This modification of the Chow test is known as the *Quandt likelihood ratio statistic.*

—Damodar N. Gujarati

REFERENCES

Chow, G. (1961). Tests of equality between sets of regression coefficients in linear regression models. *Econometrica, 28*(3), 591–605.

Gujarati, D. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

Wooldridge, J. (2002). *Introductory econometrics* (2nd ed.). Cincinnati, OH: South-Western College Publishing.

CLASSICAL INFERENCE

To make an inference is to draw a conclusion from a set of premises through some acceptable form of reasoning. Inferential reasoning is the foundation on which the scientific method is based. Classical inferences are embedded in a deterministic framework and follow either a deductive or inductive approach. Other forms of inference exist, but they are usually embedded in a probabilistic framework.

Within the classical framework, deductive reasoning draws a conclusion implicit in the stated premises. Inductive reasoning, however, argues from a series of specifics to a general statement. When well formulated, deductive arguments can be shown to be logically valid or "true." Inductive reasoning, on the other hand, does not entail logical closure; thus, inductive conclusions cannot be shown, incontrovertibly, to be valid. This is why some logicians reject inductive reasoning as a legitimate form of logic.

Deductive reasoning is best exemplified by the classical syllogism that consists of a major premise, a minor premise, and a conclusion. Symbolically, we may write the following:

$$\begin{array}{l} \text{If } p, \text{ then } q \\ p \\ \hline \text{therefore, } q. \end{array}$$

A discursive example is as follows: If one is human, then one is mortal; Diana is human; therefore, Diana is mortal. This form of syllogism is known as the *modus ponens.* Another logically acceptable form of syllogism is the *modus tolens,* and it takes the following form: If p, then q; $\sim q$; therefore, $\sim p$. Thus, we may state the following: If one is human, then one is mortal; Diana is not mortal; therefore, Diana is not human. In this instance, Diana might be the classical Greek goddess.

There are two other ways of writing the classical syllogism, but both entail errors in reasoning. The fallacy of affirming the consequent, for example, takes the following form: If p, then q; q; therefore, p. To continue with our substantive example: If one is human, then one is mortal; Diana is mortal; therefore, Diana is human. The error here is that being mortal is not a sufficient condition for being human. In fact, Diana might be one's pet cat.

The second false conclusion, or the fallacy of denying the antecedent, takes the following form: if p, then q; $\sim p$; therefore, $\sim q$. Again, discursively: If one is human, then one is mortal; Diana is not human; therefore, Diana is not mortal. Once more, Diana might be one's cat, but although cats are not human, they do share the characteristic of mortality with humans. These latter two invalid forms are mistakes commonly made in scientific reasoning.

It is the logical form of the modus tolens that forms the backbone of the scientific hypothesis. Let us assume that hypothesis H implies some outcome O. Observing O, no matter how often, does not support the hypothesis H. In fact, hypotheses cannot be confirmed directly because using O to support H is consistent with the fallacy of affirming the consequent. Not observing O, however, provides a negation of H or a rejection of the hypothesis. Thus, as the philosopher Karl Popper noted, the hallmark of the scientific method is its ability to disprove, not prove, hypotheses.

Although deduction forms a closed system of logic, induction is an open system. Induction generally argues that because all observed instances of P have characteristic C, then *all* instances of P have characteristic C. As Hume so clearly noted, however, the statement that all observed As are Bs is not inconsistent with the statement that some unobserved As are not Bs. Proponents of induction argue that although things cannot be proven inductively, induction takes on more credence as the population becomes increasingly enumerated. Thus, the closer one comes to observing all instances of P where those instances have characteristic C, then the more credence we place in the statement that all instances of P have characteristic C. This rationale is appealing when P is a closed population; it is difficult to abide when P is open or infinite.

The difficulties surrounding logical inference do not necessarily imply that it has no use in science. Often, hypotheses are generated through processes of inference—something we call the context of discovery. It is also possible to draw credible inferences by generalizing from samples to populations outside a deterministic framework. This leads to the art known as INFERENTIAL STATISTICS, in which, given certain constraints, the likelihood of a valid inference can be estimated.

—Paul S. Maxim

REFERENCES

Copi, I. M., & Cohen, C. (1990). *Introduction to logic* (8th ed.). New York: Macmillan.

Gensler, H. J. (2002). *Introduction to logic.* London: Routledge Kegan Paul.

Hempel, C. G. (1966). *Philosophy of natural science.* Englewood Cliffs, NJ: Prentice Hall.

Poundstone, W. (1988). *Labyrinths of reason: Paradoxes, puzzles and the frailty of knowledge.* New York: Anchor/Doubleday.

CLASSIFICATION TREES. *See* CART

CLINICAL RESEARCH

Clinical research is inquiry whose purpose is the creation of new knowledge and action in the context of health care. As with all forms of scientific inquiry, clinical research concerns itself with identification, description, the generation of EXPLANATIONS, the testing of explanations, and/or intervention/CONTROL. These concerns, in turn, are linked to typical methods. For the first three—exploratory investigations—qualitative methods of OBSERVATIONS, INTERVIEWS, and DOCUMENT analysis are commonly used, as are variations of surveys. Explanation testing, although part of qualitative traditions, is more commonly practiced in epidemiological research, with COHORT, CROSS-SECTIONAL, and case control designs. For example, an EXPERIMENTAL design called a RANDOMIZED CONTROL TRIAL (RCT) tests the efficacy or effectiveness of treatment (Miller & Crabtree, 1999). The health disciplinary literature, particularly in medicine, is still dominated by experimental and epidemiological designs. This distribution, in turn, reflects funding priorities for basic and applied biomedical research in developed countries. Similar models also dominated social sciences relevant to clinical practice, such as psychology and sociology, in the early years of these disciplines. However, there are now established qualitative and critical streams in these and other social sciences, and new inquiry models are applied to the questions of the clinic. In an era of "evidence-based medicine," qualitative investigators remind us that the care of patients must include many kinds of evidence—those that derive from randomized control trials and those that derive from being acquainted with the particulars of patients' lives, perspectives, and social context (Morse, Swanson, & Kuzel, 2001).

One of these newer models of clinical inquiry is PHENOMENOLOGY, which focuses on understanding and communicating the essence of experiences and intentions of individuals within their "life-world." In contrast to disease-oriented descriptions, phenomenological reports portray the illness experience. For example, Pelusi (1997) interviewed breast cancer survivors and found common themes of not only loss and an uncertain future but also growth and enlightenment. Another is GROUNDED THEORY, well suited to the evaluation of social processes over time, with a goal of generating theory that is grounded in careful observation, interview, and document analysis. An example is Lorencz's (1991) study of patients with schizophrenia who transition from inpatient to community settings—a process of "becoming ordinary." Anthropology's basic tradition of ETHNOGRAPHY has also seen wide application to the work of the clinic. This includes not only the cultural determinants of health and illness (Kleinman, Eisenberg, & Good, 1978) but also the culture of health care organizations and its impact on the delivery of clinical services. A high-profile application of an ethnographic sensibility is to the widespread problem of "medical error," resulting in a growing appreciation that a culture of "blame and shame" is a powerful deterrent to identifying and understanding the root causes of medical errors (Kohn, Corrigan, & Donaldson, 2000). A final example is PARTICIPATORY ACTION RESEARCH, with an interest in understanding and promoting social change through the "democratization" of inquiry. In this model, those who would typically be "subjects" of medical research act instead as coinvestigators. Macaulay and colleagues (1997), in a Mohawk community near Montreal, focused on diabetes prevention through healthy eating and more exercise among children. They employed both behavioral change models and knowledge of native learning styles to design interventions.

—Anton J. Kuzel

REFERENCES

Kleinman, A. L., Eisenberg, L., & Good, B. J. (1978). Culture, illness and care: Clinical lessons from anthropologic and cross-cultural research. *Annals of Internal Medicine, 88,* 251–258.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To err is human: Building a safer health system.* Washington, DC: National Academy Press.

Lorencz, B. (1991). Becoming ordinary: Leaving the psychiatric hospital. In J. M. Morse & J. Johnson (Eds.), *The illness experience: Dimensions of suffering* (pp. 140–200). Newbury Park, CA: Sage.

Macaulay, A. C., Paradis, G., Potvin, L., Cross, E. J., Saad-Haddad, C., McComber, A., et al. (1997). The Kahnawake Schools Diabetes Prevention Project: Intervention, evaluation, and baseline results of a diabetes primary prevention program with a native community in Canada. *Preventive Medicine, 26*(6), 79–90.

Miller, W. L., & Crabtree, B. F. (1999). Clinical research: A multimethod typology and qualitative roadmap. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (2nd ed., pp. 3–32). Thousand Oaks, CA: Sage.

Morse, J. M., Swanson, J. M., & Kuzel, A. J. (Eds.). (2001). *The nature of qualitative evidence.* Thousand Oaks, CA: Sage.
Pelusi, J. (1997). The lived experience of surviving breast cancer. *Oncology Nursing Forum, 24*(8), 1343–1353.

CLOSED QUESTION. *See* CLOSED-ENDED QUESTIONS

CLOSED-ENDED QUESTIONS

Survey questions come in two varieties: OPEN-ENDED QUESTIONS, in which the respondents provide their own answers, and closed-ended questions, in which specific response categories are provided in the question itself. Although there has been considerable research on the relative merits of the two types of questions, the substantial preponderance of questions that appear in any survey are closed-ended questions.

All survey research involves asking questions for which the responses are categorized to facilitate analysis. In closed-ended questions, the researcher makes prior judgments about what the appropriate categories might be and offers them immediately to the respondent in the question wording. There are a number of potential problems with this format, not the least of which is that it typically describes the "world" in dichotomous terms that sometimes reflect an oversimplification of possibilities. This kind of constraint does not exist with an open-ended question, and for this reason Schuman and Presser (1996) conclude that they often provide more valid data. However, the cost saving of not having to CODE verbatim responses and the quicker access to analysis make close-ended questions the preferred form.

When designing or using close-ended questions, a number of standard suggestions reflect considerable research on the matter. Using forced choices is preferable to asking a respondent to agree or disagree with a single statement. A middle alternative should generally be offered, except when measuring intensity, and an explicit "no opinion" option should be offered as well. Researchers should use multiple questions to assess the same topic, remaining sensitive to the effects of question order. And when in doubt about the possible effects of question wording or response options, research should include split-sample versions in their questionnaires so that comparative analyses can be run.

—Michael W. Traugott

REFERENCES

Converse, J. M., & Presser, S. (1986). *Survey questions.* Thousand Oaks, CA: Sage.
Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation.* Thousand Oaks, CA: Sage.
Schuman, H., & Presser, S. (1996). *Questions & answers in attitude surveys.* Thousand Oaks, CA: Sage.

CLUSTER ANALYSIS

Social science DATA sets usually take the form of observations on UNITS OF ANALYSIS for a set of VARIABLES. The goal of cluster analysis is to produce a simple classification of units into subgroups based on information contained in some variables. The vagueness of this statement is not accidental. Although there may be no formal definition of cluster analysis, a slightly more precise statement is possible. The *clustering problem* requires solutions to the task of establishing clusterings of the n units into r clusters (where r is much smaller than n) so that units in a cluster are similar, whereas units in distinct clusters are dissimilar. Put differently, these clusterings have homogeneous clusters that are well separated. *Cluster analysis* is a label for the diverse set of tools for solving the clustering problem (see Everitt, Landau, & Leese, 2001). Most often, these tools are used for INDUCTIVE explorations of data. The hope is that the clusterings provide insight into the structure of the data, the nature of the units, and the processes generating the variables. For example, cities can be clustered in terms of their social, economic, and demographic characteristics. People can be clustered in terms of their psychological profiles or other attributes they possess.

DEVELOPMENT OF CLUSTER ANALYSIS

Prior to 1960, many clustering problems were solved separately in different disciplines. Progress was fragmented. The early 1960s saw attempts to provide general treatments of cluster analysis, given these many developments. Sokal and Sneath (1963) provided an extensive discussion and helped set the framework for the development of cluster analysis as a data-analytic field. Specifying clustering problems is not difficult. Nor are the mathematical foundations for expressing and creating most solutions to the clustering problem. The difficulty of cluster analysis comes from the *computational complexities* in *establishing* solutions to the clustering problem. As a result, the field has been driven primarily by the evolution of computing technology. Generally, this has been beneficial, with substantive interpretations being enriched by useful clusterings. In addition, many technical developments have stemmed from exploring substantive applications in new domains. There are now many national societies of cluster analysts that are linked through the International Federation of Classification Societies.

SOLVING CLUSTERING PROBLEMS

In general, the clustering problem can be stated as establishing one (or more) clustering(s) with r clusters that have the minimized value of a well-defined criterion function over all feasible clusterings. The criterion function provides a measure of fit for all clusterings. In practice, however, the criterion function often is left implicit or ignored. In most applications, the clustering is a partition, but "fuzzy clustering" with overlapping clusters is possible. Once the units of analysis have been selected, there are five broad steps in conducting cluster analyses:

1. measuring the relevant variables (both QUANTITATIVE VARIABLES and CATEGORICAL variables can be included, and some form of standardization may be necessary),
2. creating a (dis)similarity MATRIX for an appropriate measure of (dis)similarity,
3. creating one or more clusterings via a clustering algorithm,
4. providing some assessment of the obtained clustering(s), and
5. interpreting the clustering(s) in substantive terms.

Although all steps are fraught with hazard, Steps 2 and 3 are the most hazardous, and Step 4 is ignored often. In Step 2, dissimilarity measures (e.g., Euclidean, Manhattan, and Minkowsky distances) or similarity measures (e.g., CORRELATION and matching COEFFICIENTS) can be used. The choice of a measure is critical: Different measures can lead to different clusterings. In Step 3, there are many ALGORITHMS for establishing clusterings. Each pair of choices (of measures and algorithms), in principle, can lead to different clusterings of the units.

Hierarchical clustering can take an agglomerative or a divisive form. An agglomerative clustering starts with each unit in its own cluster and systematically merges units and clusters until all units form a single cluster. Divisive hierarchical clustering proceeds in the reverse direction. Within these categories, there are multiple options. Three of the most popular are single-link, average-link, and complete-link clustering. In single-link clustering, the algorithm computes the (dis)similarity between groups as the (dis)similarity between the closest two units in the two groups. For complete-linkage clustering, the farthest pair of units in the two groups is used, and in average-linkage clustering, the algorithm uses the average (dis)similarity of units between the two groups. Ward's method is popular also for computing ways of combining clusters. The choice between these methods is critical, as shown in Figure 1.

The first panel of Figure 1 shows a BIVARIATE scattergram whose 25 units can be classified. Squared Euclidean distance was used for all three clusterings and dendrograms shown below. For this example, if (x_i, y_i) and (x_j, y_j) are the coordinates of two units, the squared Euclidian distance between them is $(x_i - x_j)^2 + (y_i - y_j)^2$. The top right panel shows the single-linkage clustering dendrogram. Those in the bottom row of Figure 1 are for the average- and complete-linkage methods. The scale on the left of the dendrogram shows the measure of dissimilarity, and the horizontal lines show when units or clusters are merged. The vertical lines keep track of the clusters. The average- and complete-linkage clusterings are the closest to each other, and both suggest clusterings with three clusters. However, the clusters differ in their details, and the single-link clustering differs markedly from the other two. Which clustering is "better" can be judged by examining the clusterings and the scattergram—with some idea of how and why units can be grouped. For real analyses, substance guides this judgment.

Because clustering tools are exploratory, there are few tools for STATISTICAL INFERENCE concerning the VALIDITY of a clustering (Step 4), and their utility is limited because of the prior exploration. Choosing the number of clusters from a dendrogram often is viewed as a matter of judgment, and ad hoc justifications for any clustering are easy to reach given a clustering. This, too, is illustrated in Figure 1.

When the number of clusters is known or assumed, some nonhierarchical clustering methods are available.

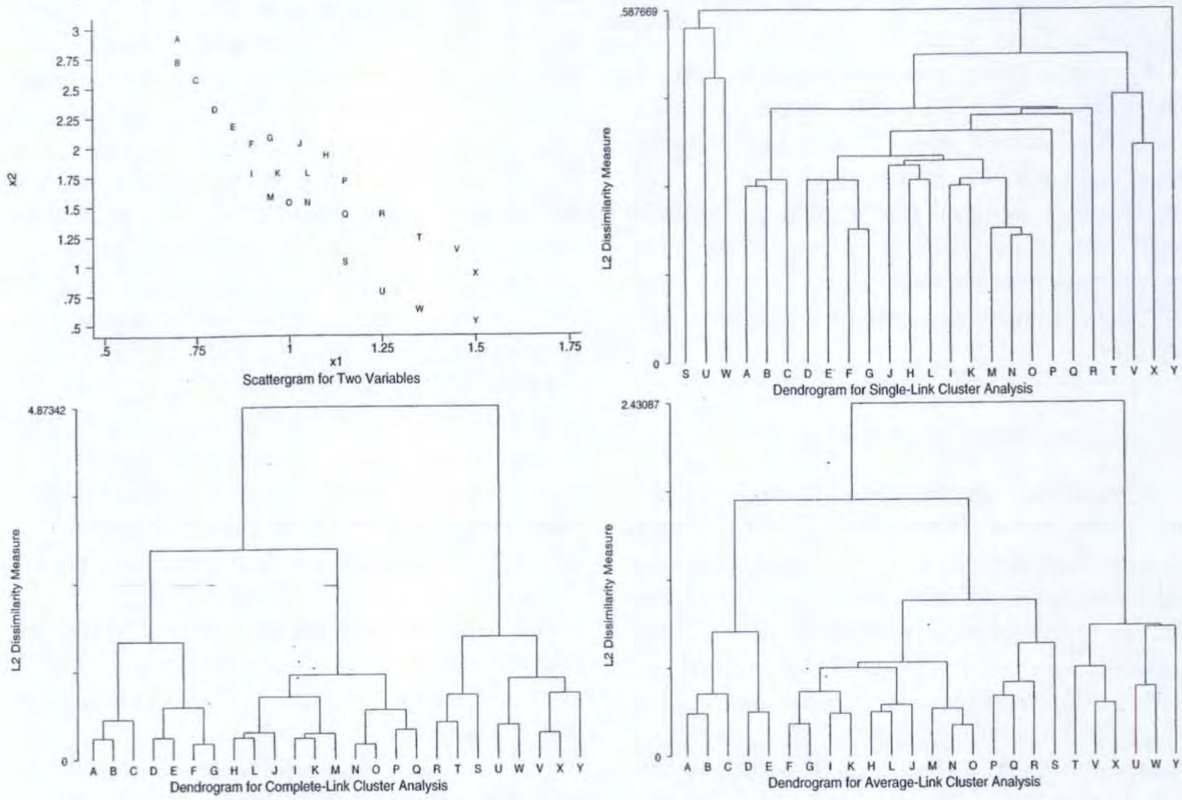

Figure 1 Three Clusterings of a Common Data Set

These include k-means and k-medians clustering. Two additional methods are the leader and relocation algorithms. Both are local optimization methods and have to be repeated many times to avoid reaching a local minimum. The relocation algorithm is at the heart of the direct clustering approach to block modeling social networks developed by Doreian, Batagelj, and Ferligoj (1994). Both network actors *and* the network ties are clustered, whereby a criterion function is defined *explicitly* in terms of substantive concerns and the NETWORK ties. As a result, the resulting solutions to the clustering problem are not ad hoc.

—Patrick Doreian

REFERENCES

Doreian, P., Batagelj, V., & Ferligoj, A. (1994). Partitioning networks based on generalized concepts of equivalence. *Journal of Mathematical Sociology, 19,* 1–27.

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). Oxford, UK: Oxford University Press.

Sokal, R., & Sneath, P. (1963). *Principles of taxonomy.* San Francisco: Freeman.

CLUSTER SAMPLING

Cluster sampling involves sorting the units in the study population into groups and selecting a number of groups. All the units in those groups are then studied. It is a special case of MULTISTAGE SAMPLING.

—Peter Lynn

COCHRAN'S *Q* TEST

W. G. Cochran (1950) developed the Q statistic for matched or WITHIN-SUBJECT DESIGNS in which each subject (r) provides a dichotomous response for each experimental condition (c). Cochran's Q tests whether the probability of a target response is equal across

Table 1 Sample Data for the Cochran's Q: Tests of Varying Levels of Complexity

	Test 1	*Test 2*	*Test 3*	*Test 4*	
Subject 1	1	0	0	1	$u_1 = 2$
Subject 2	1	1	1	0	$u_2 = 3$
Subject 3	1	0	0	0	$u_3 = 1$
Subject 4	1	1	1	1	$u_4 = 4$
Subject 5	1	0	1	0	$u_5 = 2$
Subject 6	1	1	0	0	$u_6 = 2$
	$T_1 = 6$	$T_2 = 3$	$T_3 = 3$	$T_4 = 2$	$\bar{T} = \frac{\sum_{j=1}^{c} T_j}{c} = 3.5$

$$Q = \frac{c(c-1)\sum_{j=1}^{c}(T_j - \bar{T})^2}{c(\sum_{i=1}^{r} u_i) - (\sum_{i=1}^{r} u_i^2)} = \frac{4(3)[(6-3.5)^2 + (3-3.5)^2 + (3-3.5)^2 + (2-3.5)^2]}{4(2+3+1+4+2+2) - (2^2+3^2+1^2+4^2+2^2+2^2)} = 6$$

95th percentile for $\chi^2_{(4-1)} = 7.815$

all conditions. For example, if 6 subjects ($r = 6$) take four tests ($c = 4$) that are either "failed" or "passed," Cochran's Q can assess whether the probability of passing differs across the four tests. Of course, the dependent variable can be any dichotomy, such as pass-fail, presence-absence, hit-miss, or increase-decrease.

Cochran's Q can easily be computed with data arranged in a matrix of r rows and c columns. Each row in the matrix represents a case containing c elements that are coded as "1" or "0" to represent the presence or absence of a target response. The number of target responses are summated across rows (u_i) and columns (T_j) and are used to compute Cochran's Q in the following equation:

$$Q = \frac{c(c-1)\sum_{j=1}^{c}(T_j - \bar{T})^2}{c(\sum_{i=1}^{r} u_i) - (\sum_{i=1}^{r} u_i^2)},$$

where T_j is the total of the jth column, u_i is the total of the ith row, and $\bar{T} = \frac{\sum_{j=1}^{c} T_j}{c}$.

Cochran's Q is distributed as a χ^2 with $c - 1$ DEGREES OF FREEDOM. As shown in Table 1, data from 6 subjects who each took four tests are analyzed with the Cochran's Q. The observed Q statistic is then compared to a χ^2 using the desired TYPE I ERROR rate (.05 in this case) and degrees of freedom. In the example, the Q statistic does not exceed the critical χ^2. Thus, the NULL HYPOTHESIS is retained, suggesting that the probability of passing the test does not differ across conditions.

Cochran's Q requires the assumption that cases are randomly and independently sampled from the population. Homogeneity of covariance is also required; the covariances of all possible pairs of columns are assumed to be equal. If the assumption of homogeneity of covariance is violated, Myers and Well (1995) recommended using the ε-adjustment developed by Greenhouse and Geisser (1959). The corrected Q is the following: $Q^* = \varepsilon Q$.

Degrees of freedom for the adjusted test are $\varepsilon(c-1)$. As before, Q^* is evaluated against the desired χ^2 with the adjusted degrees of freedom. The computation of ε is laborious and is omitted here. However, it is given by many statistical analysis programs when conducting repeated-measures ANALYSIS OF VARIANCE.

In using the Cochran's Q, sample sizes of 16 or more are advisable (the sample size in the table is only for illustration). Cases that exhibit no variance across conditions (i.e., all zeros or ones) do not contribute to Q. The repeated-measures analysis of variance can be considered as an alternative to Q, although large samples are needed to ensure normality of the sampling distribution of the dependent variable, and the ε-adjustment may be needed to correct for heterogeneity of covariance.

—James W. Griffith

REFERENCES

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika, 37*, 256–266.

Greenhouse, G. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 55*, 431–433.

Myers, J. L., & Well, A. D. (1995). *Research design and statistical analysis.* Hillsdale, NJ: Lawrence Erlbaum.

CODE. *See* CODING

CODEBOOK. *See* CODING FRAME

CODING

Coding is the process by which verbal data are converted into VARIABLES and categories of variables using numbers, so that the data can be entered into computers for analysis. DATA for social science research are collected using SELF-ADMINISTERED QUESTIONNAIRES and questionnaires administered through telephone or face-to-face interviews, through observation, and from records, documents, movies, pictures, or tapes. In its original "raw" form, these data comprise verbal or written language or visual images. Although many data entry programs can handle at least limited amounts of linguistic data, for purposes of most analyses, these data must eventually be converted to variables and numbers and entered into machine-readable data sets. Each variable in a data set (e.g., sex or gender) must consist of at least two categories, each with a unique code, to be a variable. It must have the possibility of varying; if it does not vary, then it is a constant. Thus, a study could include both men and women, where gender or sex is a variable, or it could include only men or only women, where gender is a CONSTANT. Numeric codes for gender can be assigned in both situations.

A single unique code can consist of a single digit or multiple digits, with the functional upper number of digits being 8 or 10. A code consisting of multiple digits may have multiple variables embedded within it. For example, respondents were asked after earthquakes whether various utilities were unavailable, including water, electricity, gas, and telephones. Four two-category variables can be created for each utility, using 1 for "yes, it was off" and 2 for "no, it was not off." But the four single-digit variables can also be combined to create a four-digit variable that creates all the possible combinations of the four variables. An example would be the code of 1221, which would indicate that water and telephones went off but electricity and gas stayed on.

The numbers assigned can be meaningful in and of themselves, or they can function simply as a shorthand representation of the verbal data. If, for example, data on age are collected, the age recorded for each person, institution, state or country, or object is meaningful, and statements can be made about the relative ages of people, states, or objects. If, in contrast, data on religious affiliations of people, households, or institutions are collected, the number assigned to each religion is arbitrary—for example, 1 for Catholics, 2 for Buddhists, and 3 for Hindus.

LEVELS OF MEASUREMENT

In developing codes for verbal data, the concept of LEVELS OF MEASUREMENT is often used (Stevens, 1968). There are four levels of measurement: NOMINAL, ORDINAL, INTERVAL, and RATIO. As we move from nominal data to ratio data, the computations and statistics that can be used during analysis become more diverse and sophisticated. When creating codes for data, it is important to understand and consider these differences.

Codes for religion are considered nominal. Nominal codes are arbitrary, shorthand space holders. In creating codes for nominal data, the researcher is concerned with creating exhaustive and mutually exclusive codes. *Exhaustive* means that a unique code number has been created for each category of the variable that may occur in the data. Thus, in creating codes or code frames for a variable on religious affiliation, it may be necessary to create unique code numbers for individuals, institutions, or groups that have no religious affiliation, as well as for those that identify as agnostics or atheists. *Mutually exclusive* means that the information being coded about each person, institution, object, or grouping can be clearly assigned to only one category of the variable. So, for example, it is clear that Buddhists always get a code of 2 and never get a code of 1 or 3.

The categories of a nominal variable have no meaningful rank order, and the distance between the categories cannot be calculated. The only measure of CENTRAL TENDENCY that can be calculated appropriately for a nominal variable is the mode. Many of the most important variables included in social science research are naturally nominal. These include sex or gender, religion, language, ethnic identification, and marital status.

Ordinal variables are considered a higher level of measurement. The categories of an ordinal variable can be meaningfully rank ordered. So, for example, a person who drinks one pint of water drinks less water than a person who drinks two quarts who, in turn, drinks less water than a person who drinks two gallons, and they can be assigned codes, respectively, of 1, 2, and 3. The person with a code of 3 clearly drinks more water than those with codes 1 or 2, but nothing can be said about the relative distance between the three people when those three codes are used in analysis. Both modes and MEDIANS can be calculated for ordinal variables, but it is technically incorrect to add or subtract scores or to calculate MEAN scores. One of the problems with assigning codes to variables that are naturally ordinal is developing codes that are both mutually exclusive and exhaustive. In the above example, the codes are mutually exclusive but are not exhaustive. Note there is no code for a person who drinks no water, three gallons of water, or one quart of water.

Age can be regarded as an instance of an interval variable, which requires a scale with equidistant (or constant) numerical categories. Consider the answer to the question, "How old were you on your last birthday?" With interval variables, the categories can be rank ordered, and one can talk about the equal distance between the categories. For instance, interval variables have an arbitrary zero, the categories can be rank ordered, and we can talk about the distance between categories. For example, the distance between 24 and 26 is the same as the distance between 33 and 35. In addition to the mode and median, statistics such as the mean, STANDARD DEVIATION, and CORRELATION are meaningful and can be interpreted.

Ratio variables differ from interval variables in that they have a meaningful zero and are completely CONTINUOUS, meaning that every point on the scale exists. Weight and height are often considered ratio variables. Ratio variables are also called continuous variables, whereas the other three types of variables are sometimes called DISCRETE variables. For purposes of most social science research, the distinction between interval and ratio is largely arbitrary because most of the statistical analyses that can be done on ratio variables can also be done on interval variables. In contrast, analysis techniques that are appropriate for interval and ratio variables technically should not be used on nominal or ordinal variables.

DETERMINING CODES FOR VARIABLES

Codes for data can be developed during the design of the study or after the data are collected. When structured questionnaires or abstraction forms are used to collect data on a finite number of variables and most variables have a finite number of possible answer categories, codes should be developed as the data collection instrument is being finalized. When OBSERVATIONS, FOCUS GROUPS, SEMI-STRUCTURED INTERVIEWS, or DOCUMENTS are the source of data, codes generally will be developed inductively using CONTENT ANALYSIS or other procedures after data collection is completed. The development of codes for CLOSED-ENDED QUESTIONS in a questionnaire will be used as an example for the first kind of code development, whereas the development of codes for OPEN-ENDED QUESTIONS will be used as an example for the second kind.

A number of things should be considered when developing codes. As noted above, codes and the associated answer categories should be both *exhaustive* and *mutually exclusive;* that is, there should be a code for each answer that might be given, and RESPONDENTS, data collectors, and data entry personnel should have no difficulty determining which answer fits the situation being described and which number or code is associated with that answer. For example, after the Loma Prieta earthquake, the authors (Bourque & Russell, 1994; Bourque, Shoaf, & Nguyen, 1997) asked respondents, "Did you turn on or find a TV or radio to get more information about the earthquake?" and provided the responses of "Yes, Regular TV," "Yes, Battery TV," "Yes, Regular Radio," "Yes, Battery Radio," and "No," with preassigned codes of 1, 2, 3, 4, and 5. Interviewers asked respondents to select a single answer.

Although the list of answers and codes appeared both exhaustive and mutually exclusive, it was not. First, many respondents sought information from more than one form of electronic media and, second, many respondents sought information from their car radio. The question-and-answer categories were changed to add car radios to the list, and respondents were allowed to select more than one answer. If that had not been done, data would have been lost.

When multiple information is recorded or coded for what is technically one variable, each response category becomes a variable with two codes or values, mentioned or not mentioned. In the example, what started out as one variable with five possible answers became six interrelated variables, each with two possible answers or codes.

Residual other categories provide another way to ensure that the CODE FRAME is exhaustive. "Residual other" is used to deal with information that the researcher did not anticipate when the code frame was created. In the same study, the author provided a list of ways in which houses might have been damaged by the earthquake, including collapsed walls, damaged roofs, and so forth. What was not anticipated were the elevated water towers that are used by residents of the hills outside Santa Cruz, California. Because a residual other was included, when respondents said a water tower had been damaged, interviewers circled the code of 19 for "Other" and wrote "water tower" in the space provided, as demonstrated here:

Other... 19
SPECIFY: ______________

When data are entered into the computer, both the code 19, which indicates that an unanticipated answer was given, and the verbal description of that answer, water tower, are typed into the data set.

Reviews of past research and PRETESTING of data collection procedures can help researchers determine if the answer options and associated codes are both mutually exclusive and exhaustive, but sometimes social change will cause well-tested code schemes to become inaccurate, incomplete, or even completely obsolete. For example, the standard categories used to code marital status are "never married," "married," "divorced," "separated," and "widowed," but increasingly, both different- and same-sex couples are cohabiting. According to the 2000 U.S. census, 5.2% of households in the United States currently describe themselves as "unmarried partner households." In a study of social networks, for example, failure to consider and account for this change may result in serious distortion of data.

Deciding on the categories and codes to use for income, education, and age is another place where problems occur. Generally, income is coded as an ordinal variable. Categories are developed such as the following: less than $10,000, $10,000–$19,999, $20,000–$29,999, and so forth. The problem comes in deciding how broad or narrow each category should be and what the highest and lowest categories should be. If categories are too broad, the top category set too low, or the bottom category set too high, heaping can occur. *Heaping* is when a substantial proportion of the data being coded falls into a single category. For example, if undergraduate college students are being studied and age is set in ordinal categories of 18–21, 22–24, and 25–29 with associated codes of 1, 2, and 3, the overwhelming majority of students will be 18 to 21, and they will be assigned a code of 1. When this happens, the variable becomes useless because it no longer has any VARIANCE.

Another problem with creating ordinal variables out of data (e.g., age, education, and income) that are naturally at least interval in underlying structure is that it may restrict data analysis. If age is coded into categories of 18–24, 25–29, 30–34, and 35–39, and so forth, the researcher can determine what the median age category is but can never calculate the mean age. Furthermore, writing out the age categories in a questionnaire or data collection form takes substantially more room than simply asking for age at last birthday or birth date. In this case, a *short open-ended* question in a data collection form may obtain just as much data—data that are just as valid and, at the same time, yield data that are more easily analyzed.

Similar problems occur when researchers use answer categories such as "none," "some," and "a lot." Although *none* is pretty clear, what do *some* and *a lot* mean, and how should they be coded? If this is a question about how many people a respondent knows who were injured in a flood, why not ask instead, "How many people in all do you know who were injured in the flood?" The information collected will result in data that more closely resemble interval data, and it will not be subject to the variety of ways in which terms such as *some* and *a lot* will be interpreted.

When coding data, it is as important to record that information is missing as it is to record the data that are there. Standard MISSING DATA conditions include "not applicable," "don't know," "refused," and "missing." Codes for "don't know" and "refused" are more frequently used when data are collected directly from people, whereas codes for "missing" are extremely important when data are being abstracted out of records or other documents. "Not applicable" codes are used in both situations. "Not applicable" is appropriately used when some data are not relevant in certain situations. Often, the need for "not applicable" occurs because the data being coded are dependent on some prior information that results in the current data being irrelevant. For example, men are not asked about how they felt when they were pregnant, and it is irrelevant to look for hemoglobin scores in medical records if no blood was drawn.

Often, when data are being abstracted out of documents or records, data instructors are instructed to look for particular kinds of information and then record it using a code. If CONTENT ANALYSIS is being used to code the content of the front page of eight major

U.S. newspapers over a period of a month, and data collectors are supposed to record all articles that reference the president and code for the "tone" of the article (e.g., whether it is complimentary, critical, analytical, or whatever), the fact that no reference to the president occurs on a given front page is important information. Hence, a code for missing is included in the code frame or range of numeric codes being used to represent the content of the page. The code for missing provides at least indirect evidence that the data collector looked for information to code and did not find it.

The timing of and extent to which codes for "don't know" should be used are one of the more controversial areas of data collection. The meaning and value of including a "don't know" category and code differs with the following factors: the researcher's objectives, how data are collected, whether factual or attitudinal information is being solicited, whether respondents are told that a "don't know" category is available, and the extent to which data collectors are trained and monitored. When data are collected directly from people and the primary objective is to collect factual or behavioral information, it is better not to include a "don't know" in the available answer categories during data collection because interviewers and respondents will be tempted to select "don't know" when other answers are more appropriate. Complicating this is the fact that in interactions such as interviews, people often use phrases such as "don't know" while they think about a question. In such cases, "don't know" is an acknowledgment that the person heard the question, but it is not necessarily intended to be the answer to the question. When attitudes or opinions are being solicited, an identified "don't know" or "no opinion" answer category may be appropriate, particularly when respondents are asked their opinion about people, policies, or practices that many know nothing about.

Some categories of answers occur repeatedly throughout data collection, such as yes, no, refused, don't know, missing information, and inapplicable. Researchers can simplify data analysis by using *consistent numeric codes* throughout the data collection process for these common response options. Table 1 shows the traditional codes generally used.

Finally, researchers should consider developing code frames that *minimize the need for transformations* during data analysis. For example, if data about immunizations are being collected off medical records that represent well-baby visits, it makes much more sense to record the actual number of immunizations given at each visit rather than creating categories for "no immunizations," "1–2 immunizations," "3–4 immunizations," and "5 or more immunizations" with codes, respectively, of 1, 2, 3, and 4.

Table 1 Examples of Consistent Codes

	Example Code		
Response	*Single Digit*	*Double Digit*	*3 + Digits*
Yes	1		
No	2		
Refused to answer	7	97	997
Don't know	8	98	998
Missing information	9	99	999
Not applicable	0	00	000

POSTCODING

Code frames for some kinds of data simply cannot be set up as the data collection procedures are being finalized. Open-ended questions that have no predetermined list of answer categories are included in interviews for a reason—namely, because the researcher is unable to anticipate the kinds and amounts of data that may be elicited or cannot create exhaustive, mutually exclusive codes that can easily be explained to and used by data collectors and data entry personnel. This is particularly true of exploratory research in which the objective is to find out whether more structured research is possible and would be of value. In such studies, interviews or other methods of data collection may be unstructured or only partially structured. Often, material is audio or visually recorded and transcribed prior to coding and analysis. When data collection is complete, the task is to determine what, if any, patterns exist in the data. This is done inductively through procedures such as content analysis. Similar procedures must be used to analyze documents, movies, and television programs. But the end objective is to come up with code frames or numeric representations of the data such that they can be analyzed and summarized.

Content analysis and related procedures are complex, and detailed information about how to do them is beyond the scope of this entry. Rather, some general guidelines can be provided here. First, the researcher must specify the objectives for which the code frame is to be used. If children have been observed at play, is the objective to count the frequency of behaviors, the duration of behaviors, the intensity of the interactions, or some combination? By identifying how the data are to be used, the researcher identifies some variables that will be extracted and others that will be ignored. The process of identifying the objectives may involve preliminary gross content analysis so that overall themes can be inductively developed from the data.

Second, in developing the code frame, a balance must be realized between too much detail and not enough detail. If clothing is being described, is it important to code the number and type of buttons on the clothing, or is it sufficient to say it is a jacket? Third, and related, is maximizing the maintenance of information within the code frame as it is created. Of particular concern here is information that appears in only a few of the documents being analyzed but that is particularly pertinent to the research questions being examined. Because content analysis is a lengthy, labor-intensive process, researchers sometimes fail to code information that is important because they are tired or bored. As long as the raw data exist, the researcher can always reexamine the data, but this often means reading many transcripts or listening to many audiotapes. Thus, every effort should be made to incorporate rare but important data into the code frame as they are being developed. This is why actively searching for information and systematically coding the fact that certain information is missing is particularly important.

Finally, researchers need to create a sufficient range of codes, variables, or dimensions so that the coder does not force data into categories that are really inappropriate. Take, for example, answers to the following statement: "Describe the vehicle(s) involved in the motor vehicle crash." First, we could code for the number of vehicles that were involved in the crash. Then, we could code for the type of vehicle(s)—sedan, sport utility vehicle (SUV), or pickup truck—that was involved. If examining safety features is an objective of the study, it probably is important to code for the make of the vehicle, the year, and whether the vehicle had air bags. The point here is that a single stimulus—whether it is a question, a paragraph in a document, a period of observation, or a sequence in a movie or tape—may yield multiple pieces of information that become multiple variables during analysis.

—Linda B. Bourque

See also CODING QUALITATIVE DATA

REFERENCES

Bourque, L. B., & Clark, V. A. (1992). *Processing data: The survey example* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–085). Newbury Park, CA: Sage.

Bourque, L. B., & Russell, L. A. (with Krauss, G. L., Riopelle, D., Goltz, J. D., Greene, M., McAffee, S., & Nathe, S.) (1994, July). *Experiences during and responses to the Loma Prieta earthquake.* Oakland: Governor's Office of Emergency Services, State of California.

Bourque, L. B., Shoaf, K. I., & Nguyen, L. H. (1997). Survey research. *International Journal of Mass Emergencies and Disasters, 15,* 71–101.

Stevens, S. S. (1968). Measurement, statistics, and the schemapiric view. *Science, 161,* 849–861.

CODING FRAME

A coding frame, code frame, or codebook shows how verbal or visual data have been converted into numeric data for purposes of analysis. It provides the link between the verbal data and the numeric data and explains what the numeric data mean.

At its simplest, a code frame refers to the range of numeric codes that are used in CODING information that represents a single VARIABLE (e.g., the state where people were born). In this example, at least 50 unique numbers would be included in the code frame, one for each state. Numbers would be assigned to states arbitrarily. Unique numbers would also be set aside for people who were not born in the United States and for people for whom no DATA were available. Once the data are entered into a data file, the code frame provides the only link between the original verbal data and the location and meaning of information in the data set. Without the code frame, the analyst will not know what the number 32 means or stands for and may not know that it represents a state.

At its most complex, a code frame or codebook shows where all data are located in a data file, what sets of numbers are associated with each piece of data or variable, and what each unique number stands for. Thus, in addition to state of birth, the data set might also include county of birth and date of birth. The code frame would describe the order in which these three variables were found in the data set, as well as what each unique code represented and the way in which date of birth was entered into the data set.

A code frame can be created prior to actual data collection, or it can be constructed afterwards through a process of CONTENT ANALYSIS and postcoding. When the number of variables and the range of probable data that will be collected for each variable are known in advance, it is advisable to precode the variables and thus create a code frame before actually collecting the data. If data are being collected through computer-assisted telephone interviews (CATI) or similar methods, code frames for most of the data must be prespecified or precoded to program the computer for data collection. A major advantage of CATI is that interviewers automatically code the data as respondents answer the questions. If the interview is dominated by questions that require interviewers to type in lengthy text answers, the value of using CATI is significantly reduced.

Code frames for some kinds of data cannot be set up prior to data collection either because the kinds and amounts of data that may be obtained cannot be anticipated, or exhaustive, mutually exclusive codes that can be easily explained to and used by data collectors and data entry personnel cannot be created. In such cases, code frames are created inductively after data collection is completed. The process by which this is done is complex and beyond the scope of this entry.

—Linda B. Bourque

REFERENCE

Bourque, L. B., & Clark, V. A. (1992). *Processing data: The survey example* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–085). Newbury Park, CA: Sage.

CODING QUALITATIVE DATA

CODING is an essential part of the many types of social research. Once data collection has been conducted, the researcher will need to begin to make sense of the data. Coding requires the researcher to interact with and to think about the data. It is a systematic way in which to condense extensive data sets into smaller analyzable units through the creation of categories and concepts derived from the data. Coding makes links between different parts of the data that are regarded as having common properties. Coding facilitates the organization, retrieval, and interpretation of data and leads to conclusions on the basis of that interpretation.

Although the process of coding may be similar for both quantitative and qualitative data, as they both include the activity of sorting the data, the objectives of coding are rather different. For example, when coding data from a social survey, values will be allocated to a category so that it can be statistically managed (frequency counts may be conducted). By contrast, the categories obtained in a SEMI-STRUCTURED INTERVIEW are not prescribed values; rather, they are explored through themes and remain embedded in their contextual position. Qualitative coding enables the data to be identified, reordered, and considered in different ways.

Whether coding open-ended questions on questionnaires, interview transcripts, or field notes, the main stage in coding is to develop a set of categories into which the data are coded. These categories may be theory driven or data driven, derived from research literature, or based on intuition. Strauss and Corbin (1990) identify three different levels of coding: open, axial, and selective. Open coding involves breaking down, comparing, and categorizing the data; axial coding puts the data back together through making connections between the categories identified after open coding; and selective coding involves selecting the core category, relating it to other categories while confirming and explaining these relationships. When a research project is designed to test a hypothesis, the categories will be generated from the theoretical framework employed, and the data will be molded into these categories. This type of coding is termed *coding down.* When data are used to generate theory, the categories are generated from the data, which involves *coding up* (Fielding, 1993).

Coffey and Atkinson (1996) highlight that coding can serve as either data simplification or data complication, although often coding endorses a combination of the two. Coding can reduce data to simple broad categories or common denominators that help the retrieval of data sharing common features. As demonstrated by Strauss (1987), coding can equally be employed to expand on, reinterpret, and open up analysis to previously unconsidered analytic possibilities and aid the generation of theories.

Over the past decade, there has been an explosion of computer software packages specifically designed for facilitating the qualitative coding process. Although computational qualitative coding can be conducted quickly and simply, be used with large numbers of coding, and open up unexpected lines of enquiry and analysis, it can offer the researcher an overwhelming range of ways in which textual data may be handled. Furthermore, research can be constrained by the computer program. It may impose unacceptable restrictions on analysis, and qualitative coding can be subject to formalization.

—Sharon Lockyer

REFERENCES

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies.* Thousand Oaks, CA: Sage.

Fielding, J. (1993). Coding and managing data. In N. Gilbert (Ed.), *Researching social life* (pp. 218–238). Thousand Oaks, CA: Sage.

Strauss, A. (1987). *Qualitative analysis for social scientists.* Cambridge, UK: Cambridge University Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

COEFFICIENT

A coefficient is the quantitative value of a RELATIONSHIP. A coefficient, once estimated, indicates the link between two or more variables. For example, a PEARSON CORRELATION COEFFICIENT between variables X and Y of 0.90 precisely quantifies that relationship, suggesting it is very strong. In a MULTIPLE REGRESSION equation, with two independent variables X and Z, a regression coefficient of 6.3 for X indicates that a unit change in X is expected to produce a 6.3-unit change in Y, holding Z constant. In general, there are many coefficients available for relating variables, with each having different and, in some cases, more desirable properties, depending on the purposes of the analyst.

—Michael S. Lewis-Beck

See also PARAMETER

COEFFICIENT OF DETERMINATION

Represented by r^2 for the bivariate case and R^2 in the multivariate case, the coefficient of determination is a measure of GOODNESS OF FIT in ORDINARY LEAST SQUARES LINEAR REGRESSION. (In the multivariate case, it is often called the coefficient of multiple determination.) The statistic measures how well the estimated regression line fits the actual data. It tells us how much of the variation in the dependent variable, y, is explained by the model by all of the INDEPENDENT VARIABLES taken together. Specifically, it is a proportion of the explained variation in y to the total variation in y.

A number of equations may be used to calculate the coefficient of determination. The first takes the explained sum of squares and divides it by the total sum of squares: $R^2 = \text{ESS/TSS}$. A second subtracts the proportion of the residual sum of squares (the unexplained sum of squares) to the total sum of squares from 1: $R^2 = 1 - (\text{RSS/TSS})$.

Because it is the proportion of the explained sum of squares over the total sum of squares, the measure must fall between 0 and 1. A value of 1 indicates a perfect fit of the linear regression line to the data; values closer to 0 suggest a poor fit. If the x and y variables are completely linearly independent, the R^2 will equal 0. It should be noted that R^2 could be a negative value ranging from 0 to –1 if the sample average accounts for more variation in the dependent variables than the independent variables explain. Multiplying the measure by 100 gives a value that allows for clearer interpretation; with this calculation, then, an R^2 of .25 would be interpreted by saying that 25% of the variation in the dependent variable is explained by the independent variables in the model.

The square root of the coefficient of determination is known as the coefficient of MULTIPLE CORRELATION, or the sample CORRELATION coefficient in the bivariate case. Conversely, R^2 is the square of the coefficient of multiple correlation, which is the measure of correlation between the estimated dependent variable calculated from the independent variables ($\hat{y}$) and the actual dependent variable (y).

Although much consideration, perhaps sometimes too much, is often paid to this measure, there are a number of important points to remember when interpreting it. First, no matter how high your R^2 is, it only is evidence of correlation; it does not provide positive support for causation. That is, a high R^2 does not allow you to state that your independent variables caused your dependent variable. Second, although, for example, 30% is not an extremely high value, your model may be performing relatively well; explaining 30% of the variation of a factor in the political environment is often still a substantial portion. Also, when conducting TIME-SERIES analysis, you may often find R^2s over .90. Third, as independent variables are added to the model, the R^2 will increase, but we should avoid trying to maximize the coefficient of correlation over theoretically sound MODELS. Finally, R^2s cannot be compared between any two given models if they do not have the same dependent variable.

When working with models with multiple variables, it is good practice to consider the ADJUSTED R^2. Because every independent variable added to a model usually increases the amount of variation explained, it is important to take into consideration the DEGREES OF FREEDOM. The adjusted R^2 does just this by adjusting for the additional parameters.

—Christine Mahoney

REFERENCES

Frankfort-Nachmias, C., & Nachmias, D. (2000). *Research methods in the social sciences* (6th ed.). New York: Worth.

Gujarati, D. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Wooldridge, J. M. (2000). *Introductory economics: A modern approach.* Cincinnati, OH: South-Western College Publishing.

COGNITIVE ANTHROPOLOGY

Cognitive anthropology is the study of human thought in the context of social relations. Some work in cognitive anthropology examines discrete cognitive processes such as memory or perception, but Goodenough (1957) set the stage for the modern development of cognitive anthropology when he defined culture as the knowledge an individual must possess to function adequately as a member of a social group. Implicit in this deceptively simple definition are questions regarding both the sharing and distribution of knowledge that are now at the core of the field.

Cognitive anthropology has been one of the most methodologically self-conscious topical areas in anthropology, driven in part by these questions of sharing and distribution. Much of this work revolves around the elicitation of culturally salient categories within a social group (be it a social network, an organization, or a community) and the determination of underlying dimensions of meaning that organize those categories. The techniques for eliciting terms that respondents use to label categories and for determining what distinctive features respondents use to define the meaning of terms, include the following: free-listing, in which terms that make up a cultural domain are generated; various forms of unconstrained and constrained pile sorts, in which the similarities and differences in the meaning of terms can be determined; and various forms of rating and ranking tasks to test in a refined way the importance of different dimensions of meaning (see Weller & Romney, 1988).

The examination of the sharing and distribution of cultural knowledge advanced dramatically with the development of the cultural consensus model by Romney, Weller, and Batchelder (1986). The cultural consensus model uses the correlations among subjects as a measure of agreement; if these associations among respondents can be accounted for by a single principal factor, then consensus exists, and it is reasonable to infer that all subjects are relying on a single, shared cultural model of that particular cultural domain. The distribution of subjects around that consensus can be examined (including the potential for alternative cultural models), as well as the specific content of the consensus.

A good example of this approach is the study of cultural models of reproductive cancers among physicians and patients from different ethnic backgrounds (Chavez, Hubbel, McMullin, Marinez, & Mishra, 1995). The tools described above were used to examine models of cancer among physicians, Anglo women, and both resident and immigrant Latinas in Southern California. Within each of these groups, there was significant consensus (i.e., groups agreed among themselves about the important causes of reproductive cancers). There was also an overlapping consensus that connected the Anglo women to the physicians, Anglo women to resident Latinas, and resident Latinas to immigrant Latinas. But immigrant Latinas were far from both Anglo women and physicians in their shared models of cancer causation. These results can be displayed graphically, demonstrating how persons and groups are distributed in a space of cultural meaning.

The systematic ethnographic techniques of cognitive anthropology enhance the traditional qualitative techniques of cultural anthropology. A primary goal of anthropological research has been to understand how other people see their worlds. Methodological innovations in both the collection and analysis of data are helping investigators to understand others' cultural models with greater precision.

—William W. Dressler

REFERENCES

Chavez, L., Hubbell, F. A., McMullin, J. M., Marinez, R. G., & Mishra, S. I. (1995). Structure and meaning in models of breast and cervical cancer risk factors: A comparison of perceptions among Latinas, Anglo women and physicians. *Medical Anthropology Quarterly, 9,* 40–74.

Goodenough, W. (1957). Cultural anthropology and linguistics. In P. L. Garvin (Ed.), *Report of the 7th Annual Roundtable on Linguistics and Language Study* (pp. 167–173). Washington, DC: Georgetown University Press.

Romney, A. K., Weller, S. C., & Batchelder, W. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist, 88,* 313–338.

Weller, S. C., & Romney, A. K. (1988). *Systematic data collection* (Qualitative Research Methods, Vol. 10). Newbury Park, CA: Sage.

COHORT ANALYSIS

Cohort analysis, a general strategy for examining data rather than a statistical technique, has become increasingly popular in the social sciences in the past few decades, especially during the last quarter of the 20th century. Useful in assessing the consequences of aging (of humans or other entities) and in understanding social and cultural change, cohort analysis has also attracted attention because it presents an unusually intriguing methodological challenge.

In its broadest sense, cohort analysis is any quantitative research that uses a measure of the concept of *cohort* and relates that measure to one or more additional variables. A cohort, in turn, consists of those individuals (human or otherwise) who experienced a particular event during a specified time. The kind of cohort most often studied is the human birth cohort—those persons born during a given year, decade, or other period of time. If the term *cohort* is used without a modifier, it is usually understood that the referent is a human birth cohort. Otherwise, the event commonly experienced by the individuals is used an as adjective to identify the kind of cohort, as in *retirement cohort.* The events that define cohorts may range from marriage to joining an organization, from entering a graduate program to becoming a parent for the first time. The individuals, if not humans, may be marriages, organizations, textbooks, movies, or other entities that came into being during a particular time.

The usual and more restricted meaning of cohort analysis is that it is any attempt to estimate age, period, and cohort effects on a dependent variable, such as some kind of behavior or attitudes, and such attempts require data from at least two cohorts for at least two points in time. An age effect results from growing older, a cohort effect from cohort membership, and a period effect from influences that vary through time. For example, an enduring preference for the kind of music that was popular when one was a teenager or young adult is a cohort effect, whereas changes in attitudes in response to an economic downturn are PERIOD EFFECTS. It is obviously important to estimate age effects in studies of human aging, and it is also important to estimate each kind of effect to understand social and cultural change.

THE AGE-PERIOD-COHORT CONUNDRUM

Estimation of age, period, and cohort effects cannot be straightforward because two of the kinds of effects are confounded with one another in the findings from any standard statistical analysis. For instance, differences between people of different ages at one point in time may be age effects, cohort effects, or both kinds of effects. Likewise, observed changes in persons studied in different waves of a panel study may be age effects, period effects, or both. This confounding of effects is an example of the IDENTIFICATION PROBLEM, which exists when three or more independent variables may affect a dependent variable when each independent variable is a perfect linear function of the others. Age is a linear function of period and cohort, cohort is a linear function of age and period, and period is a linear function of age and cohort.

A heuristic device often used to facilitate understanding of the confounding of effects in cohort data is the standard cohort table, in which two or more sets of cross-sectional data relating some dependent variable to age are juxtaposed and the age categories are equal in years to the intervals between the sets of data, as in Table 1. Period and cohort effects are confounded in the rows of the table, and age and cohort effects are confounded in the columns. Cohorts can be traced through time by reading diagonally down and to the right, and age and period effects are confounded in the cohort diagonals.

Table 1 Percentage of Respondents to the 1974, 1984, and 1994 American General Social Surveys Who Said That Extramarital Sex Relations Are "Always Wrong," by Age

	1974	*1984*	*1994*
Age	*% (n)*	*% (n)*	*% (n)*
20–29	59.2 (392)	68.3 (384)	83.8 (346)
30–39	70.9 (291)	63.9 (304)	77.7 (462)
40–49	75.3 (270)	70.8 (241)	76.7 (411)
50–59	80.8 (278)	78.6 (181)	77.2 (272)
60–69	83.0 (194)	79.7 (166)	85.8 (165)

NOTE: Data are weighted by the number of respondents in the household age 18 and older to increase representativeness.

The linear effects of age, period, and cohort are confounded with one another in such a way that no statistical analysis can separate them; an infinite number of combinations of age, period, and cohort effects could account for a linear pattern of variation of a dependent variable across ages, periods, or cohorts. There is a method that can separate nonlinear effects, but only if the effects are additive and if precisely correct assumptions can be made about some of the effects—conditions that are rarely met.

However, it is often possible to arrive at reasonable conclusions about the general nature of the effects without rigorous model testing. Some useful cohort analyses have combined theory, information about the phenomena being studied from outside the data at hand, common sense, simple statistical analyses, and informal examination of the data to arrive at tentative but defensible conclusions.

AN EXAMPLE

The utility of a simple examination of descriptive data can be illustrated by looking at the data in Table 1. The 1974 data show a positive relationship between age and restrictive attitudes concerning extramarital sex relations, and the cohort diagonals indicate that, in the aggregate, persons in each birth cohort grew more restrictive as they grew older. These data by themselves suggest a positive age effect on restrictive attitudes. However, a more careful examination of the table casts doubt on this interpretation. Restrictiveness increased at the 20 to 29 age level—an indication of period influences toward restrictive attitudes possibly strong enough to produce the increased restrictiveness within cohorts. According to theory, supported by much empirical evidence, younger adults tend to be affected more than older ones by period influences; thus, such a differential impact by age of earlier influences for permissiveness could account for the positive relationship between age and restrictiveness in 1974. Therefore, there is no convincing evidence of any age effects in the data.

AGE-PERIOD-COHORT CHARACTERISTIC MODELS

A new form of analysis, introduced by Robert O'Brien and his collaborators, can reasonably be considered a kind of cohort analysis, even though it does not provide meaningful estimates of age, period, or total cohort effects. Although it is impossible to hold constant two of the interrelated variables and let the third vary, it is possible to hold age and period constant and let certain characteristics of cohorts, such as their relative size, vary. If the cohort characteristic does not bear a strong linear relationship to cohort, the estimate of its effects should be meaningful, even though the model estimates of age and period effects are not. This promising technique has already proven useful in studying such dependent variables as crime commission and victimization.

—Norval D. Glenn

REFERENCES

Glenn, N. D. (2003). Distinguishing age, period, and cohort effects. In J. Mortimer & M. Shanahan (Eds.), *Handbook of the life course* (pp. 465–476). New York: Kluwer Academic/Plenum.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, W. K. (1973). Some methodological issues in the cohort analysis of archival data. *American Sociological Review, 38,* 242–258.

O'Brien, R. M. (2000). Age-period-cohort-characteristic models. *Social Science Research, 29,* 123–139.

COINTEGRATION

When a linear combination of nonstationary variables is stationary, the variables are said to be cointegrated, and the vector that defines the stationary linear combination is called a *cointegration vector.* A

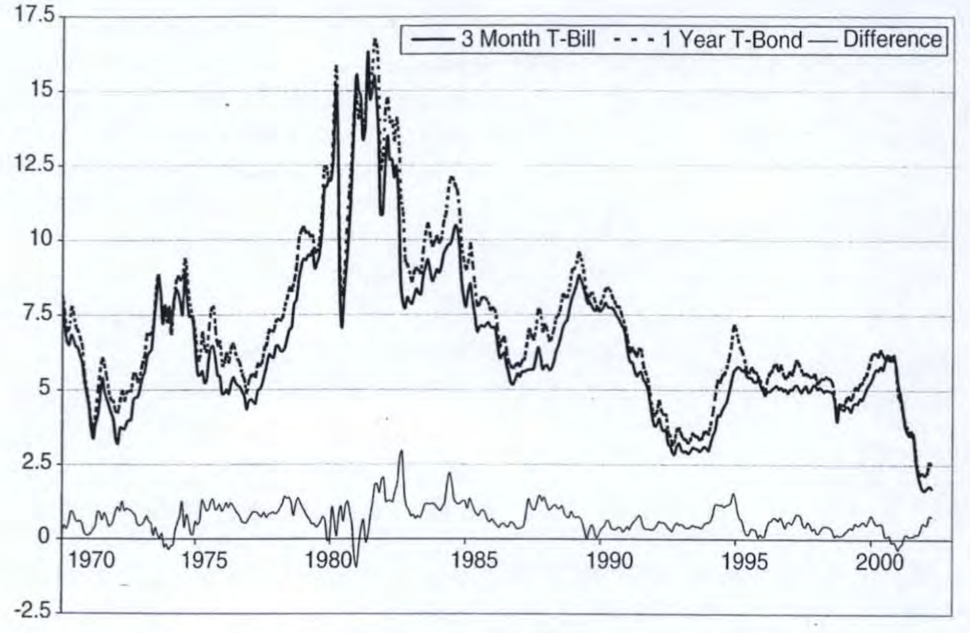

Figure 1 The Interest Rates on a 3-Month Treasury Bill and a 1-Year Treasury Bond for the Period From January 1970 to April 2002

SOURCE: Federal Reserve Bank of St. Louis.

NOTE: The two interest rates appear to be nonstationary, whereas the difference between them seems to be stationary, so the two interest rates are cointegrated.

time series is stationary if its distribution does not vary over time. The simplest example of a stationary process is $\{\varepsilon_t\} = \ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots,$ which represents a sequence of independent and identically distributed random variables. The subscript t refers to time. If the distribution of a variable depends on t, it is nonstationary. In cointegration analysis, the most common form of nonstationarity is that of the integrated variables. The random walk, $X_t = X_{t-1} + \varepsilon_t = X_0 + \sum_{i=1}^{t} \varepsilon_i$, is an example of a nonstationary variable that is integrated of order one. The word *integration* refers to the cumulation of epsilons.

The concept of cointegration is interesting because it can be applied to uncover relationships that may have theoretical interpretations. For example, the cointegration relations may be defined by the first-order conditions of an economic model, and cointegration analysis can be used to estimate economic models and test certain theoretical hypotheses.

The classical example on cointegration is about a dog that follows its drunk owner. In this example, the positions of the dog and its owner, as a function of time, are two nonstationary processes because the drunk owner is walking around, taking steps in random directions. However, the two processes are cointegrated because the distance between dog and owner is stationary.

Another example is a property of interest rates on bonds with different maturities. Individually, they may have large fluctuations, but the difference between them appears to be stationary. This is illustrated in Figure 1, which contains the interest rates on a 3-month Treasury bill and a 1-year Treasury bond for the period from January 1970 to April 2002. The thin solid line is the difference between the two interest rates.

HISTORICAL DEVELOPMENT

Cointegration was introduced by Clive W. J. Granger (1981, 1983), and the statistical analysis of cointegrated processes was formalized by Robert F. Engle and Granger (1987), Søren Johansen (1988, 1991), and Peter C. B. Phillips (1991). Cointegration is related to many concepts of ECONOMETRICS, such as unit root processes, spurious regression, and common stochastic trends (for an excellent review, see Watson, 1994). Today there is a voluminous literature on cointegration and applications of cointegration. Recent developments have been on improving and generalizing existing techniques, such as bias and Bartlett corrections; fractionally cointegrated processes; seasonal cointegration; panel cointegration; nonlinear cointegration; and cointegration in relation to processes with structural changes. Many references can be found in the book by James D. Hamilton (1994), which also contains a good introduction to cointegration. Excellent textbooks on likelihood analysis of cointegration are Johansen (1995) and the companion book by Peter R. Hansen and Johansen (1998).

ORDER OF INTEGRATION

For variables to be cointegrated, they must individually be integrated. Formally, a time series, X_t, is said to be integrated of order d if $(1-L)^d X_t = \psi(L)\varepsilon_t$ is stationary and $\psi(1) = \sum i\psi_i \neq 0$, where L is the lag operator such that $\psi(L)\varepsilon_t = \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \cdots$. The notation $X_t \sim I(d)$ is short for "X_t is integrated of order d," and the notation $\Delta^d X_t$ is sometimes used in place of $(1-L)^d X_t$. An alternative definition of the order of integration assumes that $\{\varepsilon_t\}$ is a white-noise process and substitutes *covariance-stationary* for *stationary*. A white-noise process is defined by the following: having expected value zero, $E(\varepsilon_t) = 0$; having finite variance that does not depend on time, $\text{var}(\varepsilon_t) = \sigma^2$; and being uncorrelated across time, $\text{cov}(\varepsilon_s, \varepsilon_t) = 0$ for $s \neq t$. A covariance-stationary process, $\{X_t\}$, has constant mean and variance, as well as an autocovariance function that only depends on the difference in time, $\text{cov}(X_s, X_t) = \varphi(|s-t|)$.

The order of integration, d, can be any real number, but most of the literature is concerned with the unit root processes, which corresponds to $d \in \mathbb{N}$ (the positive integers). The nonintegers are referred to as fractionally integrated processes, which is related to the long-memory processes.

The order of integration of a vector of variables, $X_t = (X_{1t}, \ldots, X_{pt})'$, is defined as the highest order of integration of the individual elements, $X_{1t}, \ldots, X_{pt}$.

If $X_t \sim I(d_x)$, if $Y_t \sim I(d_y)$, and if a and b are nonzero constants, then $aX_t + bY_t$ is, in general, integrated of order $d = \max(d_x, d_y)$. It is the special case when $aX_t + bY_t$ is integrated of order less than $\max(d_x, d_y)$, which characterizes cointegration. For two variables to be cointegrated, they must be integrated of the same order. The notation $CI(d, b)$ is used for variables that are integrated of order d with a cointegration relation, which is integrated of order $d - b$.

A more general concept of cointegration is multi-cointegration, which is sometimes called polynomial cointegration. If $(1-L)^d X_t$ and Y_t are cointegrated for some $d \neq 0$, then X_t and Y_t are multi-cointegrated. For example, $X_t \sim I(2)$ and $Y_t \sim I(1)$, then $\Delta X_t = X_t - X_{t-1}$ may cointegrate with Y_t.

EXAMPLE

Let $\{\varepsilon_t\}$ be a sequence of independent and identically distributed variables, and suppose that for $t = 1, 2, \ldots$

$$X_t = \sum_{s=1}^{t}\sum_{i=1}^{s}\varepsilon_i + \sum_{s=1}^{t}\varepsilon_s + \varepsilon,$$

$$Y_t = 2\sum_{s=1}^{t}\sum_{i=1}^{s}\varepsilon_i + \varepsilon_t,$$

$$Z_t = -\sum_{s=1}^{t}\varepsilon_s + \varepsilon_t.$$

Because $(1-L)^2 X_t = 3\varepsilon_t - 3\varepsilon_{t-1} + \varepsilon_{t-2}$ is stationary and the coefficients satisfy $\sum_i \psi_i = 3 - 3 + 1 = 1 \neq 0$, we see that X_t is integrated of order two. Similarly, it follows that $Y_t \sim I(2)$ and that $Z_t \sim I(1)$, so the vector $(X_t, Y_t, Z_t)'$ is integrated of order two. Furthermore, $2X_t - Y_t = 2\sum_{s=1}^{t}\varepsilon_s + \varepsilon_t$ is integrated of order one, so X_t and Y_t are cointegrated, $CI(2, 1)$, with cointegration vector $(2, -1)'$. Similarly, we see that $2X_t - Y_t + 2Z_t = 3\varepsilon_t$ is integrated of order zero, so the three variables—X_t, Y_t, and Z_t—are cointegrated, $CI(2, 2)$, with cointegration vector $(2, -1, 2)'$. Also, X_t and Z_t are multi-cointegrated because the two $I(1)$ variables, ΔX_t and Z_t, are cointegrated, $CI(1, 1)$, because $\Delta X_t + Z_t = 3\varepsilon_t - \varepsilon_{t-1}$ is $I(0)$. Similarly, it can be seen that Y_t and Z_t are multi-cointegrated.

ANALYSIS OF COINTEGRATED PROCESSES

The classical assumptions in the LINEAR REGRESSION model are violated in the presence of integrated variables, mainly because of the nonstationarity. This can cause the asymptotic normality of parameter estimators to fail, and the statistical analysis of cointegrated variables is therefore more complicated than standard regression analysis.

The two main approaches to cointegration analysis are the regression-based approach and the likelihood approach. The former involves standard regression techniques with various modifications, and the latter, which is based on the vector autoregressive model (VAR), is sometimes called the Johansen method.

A cointegration analysis of a system of variables will typically consist of the following:

1. *Determining the cointegration rank.* This typically involves the use of a Dickey-Fuller-type distribution. These distributions have complicated expressions that involve stochastic integrals of Brownian motions. Fortunately, most modern econometrics textbooks, including the one by Hamilton (1994), contain tables of the most common Dickey-Fuller distributions. The trace test of Johansen (1988, 1991) is a popular test for determining the cointegration rank.
2. *Estimating cointegration relations and other parameters.* The cointegration relations are not identified without additional restrictions/normalizations.
3. *Inference on the cointegration parameters and other parameters.* The analysis on cointegration relations is particularly useful because they may be interpreted as structural equations in an economic model. Identified cointegration parameters have a nonnormal limiting distribution and are extremely consistent. The latter means that small samples can be very informative about the cointegration relations. Inference on other parameters follows traditional results in the standard case.

The likelihood analysis of Johansen (1988) is based on the VAR model, $X_t = \prod_1 X_{t-1} + \cdots + \prod_k X_{t-k} + \varepsilon_t$, where X_t is a p-dimensional vector and $\{\varepsilon_t\}$ is a sequence of independent normally distributed variables with mean 0 and covariance-matrix Ω. This model is rewritten in the form of the error correction model (ECM):

$$\begin{aligned}\Delta X_t = \prod X_{t-1} + \Gamma_1 \Delta X_{t-1} \\ + \cdots + \Gamma_{k-1}\Delta X_{t-k+1} + \varepsilon_t, \\ t = 1, \ldots, T.\end{aligned}$$

If the elements of X_t are cointegrated, then the $p \times p$ matrix, $\notin \prod$, has reduced rank $r < p$, and the number of common stochastic trends in X_t is given by $p - r$. The reduced rank of $\prod$ can be made explicit by expressing the matrix as a product of two matrices, $\prod = \alpha\beta'$, where α and β are $p \times r$ matrices. This formulation leads to a regression equation that can be estimated by reduced rank regression. An important instrument for the analysis of cointegrated processes is the Granger representation, $X_t = C\sum_{i=1}^{t} \varepsilon_i + C(L)\varepsilon_t + X_0 - C(L)\varepsilon_0$. This representation divides the process into a nonstationary random walk, $C\sum_{i=1}^{t} \varepsilon_i$; a stationary component, $C(L)\varepsilon_t$; and a term that depends on initial values, $X_0 - C(L)\varepsilon_0$. An important result is that $\beta' C = 0$ (an $r \times p$ matrix of zeros), from which it follows that $\beta' X_t = \beta' C(L) \sim I(0)$. So the cointegration relations are given by the columns of β. The relation, $\Delta X_t = \alpha(\beta' X_{t-1}) + \cdots$, shows that α defines how X_t responds on a deviation in $\beta' X_{t-1}$ from its expected value, and α is therefore called the matrix of adjustment coefficients. $\beta' X_{t-1}$ can sometimes be interpreted as the deviation from an equilibrium. In this case, α informs about the variables' adjustment toward the equilibrium relations, both the direction they take and how fast the variables converge to the equilibrium.

—Peter Reinhard Hansen

REFERENCES

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica, 55,* 251–276.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric models specification. *Journal of Econometrics, 16,* 121–130.

Granger, C. W. J. (1983). *Co-integrated variables and error-correcting models.* Discussion Paper No. 1983–13, University of California, San Diego.

Hamilton, J. D. (1994). *Time series analysis.* Princeton, NJ: Princeton University Press.

Hansen, P. R., & Johansen, S. (1998). *Workbook on cointegration.* Oxford, UK: Oxford University Press.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control, 12,* 231–254.

Johansen, S. (1991). Estimation and hypotheses testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica, 59,* 1551–1580.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford, UK: Oxford University Press.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica, 59,* 283–306.

Watson, M. W. (1994). Vector autoregression and cointegration. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, pp. 2843–2915). Amsterdam: Elsevier.

COLLINEARITY. *See* MULTICOLLINEARITY

COLUMN ASSOCIATION. *See* ASSOCIATION MODEL

COMMONALITY ANALYSIS

Commonality analysis is a method used to partition the explained/predicted variance in either a measured or a latent variable into subcomponents explained (a) uniquely by each measured predictor variable and (b) in common by every possible combination of the PREDICTOR VARIABLES. The explained variance and all the subcomponents of the explained variance are each in a squared, variance-accounted-for metric. Commonality analysis was originally developed for use in REGRESSION but also can be applied in other MULTIVARIATE ANALYSES (cf. Frederick, 1999).

HEURISTIC EXAMPLE

To make the discussion concrete, presume a regression problem involving a single measured outcome variable (Y) and two predictor variables (X_1 and X_2). The results associated with the example are presented graphically in the Venn diagram in Figure 1.

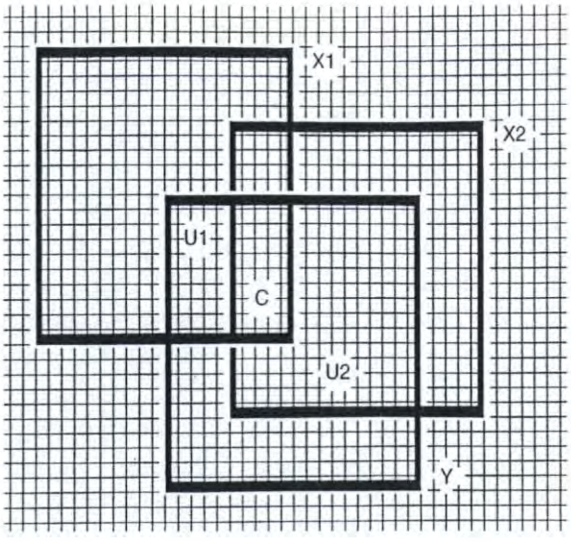

Figure 1 Venn Diagram for a Two-Predictor Regression Commonality Analysis

In this example, each of the three measured variables has a SUM OF SQUARES (SOS) of 400, corresponding to the box for each measured variable being drawn as encompassing 20×20 units of area. As regards the overlap of X_1 and Y, because 100 units of the $\text{SOS}_Y = 400$ are predicted by X_1, the $r^2_{YX_1} = 100/400 = .2500 = 25.00\%$. X_2 explains 225 of the 400 units of Y, and so the $r^2_{YX_2} = 225/400 = .5625 = 56.25\%$.

However, the MULTIPLE CORRELATION squared ($R^2_{[YX_1X_2]}$) does *not* equal 25.00% plus 56.25%, or 81.25%, because the predictors are correlated ($r^2_{X_1X_2} = 18.75\%$). Therefore, some of the explanatory power of X_1 and X_2 is common to both and should not be "double counted" in computing $R^2_{[YX_1X_2]}$. For this example, X_1 and X_2 together explain $R^2_{[YX_1X_2]} = 68.75\%$ of the variability in the Y scores (and not 81.25%).

But where does this 68.75% EFFECT SIZE originate? The following questions must be addressed:

1. How much of the observed effect size is due to the *unique* explanatory power of X_1?
2. How much of the observed effect size is due to the *unique* explanatory power of X_2?
3. How much of the observed effect size is due to the *common* explanatory power of X_1 and X_2?

These are the questions addressed in a commonality analysis.

The analysis is conducted in a series of four steps. First, the R^2 is computed using all possible combinations of the predictor variables. For this simple example, there are only three possible combinations, and $R^2_{[YX_1X_2]} = 68.75\%$, $R^2_{[YX_1]} = 25.00\%$, and $R^2_{[YX_2]} = 56.25\%$.

Second, values obtained in Step 1 are inserted into commonality analysis formulas provided in various publications (cf. Rowell, 1996). For this example, the unique predictive contribution (U_1) of X_1 is computed as

$$U_1 = R^2_{[YX_1X_2]} - R^2_{[YX_2]}$$
$$68.75\% - 56.25\% = 12.50\%.$$

The unique contribution (U_2) of X_2 is computed as

$$U_2 = R^2_{[YX_1X_2]} - R^2_{[YX_1]}$$
$$68.75\% - 25.00\% = 43.75\%.$$

The common contribution (C_{12}) of X_1 and X_2 is computed as

$$C_{12} = R^2_{[YX_1X_2]} - U_1 - U_2$$
$$68.75\% - 12.50\% - 43.75\% = 12.50\%.$$

The commonality analysis computational formulas become exponentially more complicated as there are more predictor variables (Rowell, 1996). For example, when there are three predictor variables, the formulas are as follows:

$$U_1 = R^2_{[YX_1X_2X_3]} - R^2_{[YX_2X_3]}$$
$$U_2 = R^2_{[YX_1X_2X_3]} - R^2_{[YX_1X_3]}$$
$$U_3 = R^2_{[YX_1X_2X_3]} - R^2_{[YX_1X_2]}$$
$$C_{12} = R^2_{[YX_1X_3]} + R^2_{[YX_2X_3]} - R^2_{[YX_3]} - R^2_{[YX_1X_2X_3]}$$
$$C_{13} = R^2_{[YX_1X_2]} + R^2_{[YX_2X_3]} - R^2_{[YX_2]} - R^2_{[YX_1X_2X_3]}$$
$$C_{23} = R^2_{[YX_1X_2]} + R^2_{[YX_1X_3]} - R^2_{[YX_1]} - R^2_{[YX_1X_2X_3]}$$
$$C_{123} = R^2_{[YX_1X_2X_3]} + R^2_{[YX_1]} + R^2_{[YX_2]} + R^2_{[YX_3]} - R^2_{[YX_1X_2]} + R^2_{[YX_1X_3]} - R^2_{[YX_2X_3]}.$$

Third, as a check on the computations, the $R^2_{[Y X_1X_2]}$ is recomputed as

$$R^2_{[YX_1X_2]} = U_1 + U_2 + C_{12}$$
$$12.50\% + 43.75\% + 12.50\% = 68.75\%.$$

Note, as illustrated in Figure 1, that $U_1 = 12.50\%$ and $C_{12} = 12.50\%$ are *not* the same areas.

Fourth, the variance decompositions are then organized into a table to illustrate the unique and common predictive contributions of each predictor variable individually. Table 1 presents the relevant results for this example.

Table 1 Variance Partitions of the Squared Bivariate Correlations (in percentages)

	Predictor	
Variance Partition	X_1	X_2
U_1	12.50	
U_2		43.75
C_{12}	12.50	12.50
Sum $= r^2$	25.00	56.25

COMMON MISCONCEPTION

A common misconception of persons using REGRESSION is that the predictor variables should be uncorrelated (i.e., not COLLINEAR or MULTICOLLINEAR). It is true that when predictors are correlated, sensible interpretations as regards the origins of detected EFFECT SIZES *cannot* be derived by interpreting only standardized weights ("beta weights"). It is also true that standard errors of the standardized weights become larger when predictors are correlated.

But CORRELATIONS among predictor variables are not inherently problematic as long as correct interpretation strategies are employed. Using correlated predictors also often honors the fact that, in reality, many variables are correlated. One such interpretation strategy invokes examination of both the standardized weights and the structure coefficients (Courville & Thompson, 2001).

Commonality analysis is another alternative for understanding prediction dynamics when predictors are correlated. Note that when predictor variables are perfectly uncorrelated, a commonality analysis would not be sensible because every common variance partition (e.g., C_{12}, C_{23}, C_{123}) would, by definition, be zero, and every unique contribution (e.g., U_1) would equal a corresponding squared bivariate correlation (e.g., $r^2_{YX_1}$).

NEGATIVE VARIANCE COMPONENTS

In any analysis, the results are compromised if there is excessive model specification error. So, if the wrong variables or the wrong analysis is used in a commonality analysis, the results are compromised. One indication of this possibility occurs when some of the variance components are negative because squared variance-accounted-for statistics should not be negative (Thompson, 1985).

If one or a few variance components in a commonality analysis are negative but only trivially smaller than zero, the negative values are only mildly distressing. However, if the negative values are large in magnitude in the judgment of the researcher, careful thought must be invested in whether the model has been correctly specified.

—Bruce Thompson

REFERENCES

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement, 61*, 229–248.

Frederick, B. N. (1999). Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 305–318). Stamford, CT: JAI.

Rowell, R. K. (1996). Partitioning predicted variance into constituent parts: How to conduct commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 33–44). Greenwich, CT: JAI.

Thompson, B. (1985). Alternate methods for analyzing data from experiments. *Journal of Experimental Education, 54*, 50–55.

COMMUNALITY

Communality is a squared variance-accounted-for statistic reflecting how much variance in measured variables is reproduced by the latent constructs (e.g., the factors) in a model. Conversely, communality can be conceptualized as how much of the variance of a measured/observed variable is useful in delineating the latent/composite variables in the model. The symbol typically used for a communality coefficient is h^2.

Communality coefficients are commonly used in FACTOR ANALYSIS, including both EXPLORATORY FACTOR ANALYSIS (EFA) and CONFIRMATORY FACTOR ANALYSIS (CFA) (Graham, Guthrie, & Thompson, 2003; Thompson, 1997). However, communality coefficients can also be computed in other multivariate techniques, such as CANONICAL CORRELATION ANALYSIS (Thompson, 2000). When variables are being analyzed, there will be a separate communality coefficient computed for each variable.

Communality coefficients are estimated both retrospectively, after an analysis is completed, and prospectively, as input into initial data analysis. In the retrospective application, if the factors are uncorrelated, the communality coefficient for a measured variable is computed by squaring the structure coefficients (factor loadings) for the variable on all factors and then summing these squared structure coefficients across the factors.

For example, 10 variables might be subjected to an EFA in which three factors were extracted and then subjected to ORTHOGONAL ROTATION. The squared structure coefficient for the first variable on the three factors might be .4, .3, and .2. The h^2 for this variable would be .9 or 90% (i.e., $.4 + .3 + .2 = .9$). This means that 90% of this measured variable's variance can be reproduced by the three factors. Conversely, 90% of the variance of the measured variable was useful in delineating the three factors as a set.

Communality coefficients can also be conceptualized as lower-bound estimates of the minimum RELIABILITY COEFFICIENT of a given variable. If a measured variable had an $h^2 = .9$, then whatever the score reliability of the variable is, the reliability is *at least* .9.

The input into an analysis affects the output communalities. Often, the correlation matrix is analyzed. Whatever entries are initially put on the diagonal of the matrix are the input communality estimates. Then factors are extracted. In PRINCIPAL COMPONENTS ANALYSIS, the communality coefficients are computed at this point, and the analysis is complete.

However, in analyses such as principal axis factor analysis, these communality coefficients are then substituted back into the diagonal of the correlation matrix, and factors are reextracted. This process is repeated successively (called iteration) until the communality coefficients on successive steps stabilize.

The influence of iteration on the output results is a function of (a) the score reliabilities of the variables and (b) the number of entries on the diagonal versus the number of entries in the entire matrix (e.g., for 10 variables, $10/100 = .1$; for 100 variables, $100/10{,}000 = .01$). It should also be noted that what is being estimated is reliability so as to take into account measurement ERROR, but excessive iteration may let SAMPLING ERROR distort the results too much. So some methodologists suggest no or limited iteration to balance the trade-offs between these two types of error influences.

—Bruce Thompson

See also FACTOR ANALYSIS

REFERENCES

Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling, 10*, 142–153.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement, 57*, 5–19.

Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 285–316). Washington, DC: American Psychological Association.

COMPARATIVE METHOD

The comparative method is fundamentally a case-oriented, small-N technique. It is typically used when researchers have substantial knowledge of each case included in an investigation and there is a relatively small number of such cases, often of necessity. The best way to grasp the essential features of the comparative method is to examine it in light of the contrasts it offers with textbook presentations of social science methodology. In textbook presentations, (a) the proximate goal of social research is to document general patterns that hold for a large population of observations, (b) cases and populations are typically seen as given, (c) researchers are encouraged to enlarge their number of cases whenever possible, (d) it is often presumed that researchers have well-defined theories and well-formulated hypotheses at their disposal from the very outset of their research, (e) investigators are advised to direct their focus on VARIABLES that display a healthy range of variation, and (f) researchers are instructed to assess the relative importance of competing independent variables to understand causation.

PROXIMATE GOALS

Textbook presentations of social research usually focus on the goal of documenting general patterns characterizing a large population of observations. If researchers can demonstrate a relationship between one or more independent variables and a DEPENDENT VARIABLE, then they can better predict cases' values on the dependent variable, given knowledge of their scores on the predictor variables. Typically, the study of general patterns is conducted with a sample of observations drawn from a large population. The researcher draws inferences about the larger population based on his or her analysis of the sample.

By contrast, COMPARATIVE RESEARCH focuses not on relationships between variables or on problems of inference and prediction but on the problem of making sense of a relatively small number of cases, selected because they are substantively or theoretically important in some way (Eckstein, 1975). For example, a researcher might use the comparative method to study a small number of guerilla movements in an in-depth manner. Suppose these movements were all thought to be especially successful in winning popular support. To find out how they did it, the researcher would conduct in-depth studies of the movements in question. This case-oriented comparative study would differ dramatically from a "textbook" study. In the latter, the researcher would sample guerilla movements from the population of such movements (assuming this could be established) and then characterize these movements in terms of generic variables and their relationships (e.g., the correlation between the extent of foreign financial backing and the degree of popular support).

As these examples show, the key contrast is the researcher's proximate goal: Does the researcher seek to understand specific cases or to document general patterns characterizing a population? This contrast follows a longstanding division in all of science, not just social science. Georg Henrik von Wright (1971, pp. 1–33) argues in *Explanation and Understanding* that there are two main traditions in the history of ideas regarding the conditions an explanation must satisfy to be considered scientifically respectable. One tradition, which he calls "finalistic," is anchored in the problem of making facts understandable. The other, called "causal-mechanistic," is anchored in the problem of prediction. The contrasting goals of comparative research and "textbook" social research closely parallel this fundamental division.

CONSTITUTION OF CASES AND POPULATIONS

Textbook presentations of social science methodology rarely examine the problem of constituting cases and populations. The usual case is the individual survey respondent; the usual population is demarcated by geographic, temporal, and demographic boundaries (e.g., adults in the United States in the year 2003). The key textbook problematic is how to derive a representative sample from the very large population of observations that is presumed to be at the researcher's disposal.

By contrast, comparative researchers see cases as meaningful but complex configurations of events and structures (e.g., guerilla movements), and their cases are typically macrolevel. Furthermore, they treat cases as singular, whole entities purposefully defined and selected, not as homogeneous observations drawn at random from a pool of equally plausible selections. Most comparative studies start with the seemingly simple idea that social phenomena in like settings (such as organizations, countries, regions, cultures, etc.) may parallel each other sufficiently to permit comparing and contrasting them. The clause, "may parallel each other sufficiently," is a very important part of this formulation. The comparative researcher's specification of relevant cases at the start of an investigation is really nothing more than a working hypothesis that the cases initially selected are in fact instances of the same thing (e.g., guerilla movements) and are alike enough to permit comparison. In the course of the research, the investigator may decide otherwise and drop some cases or even whole categories of cases because they do not appear to belong with what seem to be the core cases. Sometimes, this process of sifting through the cases leads to an enlargement of the set of relevant cases and a commensurate broadening of the scope of the guiding concepts (e.g., from guerilla movements to radical opposition movements) or to the development of a typology of cases (e.g., types of guerilla movements).

Usually, this sifting and sorting of cases is carried on in conjunction with concept formation and elaboration (Walton, 1992). Concepts are revised and refined as the boundary of the set of relevant cases is shifted and clarified (Ragin, 1994). Important theoretical distinctions often emerge from this dialogue of ideas and evidence. The researcher's answers to both "What are my cases?" and "What are these cases of?" may change throughout the course of the research, as the investigator learns more about the phenomenon in question and refines his or her guiding concepts and analytic schemes.

N OF CASES

One key lesson in every course in quantitative social research is that "more cases is better." More is better in two main ways. First, researchers must meet a threshold number of cases to even apply quantitative methods, usually cited as an N of 30 to 50. Second, the smaller the N, the more the data must conform to the assumptions of statistical methods, for example, the assumption that variables are normally distributed or the assumption that subgroup variances are roughly equal. However, small Ns almost guarantee that such assumptions will not be met, especially when the cases are macrolevel. This textbook bias toward large Ns dovetails with the implicit assumption that cases are empirically given, not constructed by the researcher, and that they are abundant. The only problem, in this light, is whether the researcher is willing and able to gather data on as many cases as possible, preferably hundreds if not thousands.

By contrast, comparative research is very often defined by its focus on phenomena that are of interest because they are rare—that is, precisely because the N of cases is small. Typically, these phenomena are large scale and historically delimited, not generic in any sense. The population of cases relevant to an investigation is often limited by the historical record to a mere handful. The key contrast with textbook social science derives from the simple fact that many phenomena of interest to social scientists and their audiences are historically or culturally significant. To argue that social scientists should study only cases that are generic and abundant or that can be studied in isolation from their historical and cultural contexts would severely limit both the range and value of social science.

THEORY TESTING

Conventional presentations of social science methodology place great emphasis on theory testing. In fact, its theory-testing orientation is often presented as what makes social science scientific. Researchers are advised to follow the scientific method and develop their hypotheses in isolation from the analysis of empirical evidence. It is assumed further that existing theory is sufficiently well formulated to permit the specification of testable hypotheses and that social scientific knowledge advances primarily through the rejection of theoretical ideas that consistently fail to find empirical support. Of course, the textbook view does not bar from social science those ideas that are generated in a more inductive manner, but any such idea is treated as suspect until it is subjected to an explicit test, using data that are different from those used to generate the idea.

It is without question that theory plays a central role in social research and that, in fact, almost all social research is heavily dependent on theory in some way. However, it is usually not possible to apply the theory-testing paradigm in small-N, case-oriented research. Most comparative research has a very strong inductive component. Researchers typically focus on a small number of cases and may spend many years learning about them, increasing their depth of knowledge. Countless ideas and hypotheses are generated in this process of learning about cases, and even the very definition of the population (What are these cases of?) may change many times in the course of the investigation. Furthermore, because the number of relevant cases in comparative research is usually very small and limited by the historical record, it is simply not possible to test an idea generated through in-depth study using "another sample" of cases. Very often, the cases included at the outset of an investigation comprise the entire universe of relevant cases.

The immediate objective of most comparative research is to explain the "how" of historically or culturally significant phenomena—for example, How do guerilla movements form? Theory plays an important orienting function by providing important leads and guiding concepts for empirical research, but existing theory is rarely well formulated enough to provide explicit hypotheses in comparative research. The primary scientific objective of comparative research is not theory testing, per se, but concept formation, elaboration, and refinement. In comparative research, the bulk of the research effort is often directed toward constituting the cases in the investigation and sharpening the concepts appropriate for the cases selected (Ragin & Becker, 1992).

THE DEPENDENT VARIABLE

One of the most fundamental notions in textbook presentations of social research is the idea of the *variable*—a trait or aspect that varies from one case to the next—and the associated idea of looking for patterns in how variables are related across cases. For example, do richer countries experience less political turmoil than poorer countries? If so, then social scientists might want to claim that variation in political turmoil across countries (the dependent or outcome variable) is explained in part by variation in country wealth (the independent or causal variable). Implicit in these notions about variables is the principle that the phenomena that social scientists wish to explain must *vary* across the cases they study; otherwise, there is nothing to explain. It follows that researchers who study outcomes that vary little from one case to the next have little hope of identifying the causes of those outcomes.

Comparative researchers, however, often intentionally select cases that do not differ from each other with respect to the outcome under investigation. For example, a researcher might attempt to explain how guerilla movements come about and study instances of their successful formation. From the viewpoint of textbook social research, however, this investigator has committed a grave mistake—selecting cases that vary only slightly, if at all, on the dependent variable. After all, the cases are all instances of successfully formed guerilla movements. Without a dependent variable that varies, this study seems to lack even the possibility of analysis.

These misgivings about studying instances of "the same thing" are well reasoned. However, they are based on a misunderstanding of the logic of comparative research. Rather than using variation in one variable to explain variation in another, comparative researchers often look for causal conditions that are common across their cases—that is, across all instances of the outcome (see Robinson, 1951). For example, if all cases of successfully formed guerilla movements occur in countries with both high levels of agrarian inequality and heavy involvement in export agriculture, then these structural factors might be important to the explanation of guerilla movements. In fact, the very first step in the comparative study of successfully formed guerilla movements, after constituting the category and specifying the relevant cases, would be to identify causal conditions that are common across instances. Although using constants to account for constants (i.e., searching for causal commonalities shared by similar instances) is common in comparative research, it is foreign to techniques that focus on relationships among variables—on causal conditions and outcomes that vary substantially across cases.

Sometimes, it is possible for comparative researchers to identify relevant negative cases to compare with positive cases. For example, guerilla movements that failed early on might be used as negative cases in a study of successfully established guerilla movements. Negative cases are essential when the goal is to identify the *sufficient* conditions for an outcome (Ragin, 2000). For example, if a comparative researcher argues that a specific combination of causal conditions is sufficient for the successful establishment of a guerilla movement, it is important to demonstrate that these conditions were not met in the near misses (i.e., in relevant negative cases). Note, however, that it is possible to assess the sufficiency of a set of causal conditions only if the researcher is able to constitute the set of relevant negative cases. Often, it is very difficult to do so, either because the empirical category is too broad or vague (e.g., the category "absence of a successful guerilla movement") or the evidence is impossible to locate or collect (e.g., data on aborted movements). When these obstacles can be overcome, it is important to understand that the constitution and analysis of the positive cases is a necessary preliminary for the constitution and analysis of negative cases (see also Griffin et al., 1997). Thus, even when a comparative researcher plans to look at an outcome that varies across cases (i.e., present in some and absent in others), it is often necessary first to study cases that all have the outcome (i.e., the positive cases).

CAUSAL ANALYSIS

The main goal of most analyses of social data, according to textbook presentations of the logic of social research, is to assess the relative importance of independent variables as causes of variation in the dependent variable. For example, a researcher might want to know which has the greater impact on the longevity of democratic institutions, their design or their perceived legitimacy. In this view, causal variables compete with each other to explain variation in an outcome variable. A good contender in this competition is an independent variable that is strongly correlated with the dependent variable but has only weak correlations with its competing independent variables.

Comparative researchers, by contrast, usually look at causes in terms of combinations: How did relevant causes combine in each case to produce the outcome in question? Rather than viewing causes as competitors, comparative researchers view them as raw ingredients that combine to produce the qualitative outcomes they study. The effect of any particular causal condition depends on the presence and absence of other conditions, and several different conditions may satisfy a general causal requirement—that is, two or more different causes may be equivalent at a more abstract level. After constituting and selecting relevant instances of an outcome (e.g., successfully formed guerilla movements) and, if possible, defining relevant negative cases as well, the comparative investigator's task is to address the causal forces behind each instance, with special attention to similarities and differences across cases. Each case is examined in detail, using theoretical concepts, substantive knowledge, and interests as guides to answer "how" the outcome came about in each positive case and why it did not in the negative cases (assuming they can be confidently identified). A common finding in comparative research is that different combinations of causes may produce the same outcome (Ragin, 1987; see also Mackie, 1974).

SUMMARY

The contrasts sketched in this entry present a bare outline of the distinctive features of comparative research, from the constitution of cases to the examination of the different combinations of conditions linked to an outcome across a set of cases. The key point is that comparative research does not conform well to textbook admonitions regarding the proper conduct of social research. Social scientists and their audiences are often interested in understanding social phenomena that are culturally or historically significant in some way. Of necessity, these phenomena are often rare and can be understood only through in-depth, case-oriented research.

—Charles C. Ragin

See also CASE STUDY

REFERENCES

Eckstein, H. (1975). Case study and theory in political science. In F. I. Greenstein & N. W. Polsby (Eds.), *Handbook of political science: Vol. 7. Strategies of inquiry* (pp. 79–137). Reading, MA: Addison-Wesley.

Griffin, L., Caplinger, C., Lively, K., Malcom, N. L., McDaniel, D., & Nelsen, C. (1997). Comparative-historical analysis and scientific inference: Disfranchisement in the U.S. South as a test case. *Historical Methods, 30,* 13–27.

Mackie, J. L. (1974). *The cement of the universe: A study of causation.* Oxford, UK: Oxford University Press.

Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies.* Berkeley: University of California Press.

Ragin, C. C. (1994). *Constructing social research.* Thousand Oaks, CA: Pine Forge Press.

Ragin, C. C. (2000). *Fuzzy-set social science.* Chicago: University of Chicago Press.

Ragin, C. C., & Becker, H. S. (1992). *What is a case? Exploring the foundations of social inquiry.* New York: Cambridge University Press.

Robinson, W. S. (1951). The logical structure of analytic induction. *American Sociological Review, 16,* 812–818.

von Wright, G. H. (1971). *Explanation and understanding.* Ithaca, NY: Cornell University Press.

Walton, J. (1992). Making the theoretical case. In C. C. Ragin & H. S. Becker (Eds.), *What is a case? Exploring the foundations of social inquiry* (pp. 121–137). New York: Cambridge University Press.

COMPARATIVE RESEARCH

The major aim of comparative research is to identify similarities and differences between social entities.

Comparative research seeks to compare and contrast nations, cultures, societies, and institutions. Scholars differ on their use of the terminology: To some, comparative research is strictly limited to comparing two or more nations (also known as "cross-national research"), but other scholars prefer to widen the scope to include comparison of many different types of social and/or cultural entities. Yet other scholars use the term to encompass comparisons of subcultures or other social substrata either within or across nation-states or other cultural and social boundaries.

Although scholars are far from a consensus on a definition, the trend appears to be toward defining comparative research in the social sciences as research that compares systematically two or more societies, cultures, or nations. In actual implementation, comparisons of nations prevail as the dominant practice. The intent of comparative research is more universal: Comparative research aims to develop concepts and generalizations based on identified similarities and differences among the social entities being compared, especially in their characteristic ways of thinking and acting; in their characteristic attitudes, values, and ideologies; and in the intrinsic elements of their social structures. This then serves as a means of enhancing one's understanding and awareness of other social entities.

HISTORICAL DEVELOPMENT

Sociology's founding fathers were all comparative researchers. Karl Marx, Max Weber, Emile Durkheim, and Alexis de Tocqueville, to name just a few, all firmly committed themselves to the comparative method, whether they studied roles, institutions, societies, nations, cultures, groups, or organizations. Durkheim, for example, sought to find a general "law" that would explain national and occupational suicide rate variations. In the early 20th century, though, comparative research waned because its methodological resources were deemed insufficiently rigorous during a time when scholars were insisting that social research conform to greater and greater levels of methodological precision. However, since World War II, comparative research has once again assumed its pivotal position in social science research, due in part to improved methodologies and methodological tools and in part to the international climate that emerged after World War II. That climate (the once polarized communist and noncommunist worlds, for example) and our ensuing internationalization have both contributed to the significant reemergence of comparative social science research.

APPLICATION

The purposes of comparative research are many, but one key task is to support and contribute to theory formation. Although theoretical frameworks drive the construction of comparative research endeavors, the results of such research often drive theory reformation. Another key task is to support policymaking, principally at the national level. Can Country B use the strategic policies of Country A in coping with a given social problem (drug abuse, for example)? Or are there unique differences between the two countries that render such a strategy impractical? Yet another purpose of comparative research is to ascertain whether the same dimension of a given concept (e.g., religious commitment) can be used as a common social indicator. Does a given concept generalize to all nations (or other social entities)?

Comparative research is all about perspective. Researchers in one nation need a means to adopt and understand the perspectives of their counterparts in other nations. As the world becomes ever more globalized, the need for such understanding should be clear: National policies need to consider and encompass the needs of the global partners in any way affected by policy implementation. Do policymakers have the necessary perspective to ensure such cooperation and compliance?

Comparative social science research is implemented under a variety of methodologies, ranging from strictly qualitative to rigorously quantitative. Case studies and similar kinds of analyses typify the descriptive qualitative approach, which can be conducted by researchers who are members of the social entities being studied or by researchers from outside the social entities being examined. The quantitative approaches to comparative research have assumed a more dominant position in recent years as the tools (especially computational, statistical, and mathematical) have become available to deal with the many methodological complexities of rigorous comparative research.

Empirical research surveys have become quite common, despite their often enormous complexity and expense. Complexities include the language barrier, which makes preparing comparable questionnaires difficult, and "cultural" barriers, which often make the preparation of comparable questionnaires virtually impossible: If you ask Americans, "Do you believe in God?" almost everyone says yes, whereas if you ask the same question in Japan, only about one third say yes, meaning that, beyond identifying these consistent response percentages for each country, asking the question in the two environments makes little comparative sense because no meaningful implication can be inferred. Similarly, asking about employment versus unemployment is irrelevant to a social group in which there is no such market-based distinction. In this instance, there is no "conceptual equivalence" between the two social groups in terms of the notion of "employment."

Recently, advances in statistical methods have significantly improved the social science researcher's ability to analyze primary data and/or reanalyze secondary data. Methods such as COHORT ANALYSIS have been developed that can ease the need for rigorous sample selection, for instance, and greatly facilitate the detection and explication of similarities and differences among group data. Of course, like all rigorous scientific research, comparative social science research is subject to the usual requirements for statistical validity and reliability, among others.

Examples of ongoing comparative research surveys include the Gallup Polls (since 1945), the General Social Survey (since 1972), the Eurobaromètre (since 1973), the European Community Household Study (since 1994), and the International Social Survey Program (ISSP), which, since 1984, has conducted general social attitude surveys among more than 37 nations. The Japanese National Character Study has been carried out in Japan about every 5 years since 1953. Each of these surveys is conducted periodically over time and usually in multiple countries.

Why are nations different on some characteristic parameters but the same on others? The same question can be asked of other sets of sociocultural groupings, both across regional boundaries and within national or ethnic boundaries. Because national boundaries are not always the same as ethnic boundaries, comparative research is not always amenable to simple national boundaries. This phenomenon has been especially relevant to the emerging countries of Eastern Europe in recent years.

The ultimate intent of comparative research is to contribute to theory building or rebuilding by requiring yet more and more rigor on the part of theory, by insisting that research be replicable across different social groups, by facilitating the application of generalizations across different social groups, and by explaining unique patterns.

FUTURE DIRECTIONS

If, as globalization implies, there is eventually social and cultural convergence in the world, comparative research will once again wane as there will be few distinct social entities to compare. In the meantime, however, there remains much variation and variability in the world, and comparative research promises to further explain this to us and to further enhance our understanding and awareness of one another.

—Masamichi Sasaki

REFERENCES

Hantrais, L., & Mangen, S. (Eds.). (1996). *Cross-national research methods in the social sciences.* London: Pinter.

Inkeles, A., & Sasaki, M. (Eds.). (1996). *Comparing nations and cultures.* Englewood Cliffs, NJ: Prentice Hall.

Przeworski, A., & Teune, H. (1970). *The logic of comparative social inquiry.* New York: John Wiley.

Smelser, N. (1997). *Problems of sociology.* Berkeley: University of California Press.

COMPARISON GROUP

In an experiment testing the effects of a treatment, a comparison group refers to a group of units (e.g., persons, classrooms) that receive either no TREATMENT or an alternative treatment. The purpose of a comparison group is to serve as a source of COUNTERFACTUAL causal inference. A counterfactual is something that did not happen (it is counter to fact); in this case, the counterfactual to receiving treatment is not receiving treatment. According to the theory of counterfactual causal inference, the effect of a treatment is the difference between the outcome of those who received a treatment and the outcome that would have happened had those same people simultaneously not received treatment. Unfortunately, it is not possible to observe that counterfactual, that is, to both give and not give the treatment to the same people simultaneously. Therefore, researchers often use a comparison group as a substitute for the true but unobservable counterfactual, and then they estimate the treatment effect as the difference between the outcomes in the treatment group and the outcomes in the comparison group. To do this well, researchers should make sure that the units in the comparison group are as similar as possible to the units in the treatment group. That similarity can be facilitated by RANDOMLY ASSIGNING units to either treatment or control, resulting in a randomized experiment, or it can be facilitated less well by various QUASI-EXPERIMENTAL methods such as MATCHING units on variables thought to be related to outcome (Shadish, Cook, & Campbell, 2002).

KINDS OF COMPARISON GROUPS

A treatment group can be compared to many different kinds of comparison groups, for example, (a) a no-treatment condition, sometimes called as a CONTROL GROUP, in which the participants receive no intervention to alter their condition; (b) a treatment-as-usual condition, in which the standard care or treatment is provided; (c) a PLACEBO condition, in which the intervention mimics treatment but does not contain the active ingredients of that treatment; (d) a wait-list condition, in which participants are promised the same intervention that the treatment group received after the experiment ends; or (e) an alternative treatment condition, in which participants receive an intervention other than the one of primary interest in the experiment. In the latter case, the alternative treatment is often the best treatment currently available for a problem—a "gold standard"—to see whether the experimental treatment can improve on the outcomes it produces. A further distinction is that in BETWEEN-GROUPS comparisons, the comparison group is a different set of participants than the treatment group, but in WITHIN-SUBJECTS comparisons, the same participants serve sequentially in both the comparison and treatment groups. Within-subjects comparisons require fewer units to test a treatment but can suffer from various treatment contamination effects such as practice effects, fatigue effects, and CARRYOVER of treatment effects from one condition to another (Shadish et al., 2002).

HISTORICAL DEVELOPMENT

Comparison groups may have been used to aid causal inferences as early as ancient eras. Rosenthal and Rosnow (1991) relate a story of a magistrate in ancient Egypt who sentenced criminals to be executed by poisonous snakes. When the magistrate investigated why none of the criminals died after being bitten, he learned that they had been given citron to eat. To determine whether citron had prevented their fatalities, he fed citron to only half of the prisoners and subjected all prisoners to snake bites. In this example, the participants who were not given the citron were the comparison group, all of whom died after the snake bit, whereas none of those in the treatment group who received citron died.

EXAMPLES

Antoni et al. (1991) used a no-treatment control group to examine the effects of a cognitive-behavioral stress management treatment on stress levels of homosexual men prior to notifying them of their HIV status. Those in the treatment group received training in methods to reduce stress. Their stress levels were compared to stress levels in a no-treatment control group that did not receive such training. The treatment effect was the difference in stress levels between treatment and comparison groups after HIV status notification.

Mohr et al. (2000) used a treatment-as-usual comparison group to examine the effects of a cognitive-behavioral treatment to reduce depression in patients with multiple sclerosis (MS). The treatment group was given a new experimental treatment for depression, whereas the comparison group was given the usual treatment provided to MS patients suffering from depression. The resulting treatment effect is the difference in depression scores between those receiving standard treatments for depression compared to those receiving the new treatment for depression.

Kivlahan, Marlatt, Fromme, Coppel, and Williams (1990) used a wait-list control group to examine the effects of a skills program to reduce alcohol consumption among college students. The treatment was withheld from the comparison group until the end of the experiment, after which the comparison group was offered the opportunity to receive the same treatment given to the treatment group.

—William R. Shadish and M. H. Clark

See also EXPERIMENT, TREATMENT

REFERENCES

Antoni, M. H., Baggett, L., Ironson, G., LaPerriere, A., August, S., Kilmas, N., et al. (1991). Cognitive-behavioral stress management intervention buffers distress responses and immunologic changes following notification of HIV-1 seropositivity. *Journal of Consulting and Clinical Psychology, 59*(6), 906–915.

Kivlahan, D. R., Marlatt, G. A., Fromme, K., Coppel, D. B., & Williams, E. (1990). Secondary prevention with college drinkers: Evaluation of an alcohol skills training program. *Journal of Consulting and Clinical Psychology, 58*(6), 805–810.

Mohr, D. C., Likosky, W., Bertagnolli, A., Goodkin, D. E., Van der Wende, J., Dwyer, P., et al. (2000). Telephone administered cognitive-behavioral therapy for the treatment of depressive symptoms in multiple sclerosis. *Journal of Consulting and Clinical Psychology, 68*(2), 356–361.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

COMPETING RISKS

A set of mutually exclusive, discrete outcomes or events is called *competing risks.* This term arises primarily in the literatures on SURVIVAL ANALYSIS and EVENT HISTORY ANALYSIS.

As long ago as the 18th century, scholars interested in survival and its opposite (e.g., death, failure) recognized that deaths and failures occur for a variety of reasons. For example, a person may die from a contagious disease, an organic illness (e.g., cancer), an accident, suicide, or intentional violent action (e.g., murder, war). These alternative causes of death illustrate competing risks of dying. Frequently, scholars are interested not only in mortality and failure for all reasons combined but also in the diverse reasons for death or failure and in the associated cause-specific rates of mortality or failure. That is, they want to differentiate among competing risks of some event, such as death or failure.

Although the language and application of the idea of competing risks were originally developed in the fields of health, medicine, demography, engineering, and related fields concerned with death or failure, they apply to a broad spectrum of social scientific research topics. For example, researchers studying labor market processes may wish to analyze not only exits from jobs overall but also whether people leave jobs because they have been promoted or demoted within an organization, moved laterally, quit, been fired, retired, or died. Scholars studying organizational success may want to examine not only all forms of organizational "death" but also whether organizations cease all activities, are acquired by another firm, or merge with another firm in something resembling an equal partnership. Social scientists studying international conflicts and wars may desire to examine not only the end of conflicts overall but also whether conflict ends through a negotiated peace process, by conquest, or with a stalemate. In each instance, scholars would like to distinguish among competing risks.

A recurring type of question concerns the consequences of altering the likelihood of one or more competing risks. A typical question is as follows: What would be the consequences of suppressing or completely eliminating a specific risk (e.g., a particular cause of death or failure)? To illustrate, the introduction of public sanitation and widespread immunization against contagious diseases greatly reduced the likelihood of catching contagious diseases. As a result, the risks and rates of dying from many of the most lethal contagious diseases were greatly diminished. Reduction in the rate of dying from any competing risk tends to extend life, causing an increase in the average length of life. More generally, lowering the HAZARD RATE of one competing risk prolongs the average time to the occurrence of an event for all reasons combined.

Competing risks of events are often highly age and time dependent. Consequently, the implications of a certain reduction in the hazard rate for one competing risk rather than for another can differ substantially, depending on the pattern of age and time dependence in each cause-specific hazard rate. For example, heart disease and cancer strike older people much more than children and youths, whereas contagious diseases and accidents befall infants and children more than adults. As a result, a certain reduction in the risk of dying from a contagious disease tends to lengthen life much more than a comparable reduction in the risk of dying from heart disease. Thus, improved sanitation and immunization greatly increased average life expectancy (and the length of life as a working-age adult) by causing the overall risk of dying at young ages to fall precipitously. In contrast, reducing the risk of dying from heart disease especially extends the remaining life of the elderly and particularly increases the length of life spent in retirement.

An indirect consequence of a declining rate of death from contagious diseases that especially strike children was an increase in the relative likelihood of dying from heart disease, cancer, and other diseases to which elderly people are particularly susceptible. Improved sanitation and immunization did not eliminate other competing risks of dying; indeed, they may have not directly affected the rates of dying from these other causes at all. In general, direct suppression or elimination of one competing risk tends to increase the *relative* chances of death or failure due to competing risks other than the one that was suppressed or eliminated. Contrarily, amplification of one competing risk (e.g., murder, war) tends to decrease the *relative* chances of death or failure due to other competing risks (e.g., disease).

There are two main types of models for analyzing competing risks and the effects of measured covariates on the occurrence and timing of one competing risk rather than another. The two types are differentiated on the basis of whether risks may occur only at discrete points in time or at any continuous moment in time.

In the first instance, the most common statistical models are a MULTINOMIAL LOGISTIC REGRESSION MODEL or a conditional logistic regression model (cf. Theil, 1969, & McFadden, 1973, respectively, as cited in Maddala, 1983). In the second instance, event history analysis is the standard way of analyzing empirical data on competing risks. The specific models used in event history analysis include various types of proportional or nonproportional models of hazard and transitions rates. (For an overview of the main alternative models used in event history analysis, see Tuma & Hannan, 1984.)

A key issue in both types of modeling approaches is whether the competing risks are statistically independent. To give one example, a person who has an especially serious illness such as AIDS often has a heightened risk of dying from an infectious disease. If this is true, the two causes of death (e.g., AIDS and an infectious disease) are not statistically independent. In contrast, deaths from accidents, murder, and war may plausibly be considered to be independent (or nearly independent) of the risks of dying from a chronic illness such as AIDS.

When using discrete time multinomial logistic regression models to analyze the cause-specific occurrence of various outcomes, researchers may test the independence of irrelevant alternatives (Debreu, 1960, as cited in Maddala, 1983), that is, whether distinct competing risks are statistically independent. If the competing risks are statistically independent, then estimates for a multinomial logistic regression model will be identical (within the limits of sampling variability) to the results for a series of simpler logistic regression models that exclude one or more of the competing risks. For a discussion of these tests, see Hausman and McFadden (1984).

In analyses of event histories, researchers may estimate the instantaneous transition rate (i.e., the cause-specific hazard rate) associated with each competing risk. The results of such analyses may then be used to infer the consequences of eliminating a particular competing risk, under the assumption that it is statistically independent of all other competing risks. Similarly, the same results may be used to infer the consequences of raising or lowering the transition rate associated with a certain risk (e.g., by multiplying the estimated transition rate by some hypothetical fraction greater than or less than 1). Such inferences are legitimate under the assumption that the competing risks are statistically independent (David & Moeschberg, 1978). Tsiatis (1975) proved that, even with full and complete event history data on the occurrence and timing of competing risks, it is impossible to establish conclusively that competing risks are truly statistically independent.

—Nancy Brandon Tuma

REFERENCES

David, H. A., & Moeschberg, M. L. (1978). *The theory of competing risks* (Griffin's Statistical Monograph #39). New York: Macmillan.

Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica, 52,* 1219–1240.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics.* Cambridge, UK: Cambridge University Press.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences, 72,* 20–22.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods.* Orlando, FL: Academic Press.

COMPLEX DATA SETS

There are two distinct senses in which data are complex: statistical and structural. We can explain these by contrasting them with a notional simple data set. This would consist of a set of entities corresponding to either an equal probability random sample of a population, in which each entity is independently selected, or a complete population. The data set contains the same complete set of measurements about each entity. For example, it might consist of a random sample of individuals who each respond to a set of questions about their political attitudes. Data sets are described as complex when they depart from these assumptions.

Statistical complexity arises when some of the standard assumptions about sampling do not hold, and standard statistical analyses may give misleading results without adjustment. Skinner, Holt, and Smith (1989) give an overview of this sort of statistical complexity. One example of a departure occurs if, in a survey, respondents are clustered in a limited set of areas, as is often done to reduce fieldwork costs. Observations are not then independent. This will mean that standard errors used to compute significance tests will need to be adjusted compared to the standard formulas used. For example, the British Social Attitudes Survey, which might at first sight appear to be the type of simple data set envisaged in the first paragraph, has a clustered design. Other forms of lack of independence also occur in spatial and time-series data sets. Another departure occurs if the sample is not drawn with equal probability. Estimates will then be biased and need to be adjusted, for example, by weighting. This is quite typical in surveys of employers, in which large employers tend to be selected with higher probability. Similarly, if there are variations in the probabilities of response by sampled individuals, either to the whole interview or to individual questions, then there may be biases that require adjustment.

Structural complexity arises when we have information about different sorts of entities in a single data set. In particular, many data sets have a multilevel structure, reflecting some important aspect of social reality. Examples include not only hierarchical structures, such as individuals located in households, workers in firms, and children in schools, but also multiple observations in longitudinal data of the same individual at different points in time. The collection of information about all individuals in households is quite typical in household budget or labor force surveys. There are other less clearly hierarchical structures, for example, in the study of social networks. Data sets may also be seen as complex when there are different amounts of information about each primary entity—for example, different numbers of observations in a panel study, different numbers of children born to each woman in a fertility history, or different numbers of jobs held in a work history. For instance, a survey such as the British Household Panel Survey will have a different number of waves for different individuals both because of ATTRITION and because children reach the eligible age for interview.

There are a number of consequences of this structural complexity. First, it means that analysts have to make choices about what their unit of analysis should be. It is possible with these structures to treat higher level units as the unit of analysis and attach summary information about lower level units (e.g., household-level analysis, using information about the income of individuals in the household or analyzing women by number of children born). Alternatively, it is possible to use a lower level for analysis and attach summary information about the higher level (e.g., analysis of each separate birth). Second, given these choices, the data may require substantial restructuring to create the rectangular data sets normally used by statistical software. Finally, analysts must take account of statistical consequences of the structural complexity. This might be simply to adjust for nonindependence of observations, for example, but more positively, it may involve taking advantage of the structure by using multilevel modeling techniques (see Goldstein, 1995).

—Nicholas H. Buck

REFERENCES

Goldstein, H. (1995). *Multilevel statistical models.* London: Edward Arnold.

Skinner, C., Holt, D., & Smith, T. M. F. (1989). *Analysis of complex surveys.* Chichester, UK: Wiley.

COMPONENTIAL ANALYSIS

Componential analysis (CA) assumes that the meaning of any given word is represented best by a unique bundle of meaningful features. The analytical method of CA is to compare the meanings of words from the same area of meaning to discover the minimum number of features necessary to distinguish the differences in their meanings. For example, in the area of kinship, a person's *mother* and *father, sisters* and *brothers,* and *daughters* and *sons* are distinguished by gender ("female" vs. "male") and by generation ("one generation earlier" for *mother* and *father,* "same generation" for *sister* and *brother,* and "one generation later" for *daughter* and *son*).

The insights gained through CA are regarded as revealing how the people of different societies uniquely structure their knowledge of the world and interpret life experience. CA is therefore regarded as one of the main tools for addressing the nature of cultural and linguistic relativism.

Proponents of CA purport that it provides an insightful, elegant, and rigorous method for determining lexical meanings. It is insightful in that it reveals unexpected bases for meaningful relations, elegant in that it exhibits a simplicity and regularity not available through other methods, and rigorous in that it provides precise definitions. Most dictionary definitions are composed of such features (e.g., *trudge* "to *walk* in a *laborious* and *weary* manner").

Proponents of CA also hope that it will provide a set of universal features that will allow for the analysis of any meaning in any language, thereby overcoming imprecision when translating between languages. The translator may compensate for the difference in the meanings of words from different languages by adding phrases to account for missing meaningful features. For example, when another language has several words for specific ways to *carry* something, one may express those differences by adding to *carry* phrases, such as "in the hands," "on the hip," "under the arm," "over the shoulder," and so forth.

From the mid-1950s through the 1970s, CA played a major role in the development of the discipline of COGNITIVE ANTHROPOLOGY (D'Andrade, 1995). It was subsequently rejected by the academic community, however, for a number of reasons. First, it allows for alternative analyses that cannot be evaluated in terms of how well each reflects a cultural insider's knowledge. Second, it cannot account for the meanings of a wide range of domains. For example, words that reflect a part-whole relationship, such as the *head* as a part of the body, may only be understood in relation to their position relative to the other parts of the whole rather than by a list of features (Fillmore, 1975). Third, categories that lack clearly defined boundaries defy analysis by CA. Examples include the stages of life (*infant, child, adolescent, adult,* and *elder*) and size (*sand, gravel, pebbles, stones, rocks,* and *boulders*). Fourth, the meanings of many words can only be understood in terms of the broader cultural context. For example, a *fiddle* can be discriminated from a *violin* because it may be grouped with a guitar, ukulele, and a banjo and played to accompany folk dancing at a country hoedown by men dressed in bib overalls and plaid shirts. A violin may be grouped with another violin, a viola, and a cello as part of a string quartet and played at a chamber music concert by players dressed in tuxedos and gowns.

The enduring value of CA is that it highlights the fact that people discriminate things on the basis of whether they are the same or different.

—Kenneth A. McElhanon and Thomas N. Headland

See also COGNITIVE ANTHROPOLOGY, EMIC/ETIC DISTINCTION

REFERENCES

D'Andrade, R. (1995). *The development of cognitive anthropology.* Cambridge, UK: Cambridge University Press.

Fillmore, C. (1975). An alternative to checklist theories of meaning. In C. Cogen, H. Thompson, G. Thurgood, K. Whistler, & J. Wright (Eds.), *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society* (pp. 123–131). Berkeley, CA: Berkeley Linguistics Society.

Nida, E. A. (1975). *Componential analysis of meaning: An introduction to semantic structures.* The Hague, The Netherlands: Mouton.

COMPUTER SIMULATION

A computer simulation uses a set of equations, ALGORITHMS, or electronic components to reproduce the essential components of a more complex system. Its purpose is to allow the complex system to be studied under conditions that would be difficult, expensive, or impossible to manipulate in the real world. Computer simulations also allow for mathematical equations to be studied numerically, even when they cannot be solved using algebraic techniques.

Most computer simulations are implemented on digital computers and specified using equations, if-then rules, and RANDOM VARIABLES. A variety of specialized computer languages and packages have been developed for the specification of simulations, although simulations can also be written in general-purpose programming languages such as FORTRAN or C++. The behavior of the simulation will vary depending on the PARAMETERS used in the equations and rules, the values used to initialize the simulation, and the values of the random variables during a specific run of the simulation.

Most computer simulations fall into three categories: systems dynamics simulations, finite element simulations, and agent-based simulations.

Systems dynamics simulations, pioneered by Forrester (1961), are driven by a set of difference equations that relate the variables in the system through a set of feedback loops and random disturbances; Roberts, Andersen, Deal, Grant, and Shaffer (1983) provide an extended introduction. For example, "global dynamics" simulations study highly aggregated variables such as the total level of population, resource consumption, pollution, and economic production. Pollution would increase as production and consumption increase, which might, in turn, reduce the level of population.

Finite element simulations, in contrast, compute the behavior of aggregate variables by looking at the effects of microlevel processes occurring between adjacent elements in the simulation. This approach is typically used to model physical systems, such as climate, in which the microlevel equations are known but the systems cannot be solved algebraically due to the presence of random variables and NONLINEAR equations. For example, climate change models contain equations governing the effects of greenhouse gases such as carbon dioxide on heat retention in the atmosphere, the effect of atmospheric heat on adjacent ocean temperature, the effects of ocean temperature on cloud formation in the atmosphere above the ocean, and so forth. Most finite element simulations can be efficiently implemented on massively parallel computers that contain hundreds or thousands of individual processors, and the most common application of these "supercomputers" is finite element simulation.

Finally, agent-based simulations model social systems by looking at the aggregated effects of individual agents whose behaviors are specified by rules (Axelrod, 1997). For example, in a simulation of traffic flow in a city, the "agents" would be individual cars, which would alter their speed, choice of route, and random behaviors such as the likelihood of being in an accident based on the actions of other agents. See the SIMULATION entry for further details on this approach.

Although most simulations of social systems are done with digital computers, analog computer simulations are occasionally used in MODELING in the social sciences, particularly for aspects of the nervous system and tasks such as pattern recognition. An analog computer is constructed from electronic components such as resistors, capacitors, and transistors. Like a digital simulation, an analog computer is a simplification of a more complex system, but its characteristics can be more easily and reliably manipulated than can be done with the full system.

—Philip A. Schrodt

REFERENCES

Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration.* Princeton, NJ: Princeton University Press.

Forrester, J. W. (1961). *Industrial dynamics.* Cambridge: MIT Press.

Roberts, N., Andersen, D. F., Deal, R. M., Grant, M. S., & Shaffer, W. A. (1983). *Introduction to computer simulation: A system dynamics modeling approach.* Reading, MA: Addison-Wesley.

COMPUTER-ASSISTED DATA COLLECTION

This term can be used in relation to several different kinds of context: (a) face-to-face and telephone interviews, in which the interviewer follows a series of questions from an INTERVIEW SCHEDULE that is on a computer and enters respondents' replies as they go along; (b) SELF-ADMINISTERED QUESTIONNAIRES, in which respondents read the questions on a computer screen and enter their replies as they go along; (c) INTERNET SURVEYS; and (d) online FOCUS GROUPS and personal interviews.

—Alan Bryman

COMPUTER-ASSISTED PERSONAL INTERVIEWING

Computer-assisted personal interviewing (CAPI) and its cousin, computer-assisted telephone interviewing (CATI), dominate survey data collection (see COMPUTER-ASSISTED DATA COLLECTION). There are a large number of software packages available and more coming online. In this entry, we will not attempt to evaluate the alternatives but instead discuss the major technical problems any CAPI system needs to solve and provide the outline of a solution. The advantages of computer-assisted interviewing are so substantial—such as efficient question filters, range checks, and skip pattern checks—that the only candidates for *not* using computer-assisted techniques are very short, simple questionnaires or surveys that involve few respondents or can use low-wage labor to process data and cannot afford the hardware.

Mounting a major SURVEY requires the coordination of many tasks and details that touch data collection. Devising a CAPI software system to handle a large, complex survey is difficult. Unless the CAPI system integrates well with the other aspects of the survey process, the benefits of computerized data collection can be not only offset but also outweighed by the difficulties created by a poorly integrated effort. CAPI initially promised more accurate data collection through built-in checks that prevent or correct errors and inconsistencies. It also promised faster data delivery times from the end of the field effort to the release of the data. CAPI has met these promises. On the other hand, organizations that paid for surveys hoped that CAPI would reduce their costs. In general, this has not happened. We believe that problems with cost reflect a failure to integrate CAPI software with the rest of the survey process.

In mounting a major CAPI effort, there are six closely linked tasks. First, we must design the interview to meet the needs of the project. Second, we must prepare the software that puts the question and, where appropriate, the set of allowable answers on the computer screen. Third, we prepare any input data that drives the interview to the software. The input data can range from the rudimentary—with little or no data about the respondent being fed into the survey—to large and complex data for longitudinal surveys that contain hundreds or even thousands of data points from previous rounds of the study (which can be referenced during the interview). For example, when collecting EVENT HISTORIES, the names and demographic characteristics of household members, names of employers, and dates and types of transitions figure prominently in the input file to a longitudinal survey. Input data can also include sound files for applications in which the computer "reads" the question text to the respondent, a technique used for sensitive questions about sexual practices or substance use when directly reporting to the interviewer might bias the response. The fourth task is to train interviewers to use the software and send them out to collect or "capture" the data. The fifth task is to move the data from the interviewer's portable PC to a central data repository; the sixth task is to prepare data files for analysis from the data in that central repository.

Although the balance of this entry will focus on software, any CAPI effort needs hardware to run on. Currently, there are three categories: laptops, pen-based systems, and small handheld personal digital assistants (PDAs). For large, complex surveys, laptops are currently the best choice for variety, robustness, stability, and product maturity. However, pen-based systems based on Windows/Intel technology are making another play for market share. Earlier pen-based systems did not fare well. The pen-based Windows/Intel machines are distinctly better than earlier generations, but for a notoriously cost-conscious industry, price, reliability, and physical robustness are the keys to accepting the new generation of pen-based PCs for survey use. For short surveys done standing up, the new vintage of PDAs that have more processing power and greater storage space could be strong contenders, but their relatively immature operating systems and smaller collection of software utilities are detrimental to wider use. However, for some applications in which small size, light weight, and modest cost are prerequisites, PDAs are already an excellent choice.

In the effort to write software to handle data collection, systems designers tended to focus on the data capture step. Often, they did not pay sufficient attention to how system design might affect interviewer training, as well as how the data capture software integrated with the other five tasks. With the advent of CAPI, the entire back office operation for survey work has had to be redesigned. Without an overarching conceptual framework for how the major tasks for CAPI data collection should be integrated, the costs of the redesigned survey process have often taken an unfortunate direction.

CAPI/CATI technologies have two major branches. The more common approach uses a customized programming language to present questions, capture answers, and store data. This approach requires extensive support from programmers throughout the survey process, and for every new survey, the interrelationships among these tasks must be specified anew. We will refer to this as the *programming approach.* The other approach to CAPI/CATI software uses relational database methods as a unifying mechanism. With this approach, the instrument is represented as a series of data records, each of which has a large number of attributes, including everything from the question text and skip pattern to documentation notes and instructions to the interviewer. We will refer to this as the *database approach.* (The reader can profitably read the entry on DATA MANAGEMENT, which explains relational databases, as it is germane to this section as well.) The thesis here is that the database approach provides an overarching conceptual approach to integrating all the steps in conducting a CAPI interview and allows for significant operational economies over the alternative programming approach.

In the design stage, survey specialists enter data into the various fields of a screen that populates the rows of relational database tables representing the questionnaire. The data entered include question text, edit restrictions on allowable responses, the list of acceptable answers, and skip instructions. This approach to "authoring" the CAPI questionnaire requires minimal programming skill, and junior survey specialists can enter all but the most complex passages. A set of predesigned "queries" to the database generates diagnostic checks as the data are entered, with more sophisticated checks and diagnostics run in batch mode.

Each question in the instrument is represented by joining rows from the database tables. The interviewing software performs these joins, displays the question, stores the answer, and branches as per the branching instructions. The program that reads and executes the questionnaire data is not rewritten for each survey but remains stable and is reused for different surveys. (The same records can also be used on a PDA, Web server, or local server to enable a variety of platforms for data collection.) This simplifies training for both questionnaire authors and interviewers because all interactions with the system reuse a few standard screen displays.

The survey data are transmitted to the central office and loaded into a master relational database already containing the tables that define the questionnaire. At this point, the questionnaire data become the "metadata," or data that describe the survey data, greatly facilitating the generation of documentation reports that are run as queries to the relational database. Predesigned queries generate either a printable or HTML questionnaire or a codebook. Additional information relevant to each question can be stored in the database and used to augment survey documentation.

We can extract input data files for the next round of a longitudinal survey from the relational database. One can distribute the public use data over the Web to users by employing the relational database techniques used for commercial transactions; instead of putting books in their shopping cart, researchers can search the database for the variables they need, mark them for extraction, and have a remote server e-mail them their data file.

This approach is holistic and uses commercial database software to integrate all phases of a CAPI survey. A well-designed system will handle a broad range of surveys, allowing both central office and interviewing staff to reuse standard tools. This approach offers significant opportunities to control costs and keep operations coordinated and efficient (see DATA MANAGEMENT).

—Randall J. Olsen

REFERENCES

Costigan, P., & Thomson, K. (1992). Issues in the design of CAPI questionnaires for complex surveys. In A. Westlake, R. Banks, C. Payne, & T. Orchard (Eds.), *Survey and statistical computing* (pp. 47–156). London: North Holland.

Couper, M. P., Baker, R. P., Bethlehem, J., Clark, C. Z. F., Martin, J., Nichols, W. L., et al. (Eds.). (1998). *Computer assisted survey information collection.* New York: John Wiley.

Forster, E., & McCleery, A. (1999). Computer assisted personal interviewing: A method of capturing sensitive information. *IASSIST Quarterly, 23*(2), 26–38.

Saris, W. E. (1991). *Computer-assisted interviewing.* Newbury Park, CA: Sage.

CONCEPT. *See* CONCEPTUALIZATION, OPERATIONALIZATION, AND MEASUREMENT

CONCEPTUALIZATION, OPERATIONALIZATION, AND MEASUREMENT

Research begins with a "problem" or topic. Thinking about the problem results in identifying concepts that capture the phenomenon being studied. Concepts, or CONSTRUCTS, are ideas that represent the phenomenon. Conceptualization is the process whereby these concepts are given theoretical meaning. The process typically involves defining the concepts abstractly in theoretical terms.

Describing social phenomena and testing hypotheses require that concept(s) be operationalized. Operationalization moves the researcher from the abstract level to the empirical level, where variables rather than concepts are the focus. It refers to the operations or procedures needed to measure the concept(s). Measurement is the process by which numerals (or some other labels) are attached to levels or characteristics of the variables. The actual research then involves empirically studying the variables to make statements (descriptive, relational, or causal) about the concepts.

CONCEPTUALIZATION

Although research may begin with only a few and sometimes only loosely defined concepts, the early stages usually involve defining concepts and specifying how these concepts are related. In exploratory research, a goal is often to better define a concept or identify additional important concepts and possible relationships.

Because many concepts in social science are represented by words used in everyday conversation, it is essential that the concepts be defined. For example, the concepts *norms, inequality, poverty, justice,* and *legitimacy* are a part of our everyday experiences and thus carry various meanings for different people. The definitions provided by the researchers are usually referred to as nominal definitions (i.e., concepts defined using other words). No claim is made that these definitions represent what these concepts "really" are. They are definitions whose purpose it is to communicate to others what the concept means when the word is used. A goal in social science is to standardize definitions of the key concepts. Arguably, the most fruitful research programs in social science—those that produce the most knowledge—are those in which the key concepts are agreed on and defined the same way by all.

Concepts vary in their degree of abstractness. As an example, human capital is a very abstract concept, whereas education (often used as an indicator of human capital) is less abstract. Education can vary in quality and quantity, however, and thus is more abstract than years of formal schooling. Social science theories that are more abstract are usually viewed as being the most useful for advancing knowledge. However, as concepts become more abstract, reaching agreement on appropriate measurement strategies becomes more difficult.

OPERATIONALIZATION

Operationalization of concepts involves moving from the abstract to the empirical level. Social science researchers do not use this term as much as in the past, primarily because of the negative connotation associated with its use in certain contexts. One such use has been in the study of human intelligence. Because consensus on the meaning of this concept has been hard to come by, some researchers simply argued that intelligence is what intelligence tests measure. Thus, the concept is defined by the operations used to measure it. This so-called "raw empiricism" has drawn considerable criticism, and as a consequence, few researchers define their concepts by how they are operationalized. Instead, nominal definitions are used as described above, and measurement of the concepts is viewed as a distinct and different activity. Researchers realize that measures do not perfectly capture concepts, although, as described below, the goal is to obtain measures that validly and reliably capture the concepts.

MEASUREMENT

Measurement refers to the process of assigning numerals (or other labels) to the levels or characteristics of variables. In moving from the abstract level of concepts to the empirical level of variables, the researcher must think of the concept as having characteristics that can be empirically observed or assessed. A number of concerns must be addressed. The first of these is LEVEL OF MEASUREMENT.

At the most basic level, the researcher must decide whether the underlying features of the concept allow for ordering cases (ordinal level) or allow only for categorizing cases (nominal level). Another distinction concerns whether the features of the concept are discrete or continuous, with fine gradations. Relying on these distinctions, most researchers are faced with variables that are nonordered and discrete (e.g., marital status and religious preference), ordered and discrete (e.g., social class), or ordered and continuous (e.g., income). Developments in statistics over the past several decades have dramatically increased the analysis tools associated with ordered discrete variables, but the majority of the research in social science still adheres to the dichotomy of categorical versus continuous variables, with the more "liberal" researchers assuming that ordinal-level data can be analyzed as if they were continuous. With the advances in statistical techniques, however, these decisions no longer need to be made; ordinal variables can be treated as ordinal variables in sophisticated multivariate research (see Long, 1997).

The first step in measurement, then, is to determine the level of measurement that is inherent to the concept. As a general rule, measurement should always represent the highest level of measurement possible for a concept. The reason for this is simple. A continuous variable that is measured like a category cannot be easily, if ever, converted to a continuous scale. A continuous scale, however, can always be transformed into an ordinal- or nominal-level measure. In addition, ordinal and continuous scales allow for the use of more powerful statistical tests, that is, tests that have a higher likelihood of rejecting the null hypothesis when it should be rejected.

Any measurement usually involves some type of measurement ERROR. One type of error, *measurement invalidity,* refers to the degree to which the measure incorrectly captures the concept (DeVellis, 1991). Invalidity is referred to as systematic error or bias. VALIDITY is usually thought of in terms of the concept being a target, with the measure being the arrow used to hit the target. The degree to which the measure hits the center of the target (the concept) is the measure's validity. Researchers want measures with high validity. The reason for this is simple, as shown by an example from a study of social mobility. If the researcher's goal is to determine the degree of social mobility in a society, and the measure consistently underestimates mobility, then the results with this measure are biased and misleading (i.e., they are invalid).

The other major type of measurement error is *unreliability.* Whereas invalidity refers to accuracy of the measure, unreliability refers to inconsistency in what the measure produces under repeated uses. Measures may, on average, give the correct score for a case on a variable, but if different scores are produced for that case when the measure is used again, then the measure is said to be unreliable. Fortunately, unreliability can be corrected for statistically; invalidity is not as readily corrected, however, and it is not as easily assessed as is RELIABILITY.

Validity is arguably the more important characteristic of a measure. A reliable measure that does not capture the concept is of little value, and results based on its use would be misleading. Careful conceptualization is critical in increasing measurement validity. If the concept is fuzzy and poorly defined, measurement validity is likely to suffer because the researcher is uncertain as to what should be measured. There are several ways of assessing measurement validity.

FACE VALIDITY, whether the measure "on the surface" captures the concept, is not standardized or quantifiable but is, nevertheless, a valuable first assessment that should always be made in research. For example, measuring a family's household income with the income of only the male head of household may give consistent results (be reliable), but its validity must be questioned, especially in a society with so many dual-earner households.

Content validity is similar to face validity but uses stricter standards. For a measure to have content validity, it must capture all dimensions or features of the concept as it is defined. For example, a general job satisfaction measure should include pay satisfaction, job security satisfaction, satisfaction with promotion opportunities, and so on. As with face validity, however, content validity is seldom quantified.

CRITERION-RELATED VALIDITY is the degree to which a measure correlates with some other measure accepted as an accurate indicator of the concept. This can take two forms. *Predictive validity* uses a future criterion. For example, voting preference (measured prior to the election) is correlated with actual voting behavior. A high correlation indicates that voting preference is a valid measure of voting behavior. *Concurrent validity* is assessed by obtaining a correlation between the measure and another measure that has been shown to be a valid indicator of the concept. For example, a researcher in the sociology of work area wishes to measure employee participation by asking employees how much they participated in decision making. A more time-consuming strategy—videotaping work group meetings and recording contributions of each employee—would likely be viewed as a valid measure of participation. The correlation between the two would indicate the concurrent validity of the perceptual measure.

CONSTRUCT VALIDITY of a measure refers to one of two validity assessment strategies. First, it can refer to whether the variable, when assessed with this measure, behaves as it should. For example, if the theory (and/or past research) says it should be related positively to another variable Y, then that relationship should be found when the measure is used. The second use of construct validity refers to the degree to which multiple indicators of the concept are related to the underlying construct and not to some other construct. Often, this distinction is referred to with the terms *convergent validity* and DISCRIMINANT VALIDITY. FACTOR ANALYSIS can be used to assess this. For example, if a researcher has five indicators of cultural capital and four indicators of social capital, a factor analysis should produce two lowly correlated factors, one for each set of indicators.

It is important to stress that this second strategy uses factor analysis for confirmatory purposes, not for data dredging or concept hunting, as it is used sometimes in exploratory research. If the researchers have been careful in conceptualization, then the factor analysis serves to validate their measures. Factor analysis in this instance is not used to "discover" unanticipated concepts.

As stated above, reliability refers to the consistency of the results when the measure is used repeatedly. It is viewed as random measurement error, whereas validity is thought of as measurement bias. There are clear statistical consequences of unreliable measures, with larger variances and attenuated correlations at the top of the list. Reliability is usually assessed in one of two ways. TEST-RETEST RELIABILITY is assessed with the correlation of the measure with itself at two points in time. The difficulty associated with this type of assessment is that unreliability of the measure is confounded with actual change in the variable being measured. For this reason, a second and more often used strategy is to obtain multiple indicators of the concept. Here, internal consistency measures such as the CRONBACH'S ALPHA can be used. In both instances, a normed measure of association (correlation) allows for application of certain rules of thumb. For example, any measures with reliabilities under .60 should be regarded with considerable suspicion.

ADDITIONAL ISSUES

Measurement validity should not be confused with INTERNAL VALIDITY and EXTERNAL VALIDITY, which are research design issues. If researchers claim the results support their hypothesis that X causes Y, this claim is said to have internal validity if alternative explanations for the results have been ruled out. External validity refers to whether the study results can be generalized beyond the setting, sample, or time frame for the study. Both are very important concerns in building social science knowledge, but they are not the same as measurement validity.

Another measurement issue concerns whether single or multiple indicators should be used as measures. Some concepts, almost by definition, can be measured with single indicators. Some examples are income and education. Other more abstract concepts (e.g., political conservatism and marital satisfaction) are often measured with more than one indicator. The reasons for wanting multiple indicators are fairly straightforward. First, when a concept is more abstract, finding one measure that captures it is more difficult. Second, multiple-indicator measures are usually more reliable than single-indicator measures. Third, multiple-indicator measures, when used in STRUCTURAL EQUATION MODELING (SEM) (Bollen, 1989), allow for correcting for unreliability of the measures, produce additional information that allows for making useful model tests, and allow for estimating more complicated models such as reciprocal effects models. Although SEM achieves its strengths by treating multiple indicators separately, measures based on multiple indicators are often transformed into scales or indexes, which are the simple sums or weighted sums of the indicators.

A third measurement issue concerns the source of valid and reliable measures. As has been described, there exist a number of strategies for assessing how good a measure is (i.e., how valid and reliable it is). This assessment happens after the measurement has occurred, however. If the measure is not valid or reliable and other better measures cannot be easily obtained, discontinuation of the research must be seriously considered. For this reason, considerable attention must be given to identifying valid and reliable measures at the onset of the study. If the theory in a specialty area is well established and there already exists a strong research tradition, then valid and reliable measures likely already exist. For example, in the study of organizations, key concepts such as formalization and administrative intensity are well established, as are their measures. In such instances, a careful reading of this literature identifies the measures that should be used. It is always tempting for those new to a specialty area to want to "find a better measure," but unless previously used measures have low validity and reliability, this temptation should be avoided. Accumulation of knowledge is much more likely when measures of the key concepts are standardized across studies. Exploratory research, of course, is much less likely to use standardized measures. In such instances, concepts and measures are developed "along the way." Such research, however, should not be viewed as entirely unstructured and lacking standards of evaluation and interpretation. Maxwell (1996) convincingly shows that exploratory research can still be conducted within quasi-scientific guidelines.

A final measurement issue concerns the use of available data, which now is the major source of data for social science research. Often, these data were collected for some purpose other than what the researcher has in mind. In such instances, PROXY VARIABLES are often relied on, and validity issues usually must be confronted. Reliability is usually less of a problem because proxy measures often seem to be sociodemographic variables. Fortunately, large-scale data collection organizations are now relying more on input from social scientists for measures when they collect their data. Having to locate and use proxy measures of key concepts, it is hoped, will become a thing of the past.

—Charles W. Mueller

REFERENCES

Bollen, K. (1989). *Structural equations with latent variables.* New York: John Wiley.

DeVellis, R. F. (1991). *Scale development: Theory and applications.* Newbury Park, CA: Sage.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

Maxwell, J. A. (1996). *Qualitative research design.* Thousand Oaks, CA: Sage.

CONDITIONAL LIKELIHOOD RATIO TEST

One procedure for testing a restriction on parameters estimated with CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION is the conditional likelihood ratio test. The intuition behind the conditional likelihood ratio statistic is analogous to that of the LIKELIHOOD RATIO STATISTIC. The conditional likelihood ratio statistic measures the reduction in the conditional log-likelihood function from imposing the restriction. If the restriction is valid, the reduction in the conditional log-likelihood function should be insignificant. The large sample distribution of the conditional likelihood ratio statistic is chi-squared with DEGREES OF FREEDOM equal to the number of restrictions imposed. As with the likelihood ratio statistic, the conditional likelihood ratio test has counterparts in conditional versions of the Wald and score (Lagrange multiplier) tests.

More formally, suppose that a parameter vector, $\boldsymbol{\beta}$, is estimated by maximizing the following conditional log-likelihood function, $\ln L = \sum f(y_2|\mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}, \gamma)$, where γ is a parameter vector whose values are set by theoretical assumption or, more commonly, replaced with the estimate, $\hat{\gamma}$, which is obtained by maximizing the marginal log-likelihood function, $\ln L = \sum g(y_1|\mathbf{Z}, \gamma)$. Now consider a $(r \times 1)$ vector of restrictions on $\boldsymbol{\beta}$, defined as $\mathbf{c}(\boldsymbol{\beta}) - \mathbf{q} = \mathbf{0}$. The conditional likelihood ratio test is based on the difference between $\ln \hat{L}$, the conditional log-likelihood function at the unrestricted estimate of $\boldsymbol{\beta}$, and $\ln \hat{L}_R$, the conditional log-likelihood function at the restricted estimate of $\boldsymbol{\beta}$. Note that both are conditioned on γ, which is treated as known. Then, the conditional likelihood ratio statistic is $-2(\ln \hat{L}_R - \ln \hat{L}) \sim \chi^2_r$, which tests $\mathbf{c}(\boldsymbol{\beta}) - \mathbf{q} = \mathbf{0}$ as the null hypothesis.

One application of the conditional likelihood ratio test is the test of exogeneity in the context of simultaneous PROBIT models. Consider the specific example of a probit model in which one of the regressors, $y_1 = \mathbf{Z}\gamma + v$, is a continuous ENDOGENOUS variable with a NORMAL DISTRIBUTION. Conditional maximum likelihood estimation of this system of equations would involve obtaining a maximum likelihood estimate of γ by ORDINARY LEAST SQUARES and then using this estimate to condition the maximum likelihood estimation of the probit equation. The conditioning in this case is using $\hat{\gamma}$ to derive reduced-form residuals, $\hat{v}$, that would be included on the right-hand side of the probit equation to control for simultaneity bias. The conditional likelihood ratio test would compare the log-likelihood for this unrestricted model to the log-likelihood for the restricted model that imposes the restriction of exogeneity by excluding the reduced-form residuals. In this specific example, the conditional likelihood ratio statistic would have a chi-squared distribution with one degree of freedom. Note that under the null hypothesis of exogeneity, this conditional likelihood ratio test is asymptotically equivalent to the likelihood ratio test.

–Harvey D. Palmer

REFERENCES

Alvarez, R. M., & Glasgow, G. (2000). Two-stage estimation of nonrecursive choice models. *Political Analysis, 8,* 147–165.

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics, 39,* 347–366.

CONDITIONAL LOGIT MODEL

Conditional logit is a statistical model used to study problems with unordered CATEGORICAL (NOMINAL) dependent variables that have three or more categories. Conditional logit estimates the probability that each observation will be in each category of the dependent variable. Conditional logit uses characteristics of the choice categories as independent variables and thus does not allow the effects of independent variables to vary across choice categories. This differs from the similar MULTINOMIAL LOGIT model, which uses characteristics of the individual observations as independent variables and allows the effects of the independent variables to vary across the categories of the dependent variable. Models that include both characteristics of the alternatives and characteristics of the individual observations are possible and are also usually referred to as conditional logits. Note that some researchers refer to both the multinomial logit and the conditional logit as a "multinomial logit."

There are several different ways to derive the conditional logit model. The most common approach in the social sciences is to derive the conditional logit as a DISCRETE choice model. Each observation is an actor selecting one alternative from a choice set of J alternatives, where J is the number of categories in the dependent variable. The utility an actor i would get from selecting alternative k from this choice set is given by

$$U_{ik} = X_{ik}\beta + \varepsilon_{ik}.$$

Let y be the dependent variable with J unordered categories. Thus, the probability that an actor i would select alternative j over alternative k is

$$\begin{aligned}\Pr(y_i = k) &= \Pr(U_{ik} > U_{ij}) \\ &= \Pr(X_{ik}\beta + \varepsilon_{ik} > X_{ij}\beta + \varepsilon_{ij}) \\ &= \Pr(\varepsilon_{ik} - \varepsilon_{ij} > X_{ij}\beta - X_{ik}\beta).\end{aligned}$$

McFadden (1973) demonstrated that if the εs are independent and there are identical Gumbel distributions across all observations and alternatives, this discrete choice derivation leads to a conditional logit. Multinomial logit estimates $\Pr(y_i = k|x_{ij})$, where k is one of the J categories in the dependent variable. The conditional logit model is written as

$$\Pr(y_i = k|x_{ik}) = \frac{e^{x_{ik}\beta}}{\sum_{j=1}^{J} e^{x_{ij}\beta}}.$$

Conditional logits are estimated using MAXIMUM LIKELIHOOD techniques. Unlike the multinomial logit, no identification constraints are necessary on those coefficients on variables that vary across alternatives as well as individuals. However, any coefficients on variables that do not vary across alternatives must be constrained in some way to identify the model. The most common constraint is to normalize the βs on the variables that do not vary across individuals to equal zero for one alternative.

INDEPENDENCE OF IRRELEVANT ALTERNATIVES

Conditional logits and multinomial logits have a property known as independence of irrelevant alternatives (IIA). For each observation, the ratio of the probabilities of being in any two categories is not affected by the addition or removal of other categories from the choice set. This is because the ratio of the choice probabilities for any two alternatives only depends on the variables and coefficients pertaining to those two alternatives. To see that this holds for conditional logit, observe the following:

$$\frac{\Pr(y_i = m|x_{im})}{\Pr(y_i = k|x_{ik})} = \frac{e^{x_{im}\beta} / \sum_{j=1}^{J} e^{x_{ij}\beta}}{e^{x_{ik}\beta} / \sum_{j=1}^{J} e^{x_{ij}\beta}} = \frac{e^{x_{im}\beta}}{e^{x_{ik}\beta}}.$$

Note that IIA is a property of the choice probabilities of the individual observations and is not a property that applies to the aggregate probabilities in the model. This property is often undesirable in many settings. The most commonly cited example of this is the *red bus/blue bus paradox.* Suppose that an individual has a choice of either commuting to work via car or via a bus that happens to be painted red. Let the probability this individual selects each alternative be 1/2. Now a second bus that happens to be blue is added to the choice set, identical to the red bus in every way except for color. Our intuition tells us that the choice probabilities should be 1/2 for the car and 1/4 for each bus. However, IIA necessitates that the choice probabilities for each mode of transportation be 1/3, to preserve the original ratio of choice probabilities between the car and the red bus. Note that the IIA property will hold for any discrete choice model that assumes the error term is distributed independently and identically (IID). The desire to relax the IIA property is one factor that has led to the increasing use of models such as the multinomial probit (MNP) in the study of multinomial choice problems in the social sciences.

AN EXAMPLE

As an example, consider the problem of estimating the effects of age, gender, income, and ideological distance on candidate vote choice in an election that has three candidates (A, B, and C). Age, gender, and income are variables that do not vary across alternatives and so are treated as they would be in a MULTINOMIAL LOGIT. That is, the coefficients on these variables will vary across candidates and must be constrained in some way to identify the model. In this case, the coefficients for these variables are constrained to zero for Candidate C. Ideological distance between an individual and a candidate is measured as the difference between an individual's self-placement on a 7-point ideological scale and his or her placement of the candidate on the scale. As this variable varies across both individuals and alternatives, there will be a single coefficient for the impact of ideological distance on candidate choice, and this coefficient does not need to be constrained in any way.

The results of estimating a conditional logit for this choice problem are presented in Table 1.

Table 1 Multinomial Logit Estimates, Multicandidate Election (coefficients for Candidate C normalized to zero)

Independent Variables	*Coefficient Value*	*Standard Error*
Ideological distance	−0.12*	0.04
	A/C Coefficients	
Income	−0.51*	0.20
Age	−0.01	0.01
Female	0.08	0.15
Constant	0.52	0.39
	B/C Coefficients	
Income	0.24*	0.09
Age	−0.06*	0.01
Female	0.27	0.23
Constant	0.43	0.67
Number of observations	1,000	

*Statistically significant at the 95% level.

Notice the coefficients on age, gender, and income vary across alternatives, and the coefficients for one alternative have been normalized to zero. This conditional logit tells us that as income increases, individuals are less likely to vote for Candidate A in comparison to Candidate C, but they are more likely to vote for Candidate B in comparison to Candidate C. As age increases, individuals are less likely to vote for Candidate B in comparison to Candidate C. The coefficient on ideological distance does not vary across candidates and tells us that as the ideological distance between individuals and a candidate increases, individuals are less likely to vote for that candidate.

—Garrett Glasgow

REFERENCES

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of econometrics* (pp. 105–142). New York: Academic Press.

Train, K. (1986). *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand.* Cambridge: MIT Press.

CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION

An alternative to full-information maximum likelihood (FIML) estimation is conditional maximum likelihood estimation (CMLE), which simplifies the maximization problem by treating some of the parameters as known. CMLE involves the maximization of a conditional log-likelihood function, whereby the parameters treated as known are either fixed by theoretical assumption or, more commonly, replaced by estimates. Applications of CMLE range from the iterative Cochrane-Orcutt procedure for AUTOCORRELATION to two-step estimation procedures for RECURSIVE SIMULTANEOUS EQUATIONS. CMLE produces consistent estimates that are generally less efficient than limited-information maximum likelihood (LIML) estimates, even though the two methods are ASYMPTOTICALLY equivalent under some conditions. CMLE is used instead of FIML estimation when the full log-likelihood function is difficult to derive or maximize. Consequently, the application of CMLE has declined in frequency with advancements in optimization and the speed of computer computation. CMLE is also important, however, as an intermediate step in iterative methods that produce maximum likelihood estimates, such as the expectation-maximization (EM) algorithm employed in BAYESIAN analysis.

CMLE is still often applied as a means of estimating the parameters of a recursive system of equations, such as simultaneous equations involving limited dependent variables as functions of continuous ENDOGENOUS regressors (e.g., simultaneous PROBIT and TOBIT models). Rather than maximizing the joint log-likelihood function for the system, the conditional log-likelihood function for the equation at the top of the system is maximized by replacing the parameters in equations lower in the system with maximum likelihood estimates. In this two-step procedure, the first step involves maximizing the marginal log-likelihood functions of endogenous regressors and then using the parameter estimates from this first step to condition the MAXIMUM LIKELIHOOD ESTIMATION of the parameters in the equation at the top of the system.

More formally, consider the joint distribution $h(y_1, y_2|\mathbf{X}, \beta, \mathbf{Z}, \gamma, \lambda)$ of two RANDOM VARIABLES, y_1 and y_2. Suppose that this joint distribution can be factored into a marginal distribution for y_1, $g(y_1|\mathbf{Z}, \gamma)$, and a conditional distribution for y_2, $f(y_2|\mathbf{X}, \beta, E[y_1|\mathbf{Z}, \gamma], \lambda)$, where β is the parameter vector for the exogenous regressors in $\mathbf{X}$, and λ is the parameter that defines the relationship between y_2 and $E[y_1|\mathbf{Z}, \gamma]$. CMLE of β and λ is a two-step process. The first step maximizes the marginal log-likelihood function, $\ln L = \sum g(y_1|\mathbf{Z}, \gamma)$, to obtain the estimate, $\hat{\gamma}$. The second step maximizes the conditional log-likelihood function, $\ln L = \sum f(y_2|\mathbf{X}, \beta, E[y_1|\mathbf{Z}, \gamma], \lambda)$, replacing γ with $\hat{\gamma}$.

In the specific case of the simultaneous probit model, $g(\cdot)$ would be a normal density and $f(\cdot)$ would be a probit density. The first step would apply ORDINARY LEAST SQUARES to obtain the maximum likelihood estimate of γ. The second step would use this estimate to condition the maximum likelihood estimation of the probit equation. This conditioning would use $\hat{\gamma}$ to derive reduced-form residuals that would be included on the right-hand side of the probit equation to control for simultaneity bias. One attractive characteristic of this two-stage conditional maximum likelihood (2SCML) estimator is that the estimation results can be used to test for exogeneity (i.e., $\lambda = 0$). One version of this exogeneity test is a CONDITIONAL LIKELIHOOD RATIO TEST. MONTE CARLO evidence indicates that this 2SCML estimator performs favorably against two alternatives: INSTRUMENTAL VARIABLES probit and two-stage probit least squares (Alvarez & Glasgow, 2000; Rivers & Vuong, 1988). Also, under exogeneity, this 2SCML estimator is asymptotically equivalent to the LIML estimator.

—Harvey D. Palmer

REFERENCES

Alvarez, R. M., & Glasgow, G. (2000). Two-stage estimation of nonrecursive choice models. *Political Analysis, 8,* 147–165.

Blundell, R. W., & Smith, R. J. (1989). Estimation in a class of simultaneous equation limited dependent variable models. *Review of Economic Studies, 56,* 37–58.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics, 39,* 347–366.

CONFIDENCE INTERVAL

A $100(1 - \alpha)\%$ confidence interval is an interval estimate around a POPULATION parameter θ that, under repeated RANDOM SAMPLES of size N, would be expected to include θ's true value $100(1 - \alpha)\%$ of the time. The confidence interval is a natural adjunct to PARAMETER ESTIMATION because it indicates the precision with which a POPULATION PARAMETER is estimated by a sample statistic.

The confidence level, $100(1 - \alpha)\%$, is chosen a priori. A TWO-SIDED confidence interval uses a lower limit L and upper limit U that each contain θ's true value $100(1 - \alpha/2)\%$ of the time, so that together they contain θ's true value $100(1 - \alpha)\%$ of the time. This interval often is written as $[L, U]$, and sometimes writers combine a confidence level and interval by writing $\Pr(L < \theta < U) = 1 - \alpha$. In some applications, a ONE-SIDED confidence interval is used. Confidence intervals may be computed for a large variety of population parameters such as the MEAN, VARIANCE, proportion, and R-SQUARED.

The confidence interval is closely related to the SIGNIFICANCE TEST because a $100(1 - \alpha)\%$ confidence interval includes all hypothetical values of the population parameter that cannot be rejected by a significance test using a significance criterion of α. In this respect, it provides more information than a significance test does. Confidence intervals become narrower with larger sample size and/or lower confidence levels.

The limits L and U are derived from a sample statistic (often the sample estimate of θ) and a SAMPLING DISTRIBUTION that specifies the probability of getting each value that the sample statistic can take. Thus, L and U also are sample statistics and will vary from one sample to another. Suppose, for example, that a standard IQ test has been administered to a random

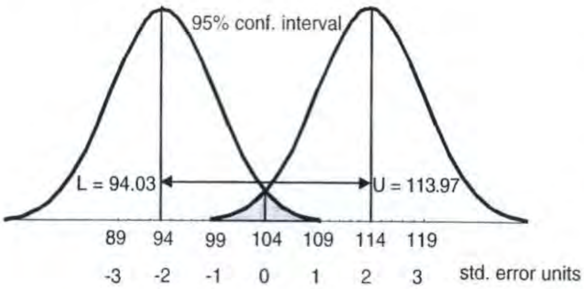

Figure 1 Lower and Upper Limits for the 95% Confidence Interval

sample of $N = 15$ adults from a large population with a sample mean of 104 and standard deviation (SD) of 18. We will construct a two-sided 95% confidence interval for the mean, μ. The limits U and L must have the property that, given a significance criterion of α, sample size of 15, mean of 104, and standard deviation of 18, we could reject the hypotheses that $\mu > 104 + U$ or $\mu < 104 - L$ but not $L < \mu < U$. Figure 1 displays those limits.

The sampling distribution of the T RATIO, where s_{err} stands for the standard error of the mean, defined by

$$t = \frac{\bar{X} - \mu}{s_{err}},$$

is a t-distribution with $df = N - 1 = 14$. When $df = 14$, the value $t_{\alpha/2} = 2.145$ standard error units above the mean cuts $\alpha/2 = .025$ from the upper tail of this t-distribution; likewise, $-t_{\alpha/2} = -2.145$ standard error units below the mean cuts $\alpha/2 = .025$ from the lower tail. The sample standard error is $s_{err} = \mathrm{SD}/\sqrt{N} = 4.648$. So a t-distribution around $U = 104 + (2.145)(4.648) = 113.97$ has .025 of its tail below 104, while a t-distribution around $L = 104 - (2.145)(4.648) = 94.03$ has .025 of its tail above 104. These limits fulfill the above required property, so the 95% confidence interval for μ is [94.03, 113.97].

—Michael Smithson

REFERENCES

Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (2000). *Statistics with confidence: Confidence intervals and statistical guidelines* (2nd ed.). London: British Medical Journal Books.

Smithson, M. (2003). *Confidence intervals* (Sage University Papers on Quantitative Applications in the Social Sciences, 140). Thousand Oaks, CA: Sage.

CONFIDENTIALITY

This ethical principle requires the researcher not to divulge any findings relating to research participants, other than in an anonymized form.

—Alan Bryman

See also ETHICAL PRINCIPLES

CONFIRMATORY FACTOR ANALYSIS

Confirmatory factor analysis is a statistical procedure for testing HYPOTHESES about the commonality among VARIABLES. As a MULTIVARIATE procedure, confirmatory factor analysis is used to simultaneously test multiple hypotheses that collectively constitute a measurement model. All such models comprise four components. At least two variables, or *indicators*, are necessary for the most rudimentary measurement model. Variability in scores on the indicators is allocated to two sources—LATENT VARIABLES and measurement ERROR. *Latent variables* are not observed; they are inferred from the commonality among indicators. *Measurement error* is variability in the indicators not attributable to the latent variables. Finally, the degree to which variability in scores on the indicators can be attributed to the latent variables is indexed by *loadings*, coefficients that are estimated in measurement equations in which the indicators are regressed on the latent variable(s) hypothesized to influence them.

SOURCES OF VARIABILITY

Confirmatory factor analysis partitions variability in scores on indicators according to sources specified in a measurement model. The basis for this partitioning is illustrated in Figure 1. On the left is a Venn diagram that shows six possible sources of variability in three indicators—V1, V2, and V3—that could be specified in a measurement model. In this informal depiction, the indicators are represented by circles, each comprising three parts. The shaded area, labeled F1, is variability in each indicator shared with the other two indicators. The crosshatched area is variability shared with only one of the remaining indicators, and the unshaded area is variability shared with neither of the remaining indicators.

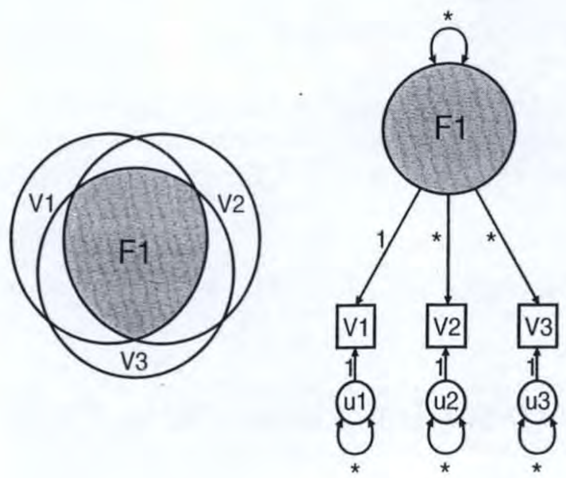

Figure 1 Venn and Path Diagram Depictions of the Partitioning of Variability in the Score on Three Variables Into Shared and Unique Components

On the right in Figure 1 is a more formal depiction of this partitioning of variability. In this PATH DIAGRAM, the indicators are represented by squares. The large shaded circle, labeled F1, corresponds to the shaded area in the Venn diagram. It is the latent variable the three indicators are purported to measure and reflects the variability they share. The directional *paths* between the latent variable and the squares represent the assumption that the latent variable is a cause of variability in the indicators. The small circles represent variability unique to each indicator and correspond to the unshaded areas in the Venn diagram. They represent measurement error because they are unintended causes of variability on the indicators. The crosshatched areas from the Venn diagram are not represented in the path diagram. This is because the typical specification of a measurement model assumes unrelated errors of measurement. If the strength of these relations is statistically nontrivial, a satisfactory measurement model will need to include covariances between the errors of measurement.

Two additional features of the path diagram warrant mention. First, notice that attached to each circle is a sharply curved two-headed arrow. This path indicates a VARIANCE and illustrates the fact that variability in each indicator has been allocated either to the latent variable or to measurement error. Second, notice that every path is accompanied by either a 1 or an asterisk. These represent PARAMETERS, numeric values that reflect the strength of the paths. Parameters indicated by a value such as 1 reflect a fixed assumption about the association between two variables or a variance. Parameters indicated by an asterisk reflect a parameter whose value is not assumed but will be estimated from the observed data. A subset of these parameters of particular interest in confirmatory factor analysis is associated with paths between the latent variable and the indicators—the loadings. The larger these values, the greater the portion of variability in the indicators that can be attributed to the latent variable(s) as opposed to measurement error.

MODEL SPECIFICATION

Confirmatory factor analysis begins with the formal SPECIFICATION of a measurement model. That is, the researcher ventures a set of hypotheses in the form of sources and paths that account for the ASSOCIATIONS among the indicators. These hypotheses are evaluated against observed data on the indicators. To illustrate, imagine that a social psychologist wishes to evaluate the performance of a new measure of students' tendency to self-handicap: engaging in counterproductive behavior prior to important evaluative activities such as tests and presentations. She proposes two independent forms of this tendency and writes three items to tap each one. She then administers the measure to a sample of 506 college students.

The measurement model that reflects the researcher's hypotheses about the associations among the items is illustrated in Figure 2. The two latent variables, F1 and F2, represent the two hypothesized forms of self-handicapping. Note that each indicator, V1 to V6, is influenced by only one of the latent variables. Also note the curved arrow between F1 and F2. This path represents a CORRELATION. Because the researcher believes the two forms of self-handicapping are independent, it will be important for her to show that the value of this parameter, as estimated from her data is, in effect, zero.

It is customary in applications of confirmatory factor analysis to compare a proposed measurement model against plausible alternatives that are either more parsimonious than the proposed model or are consistent with rival theoretical accounts of the associations among the indicators. Our hypothetical researcher might compare her model against a more parsimonious model that includes only one latent variable or a plausible but less parsimonious one that includes two latent variables that

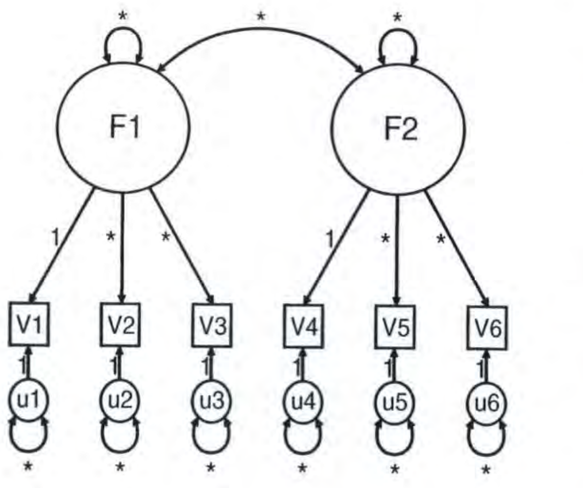

Figure 2 Path Diagram Depicting Two-Factor Model of Six-Item Measure of Self-Handicapping

are not independent. If, according to criteria presented below, the researcher's model provides a good account of the data in an absolute sense, and if it provides a better account than these two alternatives, then she has compelling evidence for the factorial VALIDITY of her measure.

Although confirmatory factor analysis allows considerable flexibility in the specification of measurement models, the mathematical requirements of estimation limit specification. For instance, a model in which each indicator in Figure 2 is influenced by both latent variables, the equivalent of an EXPLORATORY FACTOR ANALYSIS model, could not be estimated. A model in which none of the loadings was fixed at 1 could not be estimated without changing other aspects of the model. The principle that underlies these restrictions on specification is termed identification. Identification is a challenging topic that, fortunately, is rarely a concern in typical model specifications such as the one shown in Figure 2. Adherence to three basic rules will almost always result in a model that can be estimated. For each latent variable, there should be at least three indicators. The variance of, or one of the loadings on, each latent variable must be fixed at 1. In addition, there cannot be more unknown parameters in the model than the sum of the number of variances and covariances. In our example there are 6 variances, one for each indicator, and 15 covariances between the indicators, for a total of 21. (This number can be determined using the equation $p(p+1)/2$, where p is the number of indicators.) Counting asterisks in the figure reveals a total of 13 unknown parameters. Our researcher's model meets the identification criteria and can be estimated.

PARAMETER ESTIMATION

Once a measurement model has been specified, it can be estimated. That is, values for unknown parameters are derived using data on the indicators. A number of ESTIMATORS could be used for this purpose, but MAXIMUM LIKELIHOOD ESTIMATION is virtually always the estimator of choice. The goal of maximum likelihood estimation is to determine the values of the unknown parameters that are most likely given the data. This determination is reached through an iterative process of substituting values for the unknown parameters, evaluating the likelihood of the parameters given the data, then attempting to improve on the likelihood by adjusting the values. Each evaluation is an *iteration,* and when it is not possible to improve significantly on the likelihood, the estimation process has *converged.* The parameter values at convergence are those on which hypothesis tests are based.

The iterative process cannot begin without a tentative value assigned to each unknown parameter. The simplest solution to this problem is to begin estimation with a value of 1 assigned to every unknown parameter. At the first iteration, these *starting values* will be adjusted, some upward and some downward, typically resulting in a substantial increase in the likelihood of the parameters given the data. After the first 2 or 3 iterations, improvement in the likelihood usually is slight, and convergence typically is reached in 10 or fewer iterations.

The parameter estimates for our example are shown in the top portion of Table 1. The estimates on which hypothesis tests are based are the UNSTANDARDIZED estimates. Notice that, among the loadings, V1, F1 and V4, F2 are 1, as prescribed in our model specification. Note that there are 15 entries under the "Parameter" heading. If we exclude the 2 loadings that were fixed, we are left with 13, each corresponding to one of the asterisks in Figure 2.

EVALUATION OF FIT

Having obtained the optimal set of parameter estimates for her model given the data she gathered, our researcher must now address two questions. First, does the model provide a satisfactory account of the data? Second, if the model provides a satisfactory account

Table 1 Parameter Estimates and Omnibus Fit Information

	Parameter Estimates			
Parameter	*Unstandardized Estimate*	*Standard Error*	*Critical Ratio*	*Standardized Estimate*
Loadings				
V1, F1	1.000[a]			.516
V2, F1	1.236	.257	4.804	.437
V3, F1	1.690	.380	4.445	.565
V4, F2	1.000[a]			.749
V5, F2	.578	.106	5.472	.487
V5, F2	.694	.127	5.472	.487
Variances				
F1	.136	.038	3.572	
F2	.870	.175	4.690	
U1	.377	.039	9.594	
U2	.881	.073	12.034	
U3	.831	.104	8.027	
U4	.683	.158	4.326	
U5	.938	.078	12.042	
U6	1.347	.112	12.024	
Covariances				
F1, F2	−.004	.026	−.161	−.012

	Omnibus Fit Indices					
Model	χ^2	*df*	*p*	*CFI*	*RMSEA*	$CL_{(.90)}$
Two-factor model	13.820	8	.087	.976	.038	.000, .071
Drop F1, F2	13.845	9	.128	.980	.033	.000, .065
Drop F1, F2, add U1, U6	8.018	8	.432	1.000	.003	.000, .052
One-factor model	100.149	9	< .001	.622	.142	.117, .167

NOTE: CFI = comparative fit index; RMSEA = root mean square error of approximation. a. Fixed parameter.

of the data, do the parameter estimates conform to the predicted pattern? The first question concerns omnibus fit and the second component fit.

The *omnibus fit* of a measurement model is the degree to which the parameter estimates given the model imply a set of data that approximates the observed data. The word *data* in this case refers not to the individual responses of the 506 participants to the six self-report items. Rather, the data in question are elements of the observed covariance matrix. The observed covariance matrix for our example is shown in the top panel of Table 2. The matrix comprises 6 variances, one for each indicator, and 15 covariances, which index the pairwise associations between the indicators.

There are numerous inferential and descriptive statistics for indexing fit, each with its own rationale and interpretation. At the most basic level, however, the evaluation of omnibus fit is an exercise in comparison. The most straightforward comparison is between the observed covariance matrix and the covariance matrix implied by the specified model given the parameters estimated from the data. To obtain the *implied covariance matrix,* one could substitute the estimated and fixed values of the parameters into the measurement equations and derive a covariance matrix. The implied covariance matrix for our example model is shown in the second panel of Table 2. The specified model fits this theoretical matrix exactly. To the extent that this matrix matches the observed covariance matrix, the specified model provides a satisfactory account of the observed data.

This comparison between the observed and implied covariance matrices is accomplished by subtracting each element of the implied covariance matrix from the corresponding element in the observed matrix, yielding the *residual matrix.* The residual matrix for our example is shown in the third panel of

Table 2 Observed, Implied, and Residual Covariance Matrices

Observed Covariance Matrix

V1	0.513					
V2	0.169	1.090				
V3	0.231	0.284	1.221			
V4	0.031	−0.100	−0.066	1.553		
V5	0.035	−0.031	0.052	0.503	1.229	
V6	0.106	−0.017	−0.020	0.604	0.351	1.767

Implied Covariance Matrix

V1	0.513					
V2	0.169	1.090				
V3	0.231	0.285	1.221			
V4	−0.004	−0.005	−0.007	1.553		
V5	−0.002	−0.003	−0.004	0.503	1.229	
V6	−0.003	−0.004	−0.005	0.604	0.350	1.767

Residual Matrix

V1	0.000					
V2	0.000	0.000				
V3	0.000	−0.001	0.000			
V4	0.035	−0.095	−0.059	0.000		
V5	0.037	−0.028	0.056	0.000	0.000	
V6	0.109	−0.013	−0.015	0.000	0.001	0.000

Standardized Residual Matrix

V1	0.000					
V2	0.000	0.000				
V3	0.000	−0.001	0.000			
V4	0.039	−0.073	−0.043	0.000		
V5	0.047	−0.024	0.046	0.000	0.000	
V6	0.114	−0.010	−0.010	0.000	0.001	0.000

Table 2. This matrix includes a number of zeros, which indicate no additional variance or covariance to be explained. The zeros on the diagonal are to be expected because all of the variance in the indicators is accounted for in the model either by the latent variables or measurement error. The zeros off the diagonal indicate that the researcher's model fully accounts for some associations between the indicators. Among the residuals that are not zero, some are positive and some are negative, indicating that the model underestimates some associations and overestimates others. Although the residual matrix is informative, particularly if respecification of the model is necessary, it does not provide the basis for formally judging the degree to which the specified model provides a satisfactory account of the data.

A number of omnibus fit indices are available for this purpose. The most basic is the chi-square statistic, the basis for an inferential test of GOODNESS OF FIT. In our example, the value of chi-square evaluated on 8 degrees of freedom—the number of variances and covariances, 21, minus the number of free parameters, 13—is 13.82 with a p value of .09, indicating that the observed and implied covariance matrices are not significantly different. Two widely used indices are the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). Values of the former range from 0 to 1, and values greater than .90 indicate satisfactory fit; the value of CFI in our example is .976. RMSEA originates at 0, and smaller values indicate better fit. Typically, a 90% confidence interval is put on the point estimate of RMSEA, and satisfactory fit is indicated by an upper limit less than .08. In our example, the upper limit is .071. This information in a format typical of reports of omnibus fit is shown in the first line of the lower portion of Table 1.

A second level of evaluation of fit concerns *component fit,* the degree to which the magnitude and sign of the parameter estimates correspond to predictions. In our example, we expect the factor loadings to be positive and significantly greater than zero. We expect the variances of the latent variables to be greater than zero as well. We venture no hypotheses regarding the measurement errors; however, smaller values indicate less variance in indicators attributable to measurement error. Finally, the researcher posited independent forms of self-handicapping; therefore, we predict that the covariance between F1 and F2 would not differ from zero.

Returning to the top portion of Table 1, we see that each of the 13 parameters estimated from the data is accompanied by a standard error. The ratio of the unstandardized parameter estimate to its standard error is termed the *critical ratio* and is distributed as a z statistic. As such, values greater than 1.96 indicate that the estimate is significantly different from zero. Only 1 of the 13 estimates is not significant, and it is the parameter our hypothetical researcher predicted to be zero: the F1, F2 covariance.

It is customary to interpret standardized factor loadings, which, because they are estimated in measurement equations, are STANDARDIZED REGRESSION COEFFICIENTS. A rule of thumb is that values greater than .30 represent a significant contribution of the latent variable to the explanation of variability in the indicator. All six standardized loadings meet this criterion.

Our researcher's model provides a satisfactory account of her observed data at both the omnibus and component levels. To further strengthen this inference, she might compare this model against the two alternatives we described earlier. We established that the correlation between the latent variables is not different from zero; therefore, we know that a model in which this parameter is set to zero would fit the data as well as the initial model and, for reasons of parsimony, would be favored over it. The second alternative is a one-factor model. Omnibus fit indices for this model are presented in the last line of Table 1. Note that the value of chi-square is highly significant and substantially larger than the value for the two-factor model. The value of CFI is substantially less than .90, and both the upper and lower limits of the confidence interval exceed .08. Thus, there is strong support for our researcher's model in both an absolute and a comparative sense.

RESPECIFICATION

In practice, the initially specified measurement model does not yield satisfactory omnibus fit. In such instances, the researcher might attempt to *respecify* the model so as to find a defensible account of the data. At this point, the analysis shifts from hypothesis testing to hypothesis generating. That is, statistical tests are no longer, in a strict sense, valid. Nonetheless, respecification can be highly informative when evaluations of an initially specified model do not yield support for it.

Although the initially specified model in our example fit well, we might ask, for purposes of illustration, whether the fit can be improved. We already recognized that the model could be simplified by dropping the F1, F2 path, and the omnibus fit indices for such a model, shown in Table 1, support this assumption. Beyond this rather obvious adjustment to the model, it is not immediately evident how omnibus fit might be improved. In such cases, the researcher would resort to one of two *specification search* strategies. One uses one of the empirical strategies available in virtually all computer programs for confirmatory factor analysis. The other is a manual search strategy that focuses on the residual matrix. Because the elements of the residual matrix are covariances, which are not readily interpretable, we standardize them to produce the standardized residual matrix shown in the bottom panel of Table 2. The values in this matrix are CORRELATION COEFFICIENTS and represent unexplained covariation between indicators. The largest of these is the correlation between V1 and V6. The simplest means of accounting for this association would be to allow the corresponding measurement errors, U1 and U6, to covary. When we respecify the model to include this parameter, the standardized residual drops to .024, and the omnibus fit indices, shown at the bottom of Table 1, approach their maxima. Were there no support for the model as initially specified, modifications such as this might be sufficient to achieve statistical support. It is important to keep in mind, however, that because such increments in fit are achieved using knowledge about the observed data, they might be sample specific and, therefore, must be confirmed through evaluation of the respecified model when specified a priori and estimated from new data.

ORIGIN AND CURRENT DIRECTIONS

Confirmatory factor analysis traces its roots to the work of Swedish psychometrician Karl Jöreskog. During the early 1970s, Jöreskog and associates developed algorithms and, importantly, computer software for analyzing the structure of covariance matrices. Their LISREL (*li*near *s*tructural *rel*ations) model allowed for the formal evaluation of COVARIANCE STRUCTURE MODELS, of which measurement models are a specific instance.

Most advances in confirmatory factor analysis are a by-product of advances in the general analysis of covariance structures. These include additional computer programs, some permitting model specification by sketching a path diagram; a host of omnibus fit indices that address documented shortcomings of the chi-square test; and estimation procedures that do not impose the demands on data that attend maximum likelihood estimation. Advances specific to confirmatory factor analysis typically manifest as solutions to specific research problems. Examples include tests of INTERACTION EFFECTS involving latent variables, latent GROWTH CURVE MODELS, and state-trait models.

—Rick H. Hoyle

REFERENCES

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsely & S. D. Brown (Eds.), *Handbook of* *applied multivariate statistics and mathematical modeling* (pp. 465–497). New York: Academic Press.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

CONFOUNDING

The word *confounding* has been used to refer to at least three distinct concepts. In the oldest and most widespread usage, confounding is a source of bias in estimating causal effects. This bias is sometimes informally described as a mixing of effects of extraneous factors (called confounders) with the effect of interest. This usage predominates in nonexperimental research, especially in epidemiology and sociology. In a second and more recent usage originating in statistics, *confounding* is a synonym for a change in an effect measure on stratification or adjustment for extraneous factors (a phenomenon called *noncollapsibility* or *Simpson's paradox*). In a third usage, originating in the EXPERIMENTAL DESIGN literature, confounding refers to inseparability of main effects and interactions under a particular design. The three concepts are closely related and are not always distinguished from one another. In particular, the concepts of confounding as a bias in effect estimation and as noncollapsibility are often treated as equivalent, even though they are not. Only the former concept will be described here; for more detailed coverage and comparisons of concepts, see the following: Rothman and Greenland (1998, chaps. 4, 20); Greenland, Robins, and Pearl (1999); Pearl (2000); and Greenland and Brumback (2002).

CONFOUNDING AS A BIAS IN EFFECT ESTIMATION

A classic discussion of confounding in which explicit reference is made to "confounded effects" is Mill (1843/1956, chap. 10), although in Chapter 3, Mill lays out the primary issues and acknowledges Francis Bacon as a forerunner in dealing with them. There, he lists a requirement for an EXPERIMENT intended to determine causal relations: "None of the circumstances [of the experiment] that we do know shall have effects susceptible of being *confounded* with those of the agents whose properties we wish to study" (emphasis added).

In Mill's time, the word *experiment* referred to an observation in which some circumstances were under the control of the observer, as it still is used in ordinary English, rather than to the notion of a comparative trial. Nonetheless, Mill's requirement suggests that a comparison is to be made between the outcome of our "experiment" (which is, essentially, an uncontrolled trial) and what we would expect the outcome to be if the agents we wish to study had been absent. If the outcome is not as one would expect in the absence of the study agents, then Mill's requirement ensures that the unexpected outcome was not brought about by extraneous "circumstances" (factors). If, however, those circumstances do bring about the unexpected outcome and that outcome is mistakenly attributed to effects of the study agents, then the mistake is one of confounding (or confusion) of the extraneous effects with the agent effects.

Much of the modern literature follows the same informal conceptualization given by Mill. Terminology is now more specific, with *treatment* used to refer to an agent administered by the investigator and *exposure* often used to denote an unmanipulated agent. The chief development beyond Mill is that the expectation for the outcome in the absence of the study exposure is now almost always explicitly derived from observation of a CONTROL GROUP that is untreated or unexposed. Confounding typically occurs when natural or social forces or personal preferences affect whether a person ends up in the treated or control group, and these forces or preferences also affect the outcome variable. Although such confounding is common in observational studies, it can also occur in randomized experiments when there are systematic improprieties in TREATMENT allocation, administration, and compliance. A further and somewhat controversial point is that confounding (as per Mill's original definition) can also occur in perfect randomized trials due to *random* differences between comparison groups (Fisher, 1935; Rothman, 1977).

THE POTENTIAL OUTCOME MODEL

Various models of confounding have been proposed for use in statistical analyses. Perhaps the one closest to Mill's concept is based on the *potential outcome* or COUNTERFACTUAL model for causal effects. Suppose we wish to consider how a health status (outcome) measure of a population would change in response to an intervention (population treatment). More precisely, suppose our objective is to determine the effect that applying a treatment x_1 had or would have on an outcome measure μ relative to applying treatment x_0 to a specific target Population A. For example, Cohort A could be a cohort of breast cancer patients, treatment x_1 could be a new hormone therapy, x_0 could be a placebo therapy, and the measure μ could be the 5-year survival probability. The treatment x_1 is sometimes called the *index* treatment, and x_0 is sometimes called the *control* or *reference* treatment (which is often a standard or placebo treatment).

The potential outcome model posits that, in Population A, μ will equal μ_{A1} if x_1 is applied and μ_{A0} if x_0 is applied; the causal effect of x_1 relative to x_0 is defined as the change from μ_{A0} to μ_{A1}, which might be measured as $\mu_{A1} - \mu_{A0}$ or μ_{A1}/μ_{A0}. If A is given treatment x_1, then μ will equal μ_{A1} and μ_{A1} will be observable, but μ_{A0} will be unobserved. Suppose, however, we expect μ_{A0} to equal μ_{B0}, where μ_{B0} is the value of the outcome μ observed or estimated for a Population B that was administered treatment x_0. The latter population is sometimes called the control or reference population. Confounding is said to be present if $\mu_{A0} \neq \mu_{B0}$, for then there must be some difference between Populations A and B (other than treatment) that is affecting μ.

If confounding is present, a naive (crude) ASSOCIATION measure obtained by substituting μ_{B0} for μ_{A0} in an effect measure will not equal the effect measure, and the association measure is said to be *confounded.* For example, if $\mu_{A0} \neq \mu_{B0}$, then $\mu_{A1} - \mu_{B0}$, which measures the association of treatments with outcomes *across* the populations, is confounded for $\mu_{A1} - \mu_{A0}$, which measures the effect of treatment x_1 on Population A. Thus, to say an association measure $\mu_{A1} - \mu_{B0}$ is confounded for an effect measure $\mu_{A1} - \mu_{A0}$ is to say that these two measures are not equal.

A noteworthy aspect of this view is that confounding depends on the outcome measure. For example, suppose Populations A and B have a different 5-year survival probability μ under placebo treatment x_0; that is, suppose $\mu_{B0} \neq \mu_{A0}$, so that $\mu_{A1} - \mu_{B0}$ is confounded for the actual effect $\mu_{A1} - \mu_{A0}$ of treatment on 5-year survival. It is then still possible that 10-year survival, ν, under the placebo would be identical in both populations; that is, ν_{A0} could still equal ν_{B0}, so that $\nu_{A1} - \nu_{B0}$ is not confounded for the actual effect of treatment on 10-year survival. (We should generally expect no confounding for 200-year survival because no treatment is likely to raise the 200-year survival probability of human patients above zero.)

A second noteworthy point is that confounding depends on the target population of INFERENCE. The preceding example, with A as the target, had different 5-year survivals μ_{A0} and μ_{B0} for A and B under placebo therapy, and hence $\mu_{A1} - \mu_{B0}$ was confounded for the effect $\mu_{A1} - \mu_{A0}$ of treatment on Population A. A lawyer or ethicist may also be interested in what effect the hormone treatment would have had on Population B. Writing μ_{B1} for the (unobserved) outcome under treatment, this effect on B may be measured by $\mu_{B1} - \mu_{B0}$. Substituting μ_{A1} for the unobserved μ_{B1} yields $\mu_{A1} - \mu_{B0}$. This measure of association is confounded for $\mu_{B1} - \mu_{B0}$ (the effect of treatment x_1 on 5-year survival in Population B) if and only if $\mu_{A1} \neq \mu_{B1}$. Thus, the same measure of association, $\mu_{A1} - \mu_{B0}$, may be confounded for the effect of treatment on neither, one, or both of Populations A and B and may or may not be confounded for the effect of treatment on other targets.

Confounders (Confounding Factors)

A third noteworthy aspect of the potential outcome model is that it invokes no explicit differences (imbalances) between Populations A and B with respect to circumstances or covariates that might influence μ (Greenland & Robins, 1986). Clearly, if μ_{A0} and μ_{B0} differ, then A and B must differ with respect to factors that influence μ. This observation has led some authors to define confounding as the presence of such covariate differences between the compared populations. Nonetheless, confounding is only a consequence of these covariate differences. In fact, A and B may differ profoundly with respect to covariates that influence μ, and yet confounding may be absent. In other words, a covariate difference between A and B is a necessary but not sufficient condition for confounding. For example, the confounding effects of covariate differences may balance each other out, leaving no confounding.

Suppose now that Populations A and B differ with respect to certain covariates and that these differences have led to confounding of an association measure for the effect measure of interest. The responsible covariates are then termed *confounders* of the association measure. In the above example, with $\mu_{A1} - \mu_{B0}$ confounded for the effect $\mu_{A1} - \mu_{A0}$, the factors responsible for the confounding (i.e., the factors that led to $\mu_{A0} \neq \mu_{B0}$) are the confounders. It can be deduced that a variable cannot be a confounder unless it can affect the outcome parameter μ within treatment groups and it is distributed differently among the compared populations (e.g., see Yule, 1903, who however uses terms such as *fictitious association* rather than *confounding*). These two necessary conditions are sometimes offered together as a definition of a confounder.

Nonetheless, counterexamples show that the two conditions are not sufficient for a variable with more than two levels to be a confounder (Greenland et al., 1999).

PREVENTION OF CONFOUNDING

Perhaps the most obvious way to avoid confounding is estimating $\mu_{A1} - \mu_{A0}$ to obtain a reference Population B for which μ_{B0} is known to equal μ_{A0}. Such a population is sometimes said to be *comparable* to or *exchangeable* with A with respect to the outcome under the reference treatment. In practice, such a population may be difficult or impossible to find. Thus, an investigator may attempt to construct such a population or an exchangeable index and reference populations. These constructions may be viewed as design-based methods for the control of confounding.

Perhaps no approach is more effective for preventing confounding by a known factor than *restriction*. For example, gender imbalances cannot confound a study restricted to women. However, there are several drawbacks: Restriction on enough factors can reduce the number of available subjects to unacceptably low levels and may greatly reduce the generalizability of results as well. Matching the treatment populations on confounders overcomes these drawbacks and, if successful, can be as effective as restriction. For example, gender imbalances cannot confound a study in which the compared groups have identical proportions of women.

Unfortunately, differential losses to observation may undo the initial covariate balances produced by matching. Neither restriction nor matching prevents (although it may diminish) imbalances on unrestricted, unmatched, or unmeasured covariates. In contrast, *randomization* offers a means of dealing with confounding by covariates not accounted for by the design. It must be emphasized, however, that this solution is only probabilistic and subject to severe constraints in practice.

Randomization is not always feasible or ethical, and many practical problems, such as differential loss and noncompliance, can lead to confounding in comparisons of the groups actually receiving treatments x_1 and x_0. One somewhat controversial solution to noncompliance problems is *intent-to-treat analysis*, which defines the comparison groups A and B by treatment assigned rather than treatment received. Confounding may, however, affect even intent-to-treat analyses, and (contrary to widespread misperceptions) the bias in those analyses can exaggerate the apparent treatment effect (Robins, 1998). For example, the assignments may not always be random, as when blinding is insufficient to prevent the treatment providers from violating the assignment protocol. Also, purely by bad luck, randomization may itself produce allocations with severe covariate imbalances between the groups (and consequent confounding), especially if the study size is small (Fisher, 1935; Rothman, 1977). *Blocked* (matched) randomization can help ensure that random imbalances on the blocking factors will not occur, but it does not guarantee balance of unblocked factors.

ADJUSTMENT FOR CONFOUNDING

Design-based methods are often infeasible or insufficient to prevent confounding. Thus, an enormous amount of work has been devoted to analytic adjustments for confounding. With a few exceptions, these methods are based on observed covariate distributions in the compared populations. Such methods can successfully control confounding only to the extent that enough confounders are adequately measured. Then, too, many methods employ parametric models at some stage, and their success may thus depend on the faithfulness of the model to reality. These issues cannot be covered in depth here, but a few basic points are worth noting.

The simplest and most widely trusted methods of adjustment begin with *stratification* on confounders. A covariate cannot be responsible for confounding within internally homogeneous strata of the covariate. For example, gender imbalances cannot confound observations within a stratum composed solely of women. More generally, comparisons within strata cannot be confounded by a covariate that is unassociated with treatment within strata. This is so whether or not the covariate was used to define the strata. Thus, one need not stratify on all confounders to control confounding. Furthermore, if one has accurate background information on relations among the confounders, one may use this information to identify sets of covariates sufficient for adjustment (Pearl, 2000). Nonetheless, if the stratification on the confounders is too coarse (e.g., because categories are too broadly defined), stratification may fail to adjust for much of the confounding by the adjustment variables.

One of the most common adjustment approaches today is to enter suspected confounders into a model for the outcome parameter μ. For example, let μ be the mean (expectation) of an outcome variable of interest, Y; let X be the treatment variable of interest; and let Z be a suspected confounder of the X–Y relation. Adjustment for Z is often made by fitting a GENERALIZED LINEAR MODEL $g(\mu) = \alpha + \beta x + \gamma z$ or some variant, where x and z are specific values of X and Z, and $g(\mu)$ is a strictly increasing function such as the natural log $\ln(\mu)$, as in LOG-LINEAR MODELING, or the logit function $\ln\{\mu/(1-\mu)\}$, as in LOGISTIC REGRESSION; the estimate of β that results is then taken as the Z-adjusted estimate of the effect on $g(\mu)$ of a one-unit increase in X.

An oft-cited advantage of model-based adjustment methods is that they allow adjustment for more variables and in finer detail than stratification. If the form of the fitted model cannot adapt well to the true dependence of Y on X and Z, however, such model-based adjustments may fail to adjust for confounding by Z. For example, suppose Z is symmetrically distributed around zero within X levels, and the true dependence is $g(\mu) = g(\alpha + \beta x + \gamma z^2)$; then, using the model $g(\mu) = g(\alpha + \beta x + \gamma z)$ will produce little or no adjustment for Z. Similar failures can arise in adjustments based on *propensity scores*. Such failures can be minimized or avoided by using reasonably flexible models and by carefully checking each fitted model against the data.

Some controversy has occurred about adjustment for covariates in randomized trials. Although Fisher (1935, p. 49) asserted that randomized comparisons were "unbiased," he also pointed out that they could be confounded in the sense used here. Resolution comes from noting that Fisher's use of the word *unbiased* referred to the design and was not meant to guide analysis of a given trial. Once the trial is under way and the actual treatment allocation is completed, the unadjusted treatment effect estimate will be biased if the covariate is associated with treatment, and this bias can be removed by adjustment for the covariate (Greenland & Robins, 1986; Rothman, 1977).

Finally, it is important to note that if a variable used for adjustment is not a confounder, bias may be introduced by the adjustment (Pearl, 2000); this bias tends to be especially severe if the variable is affected by both the treatment and the outcome under study (Greenland, 2003).

—Sander Greenland

REFERENCES

Fisher, R. A. (1935). *The design of experiments.* Edinburgh, UK: Oliver & Boyd.

Greenland, S. (2003). Quantifying biases in causal models. *Epidemiology, 14,* 300–306.

Greenland, S., & Brumback, B. A. (2002). An overview of relations among causal modelling methods. *International Journal of Epidemiology, 31,* 1030–1037.

Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology, 15,* 413–419.

Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science, 14,* 29–46.

Mill, J. S. (1956). *A system of logic, ratiocinative and inductive.* London: Longmans, Green & Company. (Original work published 1843)

Pearl, J. (2000). *Causality.* New York: Cambridge University Press.

Robins, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine, 17,* 269–302.

Rothman, K. J. (1977). Epidemiologic methods in clinical trials. *Cancer, 39,* 1771–1775.

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika, 2,* 121–134.

CONSEQUENTIAL VALIDITY

Consequential VALIDITY emerged from concerns related to societal ramifications of assessments. The *Standards for Educational and Psychological Testing* (American Educational Research Association, 1999) views validity as a unitary concept that measures the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose. Thus, consequential validity is that type of validity evidence that addresses the intended and unintended consequences of test interpretation and use (Messick, 1989, 1995).

Examples of harmful and unintended consequences that relate to the need to determine the consequential validity of an instrument include the following:

1. High-stakes testing may negatively affect present and future educational and employment opportunities when a single test score is used to make a decision.
2. Lower pass rates may occur on tests by individuals of certain subgroups, such as people with disabilities or racial/ethnic minorities, because of unequal opportunities or inappropriate demands of the testing situation in terms of written or spoken language.
3. A less diverse workforce may result because of employment tests that increase reluctance to seek employment or complete the required assessments.
4. Teachers may narrow the curriculum to teach only what will be on the test.

Messick (1995) suggests that researchers should distinguish between adverse consequences that stem from valid descriptions of individual and group differences from adverse consequences that derive from sources of test invalidity such as CONSTRUCT underrepresentation and construct-irrelevant variance. If differences are related to the latter, then test invalidity presents a measurement problem that needs to be investigated in the validation process. However, if the differences are the result of valid group differences, then the adverse consequences represent problems of social policy. It is the researcher's responsibility to ensure that low scores do not occur because the assessment is missing something relevant to that construct that, if it were present, would have permitted the affected persons to display their competence. In addition, causes of low scores because of irrelevant demands in the testing process can prevent an individual from demonstrating his or her competence.

The *Standards for Educational and Psychological Testing* also emphasize the need to distinguish between those aspects of consequences of test use that relate to the construct validity of a test and issues of social policy. Standard 1.24 addresses the issue of consequential validity:

> When unintended consequences result from test use, an attempt should be made to investigate whether such consequences arise from the test's sensitivity to characteristics other than those it is intended to assess or to the test's failure fully to represent the intended construct. (American Educational Research Association, 1999, p. 23)

Group differences do not in themselves call into question the validity of a proposed interpretation; however, they do increase the need to investigate rival hypotheses that may be related to problems with the validity of the measurement.

—Donna M. Mertens

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Educational Measurement. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50*(9), 741–749.

CONSISTENCY. *See* PARAMETER ESTIMATION

CONSTANT

The term *constant* simply refers to something that is not variable. In statistics, responses are typically described as RANDOM VARIABLES, roughly meaning that the responses cannot be predicted with certainty. For example, how much weight will the typical adult lose following some particular diet? How will individuals respond to a new medication? If we randomly pick some individual, will he or she approve of the death penalty?

Although at some level, the difference between a constant and a random variable is clear, the distinction between the two often becomes blurred. Consider, for example, the population mean, μ. That is, μ is the average of all individuals of interest in a particular study if they could be measured. The so-called frequentist approach to statistical problems views μ as a constant. It is some fixed but unknown value. However, an alternative view, reflected by a Bayesian approach to statistics, does not view μ as a constant but rather as a quantity that has some distribution. The distribution might reflect prior beliefs about the likelihood that μ has some particular value. As another example, p might represent the probability that an individual responds yes when asked if he or she is happily married. In some sense, this is a constant: At a particular moment in time, one could view p as fixed among all married couples. Simultaneously, p could be viewed as a random variable, either in the sense of prior beliefs held by the investigator or perhaps as varying over time.

Another general context in which the notion of constant plays a fundamental role has to do with assumptions made when analyzing data. Often, it is assumed that certain features of the data are constant to simplify technical issues. Perhaps the best-known example is HOMOSKEDASTICITY. This refers to the frequently made assumption that the VARIANCE among groups of individuals is constant. In REGRESSION, constant variance means that when trying to predict Y based on some variable X, the (conditional) variance of Y, given X, does not vary. So, for example, if X is aggression in the home and Y is a measure of cognitive functioning, constant variance means that the variance of Y among homes with an aggression score of, for example, $X = 10$ is the same as homes with $X = 15$, $X = 20$, and any other value for X we might choose.

—Rand R. Wilcox

REFERENCES

Conover, W. J. (1980). *Practical nonparametric statistics.* New York: John Wiley.

Doksum, K. A., & Sievers, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika, 63,* 421–434.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques.* San Diego: Academic Press.

CONSTANT COMPARISON

Constant comparison is the data-analytic process whereby each interpretation and finding is compared with existing findings as it emerges from the data analysis. It is associated with QUALITATIVE RESEARCH more than with QUANTITATIVE RESEARCH. It is normally associated with the GROUNDED THEORY data-analytic method, within which Glaser and Strauss (1967) referred to it as the "constant comparative method of qualitative analysis." Qualitative and quantitative data can be subject to constant comparison, but the analysis of those data is invariably qualitative.

Each comparison is usually called an iteration and is normally associated with INDUCTIVE reasoning rather than deductive reasoning; as a result, it is also referred to as "analytic induction" (Silverman, 1993). However, hypothetico-deductive reasoning will often occur within each iteration of the constant comparison method. Constant comparison is normally associated with the IDIOGRAPHIC philosophy and approach to research rather than the nomothetic philosophy and approach. Methodologies that normally employ constant comparison include ETHNOGRAPHY, PHENOMENOLOGY, SYMBOLIC INTERACTIONISM, and ETHNOMETHODOLOGY. Constant comparison contributes to the validity of research.

An example of constant comparison might be apparent when a researcher is researching the phenomenon of leadership within an organization. The methodology might employ interview data supported by observation and document data. The initial analysis of those data might involve CODING of interview transcripts to identify the variables, or categories, that seem to be present within the manifestation of the phenomenon. After analysis of the initial interviews, a number of categories might emerge, and relationships between those categories might be indicated. With each subsequent interview, each emerging category is compared with the extant categories to determine if the emerging category is a discrete category, a property of an existing category, or representative of a category at a higher level of abstraction.

For example, Kan (2002) used the full grounded theory method to research nursing leadership within a public hospital. She determined the presence of a number of lower order categories from the constant comparison of interviews and observations tracked over a 14-month period. In addition, Kan administered a questionnaire over this time frame to measure the leadership demonstrated by the managers in question. The comparison of the questionnaire results with the categories that emerged from observation and interview secured the identification of two higher order categories called *multiple realities* and *repressing leadership.*

This ongoing, or constant, comparison continues throughout the analysis of all data until the properties of all categories are clear and the relationships between categories are clear to the researcher. In the case of the work of Kan (2002), probing into the low reliabilities of some factors and the written comments provided on the questionnaires provided insights that highlighted the characteristics of the higher order categories and ultimately confirmed the basic social process of "identifying paradox."

—Ken W. Parry

REFERENCES

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory.* Chicago: Aldine.

Kan, M. (2002). Reinterpreting the multifactor leadership questionnaire. In K. W. Parry & J. R. Meindl (Eds.), *Grounding leadership theory and research: Issues, perspectives and methods* (pp. 159–173). Greenwich, CT: Information Age Publishing.

Silverman, D. (1993). *Interpreting qualitative data: Methods for analysing talk, text, and interaction.* London: Sage.

CONSTRUCT

A construct is a concept used by social scientists to help explain empirical data on a phenomenon or to conceptualize unobservable or unmeasurable elements of a domain of study to formulate a theory. Specifically, it can mean (a) an idea or a concept not directly observable or measurable (i.e., a facet of a theory), (b) a concept inferred or constructed from empirical data to explain a phenomenon in an integrated way, or (c) an abstract (instead of operational) definition of a phenomenon or concept that is manifested in other observable or empirical concepts or consequences.

For example, logical reasoning is a construct that, according to some intelligence theories, represents a facet of intelligence. Logical reasoning is not directly observable; it is inferred from observable indices such as problem-solving ability, critical thinking ability, and others. Motivation is a construct that can be inferred from empirical data on learning; it helps to integrate diverse data to construct a theory of learning. Mathematical aptitude is an abstract definition of a behavior; its existence is inferred from such manifestations as mathematical achievement, mathematical anxiety, and the number of required or elective math courses a person ever completed.

In formulating personality and social intelligence theories, psychologists have used verbal and nonverbal judgmental data, provided by ordinary people, and MULTIDIMENSIONAL SCALING to identify and confirm constructs (Jackson, 2002). This approach was equally successfully applied to study clinical judgments of psychopathology (Jackson, 2002). Still others (e.g., Digman, 1990) used FACTOR ANALYSIS to study personality dimensions such as *agreeableness, will, emotional stability,* and *intellect* (or openness). Thes constructs (i.e., factors) were extracted from CORRELATION or COVARIANCE matrices of numerous instruments for measuring personality. With advanced statistical inferential techniques such as structural equating modeling, Carroll (2002) showed that HIGHER-ORDER factors could be further extracted from these five factors to account for individual differences in personality and to help diagnose and treat personality disorders. In educational assessments, construct is a type of evidence that needs to be gathered to support the validity claim of an assessment instrument, according to the 1999 *Standards for Educational and Psychological Testing* (American Educational Research Association, 1999). Messick also (1992) asserted that "the construct validity of score interpretation comes to undergird all score-based inferences—not just those related to interpretive meaningfulness, but including the content- and criterion-related inferences specific to applied decisions and actions based on test scores." (p. 1493). For standards-based assessment programs, such as the National Assessment of Educational Progress (NAEP), measurement constructs are shaped by content standards and differential emphasis on different aspects of student understandings and performance.

For a construct to be useful for studying social phenomena, (a) it must be clearly defined, and (b) its relationships to similar or dissimilar concepts or constructs in the same domain of study must be specified. Together, the definition and the articulation of a construct's relationship with other elements in a domain of study are sometimes referred to as "the theory of the construct" or its "nomological net" (Mueller, 1986). Even though a construct may be abstract, unobservable, or unmeasurable, its manifestation or consequence must be observable or measurable by some means for social scientists to study it, formulate theories, or test hypotheses about it.

—Chao-Ying Joanne Peng

See also CONSTRUCT VALIDITY

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Carroll, J. B. (2002). The five-factor personality model: How complete and satisfactory is it? In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 97–126). Mahwah, NJ: Lawrence Erlbaum.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41,* 417–440.

Jackson, D. N. (2002). The constructs in people's heads. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 3–18). Mahwah, NJ: Lawrence Erlbaum.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., p. 1493). New York: Macmillan.

Mueller, D. J. (1986). *Measuring social attitudes: A handbook for researchers and practitioners.* New York: Teachers College, Columbia University.

CONSTRUCT VALIDITY

In educational assessment and social science measurement, construct validity is defined within the context of test validation. Test validation is the process whereby data are collected to establish the credibility of inferences to be made from test scores—both inferences about use of scores (as in decision making) and inferences about interpretation (meaning) of scores. Prior to the 1985 *Standards for Educational and Psychological Testing* (American Educational Research Association, 1985) validity was conceptualized, in the behaviorist tradition, as a taxonomy of three major and largely distinct categories (i.e., content validity, CRITERION RELATED validity, and construct validity). Construct validity is the extent to which the test is shown to measure a theoretical construct or trait. Constructs, such as creativity, honesty, and intelligence, are concepts used by social scientists to help explain empirical data about a phenomenon or to construct a theory. To execute the gathering of validity evidence, researchers must first define the construct in question. Theoretically based hypotheses ("validity hypotheses") are then generated regarding the structure of the construct and its relationship to other constructs, as well as to any other human characteristics. Last, these validity hypotheses are supported or refuted with empirical data. A thorough validation requires the testing of many validity hypotheses. Thus, test validation is a cumulative process.

Many techniques have been proposed to support construct validity, including FACTOR ANALYSIS, convergent and discriminant validation (multitrait-multimethod matrix proposed by Campbell & Fiske, 1959), examination of developmental changes (e.g., age differences on an intelligence test), internal consistency, experimental intervention, known-group differences, and correlation with other measures or tests that tap traits or behavior domains that are hypothesized to be related to the construct in question (Anastasi & Urbina, 1997; Cronbach & Meehl, 1955).

Loevinger (1957) argued that construct validity encompassed both content and criterion-related validity. This argument and others moved the 1985 edition and the most recent (1999) *Standards* to abandon the notion of three *types of validity.* Instead, validity is construed to be "a unified though faceted concept" (Messick, 1989) for which different *types of evidence* must be gathered to support inferences about the meaning of test scores and their proper use. It must be emphasized that inferences regarding the interpretation and use of the test scores are validated, not the test itself. Thus, construct validity is at the heart of the validation process—an evidence-gathering and inference-making progression. Professional organizations, such as the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), recommend the consideration of different *sources of validity evidence,* including, but not limited to, the following:

- test criterion evidence,
- evidence based on relations to other variables (such as correlation of test scores to external variables),
- evidence based on test content,
- convergent and discriminant evidence,
- evidence based on internal structure (i.e., test items and test components conform to a theoretical construct), and
- generalizability (the extent to which the test criterion can be generalized to a new context without additional study of validity in the new context).

Messick (1989, 1992) proposed the matrix in Table 1, which summarizes key concepts in his conceptualization of the test validity model.

Table 1 Facets of Validity as a Progressive Matrix

	Test Interpretation	*Test Use*
Evidential basis	Construct validity (CV)	CV + relevance/utility (R/U)
Consequential basis	CV + value implications (VI)	CV + VI + R/U + social consequences

It should be noted that the field of research designs uses a slightly different nuance in its definition of construct validity. In this definition, the concern is with the veracity of the treatment/manipulation. Does the treatment effect a manipulation of the intended variable? Is the treatment true to the theoretical definition of the construct represented by an independent variable? In this context, the term is sometimes called the *construct validity of causes and effects* (Kirk, 1995).

—Chao-Ying Joanne Peng
and Daniel J. Mueller

See also CONSTRUCT, VALIDITY

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New York: Prentice Hall.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod validity. *American Psychologist, 15,* 546–553.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(Suppl. 9), 635–694.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5–11.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., p. 1487–1495). New York: Macmillan.

CONSTRUCTIONISM, SOCIAL

Social constructionism is, first of all, an account of knowledge-generating practices—both scientific and otherwise. At this level, constructionist theory offers an orientation toward knowledge making in the sciences, a standpoint at considerable variance with the empiricist tradition. At the same time, social constructionism contains the ingredients of a theory of human functioning; at this level, it offers an alternative to traditional views of individual, psychological processes. Constructionist premises have also been extended to a variety of practical domains, opening new departures in such fields as therapy, organizational management, and education. (For more complete accounts, see Gergen, 1994, 1999.) Of special relevance, they have contributed to the flourishing of many new forms of research methods in the social sciences.

SOCIAL CONSTRUCTIONIST ASSUMPTIONS

Social constructionism cannot be reduced to a fixed set of principles but is more properly considered a continuously unfolding conversation about the nature of knowledge and our understanding of the world. However, several themes are typically located in writings that identify themselves as constructionist. At the outset, it is typically assumed that our accounts of the world—scientific and otherwise—are not dictated or determined in any principled way by what there is. Rather, the terms in which the world is understood are generally held to be social artifacts, products of historically situated interchanges among people. Thus, the extent to which a given form of understanding prevails within a culture is not fundamentally dependent on the empirical validity of the perspective in question but rather on the vicissitudes of social process (e.g., communication, negotiation, communal conflict, rhetoric). This line of reasoning does not at all detract from the significance of various forms of cultural understanding, whether scientific or otherwise. People's constructions of the world and self are essential to the broader practices of a culture—justifying, sustaining, and transforming various forms of conduct. In addition, different communities of meaning making may contribute differentially to the resources available to humankind—whether it be "medical cures," "moral intelligibilities," institutions of law, or "reasons to live." However, constructionism does challenge the warrant of any group—science included—to proclaim "truth" beyond its perimeters. What is true, real, and good within one tradition may not be within another, and there are no criteria for judging among traditions that are themselves free of traditions, their values, goals, and way of life.

SOCIAL CONSTRUCTION AND SOCIAL SCIENCE

The social constructionist views favored by this composite of developments begin to furnish a replacement for traditional empiricist accounts of social science. In the process of this replacement, one may discriminate between two phases, deconstruction and reconstruction. In the former phase, pivotal assumptions of scientific rationality, along with bodies of empirically justified knowledge claims, are placed in question. This work essentially represents an elaboration and extension of the early anti-foundationalist arguments, now informed by the additional developments within the literary and critical domains. Thus, an extensive body of literature has emerged, questioning the warrant and the ideological implications of claims to truth, empirical hypothesis testing, universal rationality, laws of human functioning, the value neutrality of science, the exportation of Western scientific practices, and so on.

Immersion in this literature alone would lead to the conclusion that social constructionism is nihilistic in its aims. However, as many believe, the deconstructive process is only a necessary prolegomenon to a reconstructive enterprise. Within the reconstructive phase, the chief focus is on ways in which scientific inquiry, informed by constructionist views, can more effectively serve the society of which it is a part. From this emerging sensibility, several developments are noteworthy. First, constructionst ideas place a strong emphasis on theoretical creativity; rather than "mapping the world as it is," the invitation is to create intelligibilities that may help us to build new futures. Theories of collaborative cognition, cyborg politics, and actor networks are illustrative. Second, constructionism has stimulated much work in cultural study, the critical and illuminating examination of everyday life practices and artifacts. Third, constructionist ideas have helped to generate a range of new practices in therapy, organizational change, and education in particular. Many scholars also find that in challenging disciplinary boundaries to knowledge, constructionist ideas invite broad-ranging dialogue. Thus, new areas of interest have been spawned, linking for example, theology and costructionism, literary theory and social movements, and personality study and ethical theory.

SOCIAL CONSTRUCTION AND RESEARCH METHODS

Although much constructionist writing is critical of traditional empirical methods in the social sciences, these criticisms are not lethal. There is nothing about constructionist ideas that demands one kind of research method as opposed to another; every method has its ways of constructing the world. Thus, although traditional empiricist methods may be viewed as limited and ideologically problematic, they do have important uses. However, the major importance of constructionist ideas in the domain of methodology has been to incite discussion of new methods of inquiry. Although these new methods tend to be viewed as "qualitative" (Denzin & Lincoln, 2000), constructionists do not subscribe to the traditional qualitative/quantitative distinction that holds the former as preliminary and inferior to the latter. Most qualitative inquiry has different aims, different values, and a different politics than those inherent in quantitative inquiry. Thus far, constructionist work has functioned, in particular, to support research methods emphasizing the following:

- *Value reflection:* Who is advantaged by the research methods and who may be discredited? Is the research subject exploited by the research or treated as a mere object?
- *Subject voice:* Is the voice of the subject of research heard or legitimated by the method or obliterated by the research procedure?
- *Collaborative participation:* Can the subjects of research participate with the researcher in the generation of knowledge? Can they share in or benefit from the outcomes?
- *Multiple standpoints:* Are multiple viewpoints and values represented in the research, or does one standpoint dominate?
- *Representational creativity:* Must the representation of research be limited to formal writing, or can more populist and richly compelling forms of representation be located?

In its research emphases, constructionist assumptions are particularly evident in PARTICIPATORY ACTION

RESEARCH (Reason & Bradbury, 2000), discourse analysis (Wetherell, Taylor, & Yates, 2001), narrative inquiry (see, e.g., Josselsyn, 1996), participatory ethnography, and literary and performative approaches to representation (see, e.g., Ellis & Bochner, 1996).

—Kenneth J. Gergen

REFERENCES

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Ellis, C., & Bochner, A. P. (Eds.). (1996). *Composing ethnography.* Walnut Creek, CA: AltaMira.

Gergen, K. J. (1994). *Realities and relationships.* Cambridge, MA: Harvard University Press.

Gergen, K. J. (1999). *An invitation to social construction.* London: Sage.

Josselsyn, R. (Ed.). (1996). *Ethics and process in the narrative study of lives.* Thousand Oaks, CA: Sage.

Reason, P., & Bradbury, H. (Eds.). (2000). *Handbook of action research, participative inquiry and practice.* Sage: London.

Wetherell, M., Taylor, S., & Yates, S. J. (Eds.). (2001). *Discourse theory and practice.* London: Sage.

CONSTRUCTIVISM

Although found in various forms throughout philosophy and science, constructivism emerged in psychology as a theory of learning. It was based on Jean Piaget's (1929, 1937/1955) research with children, which indicated that children actively construct their own knowledge out of their experience rather than simply absorb what they are told by teachers. Jerome Bruner (1960, 1966) later took up the theory and applied it to education. He recommended that teaching should be designed in such a way that it encourages children to reflect on and use their existing knowledge and capabilities. In this way, they come to understand new concepts by working things out for themselves.

However, constructivism is also used more broadly to refer to a number of theories and approaches in psychology that see the person as actively engaged in the construction of his or her own subjective world. This is in opposition to views that regard objects and events as having an essential nature and universal meaning, where perception is a matter of internalizing a truthful representation of the world. Ernst von Glasersfeld (1984) coined the term *radical constructivism,* arguing that all our knowledge of the world is constructed rather than gained through perception. We are therefore the constructors of our own subjective realities, the world that we each personally inhabit. For example, it is commonplace for those who have witnessed the same event to have quite different memories and descriptions of it. Radical constructivism would say that this is not because perception and memory are unreliable but because each person has constructed the event in a different way. Radical constructivism makes no claim about the existence or nature of an objective reality outside of our constructions.

A similar position is espoused by George Kelly, the founder of personal construct psychology (PCP). Kelly (1955) argues that each of us develops a system of dimensions of meaning, called constructs. We perceive the world in terms of these constructs, and our conduct can be understood in light of our own idiosyncratic construal of the world. Kelly, a clinical psychologist, devised a flexible method for exploring his clients' constructions that he called the REPERTORY GRID TECHNIQUE. This has since been adapted as a research tool in a diversity of settings outside the clinical field.

Constructivism shares some basic assumptions with social constructionism but differs from it in the extent to which the individual is seen as an agent who is in control of the construction process and in the extent to which our constructions are the product of social forces. Chiari and Nuzzo (1996) have attempted to provide a basis on which to differentiate the various constructivisms and suggest a distinction between epistemological and hermeneutic constructivism. Within hermeneutic constructivism, which would include social constructionism, reality is constituted through language and discourse. Epistemological constructivism, which would include PCP, emphasizes the construction of our personal worlds through the ordering and organization of our experience.

—Vivien Burr

REFERENCES

Bruner, J. (1960). *The process of education.* Cambridge, MA: Harvard University Press.

Bruner, J. (1966). *Toward a theory of instruction.* Cambridge, MA: Harvard University Press.

Chiari, G., & Nuzzo, M. L. (1996). Psychological constructivisms: A metatheoretical differentiation. *Journal of Constructivist Psychology, 9,* 163–184.

Kelly, G. (1955). *The psychology of personal constructs.* New York: Norton.

Piaget, J. (1929). *The child's conception of the world.* New York: Harcourt Brace Jovanovich.

Piaget, J. (1955). *The construction of reality in the child.* New York: Routledge Kegan Paul. (Original work published 1937)

von Glasersfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality.* New York: Norton.

CONTENT ANALYSIS

Among the earliest definitions of content analysis are the following: "The technique known as content analysis . . . attempts to characterize the meanings in a given body of discourse in a systematic and quantitative fashion" (Kaplan, 1943, p. 230). "Content analysis is a research technique for the objective, systematic, and quantitative description of the manifest content of communication" (Berelson, 1954, p. 489). Later definitions abandoned the requirement of quantification. Emphasis shifted to "inference," "objectivity," and "systematization." Holsti, in one of the early textbooks, wrote, "Content analysis is any technique for making inferences by objectively and systematically identifying specified characteristics of messages. . . . Our definition does not include any reference to quantification" (Holsti, 1969, p. 14), a point later reiterated by other contributors to the development of the technique (e.g., Krippendorf, 1980). Whether quantitative or simply inferential, content analysis approaches the study of text as a scientific, rather than simply interpretive, activity.

A COLLECTION OF DIFFERENT APPROACHES

What goes under the common label of content analysis is not a single technique; rather, it is a collection of different approaches to the analysis of texts or, more generally, of messages of any kind—from the word counts of the simplest forms of syntactical analysis to thematic analysis, referential analysis, and prepositional analysis (Krippendorf, 1980, pp. 60–63). To this list we must add a more recent approach based on story grammars and concerned with the study of social relations.

Word Counts and Key Word Counts in Context

The advance of computers has made computing the frequencies of the occurrence of individual words and concordances easy (i.e., words in the context of other words or key word in context [KWIC]) (Weber, 1990, pp. 44–53). This type of analysis (also known as corpus analysis among linguists) can provide a valuable first step in the process of data analysis of texts (using such corpus software as WordSmith).

Thematic Content Analysis

Thematic analysis is the most common approach in content analysis. In thematic analysis, the coding scheme is based on categories designed to capture the dominant themes in a text. Unfortunately, there cannot be a universal coding scheme: Different texts emphasize different things, and different investigators could be looking for different things in the same texts. That is why the development of good thematic analysis requires intimate familiarity with the input text and its characteristics. It also requires extensive pretesting of the coding scheme.

Referential Content Analysis

Thematic content analysis is a good tool for teasing out the main themes expressed in a text. But meaning is also the result of other kinds of language games: backgrounding and foregrounding information, silence and emphasis, or different ways of describing the same thing. Referential content analysis is a tool better suited than thematic analysis to capture the complexity of language in the production of meaning. Referential analysis, Krippendorf (1980) tells us, is used when "the task is to ascertain how an existing phenomenon is portrayed" (p. 62). That portrayal is affected by the choice of nouns and adjectives or even of different syntactical constructs, such as passive or active forms (Franzosi, 2003).

"Story Grammars" and the Structure of Narrative

From the very early days of the technique, researchers have pleaded for an approach to content analysis grounded on the linguistic properties of the text (e.g., de Sola Pool, Hays, Markoff, Shapiro, and Weitman; see Franzosi, 2003, chap. 1). In recent years, several authors have attempted to bridge the gap between linguistics and content analysis, proposing some basic variants of "story grammars" (or text grammars or semantic grammars, as they are also known) for the analysis of narrative texts (Abell, 1987; Franzosi, 2003; Shapiro & Markoff, 1998). Basically, a story grammar is nothing but the simple structure of the five Ws of journalism: who, what, where, when, and why (and how); someone doing something, for or against someone else, in time and space; or subject-action-object or, better, agent-action-patient/beneficiary and their modifiers (e.g., type and number of subjects and objects, time and space, reasons and outcomes of actions). The technique has been used to great effect in the study of social and protest movements.

HOW CONTENT ANALYSIS WORKS

Quantification may not necessarily be part of any formal definition of content analysis, but what the different approaches to content analysis have in common is a concern with numbers—which distinguishes them from typically qualitative approaches to the analysis of texts and symbolic material, such as NARRATIVE ANALYSIS, CONVERSATION ANALYSIS, DISCOURSE ANALYSIS, or SEMIOTICS. But, if so, how does content analysis transform words into numbers? The engine of this transformation is certainly the "coding scheme," although other "scientific" procedures play a vital role in the process: SAMPLING, VALIDITY, and RELIABILITY.

Coding Scheme

The coding scheme is the set of all coding categories applied to a collection of texts, in which a "coding category" identifies each characteristic of interest to an investigator. The scheme is systematically applied to all selected texts for the purpose of extracting uniform and standardized data: If a text contains information on any of the coding categories of the coding scheme, the relevant coding category is "ticked off" by a human coder (a process known as CODING in content analysis; modern qualitative software such as NUD*IST/N6 or Atlas.ti allow users to code text directly on the computer and to make up coding categories as they go along). When coding is completed, the ticks are added up for each coding category. In the end, the alchemic recipe of content analysis for getting numbers out of words is simple: You count. Whether it is words, themes, references, or actors and their actions, depending on the specific type of content analysis, the numbers are the result of counting.

Sampling

Sampling provides an efficient and cost-effective way to achieve research results. Rather than working with all the possible available sources and documents, investigators may decide to sample an appropriate number of sources and an appropriate number of documents. Sampling may be particularly appealing in content analysis given that it is a very labor-intensive and, therefore, expensive technique (and sampling can also be used to check the reliability of coded data through acceptance sampling schemes) (Franzosi, 2003). By sampling, you may focus on a subset of newspapers, letters, or other documents of interest. You may focus on selected years. Even for those years, you may read one or more days a week of a newspaper, one or more weeks a month, or one or more months a year. Setting up a sampling frame that will not bias the data requires a sound knowledge of the input material. For example, how would relying on only the Sunday issues of a newspaper affect data? To answer this and similar questions, you should base any choice of sources and sampling frame on systematic comparative analyses. The validity of your data (i.e., the basic correspondence between a concept that you are trying to measure and your actual measurements) will be a function of the sources and sampling frame adopted.

Intercoder Reliability

In content analysis, issues of reliability (i.e., of the repeatability of measurement) are known as *intercoder reliability*. Would different coders, confronted with the same text and using the same coding scheme, "tick off" the same coding categories? The answer to that question partly depends on the design of the coding scheme and coding categories: The more abstract and theoretically defined the categories, the more likely that different coders will come up with different results. It is good practice to test the reliability of each coding category by having different coders code the same material.

AN EXAMPLE

Content analysis can be applied to a variety of texts (e.g., newspaper editorials, speeches, documents, letters, ethnographic field notes, transcripts from in-depth interviews or focus groups) or images (e.g., advertisements, photographs). Below is a transcript of a focus group of young British-born Pakistani women who married men from Pakistan. The participants are as follows: Nadia (N), Laila (L), Ayesha (A), and the moderator (M: Maria Zubair, to whom I am grateful for the transcripts).

M: How did your husbands find life here—when they got here?
L: I think it depends really on how they live ... when they're in Pakistan. I mean some people, you know—they're less fortunate than others, they live in villages or whatever, you know, they don't really go to college, but then there are those who live in towns and they have decent upbringing—you know so when they come here they don't—they are obviously—they're quite different but not so much. . . .

L: Yeah, I mean with my husband ... when he came over here, he was the only child—um—he'd been to college, university and everything so he's educated—so when he came here he didn't find it too much of a difference.... So of course they find it different because it's ... a lot different, isn't it!

L: You know going out—they find it difficult to make friends, I think that's the main thing.
A: Yeah.
L: To make friends—or places to go—you know sometimes in the evenings and weekends we say oh where shall we go, but there's nowhere to go you know—clubs and pubs they're for English people they're not for us.

L: They find it difficult anyway because for them to come and mix with—um British-born Asians is—it's a lot difficult—it is more difficult because they're just not on the same wavelength, you know, I mean my husband—he thinks Pakistaniwise and the boys from here they think differently—don't they!
A: I find that—I find that with my husband as well.
L: Yeah.
A: Sometimes I just sit there and um—we don't (laughs)—we're not on the same wavelength ... he sees things the Pakistani way.... The time he comes home, he just sits there in front of it [TV]—on sofa—right there. He won't move then. He just sits and that's it, you know. (laughs)

L: My husband used to say to me when he came over here "oh I don't like it"... "oh I don't like you wearing English clothes and stuff—oh why don't you wear hijab [head scarf]," and you know, all that stuff. "Why can't you wear Asian clothes to work."... now he never says that, you know (N: yeah). So he's adapted to it ... but when they first come here they do get a cultural shock—I mean everything is different (M: yeah) from there, you know, so—over there it is, isn't it. It's totally different.

A: (laughs) My husband—he, you know, he just winds me up. He used to wear shalwar qameez [Pakistani clothes] when he first got here—and I said to him, look when you go out ... you're not going in these.

Even a simple word count reveals the dichotomous nature of the social context (perhaps, not unsurprisingly, given the moderator's leading question). The words occurring most frequently are *they* (21), *he* (16), *I* (12), *here* (9), and *there* (6). The relative high frequency of such words as *different* (5) and *difficult* (5) also goes a long way in highlighting the complex social relations of a marriage between a British-born Pakistani woman and a Pakistani man. After all, such words as *same* and *easy* never appear in the text. Thematic coding categories would also help to tease out the peculiarities of these different worlds. Whether done with the help of a specialized computer software or simply on paper, a coding scheme based on thematic categories would have to tap the following spheres of social life: family, friendships, jobs, and leisure. But these categories are too broad. The coding scheme would have to zoom in on the "difficulty/easiness" in "marital relationships," "own family and in-laws," "male and female friends," and "friends in Pakistan and in England." Frequency distributions of these categories would offer further insights on the question, "How did your husbands find life here?"

Yet there is more to the meaning of a text. Take Laila's words: "some people ... less fortunate than others." The choice of words is not random. Consider the alternatives: "less brave," "less driven," or "less adventurous." No—they are simply "less fortunate." There is no blame on them. Just the luck of the draw.

Conservative meritocratic views of the world would have us believe that we are all dealt a fair deck of cards. What we do with them is up to us, to the brave, the daring, the entrepreneurial. Implicitly, Laila frames her discourse along more liberal rather than conservative lines. But she also frames her discourse on the basis of implicit social scientific models that we can easily recognize as such. People who grow up in villages do not go to college. People in town have a decent upbringing. They get an education. They are not so different. Social scientific models of urbanization and of the relation between education and tolerance lurk behind Laila's words. Finally, Laila frames her discourse in very dichotomous terms of "us" versus "them": "clubs and pubs they're for English people they're not for us," "he thinks Pakistaniwise," and "we are not on the same wavelength." Laila and the other women in the group, as well as their husbands, are different from the English. But they are also different from their husbands. And Laila's educated husband is different from some of the other husbands.

These different modes of argumentation and framing could be highlighted through such coding categories as the following: "Conservative Political Views" and "Liberal Political Views," "Inclusive Worldviews" and "Exclusive/Dichotomous Worldviews," and "Mode of Reasoning (or Argumentation)," with such subcategories as "Emotive" and "Social Scientific." An emphasis on rhetoric and modes of argumentation would lead us to include such categories as "Appeal to Religion," "Appeal to Science," "Appeal to Emotion," "Appeal to Reason," "Appeal to Values of Equality," and so forth.

Referential analysis could be used to tap the differential portrayal of social actors and situations. For each main actor (e.g., the English, the husbands, the wives, the friends, the relatives back in Pakistan), event (e.g., visiting Pakistan, getting married, moving to England), or situations (e.g., an evening at home, time spent on weekends, dressing and eating habits), we could have three basic coding referential categories: positive, negative, or neutral.

CONCLUSIONS

Content analysis, as a typically quantitative approach to the study of texts, offers invaluable tools for teasing out meaning from texts or any other symbolic material. If confronted with the analysis of hundreds of pages of transcripts of the kind illustrated above (or of any other types of documents/symbolic material), the variety of content analysis techniques would certainly help us reveal patterns in the data. One advantage of quantitative analysis is that it is efficient: It helps to deal with large quantities of data without getting buried under the sheer volume of material. Its disadvantage is that it may miss out on subtle nuances in the production of meaning. Even the few examples discussed above should have made that very clear. To think in terms of a quantitative versus a qualitative approach to texts is a misguided approach. Each has strengths and weaknesses. (Consider this: How long would it take to do a qualitative analysis of hundreds of pages of transcripts? How would you ensure that you are not relying on the best sound bytes or that you have not forgotten anything?) Theoretical/methodological developments in the analysis of texts coupled with technological developments in computer software are increasingly blurring the lines between quantitative and qualitative approaches (e.g., computer-aided qualitative data analysis software [CAQDAS]) (Fielding & Lee, 1998; Kelle, 1995). At the current state of the art, the best approach is probably one that combines a broad picture of a social phenomenon through quantitative content analysis with sporadic and deep incursions into selected texts using more qualitative approaches.

—Roberto Franzosi

REFERENCES

Abell, P. (1987). *The syntax of social life: The theory and method of comparative narratives.* Oxford, UK: Clarendon.

Berelson, B. (1954). Content analysis. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 1, pp. 488–522). Reading, MA: Addison-Wesley.

Fielding, N., & Lee, R. (1998). *Computer-analysis and qualitative research.* London: Sage.

Franzosi, R. (2003). *From words to numbers.* Cambridge, UK: Cambridge University Press.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities.* Reading, MA: Addison-Wesley.

Kaplan, A. (1943). Content analysis and the theory of signs. *Philosophy of Science, 10,* 230–247.

Kelle, U. (Ed.). (1995). *Computer-aided qualitative data analysis.* London: Sage.

Krippendorf, K. (1980). *Content analysis: An introduction to its methodology.* Beverly Hills, CA: Sage.

Shapiro, G., & Markoff, J. (1998). *Revolutionary demands: A content analysis of the Cahier de Doléances of 1789.* Stanford, CA: Stanford University Press.

Weber, R. P. (1990). *Basic content analysis.* Newbury Park, CA: Sage.

CONTEXT EFFECTS. *See* ORDER EFFECTS

CONTEXTUAL EFFECTS

The notion of groups plays a major role in social science. Does membership of an individual in a particular group have any impact on thoughts or actions of the individual? Does living in a neighborhood consisting of many Democrats mean than an individual has a higher probability of being a Democrat than otherwise would have been the case? Such an impact by the group is known as a *contextual effect.* The study of these effects is called contextual analysis or MULTILEVEL ANALYSIS. It is also known as the study of HIERARCHICAL models.

If contextual effects exist, it raises the following questions: Why does belonging to a group have an impact, how does this mechanism work, and how should we measure the impact? The first two questions may best be left to psychologists, but the third question has sparked much interest among statisticians and others. In particular, in education, a classroom makes an obvious group for contextual analyses.

When we have data on two variables on individuals from several groups, within each group we can analyze the relationship of a dependent variable Y and an independent variable X. Regression analysis gives a SLOPE and an INTERCEPT for the regression line for each group. It may now be that the slopes and intercepts vary from one group to the next. If this variation in the regression lines is larger than can be explained by randomness alone, then there must be something about the groups themselves that affects the lines.

There can be several reasons why the slopes and intercepts vary. It may be that both depend on the level of the X variable in the group through the group mean. Thus, one possible contextual model can be expressed in the following equations:

$$a_j = c_0 + c_2\bar{x}_j + u_j,$$
$$b_j = c_1 + c_3\bar{x}_j + v_j,$$

where a and b are the intercept and slope in the jth group, and the u and v are residual terms.

It is also possible to substitute for a and b in the equation for the line in the jth group. That gives the following equation:

$$y_{ij} = c_0 + c_1x_{ij} + c_2\bar{x}_j + c_3(x_{ij}\bar{x}_j) + e_{ij}$$

for the ith individual in the jth group. Thus, we can now run a MULTIPLE REGRESSION on data from all the groups and get the c coefficients. For this model, c_1 measures the effect of X on the level of the individual, c_2 measures the effect of X on the level of the group, and c_3 measures the effect on Y of the interaction of the individual and the group-level effects. One problem with this analysis is that the RESIDUAL E depends on the group means of X; another is the COLLINEARITY that exists in the three explanatory variables.

Contextual analysis is not limited to the case in which the group intercepts and slopes depend on the group means of X. They could also depend on a variety of other variables.

Bayesian statistics lends itself particularly well to multilevel analysis because the slopes and intercepts vary from group to group. It may then be possible to specify how the intercepts and slopes within the groups come from specific distributions.

—Gudmund R. Iversen

REFERENCES

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

Goldstein, H. (1987). *Multilevel models in educational and social research.* London: Griffin.

Iversen, G. R. (1991). *Contextual analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–081). Newbury Park, CA: Sage.

CONTINGENCY COEFFICIENT. *See* SYMMETRIC MEASURES

CONTINGENCY TABLE

A contingency table is a statistical table classifying observed data frequencies according to the categories in two or more variables. A table formed by the cross-classification of two variables is called a *two-way contingency table* (for an example, see

Table 1), a table of three variables is termed a *three-way contingency table,* and, in general, a table of three or more variables is known as a *multiway contingency table.* The analysis of multiway contingency tables is sometimes known as multivariate contingency analysis.

All cells in a table may not have observed counts. Sometimes, cells in a contingency table contain zero counts, known as empty cells. A zero count may be generated by two distinctive mechanisms. If sample size causes a zero count (e.g., Chinese American farmers in Illinois), it is known as a sampling zero. If, for a cell, it is theoretically impossible to have observations (e.g., male [contraceptive] diaphragm users), it is called a structural zero. Contingency tables with at least one structural zero are termed *incomplete tables.* When many cells in a contingency table have lower observed frequencies, be they zero or not, such a table is known as a SPARSE TABLE.

HISTORICAL DEVELOPMENT

Early work on CATEGORICAL DATA ANALYSIS in the beginning of the 20th century was primarily concerned with the analysis of contingency tables. The well-known debate between Karl Pearson and G. Udny Yule sparked interest in contingency table analysis. How would we analyze a 2×2 contingency table? Pearson was a firm believer in continuous bivariate distributions underlying the observed counts in the cross-classified table; Yule was of the opinion that variables containing discrete categories such as inoculation versus no inoculation would be best treated as discrete without assuming underlying distributions. Pearson's contingency coefficient was calculated based on the approximated underlying correlation of bivariate normal distributions collapsed into discrete cross-classifications, whereas Yule's Q was computed using a function of the ODDS RATIO of the observed discrete data directly.

Contributions throughout the 20th century can often be traced by the names of those inventing the statistics or tests for contingency table analysis. Pearson chi-square test, Yule's Q, Fisher's exact test, Kendall's tau, Cramér's V, Goodman and Kruskal's tau, and the Cochran-Mantel-Haenszel test (or the Mantel-Haenszel test) are just some examples. These statistics are commonly included in today's statistical software packages (e.g., proc freq in SAS generates most of the above).

Table 1 A Contingency Table of Parents' Socioeconomic Status (SES) and Mental Health Status (expected values in parentheses)

	Mental Health Status			
Parent's SES	*Well*	*Mild Symptom*	*Moderate Symptom*	*Impaired*
A (high)	64 (48.5)	94 (95.0)	58 (57.1)	46 (61.4)
B	57 (45.3)	94 (88.8)	54 (53.4)	40 (57.4)
C	57 (53.1)	105 (104.1)	65 (62.6)	60 (67.3)
D	72 (71.0)	141 (139.3)	77 (83.7)	94 (90.0)
E	36 (49.0)	97 (96.1)	54 (57.8)	78 (62.1)
F (low)	21 (40.1)	71 (78.7)	54 (47.3)	71 (50.9)

Contingency table analysis can be viewed as the foundation for contemporary categorical data analysis. We may see the series of developments in LOG-LINEAR MODELING in the latter half of the 20th century as refinements in analyzing contingency tables. Similarly, the advancement in latent trait and latent class analysis is inseparable from the statistical foundation of contingency table analysis.

EXAMPLE

We present a classic data table, the midtown Manhattan data of mental health and parents' socioeconomic status (SES), analyzed by L. Goodman, C. C. Clogg, and many others (see Table 1).

To analyze contingency tables, we need expected frequencies. Let F_{ij} indicate the expected value for the observed frequency f_{ij} in row i and column j in the table. The F_{ij} under the model of independence (no association between parents' SES and mental status) is given by

$$F_{ij} = \frac{f_{i+}f_{+j}}{f_{++}},$$

where f_{i+} gives the column total for the ith row, f_{+j} gives the row total for the jth column, and f_{++} gives the grand total of the entire table. For example, $f_{1+} = 64 + 94 + 58 + 46 = 262$, $f_{+1} = 64 + 57 + 57 + 36 + 21 = 307$, and $F_{11} = (262 \times 307)/1,660 = 48.5$. We compute the expected values for the entire table accordingly and present them in parentheses in Table 1.

The null hypothesis is that the two variables are unrelated or independent of each other. To test the hypothesis, we compute the Pearson chi-square statistic and the LIKELIHOOD RATIO STATISTIC L^2 (or G^2):

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - F_{ij})^2}{F_{ij}} \quad \text{and}$$

$$L^2 = 2 \sum_i \sum_j f_{ij} \log\left(\frac{f_{ij}}{F_{ij}}\right).$$

Applying these formulas, we obtain a Pearson chi-square statistic of 45.99 and an L^2 of 47.42. With 15 degrees of freedom (= the number of rows minus 1 times the number of columns minus 1), we reject the null hypothesis of independence at any conventional significance level and conclude that mental health status and parents' SES are associated.

The model of independence can also be understood from the odds ratios computed from the expected frequencies. One may calculate the cross-product ratio for every adjacent 2×2 table contained in the contingency table to understand the independence assumption. That is, $(48.5 \times 88.8)/(95.0 \times 45.3) \approx 1.0$, $(95.0 \times 53.4)/(57.1 \times 88.8) \approx 1.0, \ldots, (57.8 \times 50.9)/(62.1 \times 47.3) \approx 1.0$. One quickly discovers that all such odds ratios are approximately 1! An ODDS RATIO of 1 indicates that there is no relationship between the two variables forming the table.

Both variables are ordinal in nature, but that information is not considered in the current analysis. Further fine-tuning via log-linear modeling will be necessary to analyze the type of association between the two variables.

COLLAPSIBILITY OF CONTINGENCY TABLES

A prominent issue in contingency table analysis is that of collapsibility. Simply put, a contingency table is collapsible if the fit of a particular statistical model is not significantly affected by combining its dimensions (i.e., removing variables), combining some categories in a variable (i.e., reducing the number of response categories), or both. In the current example, someone may suspect that there exists little difference between the mild and the moderate symptoms, noticing that the differences between the observed and the expected counts in the two middle columns in Table 1 are uniformly below 1. Table 2 presents the observed data (and the expected counts) with those two categories collapsed.

Table 2 A Contingency Table of Parents' Socioeconomic Status (SES) and Mental Health Status With Collapsed Middle Symptom Categories (expected values in parentheses)

	Mental Health Status		
Parent's SES	*Well*	*Mild or Moderate Symptom*	*Impaired*
A (high)	64 (48.5)	152 (95.0)	46 (61.4)
B	57 (45.3)	148 (88.8)	40 (57.4)
C	57 (53.1)	170 (104.1)	60 (67.3)
D	72 (71.0)	218 (139.3)	94 (90.0)
E	36 (49.0)	151 (96.1)	78 (62.1)
F (low)	21 (40.1)	125 (78.7)	71 (50.9)

From Table 2, we arrive at a new set of test statistics: a Pearson chi-square of 43.51 and an L^2 of 44.94, with 10 degrees of freedom based on six rows and three columns (a still significant result at the .05 level). A likelihood ratio test based on either the Pearson chi-square statistics or the L^2s by taking the difference of the new and the old values gives a test statistic of 2.47 with $5 (= 15 - 10)$ degrees of freedom. This result, which follows a chi-square distribution, is highly insignificant, suggesting that the two middle symptom categories can be collapsed without losing much information. The reader is encouraged to find out whether the parents' SES categories can be collapsed.

—Tim Futing Liao

REFERENCES

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley.

Fienberg, S. E. (1981). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: MIT Press.

Rudas, T. (1997). *Odds ratios in the analysis of contingency tables.* Thousand Oaks, CA: Sage.

CONTINGENT EFFECTS

In theorizing about the relationship between an INDEPENDENT VARIABLE and a DEPENDENT VARIABLE, it may be the case that this relationship is contingent on the value of another independent variable. This type of relationship implies a multiplicative term in any statistical model designed to empirically estimate the relationships of interest. Formally, this would take the following form:

$$Y = B_{12}(X_1 \cdot X_2),$$

where Y is the dependent variable, X_1 is an independent variable, X_2 is the variable on which the relationship between X_1 and Y is contingent, and B_{12} is the parameter that defines this relationship. Thus, the relationship between X_1 and Y is contingent on the value of X_2. An example of this comes from Douglas Hibbs's (1982) classic work on presidential popularity in the United States during the Vietnam War era. Hibbs theorized that individuals' opinions toward the job that the president of the United States was doing would be directly affected by the number of U.S. military casualties that had occurred in recent weeks. In addition, he theorized that the magnitude of this effect would be contingent on individuals' social class because U.S. military casualties during the Vietnam era were disproportionately concentrated among the lower social classes.

A secondary definition for contingent effects comes from the standard inferences of multivariate models. For instance, consider the following nonstochastic component of a multivariate ORDINARY LEAST SQUARES regression with two independent variables:

$$Y = B_0 + B_1X_1 + B_2X_2,$$

where Y is the dependent variable, X_1 and X_2 are independent variables, and B_0 is the intercept term to be estimated along with the other parameters of interest. B_1 is the parameter (to be estimated) of the effect of X_1 on Y contingent on the effect of X_2 on Y. B_2 is the parameter (to be estimated) of the effect of X_2 on Y contingent on the effect of X_1 on Y. Another way that this is often expressed is that each slope parameter (B_1 and B_2 are the slope parameters in this example model) is the effect of one independent variable on the dependent variable, "controlling for" all other independent variables in the model.

—Guy Whitten

See also INTERACTION EFFECT, MULTIPLICATIVE

REFERENCE

Hibbs, D. A. (1982). The dynamics of political support for American presidents among occupational and partisan groups. *American Journal of Political Science, 26,* 312–332.

CONTINUOUS VARIABLE

A continuous variable is one within whose range of values any value is possible. Sometimes, it is also known as a metric variable or a quantitative variable. In practice, integer variables are often considered continuous.

—Tim Futing Liao

See also LEVEL OF MEASUREMENT

CONTRAST CODING

Contrast coding is a technique for coding $j - 1$ DUMMY VARIABLES to fully represent the information contained in a j-category NOMINAL classification. Contrast-coded dummy variables allow the researcher to aggregate categories into meaningful groupings for comparison purposes and, at the same time, test to determine whether component categories of each group are significantly different from each other. Coefficients for contrast-coded dummy variables provide comparisons of subgroup means sequentially disaggregated from *theoretically interesting* group combinations.

To illustrate, consider the relationship between marital status (married, separated, divorced, widowed, never married) and hourly wages (in U.S. dollars). In this example, marital status consists of five (J) subgroups, represented by four contrasts ($J - 1$). Although several initial aggregations are reasonable, we construct the first contrast as one between people who are currently legally married (married and separated) and those who are not (never married, divorced, and widowed), as illustrated in Table 1. Each successive contrast will disaggregate these two groups into component subgroups. Contrast 2 compares those separated to the married respondents. Because the "not married" aggregate contains three subgroups, we choose to compare the never-marrieds to a combination of respondents who are divorced or widowed in Contrast 3, with those divorced compared to the widowed in Contrast 4.

Two additional rules apply when using contrast codes for dummy variables. First, the codes within any single contrast-coded variable must sum to zero.

Table 1

Groups of Interest	Contrast 1	Contrast 2	Contrast 3	Contrast 4	Means (U.S.$)
	New Variables Status				
Married	$-1/2$	-1	0	0	15.33
Never married	$1/3$	0	-1	0	12.95
Separated	$-1/2$	1	0	0	11.64
Divorced	$1/3$	0	$1/2$	1	12.94
Widowed	$1/3$	0	$1/2$	-1	11.93
Sum	$-1/2+1/3-1/2+1/3+1/3=0$	$-1+0+1+0+0=0$	$0-1+0+1/2+1/2=0$	$0+0+0+1-1=0$	Grand mean 14.26
Independence	C1 and C2 $= (-1/2)(-1) + 1/3(0) + (-1/2)(1) + (1/3)(0) + (1/3)(0) = 0$				
	C1 and C3 $= (-1/2)(0) + (1/3)(-1) + (-1/2)(0) + (1/3)(1/2) + (1/3)(1/2) = 0$				
	C1 and C4 $= (-1/2)(0) + (1/3)(0) + (-1/2)(0) + (1/3)(1) + (1/3)(-1) = 0$				
	C2 and C3 $= (-1)(0) + (0)(-1) + (1)(0) + (0)(1/2) + (0)(1/2) = 0$				
	C2 and C4 $= (-1)(0) + (0)(0) + (1)(0) + (0)(1) + (0)(-1) = 0$				
	C3 and C4 $= (0)(0) + (-1)(0) + (0)(0) + (1/2)(1) + (1/2)(-1) = 0$				

Second, the codes across dummy variables must be orthogonal. To test independence, we sum the products of successive pairs of codes, as illustrated in the bottom panel of the table, titled "Independence."

In general, descriptive statistics and zero-order correlations involving these dummy variables are not useful (Hardy, 1993). When contrast-coded dummy variables are used as independent variables in ORDINARY LEAST SQUARES REGRESSION, the constant reports the unweighted mean of the subgroup means, calculated as $[\sum_{j=1}^{J} \bar{X}_j] \div J$, the same as in regression with effects-coded dummy variables but unlike regression with BINARY coded dummy variables. The REGRESSION COEFFICIENTS can be used to calculate the difference in means within each contrast using the following formula: $C_j = B((n_{g1} + n_{g2})/(n_{g1})(n_{g2}))$, in which n_{g1} is the number of subgroups included in the first group, n_{g2} is the number of subgroups included in the second group, and B is the regression coefficient for the dummy variable of interest.

Results from a regression of hourly wage on the four contrast-coded dummy variables plus a constant are reported in Table 2. The value of the constant is 12.96, which one can verify as the unweighted mean of subgroup means by summing the category means provided in the previous table and dividing by 5. To evaluate the regression coefficients, consider the coefficient for Contrast 1. Substituting into the above equation, we have the following: Contrast 1 $= -1.054((2 + 3)/(2 \cdot 3)) = -.878$. This value is the difference between the unweighted mean of the subgroups of legally married couples ($[15.33 + 11.64]/2 = 13.49$) and the unweighted mean of the subgroups of people who are not legally married ($[12.95 + 12.94 + 11.93]/3 = 12.61$). People who are not married legally make, on average, \$.88 less than people who are legally married. However, the t-test for this coefficient is not statistically significant at conventional levels; therefore, the difference of \$.88 is not reliable. The appropriate inferential test to determine whether hourly wages differ by marital status is the F-test, not the T-TEST, because including information about the different marital statuses requires all four dummy variables.

Table 2

Model	Unstandardized Coefficients: B	Unstandardized Coefficients: Standard Error	t	Significance
Constant	12.96	.471	27.527	.000
Contrast 1	−1.054	.993	−1.061	.289
Contrast 2	−1.844	.353	−5.226	.000
Contrast 3	−.346	.772	−.448	.654
Contrast 4	.502	1.111	.452	.651

—Melissa A. Hardy and Chardie L. Baird

See also DUMMY VARIABLE, EFFECTS CODING

REFERENCE

Hardy, M. A. (1993). *Regression with dummy variables.* Newbury Park, CA: Sage.

CONTROL

Control is concerned with a number of functions by which the existence of a relationship between two main variables, an independent (or explanatory) variable and a dependent (or response) variable, can be ascertained without the interference of other factors. The four main functions of control are to fix, or "hold constant," CONFOUNDING effects; to test for spurious effects between the variables; to serve as a reference point in experimental tests; and to ascertain the *net* effect of the independent variable when other variables also have an effect.

Confounding effects could be ones that intervene in the bivariate relationship or ones that precede it. Alternatively, the relationship could be a spurious one, that is, when a third variable affects both the independent and dependent variables. To check for the effect of such confounders, we control for them, so that we are comparing like with like (see CETERIS PARIBUS). For example, we may notice that certain minority ethnic groups in Britain achieve less well at school than White children. We might, therefore, posit that cultural or genetic factors associated with ethnicity cause some groups to fare worse than others at school. However, if we consider that there is also known to be an impact of income and deprivation on school achievement levels and that certain minority groups tend to be concentrated at the bottom of the income bracket, we might wish to control for income levels. We would then find that the relationship is not sustained and that, controlling for income (i.e., comparing poor kids with poor kids and rich kids with rich kids), children from Bangladeshi backgrounds fare at least as well as their White counterparts. The apparent relationship between ethnicity itself and low achievement is thus not sustained when background is controlled for. This example of an intervening effect is illustrated in Figure 1. Bangladeshi children still have poor educational outcomes, but the process of causal reasoning has been modified, placing the onus of explanation on their poverty rather than their ethnicity.

An alternative example of an intervening relationship could deal with women's lower average earnings rather than men's. An initial relationship might be posited between sex and lower earnings, which focuses on discrimination within the workplace. It is relevant to consider that women tend to be concentrated in the lower echelons of the employment hierarchy. If we hold occupational class constant across men and women, do we still see a gender effect? What we find is that the effect remains, but it is much reduced. Women's occupational structure explains in part their lower earnings rather than their sex per se, and this is revealed by comparing men and women of the same occupational class. See Figure 2 for an illustration of this relationship.

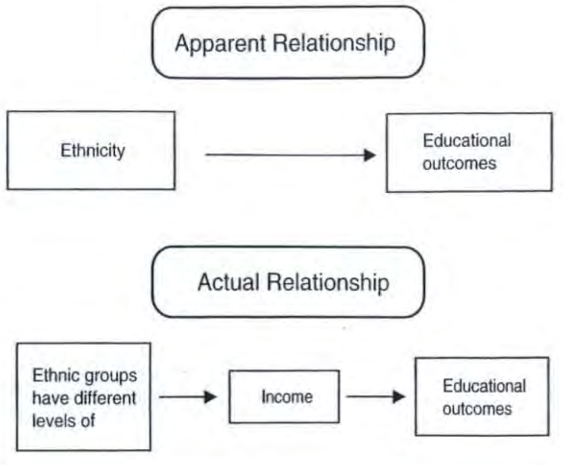

Figure 1 The Effect of Controlling for Income in Considering the Relationship Between Ethnicity and Educational Outcomes

Additional relevant factors to those of interest can be controlled for through experimental design or by using statistical methods. In experimental design, the effect of an intervention can be tested by randomly applying the intervention to one subset of test cases and withholding it from the remaining CONTROL GROUP. The control group can also be selected to match the "experimental group" in critical ways (such as age, sex, etc.). For example, medical interventions are typically tested using RANDOMIZED CONTROLLED TRIALS. *Random* refers to the fact that those who are experimented on are randomly selected from within the population of interest, and *controlled* means that for every case that is selected for EXPERIMENT, a parallel case will not receive the intervention but will be treated in every other way as the experimental group, including, possibly, being given a placebo treatment. Such trials are double blind if not only the experiment and control groups do not know which group they belong to but also even the practitioner administering the intervention (or placebo) is unaware, thus minimizing the impact of researcher bias. In such experiments, given random allocation of the intervention, any differences

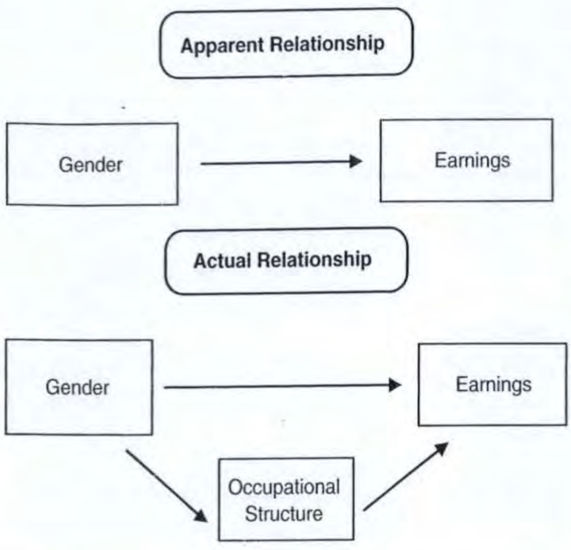

Figure 2 The Effect of Holding Occupational Class Constant When Assessing the Relationship Between Gender and Earnings

in outcomes between the experimental group and the control group can then be attributed to the intervention.

Although experimental techniques are regularly used in psychology, more commonly in the social sciences, it is not practical and would not be meaningful to run experiments. For example, it would not be feasible to establish an experiment on the effects of parental separation on child outcomes, which would require a random sample of the experimental group to separate from their partners and the matched control group to remain with them. Instead, confounding effects are controlled for using statistical methods. There are a variety of multivariate techniques for assessing the effect of a number of different variables simultaneously and thus being able to distinguish the net effect of one when another (or several others) is controlled for. Regression analysis, such as MULTIPLE REGRESSION ANALYSIS or LOGISTIC REGRESSION, is the most commonly used way of ascertaining such net effects. In such analysis, it may be that additional variables are included in a model that do not show any significant effect themselves, simply to ensure that all relevant factors have been controlled for.

—Lucinda Platt

See also CETERIS PARIBUS, LOGISTIC REGRESSION, MULTIVARIATE ANALYSIS, MULTIPLE REGRESSION ANALYSIS

REFERENCE

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

CONTROL GROUP

A control group is the group with which an experimental group (or treatment group) is contrasted. It consists of units of study that either did not receive the treatment, whose effect is under investigation, or units of study that did receive a placebo, whose effect is not being studied. Control groups may be alternatively called baseline groups or contrast groups. They are formed through random assignment of units of study, as in between-subject designs, or from the units of study themselves, as in WITHIN-SUBJECT DESIGNS. In true-experimental studies and between-subject designs, units of study (such as children, clinical patients, schools, hospitals, communities, etc.) are first randomly selected from their respective populations; then they are randomly assigned into either a control group or an experimental group. At the conclusion of the study, outcome measures (such as scores on one or more dependent variables) are compared between those in the control group and those in the experimental group (or groups). The effect of a treatment is assessed on the basis of the difference (or differences) observed between the control group and one or more experimental group. In true-experimental studies and within-subject designs, units of study are also randomly selected from their respective populations, but they are not randomly assigned into control versus experimental groups. Instead, baseline data are gathered from units of study themselves. These data are treated as "control data," to be compared with outcome measures that are hypothesized to be the result of a treatment. Thus, units of study act as their own "control group" in within-subject designs. For QUASI-EXPERIMENTAL studies, treatments are not administered by an experimenter to units of study or participants, as in true-experimental studies. Treatments are broadly construed to be characteristics of participants, such as gender, age, and socioeconomic status (SES), or features of their settings, such as public versus private school and single- or intact-family background. Quasi-experimental studies may also be necessary when researchers have little or no control over randomly assigning participants into different treatment conditions. In this case, a nested design is appropriate and necessary.

—Chao-Ying Joanne Peng

See also EXPERIMENT

REFERENCES

Huck, S. (2000). *Reading statistics and research* (3rd ed.). New York: Addison-Wesley Longman.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

CONVENIENCE SAMPLE

Convenience sampling, as its name suggests, involves selecting sample units that are readily accessible to the researcher. It is also sometimes called accidental sampling and is a form of NONPROBABILITY SAMPLING; that is, each member of a population has an unknown and unequal probability of being selected.

The advantages of convenience samples are that they are relatively inexpensive and, by definition, easy to access. Sometimes, this form of sampling may be the most efficient way to access a hard-to-reach population, such as sex workers or illicit drug users. Convenience sampling becomes more appropriate and acceptable when the POPULATION of interest is difficult to define to allow any reliable form of random sampling. It is also deemed more legitimate when gaining access to the population of interest is difficult.

Although studies using convenience samples may yield intriguing findings, they suffer from the inability to generalize beyond the samples. Researchers will not know whether the sample is representative of the population under study. On the basis of only a convenience sample, researchers will not know whether their sample is typical or atypical of other groups. However, this shortfall is common to all nonprobability samples. Researchers using this form of sampling should also consider the potential BIAS involved.

The use of convenience sampling is not limited to any particular subject area. Researchers have used convenience samples to address issues of ageism, examine sexual decisions made by inner-city Black adolescents, and explore the stigma of wheelchair use and how the public reacts to wheelchair users.

Convenience sampling is also employed in both qualitative and quantitative data collection. A typical example of employing convenience sampling in quantitative data collection is surveying students in a particular course, generally a course taught by the researcher or a friend of the researcher. The students are essentially "captured" participants in a class setting, easily accessible by the researcher. A typical example of using convenience sampling in collecting qualitative information is when researchers conduct in-depth interviews with family members. Most family members, especially those living in the same household, are likely to participate in the interviews as they are likely to have a more trusting relationship with the interviewer than with a stranger. Most of all, they are also readily available for the interview.

—VoonChin Phua

CONVERSATION ANALYSIS

When people talk with one another, they are not merely communicating thoughts, information, or knowledge: In conversation, as in all forms of interaction, people are *doing* things in talk. They are engaged in social activities with one another, such as blaming, complaining, inviting, instructing, telling their troubles, and so on. Conversation analysis (CA) is the study of how participants in talk-in-interaction organize their verbal and nonverbal behavior so as to conduct such social actions and, more broadly, their social affairs. It is therefore primarily an approach to social action (Schegloff, 1996). Methodologically, it seeks to uncover the practices, patterns, and generally the methods through which participants perform and interpret social action.

CA originated with the work that Harvey Sacks (1935–1975) undertook at the Center for the Scientific Study of Suicide, in Los Angeles, 1963–1964. Drawn by his interests both in the ethnomethodological concern with members' methods of practical reasoning (arising from his association with Harold Garfinkel) and in the study of interaction (stimulated by his having been taught as a graduate student at Berkeley by Erving Goffman), Sacks began to analyze telephone calls made to the Suicide Prevention Center (SPC). Without any diminished sensitivity to the plight of persons calling the SPC, Sacks investigated how callers' accounts of their troubles were produced in the course of their conversations over the telephone with SPC counselors. This led him to explore the more generic "machineries" of conversational turn taking, along with the sequential patterns or structures associated with the management of activities in conversation. Through the collection of a broader corpus of interactions, including group therapy sessions and mundane telephone conversations, and in collaboration with Gail Jefferson (1938–) and Emanuel Schegloff (1937–), Sacks began to lay out a quite comprehensive picture of the conversational organization of turn taking; overlapping talk; repair; topic initiation and closing; greetings, questions, invitations, requests, and so forth and their associated sequences (adjacency pairs); agreement and disagreement; storytelling; and integration of speech with nonvocal activities (Sacks, 1992; for a comprehensive review of CA's topics and methodology, see ten Have, 1999). Subsequent research in CA over the past 30 years has shown how these and other technical aspects of talk-in-interaction are structured, socially organized resources—or methods—whereby participants perform and coordinate activities through talking together. Thus, they are the technical bedrock on which people build their social lives or, in other words, construct their sense of sociality with one another.

In many respects, CA lies at the intersection between sociology and other cognate disciplines, especially linguistics and social psychology. Certainly, research in CA has paralleled developments within sociolinguistics, pragmatics, discourse analysis, and so forth toward a naturalistic, observation-based science of actual verbal behavior, which uses recordings of naturally occurring interactions as the basic form of data (Heritage, 1984). Despite the connections between CA and other disciplines, CA is distinctively sociological in its approach in the following kinds of ways. First, in its focus on how participants understand and respond to one another in their turns at talk, CA explores the social and interactional underpinnings of intersubjectivity—the maintenance of common, shared, and even "collective" understandings between social actors. Second, CA develops empirically Goffman's (1983) insight that social interaction embodies a distinct moral and institutional order that can be treated like other social institutions, such as the family, economy, religion, and so on. By the "interaction order," Goffman meant the institutional order of interaction, and CA studies the practices that make up this institution as a topic in its own right. Third, all levels of linguistic production (including syntactic, prosodic, and phonetic) can be related to the actions (such as greetings, invitations, requests) or activities (instructing, cross-examining, performing a medical examination and diagnosing, etc.) in which people are engaged when interacting with one another. In this way, conversational organizations underlie social action (Atkinson & Heritage, 1984); hence, CA offers a methodology, based on analyzing sequences in which actions are produced and embedded, for investigating how we accomplish social actions. This is applicable equally to both verbal and nonverbal conduct, as well as the integration between them, in face-to-face interaction (e.g., Goodwin, 1981). Fourth, it is evident that the performance by one participant of certain kinds of actions (e.g., a greeting, question, invitation, etc.) sets up certain expectations concerning what the other, the recipient, should do in response; that is, he or she may be expected to return a greeting, answer the question, accept or decline the invitation, and so on. Thus, such pairs of actions, called *adjacency pairs* in CA, are normative frameworks within which certain actions should properly or accountably be done: The normative character of action and the associated accountability of acting in accordance with normative expectations are vitally germane to sociology's sense of the moral order of social life, including ascriptions of deviance. Fifth, CA relates talk to social context. CA's approach to context is distinctive, partly because the most proximate context for any turn at talk is regarded as being the (action) sequence of which it is a part, particularly the immediately prior turn. Also, CA takes the position that the "context" of an interaction cannot be exhaustively defined by the analyst a priori; rather, participants display their sense of relevant context (mutual knowledge, including what each knows about the other; the setting; relevant biographical information; their relevant identities or relationship, etc.) in the particular ways in which they design their talk—that is, in the recipient design of their talk.

Finally, as a method of enquiry, CA is equally as applicable to institutional interactions as it is to mundane conversation (Drew & Heritage, 1992). The social worlds of the law, medicine, corporations and business negotiations, counseling, education, and public broadcast media (especially news interviews) and other such institutional and workplace settings are conducted through talk-in-interaction. CA has increasingly come to focus on how participants in such settings manage their respective institutional activities, for instance, as professionals (doctors, lawyers, etc.) or as laypersons (clients, patients, students, etc.). Methodologically, the objective of CA research into institutional interactions is to reveal the practices through which participants manage their interactions as a specialized form of interaction—for instance, as a news interview (Clayman & Heritage, 2002), as doing counseling, as courtroom cross-examination, and so forth. In such research, CA connects very directly with ethnographic sociology and offers a distinctive, rigorous, and fruitful methodology for investigating the organizational and institutional sites of social life.

—Paul Drew

REFERENCES

Atkinson, J. M., & Heritage, J. (Eds.). (1984). *Structures of social action: Studies in conversation analysis*. Cambridge, UK: Cambridge University Press.

Clayman, S., & Heritage, J. (2002). *The news interview: Journalists and public figures on the air.* Cambridge, UK: Cambridge University Press.

Drew, P., & Heritage, J. (Eds.). (1992). *Talk at work: Interaction in institutional settings*. Cambridge, UK: Cambridge University Press.

Goffman, E. (1983). The interaction order. *American Sociological Review, 48,* 1–17.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York: Academic Press.

Heritage, J. (1984). *Garfinkel and ethnomethodology.* Cambridge, UK: Polity.

Sacks, H. (1992). *Lectures on conversation* (2 vols.). Oxford, UK: Blackwell.

Schegloff, E. A. (1996). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology, 2,* 161–216.

ten Have, P. (1999). *Doing conversation analysis.* Thousand Oaks, CA: Sage.

CORRELATION

Generally speaking, the notion of correlation is very close to that of ASSOCIATION. It is the statistical association between two variables of interval or ratio measurement level.

To begin with, correlation should not be confused with causal effect. Indeed, statistical research into causal effect for only two variables happens to be impossible, at least in observational research. Even in extreme cases of so-called unicausal effects, such as contamination with the *Treponema* pallidum bacteria and contracting syphilis, there are always other variables that come into play, contributing, modifying, or counteracting the effect. The presence of a causal effect can only be sorted out in a multivariate context and is consequently more complex than correlation analysis.

Let us limit ourselves to two interval variables, denoted as X and Y, and let us leave CAUSALITY aside. We assume for the moment a linear model. It is important to note that there is a difference between the correlation coefficient, often indicated as r, and the UNSTANDARDIZED regression coefficient b (computer output: B). The latter merely indicates the slope of the REGRESSION line and is computed as the tangent of the angle formed by the regression line and the x-axis. With income (X) and consumption (Y) as variables, we now have the consumption quote, which is the additional amount one spends after obtaining one unit of extra income, so the change in Y per additional unit in X is $B = \Delta Y/\Delta X$. We will see below that there are, in fact, two such regression coefficients and that the correlation coefficient is the geometrical MEAN of them.

FIVE MAIN FEATURES OF A CORRELATION

Starting from probabilistic correlations, each correlation has five main features:

1. nature,
2. direction,
3. sign,
4. strength,
5. statistical generalization capacity.

1. Nature of the Correlation

The nature of the correlation is linear for the simple correlation computation suggested above. This means that through the scatterplot, a linear function of the form $E(Y) = b_0 + b_{y1}X_1$ is estimated. Behind the correlation coefficient of, for example, $r = 0.40$ is a linear model. Many researchers are fixated on the number between 0 and 1 or between 0 and -1, and they tend to forget this background model. Often, they unconsciously use the linear model as a tacit obviousness. They do not seem to realize that nonlinearity occurs frequently.

An example of nonlinearity is the correlation between the percentage of Catholics and the percentage of CDU/CSU voters in former West Germany (Christlich-Demokratische Union/Christlich-Soziale

Union). One might expect a linear correlation: the more Catholics, the more voters of CDU/CSU according to a fixed pattern. However, this "the more, the more" pattern only seems to be valid for communes with many Catholics. For communes with few Catholics, the correlation turns out to be fairly negative: The more Catholics, the fewer voters for CDU/CSU. Consequently, the overall scatterplot displays a U pattern. At first, it drops, and from a certain percentage of Catholics onwards, it starts to rise. The quadratic function that describes a parabola therefore shows a better fit with the scatterplot than the linear function.

Many other nonlinear functions describe reality, whether exponential, logistic, discontinuous, or other. The exponential function, for example, was used in the reports by the Club of Rome, in which increased environmental pollution (correlated with time) was described in the 1970s. The logistic function was used by those who reacted against the Club of Rome, with the objection that the unchecked pollution would reach a saturation level.

2. The Direction of the Correlation

When dealing with two variables, X_1 and Y, one of them is the INDEPENDENT VARIABLE and the other is the DEPENDENT VARIABLE. In the case of income (X_1) and consumption (Y) the direction $X_1 \rightarrow Y$ is obvious. Here, a distinction between the correlation coefficient and the regression coefficient can already be elucidated. The regression coefficient indicates b_{y1} for the direction $X_1 \rightarrow Y$ and b_{1y} for the direction $Y \rightarrow X_1$. On the other hand, the correlation coefficient, which can be calculated as the geometrical mean of b_{y1} and b_{1y} (= the square root of the product of these coefficients) incorporates both directions and therefore is a nondirected symmetrical coefficient.

A novice might conclude that causal analysis is the privilege of regression coefficients and that correlation coefficients are unsuitable for that purpose. This would be an understandable but premature conclusion. Indeed, the direction in which we calculate mathematically is not the same as the causality direction. Each function $Y = aX + b$, with Y as a dependent variable, is convertible into its inverted function $X = Y/a - b/a$, with X as a dependent variable. The direction here is of relative value. However, when a stone is thrown into a pond, the expanding ripples are explained by the falling stone, but the opposite process with shrinking ripples would imply an enormous number of ripple generators, the coherence of which would need to be explained. It would also be imperative to demonstrate that they come from one central point (the improbability of an implosion). The direction here is not of relative value.

As was said before, causality cannot be equated to the statistical computation of a regression coefficient. True, it is fortunate that this coefficient represents a direction, and Hubert Blalock (1972) gratefully used this in his CAUSAL MODEL approach. However, this direction in terms of causality is hopelessly relative in statistical computation. This is admitted with perfect honesty by Sewell Wright (1934) in his article "The Method of Path Coefficients" and in earlier articles, when he claimed to have found a method to link the (a priori) causal knowledge with the (a posteriori) empiric correlations in the correlation matrix.

3. The Sign of the Correlation

A correlation can be positive (the more of X_1, the more of Y) or negative (the more of X_1, the less of Y). In the case of income and consumption, there is a positive correlation (the sign is +): The higher the income, the higher the consumption.

An example of negative correlation is the relationship between social class and number of children at the beginning of the past century. Higher social classes displayed the mechanism that was known as *social capillarity*. This was strikingly described by Friedrich Engels in *Der Ursprung der Familie, des Privateigentums und des Staates*. As soon as a son was born in the higher social classes, the parents started applying birth control because the estate could be passed on from father to son. In the lower social classes, however, this mechanism was lacking. A family of 10 children was the rule, and sometimes there were as many as 20 children. The sign here is abundantly clear—the higher the social class, the lower the number of children.

4. The Strength of the Correlation

A correlation can be weak or strong. A correlation coefficient of 0.90 or −0.90 is strong. A coefficient of 0.10 or −0.10 is weak. A coefficient 0 signals lack of correlation. A coefficient 1 or −1 indicates maximal deterministic correlation.

It is convenient to keep the vector model in mind. The vector model is the coordinate system in which the roles of variables and units have been changed:

Units are now the axes, and variables are the points forming the scatterplot. These variables are conceived as the end points of vectors starting in the origin—hence the name. In such a vector model, the correlation coefficient between two variables is equal to the cosine of the angle formed by the two vectors. We know that the cosine of 90° equals 0: The vectors then are at perpendicular angles to each other, and the correlation is 0. On the other hand, the cosine of 0° equals 1: The vectors then coincide and are optimally correlated.

Mostly, a strong correlation coefficient (between two variables, not in a multivariate context) will coincide with a strong regression coefficient, but this is not always the case. Many situations are possible, and this is related to the dispersions of the variables, which will be dealt with below.

In addition, a strong correlation will tend to be more significant than a weak correlation, but it does happen that a strong correlation is not significant. Moreover, a significant correlation can be weak. It follows that the strength of a correlation (correlation or regression coefficient) cannot be judged as such because it should be studied in combination with other criteria, the first one being the significance, which is discussed next.

5. The Statistical Potential for Generalization

A correlation calculated on the basis of a random sample may or may not be significant (i.e., it can or cannot be statistically generalized with respect to the population from which the sample was taken). Here we have to go against the postmodern range of ideas and take a traditional rigorous point of view: Only those correlations that can be statistically generalized (and are therefore statistically significant) are appropriate for interpretation. Nonsignificant correlations are the result of mere randomness and should therefore be considered nonexistent for the population.

Consequently, a strong coefficient (e.g., $r = 0.90$), which is not significant because the sample size is much too small (e.g., $n = 5$), does not stand the test and will be rejected by any social science researcher. A discerning mind will object that a weak correlation (e.g., $r = 0.10$) in a sufficiently large sample (e.g., sample size $n = 5{,}000$) is still significant and that this is only due to the size of the sample. Take a sufficiently large sample and you can prove anything with statistics, so it would appear. However, this argument is only partly correct. When dealing with large numbers, weak correlations or small changes such as changes in fertility can be extremely significant (in the sense of important) for policymaking. Any demographer knows how irreversible such processes are. Besides, for large numbers, it is also still possible that the correlation is not statistically significant. Therefore, it does not hold to pretend that the result is only due to the size of the sample.

Admittedly, there is still a grey area between two extremes: a significant result for a small sample and a nonsignificant one for a large sample. In the latter case, we know with certainty that the correlation is not significant. In the former case, we are certain that a larger sample survey will also yield a significant result. However, in general and certainly in the grey area, nonstatistical criteria, such as social relevance, will always contribute to the interpretation of correlation results.

So far, we have looked at some properties of correlation: nature, direction, sign, strength, and significance. As was already mentioned, one should not be blinded here by the number representing the correlation coefficient; indeed, all these features need to be taken into account. In this sense, major META-ANALYSES in which correlation coefficients from different studies are all mixed together should not be trusted. Indeed, in each separate study, one should raise many questions. Are we dealing with a linear correlation? If so, is the regression coefficient different from the correlation coefficient, and how large is the intercept? Are the variables standardized? Is the correlation significant? How large is the confidence interval? What is the probability of TYPE I ERROR α? What is the size of the sample? What is the probability of TYPE II ERROR β and the power?

THE DISPERSION OF THE VARIABLES: STANDARDIZED COEFFICIENTS?

The question about the difference between correlation and regression effect actually also refers to the DISPERSION of the variables. Indeed, we have already mentioned that a strong correlation coefficient does not always coincide with a strong regression coefficient and that this is linked to the dispersion of the variables. This can be derived from the formulas of the coefficients: r_{y1} (or r_{1y}) is the covariance divided by the geometrical mean of the variances, b_{y1} is the covariance divided by the variance of X_1, and b_{1y} is the covariance divided by the variance of Y. In the case of equal dispersions, all three coefficients will be equal. A greater dispersion of Y will have a decreasing effect on the correlation coefficient and on b_{1y}, but not on b_{y1}.

With respect to this, one may wonder if the standardized coefficients (betas) should be preferred. (Standardization is a procedure of subtracting the mean and dividing by the STANDARD DEVIATION.) Dividing by the standard deviation happens to be a rather drastic operation, which causes the variance of the variables to be artificially set equal to 1, with a view to obtaining comparability. But some have argued that a similar difference in education has a different meaning in a society with a great deal of educational inequality compared to a society with a small degree of such inequality, and therefore they advocate nonstandardized coefficients.

All this is not just abstract thinking; it does occur in specific processes in society. Examples may include the processes of efficient birth control, underemployment, the raising of political public opinion, and environmental awareness. From all these examples, we learn that society introduces changes in the dispersion of its features.

CALCULATION OF THE CORRELATION COEFFICIENT

We would like to illustrate this importance of dispersions with a small example in which the correlation coefficient between education (X) and income (Y) is calculated. The formula is $r_{xy} = [\sum(X-\bar{X})(Y-\bar{Y})]/[\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2]^{1/2}$.

Let us take a nation in which few of its citizens go to school, and those who do happen to be educated have high incomes and the rest are poor. Let the data matrix be as below. Education is measured on a scale from 1 to 9, and income is measured in thousands of Euros.

X 1 1 1 2 2 3 3 3 4 4 6 6 6 7 7 8 8 8 9 9

Y 2 4 5 3 5 3 4 6 4 5 5 7 8 6 8 6 7 9 7 8

The scatterplot is shown in Figure 1.

The standard deviations are 2.81 for education and 1.93 for income.

The correlation coefficient is 0.798. The model is linear, the coefficient is symmetric, the relationship is positive, the strength is substantial, and the coefficient is statistically significant ($p < .001$).

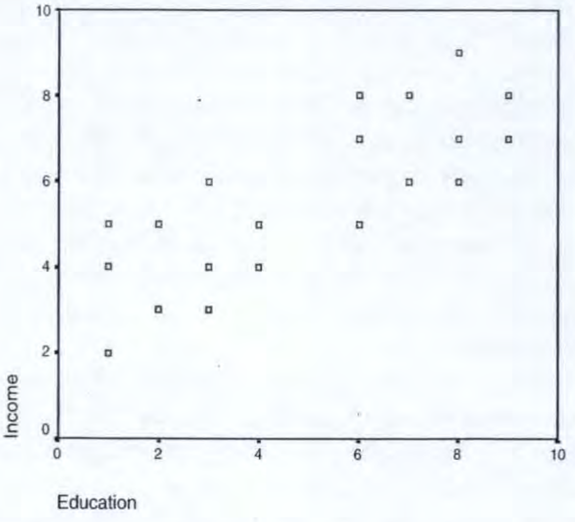

Figure 1

Now suppose that in this nation, the more ambitious among the poor demand that their children be admitted to the schools and that the state yields to this pressure and begins to subsidize education. Large numbers now go to school and college, average income rises as education raises the general level of human capability, and the variation of incomes declines. But once subsidized education becomes widespread, the individuals who are more schooled can no longer capture for themselves much of the benefit of that schooling, and the correlation between income and education diminishes.

Now the data matrix is the following.

X 6 6 6 6 6 6 7 7 7 7 8 8 8 8 8 8 9 9 9 9

Y 5 7 8 5 7 8 6 8 6 8 6 7 9 6 7 9 7 8 7 8

The scatterplot is shown in Figure 2.

The standard deviations are much smaller: 1.14 for education and 1.17 for income. The correlation coefficient has diminished from 0.798 to 0.285. The relationship is positive but small and no longer statistically significant.

One could also suppose that in this nation, the variation of education remains the same because only some of the citizens go to school, but the variation of incomes has declined because there is a lot of underemployment (people with high diplomas having low-salary jobs). This would result in the following data matrix:

X 1 1 1 2 2 3 3 3 4 4 6 6 6 7 7 8 8 8 9 9

Y 2 4 5 3 5 3 4 6 4 5 2 4 5 3 5 3 4 6 4 5

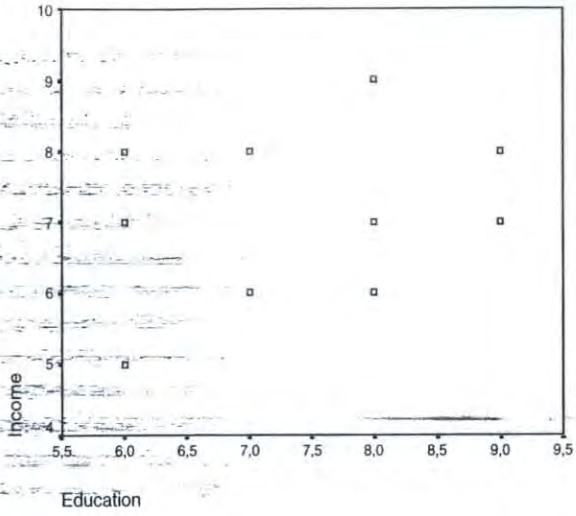

Figure 2

The scatterplot is now as follows, in Figure 3:

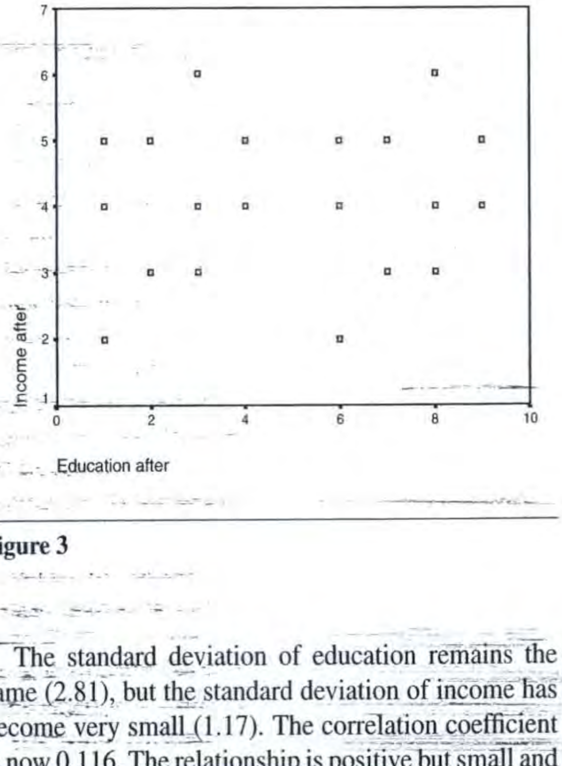

Figure 3

The standard deviation of education remains the same (2.81), but the standard deviation of income has become very small (1.17). The correlation coefficient is now 0.116. The relationship is positive but small and nonsignificant.

—Jacques Tacq

REFERENCES

Blalock, H. M. (1972). *Social statistics*. Tokyo: McGraw-Hill Kogakusha.

Brownlee, K. (1965). *Statistical theory and methodology in science and engineering*. New York: John Wiley.

Hays, W. L. (1972). *Statistics for the social sciences*. New York: Holt.

Kendall, M., & Stuart, A. (1969). *The advanced theory of statistics*. London: Griffin.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis*. London: Sage.

CORRESPONDENCE ANALYSIS

Correspondence analysis is a way of *seeing* the ASSOCIATION in a two-way cross-tabulation rather than *measuring* it. Consider, for example, the cross-tabulation in Table 1, taken from the 1994 International Social Survey Programme (ISSP) on Family and Changing Gender Roles. This table shows, for 24 different countries, how many respondents think that the ideal number of children in a family is 0, 1, 2, 3, 4, or 5 or more. Response percentages for each country are given in parentheses: For example, in Australia, 833 of the 1,529 respondents, or 54.5%, considered two children to be ideal.

The usual Pearson CHI-SQUARE statistic for this 24×6 table has the value 6,734, and CRAMÉR'S V is 0.206, both of which are highly significant. In fact, any MEASURE OF ASSOCIATION, even with a low value, will almost always be highly significant for a table with such high frequencies. Clearly, there are significant differences between countries, but what are these differences? We can see from the table that the United States and Canada are quite similar, but how can we compare their positions on this issue with the European countries? Correspondence analysis can answer this question because it aims to show the similarities and differences between the countries in a compact, graphical way.

The correspondence analysis of Table 1 is shown in Figure 1. This is a two-dimensional "map" of the positions of the row and column points, showing the main features of the contingency table. The points representing the columns form an arch pattern, starting from the bottom right and ending at the bottom left. This indicates a monotonic ordination of the countries in terms of their opinions on this issue, from those wanting fewer children (e.g., [former] East Germany, Bulgaria, and the Czech Republic) to those wanting more children (e.g., Ireland, Israel, and the Philippines). The horizontal axis, or first principal axis, captures this

Table 1 Ideal Number of Children in Family: Response Frequencies in 24 Countries (percentages in parentheses)

Country	*0*	*1*	*2*	*3*	*4*	*≥ 5*	*Total*
Australia	4 (0.3)	19 (1.2)	833 (54.5)	448 (29.3)	203 (13.3)	22 (1.4)	1,529
West Germany	18 (0.8)	116 (5.2)	1551 (69.5)	441 (19.8)	79 (3.6)	26 (1.2)	2,231
East Germany	6 (0.6)	108 (10.1)	831 (77.7)	115 (10.7)	10 (3.6)	0 (0.0)	1,070
Great Britain	5 (0.6)	16 (1.8)	679 (74.9)	145 (16.0)	55 (6.1)	6 (0.7)	906
Northern Ireland	0 (0.0)	7 (1.2)	271 (45.5)	157 (26.4)	142 (23.9)	18 (3.0)	595
United States	11 (0.8)	34 (2.6)	783 (60.0)	295 (22.6)	159 (12.2)	22 (1.7)	1,304
Austria	5 (0.5)	51 (5.4)	667 (70.4)	195 (20.6)	28 (3.0)	2 (0.2)	948
Hungary	8 (0.6)	70 (4.8)	839 (57.7)	485 (33.4)	41 (2.8)	11 (0.8)	1,454
Italy	4 (0.4)	46 (4.5)	678 (67.1)	254 (25.1)	22 (2.2)	7 (0.7)	1,011
Ireland	2 (0.2)	7 (0.8)	276 (31.3)	278 (31.5)	270 (30.6)	49 (5.6)	882
Netherlands	41 (2.1)	42 (2.1)	1050 (53.4)	575 (29.2)	199 (10.1)	61 (3.1)	1,968
Norway	3 (0.2)	10 (0.5)	939 (48.2)	824 (42.3)	145 (7.4)	29 (1.5)	1,950
Sweden	0 (0.0)	8 (0.7)	741 (63.9)	329 (28.4)	65 (5.6)	16 (1.4)	1,159
Czech Republic	7 (0.7)	109 (10.7)	688 (67.3)	192 (18.8)	16 (1.6)	11 (1.1)	1,023
Slovenia	9 (0.9)	40 (3.9)	601 (59.2)	311 (30.6)	45 (4.4)	9 (0.9)	1,015
Poland	0 (0.0)	27 (1.9)	756 (54.1)	483 (34.6)	99 (7.1)	32 (2.3)	1,397
Bulgaria	55 (4.9)	80 (7.1)	742 (65.9)	216 (19.2)	29 (2.6)	4 (0.4)	1,126
Russia	0 (0.0)	188 (9.6)	1,219 (62.5)	463 (23.7)	50 (2.6)	30 (1.5)	1,950
New Zealand	8 (0.9)	11 (1.2)	497 (55.7)	258 (28.9)	106 (11.9)	12 (2.3)	892
Canada	5 (0.4)	25 (2.2)	674 (59.3)	308 (27.1)	107 (9.4)	17 (1.5)	1,136
Philippines	0 (0.0)	10 (0.8)	229 (19.1)	561 (46.8)	282 (23.5)	117 (9.8)	1,199
Israel	1 (0.1)	9 (0.7)	204 (16.7)	525 (42.9)	384 (31.4)	100 (8.2)	1,223
Japan	0 (0.0)	9 (0.7)	483 (37.2)	737 (56.7)	56 (4.3)	15 (1.2)	1,300
Spain	0 (0.0)	140 (6.0)	1,441 (61.7)	593 (25.4)	108 (4.6)	55 (2.4)	2,337
Total	192 (0.6)	1,182 (3.7)	17,672 (55.9)	9,188 (29.1)	2,700 (8.5)	671 (2.1)	31,605

NOTE: The ISSP data on which Table 1 is based were made available by the *Zentralarchiv für Empirische Sozialforschung* in Cologne (www.gesis.org/en/data_service/issp/index.htm).

general scaling of the countries, whereas the vertical second principal axis contrasts the extreme categories with the middle categories. Such an arch pattern is frequently encountered in correspondence analysis maps. Most of the highly developed countries are around the center of the map, which represents the average, but there are interesting differences among them. The United States, for example, lies lower down on the second axis, mostly because of its relatively high frequencies of "4 children" compared to other countries around the center. Japan, by contrast, lies high up on the second axis because of its very high frequency of "3 children."

Each country finds its position in the map according to its *profile,* or set of relative frequencies, shown as percentages in Table 1. The center of the map represents the *average profile*—that is, 0.6%, 3.7%, 55.9%, 29.1%, 8.5%, and 2.1% for all the countries (see last row of Table 1)—and a particular country will be attracted in the direction of a category if its percentage is higher than the corresponding average value. The method can be equivalently defined as a variant of METRIC VARIABLE multidimensional scaling, in which distances between countries are measured by the so-called chi-square distances (Benzécri, 1973), and each country is weighted proportional to its respective sample size.

The main issue in constructing the map is to decide on which axis scaling to choose from several different options. Figure 1 is called a *symmetric map* because it represents the rows and columns in an identical way using their respective *principal coordinates,* so that each set of points has the same dispersion in the map. An *asymmetric map* is also possible when each country is located at the weighted average position of the response category points, with the latter being represented in *standard coordinates* (Greenacre, 1993). The VARIANCE in the data table is measured by the *inertia,* equal to the Pearson chi-square statistic divided by the grand total of the table. The quality of the display is measured by the percentages of inertia accounted for by each principal axis, similar to PRINCIPAL COMPONENTS

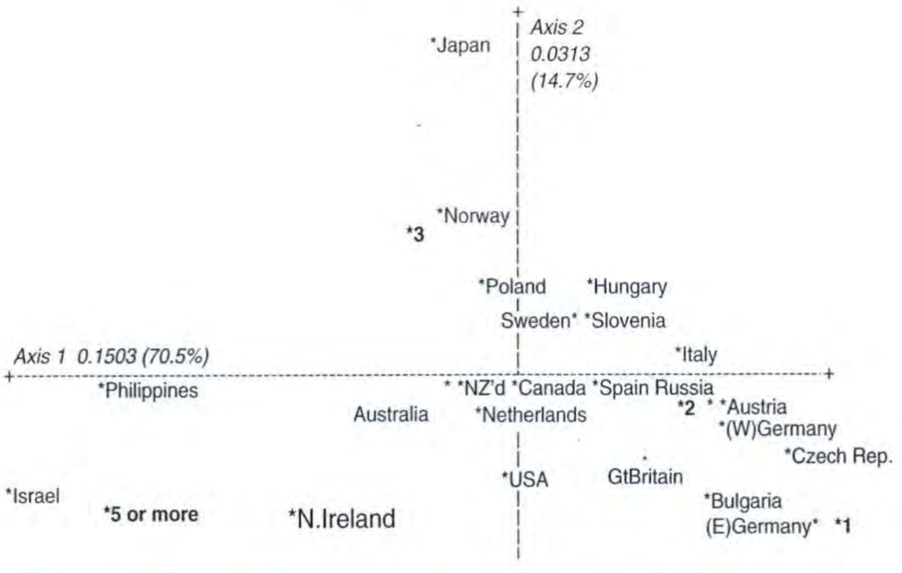

Figure 1 Correspondence Analysis of Table 1: Symmetric Map in Two Dimensions

ANALYSIS. In Figure 1, the two axes explain a combined $70.5 + 14.7 = 85.2\%$ of the inertia, leaving 14.8% along further principal axes.

Although correspondence analysis applies primarily to CONTINGENCY TABLES, the method can also be used to visually interpret patterns in multivariate categorical data (multiple correspondence analysis), preferences, rankings, and ratings. This is achieved by appropriate transformation of the original data prior to applying the correspondence analysis procedure.

—Michael Greenacre

REFERENCES

Benzécri, J.-P. (1973). *Analyse des Données. Tome 2: Analyse des Correspondances* [Data analysis: Vol. 2. Correspondence analysis]. Paris: Dunod.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Greenacre, M. J. (1993). *Correspondence analysis in practice.* London: Academic Press.

Greenacre, M. J., & Blasius, J. (1994). *Correspondence analysis in the social sciences.* London: Academic Press.

COUNTERBALANCING

Participants are exposed to all treatment conditions in a REPEATED-MEASURES DESIGN, which leads to concerns about CARRYOVER EFFECTS and ORDER EFFECTS. Counterbalancing refers to exposing participants to different orders of the treatment conditions to ensure that such carryover effects and order effects fall equally on the experimental conditions.

Counterbalancing can be either complete or incomplete. With complete counterbalancing, every possible order of treatment is used equally often. Thus, with k treatments, there would be $k!$ different orders. As seen below, for three treatment conditions (A, B, and C), each treatment would occur equally often in each position. Moreover, each treatment would be preceded by every other treatment equally often (see Table 1).

Table 1

	1st Position	*2nd Position*	*3rd Position*
Order 1	A	B	C
Order 2	A	C	B
Order 3	B	A	C
Order 4	B	C	A
Order 5	C	A	B
Order 6	C	B	A

The advantage of counterbalancing is best understood by considering what might happen in the above experiment without counterbalancing. Suppose that all participants receive the treatments in the same order (A, B, C) and that mean performance under C is found to be significantly better than mean performance under A. Such an outcome cannot be interpreted

Table 2

Order	*P1*	*P2*	*P3*	*P4*	*P5*	*Order*	*P1*	*P2*	*P3*	*P4*	*P5*
1	A	B	E	C	D	6	D	C	E	B	A
2	B	C	A	D	E	7	E	D	A	C	B
3	C	D	B	E	A	8	A	E	B	D	C
4	D	E	C	A	B	9	B	A	C	E	D
5	E	A	D	B	C	10	C	B	D	A	E

unambiguously because of the CONFOUNDING of treatment and position. That is, the difference may be due to the treatment (A vs. C) *or* due to a practice effect (first position vs. third position), in which a participant's scores will improve over time regardless of treatment.

Counterbalancing the order of exposure to the treatment will eliminate the confounding of treatment and order. However, doing so will also inflate the error term in an analysis because any systematic order effects will increase variability of scores within each treatment condition. Order effects may be estimated and removed from the error term, as detailed elsewhere (cf. Keppel, 1991, pp. 361–373).

Complete counterbalancing is ideal because it ensures not only that each treatment occurs equally often in each position but also that each treatment occurs equally often in *all* of the preceding positions (P1, P2, etc.). However, with larger numbers of treatments, complete counterbalancing may be impractical. Under those circumstances, incomplete counterbalancing should be used.

One useful incomplete counterbalancing approach is referred to as a digram-balanced LATIN SQUARE. With this approach, when the number of treatments (k) is even, then only k orders are needed. When the number of treatments is odd, then $2 \times k$ orders are needed.

Each order is constructed in the following way. Lay out the treatment conditions in a row (ABCDE). The first order would be first, second, last, third, next to last, fourth, and so forth. To create the second order, first rearrange your row of conditions by moving the first treatment to the last position, with all other conditions remaining in the same order (BCDEA). Then select your order in the same way as before (first, second, last, etc.). Continue to rearrange and select in the same fashion until you have obtained k orders. When k is even, you would be finished when you had determined k orders.

The five orders on the left would result from this procedure for $k = 5$ (see Table 2). Every treatment occurs in every position equally often, which is important. However, because k is odd, treatments are not preceded by every other treatment. For example, as seen in Table 2 for Orders 1 through 5, B is never preceded by C or E. For that reason, when k is odd, it is essential that each order be mirrored, as seen in Orders 6 through 10. As you can see with this example, the digram-balanced approach ensures that every condition occurs in every position equally often and that every condition is preceded by every other condition equally often.

—Hugh J. Foley

REFERENCES

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Maxwell, S. E., & Delaney, H. D. (2000). *Designing experiments and analyzing data: A model comparison perspective.* Mahwah, NJ: Lawrence Erlbaum.

COUNTERFACTUAL

A counterfactual assertion is a conditional whose antecedent is false and whose consequent describes how the world *would have been* if the antecedent had obtained. The counterfactual takes the form of a subjunctive conditional: "If P had obtained, then Q would have obtained." In understanding and assessing such a statement, we are asked to consider how the world would have been if the antecedent condition had obtained. For example, "If the wind had not reached 50 miles per hour, the bridge would not have collapsed" or "If the Security Council had acted, the war would have been averted." We can ask two types of questions about counterfactual conditionals: What is the meaning of the statement, and how do we determine whether it is true or false? A counterfactual conditional cannot be evaluated as a truth-functional conditional because a truth-functional conditional with false antecedent is ipso facto true (i.e., "if P then Q" is equivalent to "either not P or Q"). So is it necessary to provide a logical analysis of the truth conditions of counterfactuals if they are to be useful in rigorous thought.

There is a close relationship between counterfactual reasoning and causal reasoning. If we assert that "P caused Q (in the circumstances C_i)," it is implied that we would assert the following: "If P had not occurred (in circumstances C_i), then Q would not have occurred." So a causal judgment implies a set of counterfactual judgments. Symmetrically, a counterfactual judgment is commonly supported by reference to one or more causal processes that would have conveyed the world from the situation of the antecedent to the consequent. When we judge that the Tacoma Narrows Bridge would not have collapsed had the wind not reached 50 miles per hour, we rely on a causal theory of the structural dynamics of the bridge and the effects of the wind in reaching the consequent.

How do we assign a truth value to a counterfactual statement? The most systematic answer is to appeal to causal relations and causal laws. If we believe that we have a true causal analysis of the occurrence of Q, and if P is a necessary part of the set of sufficient conditions that bring Q to pass, then we can argue that, had P occurred, Q would have occurred. David Lewis (1973) analyzes the truth conditions and logic of counterfactuals in terms of possible worlds (possible world semantics). A counterfactual is interpreted as a statement about how things occur in other possible worlds governed by the same laws of nature. Roughly, "in every possible world that is relevantly similar to the existing world but in which the wind does not reach 50 miles per hour, the bridge does not collapse." What constitutes "relevant similarity" between worlds is explained in terms of "being subject to the same laws of nature." On this approach, we understand the counterfactual "If P had occurred, Q would have occurred" as a statement along these lines: "P & {laws of nature} entail Q." This construction introduces a notion of "physical necessity" to the rendering of counterfactuals: Given P, it is physically necessary that Q occurs.

—Daniel Little

REFERENCE

Lewis, D. K. (1973). *Counterfactuals.* Cambridge, MA: Harvard University Press.

COVARIANCE

Covariance represents the fundamental measure of association between two random variables. Here we will discuss the concept of covariance in terms of the bivariate NORMAL DISTRIBUTION.

To begin, consider two random variables X and Y, representing, for example, a measure of reading proficiency and a measure of mathematics proficiency. Assume also that the two variables follow a bivariate normal distribution. Define $E(X)$ and $E(Y)$ to be the expectations of the random variables X and Y. Then, the covariance of X and Y, denoted as $\mathrm{Cov}(X, Y)$ is

$$\mathrm{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)], \qquad (1)$$

where $\mu_x = E(X)$ and $\mu_y = E(Y)$. A computation formula for the unbiased estimator of the population covariance can be written as

$$\frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{N - 1}, \qquad (2)$$

where N is the sample size, $\bar{X}$ is the sample mean of X, and $\bar{Y}$ is the sample mean of Y.

Returning to equation (1), the means of the bivariate normal distribution are contained in the mean vector μ, and the covariance of X and Y is contained in the *covariance matrix* Σ. The covariance matrix is a symmetric matrix, with the variances of X and Y on the diagonal and the covariance of X and Y on the off-diagonal.

The covariance represents the linear relationship between the random variables X and Y; positive values of the covariance are likely to occur when $X - \mu_x$ and $Y - \mu_Y$ are of the same sign, and negative values of the covariance are likely to occur when $X - \mu_x$ and $Y - \mu_Y$ are of opposite sign. However, regardless of the sign of the covariance, its value is meaningless because it depends on the variances and hence on the metrics of the random variables themselves. To mitigate this problem, the CORRELATION coefficient standardizes the covariance to a mean of zero and a variance of 1 and hence puts the random variables on a common metric. This yields a measure of linear association that ranges from –1 to +1, with zero representing the absence of a linear relationship.

The covariance is a statistic and hence has a SAMPLING DISTRIBUTION. Imagine repeated draws of the two random variables from a population characterized by a bivariate normal distribution. For each draw, we can calculate the means, variances, and the covariance of the two variables. The sampling distribution of the means is known to be normally distributed, and the sampling distributions of the variances are known to be CHI-SQUARE distributed. Concentrating on the covariance matrix, the sampling distribution of the covariance matrix is referred to as the *Wishart distribution.* The Wishart distribution is the multivariate analog of the chi-square distribution (see, e.g., Timm, 1975). It is interesting to note that the MAXIMUM LIKELIHOOD fitting function used for procedures such as FACTOR ANALYSIS and STRUCTURAL EQUATION MODELING can be derived from the Wishart distribution.

Having defined the covariance between two random variables and its sampling distribution, it is instructive to consider its use in applied settings. As noted earlier, the covariance provides a measure of linear association between two random variables. Also as noted, it is more common to transform the covariance into a correlation coefficient for ease of interpretation. Nevertheless, the covariance—particularly the covariance matrix—is a very important measure for other procedures. A large number of multivariate statistical procedures have been defined based on expressing the covariances in terms of more fundamental parameters. Such procedures as REGRESSION analysis, PATH ANALYSIS, factor analysis, and structural equation modeling can be considered as models for the covariances of the variables.

—David W. Kaplan

REFERENCE

Timm, N. H. (1975). *Multivariate analysis with applications in education and psychology.* Monterey, CA: Brooks/Cole.

COVARIANCE STRUCTURE

The term *covariance structure* is used to designate a large class of statistical models that seek to represent the variances and covariances of a set of variables in terms of a smaller number of parameters. The premise here is that the sample variances and covariances are sufficient for the estimation of model parameters. The term *mean structure* is used to denote a class of models that represent the means of the sample data in terms of a smaller number of parameters. The general problem of covariance structure modeling has been taken up in Browne (1982).

Models that are subsumed under the term *covariance structure* include but are not limited to regression analysis, path analysis, factor analysis, and structural equation modeling. The general form of the model can be written as

$$\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{\Theta}), \tag{1}$$

where $\mathbf{\Sigma}$ is a $p \times p$ population covariance matrix with p being the number of variables, and $\mathbf{\Theta}$ is a q-dimensional vector of model parameters with $q \leq p$. The function in equation (1) states that the covariance matrix can be expressed as a matrix-valued function of a parameter vector.

As stated above, the covariance structure model in equation (1) subsumes a number of commonly used statistical procedures. What is required is that the parameters of the model can be expressed in terms of elements of the covariance matrix elements. Consider the simple example of a two-variable linear regression model. Denote the model as

$$y = x\beta + u, \tag{2}$$

where y and x are the outcome and regressor variables, respectively; β is the regression coefficient; and u is the regression disturbance term. Equation (2) can be seen as a covariance structure when the regression coefficient β can be estimated as

$$\hat{\beta} = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}, \tag{3}$$

where $\mathrm{Cov}(x, y)$ is the covariance of x and y, and $\mathrm{Var}(x)$ is the variance of x. Similarly, the variance of the disturbance term u (denoted as σ_u^2) can be solved in terms of elements of $\mathbf{\Sigma}$. Thus, we see that regression analysis is a special case of covariance structure modeling.

To take another example of the generality of the covariance structure representation, consider the factor analysis model, written in linear form as

$$\mathbf{y} = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varsigma}, \tag{4}$$

where $\mathbf{y}$ is a $p \times 1$ vector of variables, $\mathbf{\Lambda}$ is a $p \times k$ matrix of factor loadings, $\boldsymbol{\eta}$ is a $k \times 1$ vector of common factors, and $\boldsymbol{\varsigma}$ is a $p \times 1$ vector of uniquenesses (composed of measurement error and specific error). Assuming that the mean of the factors and the uniquenesses are zero and that the covariance of the factors and the uniquenesses are zero, the covariance matrix for $\mathbf{y}$ can be expressed as

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}, \tag{5}$$

where $\boldsymbol{\Phi}$ is the matrix of covariances among the factors, and $\boldsymbol{\Psi}$ is a diagonal matrix of unique variances. It can be seen from equation (5) that the covariance matrix is expressed in terms of the fundamental parameters of the factor analysis model and hence is, by definition, a covariance structure.

Finally, it may be of interest to see how structural equation modeling fits into the general covariance structure framework. The system of structural equations can be written as

$$\mathbf{y} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\varsigma}, \tag{6}$$

where $\mathbf{y}$ is a $p \times 1$ vector of observed endogenous variables, $\mathbf{x}$ is a $q \times 1$ vector of observed exogenous variables, $\boldsymbol{\alpha}$ is a $p \times 1$ vector of structural intercepts, $\mathbf{B}$ is a $p \times p$ coefficient matrix that relates endogenous variables to each other, $\boldsymbol{\Gamma}$ is a $p \times q$ coefficient matrix that relates endogenous variables to exogenous variables, and $\boldsymbol{\varsigma}$ is a $p \times 1$ vector of disturbance terms where $\mathrm{Var}(\boldsymbol{\varsigma}) = \boldsymbol{\Psi}$ is the $p \times p$ covariance matrix of the disturbance terms, where $\mathrm{Var}(\cdot)$ is the variance operator. Finally, let $\mathrm{Var}(\mathbf{x}) = \boldsymbol{\Phi}$ be the $q \times q$ covariance matrix for the exogenous variables.

The system described in equation (6) is referred to as the *structural form* of the model. It is convenient to rewrite the structural model so that the endogenous variables are on one side of the equation and the exogenous variables on the other side. Thus, equation (6) can be rewritten as

$$(\mathbf{I} - \mathbf{B})\mathbf{y} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\varsigma}. \tag{7}$$

Assuming that $(\mathbf{I} - \mathbf{B})$ is nonsingular so that its inverse exists, equation (7) can be written as

$$\begin{aligned}\mathbf{y} &= (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{x} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\varsigma} \\ &= \boldsymbol{\Pi}_0 + \boldsymbol{\Pi}_1\mathbf{x} + \boldsymbol{\varsigma}^*,\end{aligned} \tag{8}$$

where $\boldsymbol{\Pi}_0$ is the vector of reduced-form intercepts, $\boldsymbol{\Pi}_1$ is the vector of reduced-form slopes, and $\boldsymbol{\varsigma}^*$ is the vector of reduced-form disturbances with $\mathrm{Var}(\boldsymbol{\varsigma}^*) = \boldsymbol{\Psi}^*$. This specification is referred to as the *reduced form* of the model. Note that equation (8) is a straightforward multivariate regression of $\mathbf{y}$ on $\mathbf{x}$, and so this model also falls into the general form of a covariance structure.

Next, let $\boldsymbol{\Theta}$ be a vector that contains the structural parameters of the model—so in this case, $\boldsymbol{\Theta} = (\mathbf{B}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\Phi})$. Furthermore, let $E(\mathbf{x}) = \boldsymbol{\mu}_{\mathbf{x}}$ be the vector of means for $\mathbf{x}$, $\mathrm{Var}(\mathbf{x}) = E(\mathbf{x}'\mathbf{x}) = \boldsymbol{\Phi}$, and $E(\boldsymbol{\varsigma}) = \mathbf{0}$, where $E(\cdot)$ is the expectation operator. Then,

$$\begin{aligned}E(\mathbf{y}) &= (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}E(\mathbf{x}) \\ &= (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\mu}_{\mathbf{x}}\end{aligned} \tag{9}$$

and

$$\begin{aligned}E(\mathbf{y}, \mathbf{x}) = \boldsymbol{\Sigma}_{\mathbf{yx}'} &= \begin{bmatrix} E(\mathbf{yy}') & E(\mathbf{yx}') \\ E(\mathbf{x}'\mathbf{y}) & E(\mathbf{x}'\mathbf{x}) \end{bmatrix}, \\ &= \begin{bmatrix} (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})(\mathbf{I} - \mathbf{B})'^{-1} & (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}'(\mathbf{I} - \mathbf{B})'^{-1} & \boldsymbol{\Phi} \end{bmatrix}.\end{aligned} \tag{10}$$

Thus, equations (9) and (10) show that structural equation modeling represents a structure on the mean vector and covariance matrix. The structure is in terms of the parameters of the model. Covariance structure modeling concerns the elements of equation (10).

It is clear from the above discussion that the covariance structure representation of statistical modeling has wide generality. The covariance structure representation also provides a convenient way to consider problems in model identification. Although model identification is taken up elsewhere in this encyclopedia, a necessary condition for the identification of model parameters is that the number of model parameters be less than or equal to the number of elements in the covariance matrix. Specifically, arrange the unknown parameters of the model in the vector $\boldsymbol{\Theta}$. Consider next a population covariance matrix $\boldsymbol{\Sigma}$ whose elements are population variances and covariances. It is assumed that an underlying, substantive model can be specified to explain the variances and covariances in $\boldsymbol{\Sigma}$. We know that the variances and covariances in $\boldsymbol{\Sigma}$ can be estimated by their sample counterparts in the sample covariance matrix $\mathbf{S}$ using straightforward formulas for the calculation of sample variances and covariances. Thus, the parameters in $\boldsymbol{\Sigma}$ are identified.

Having established that the elements in $\boldsymbol{\Sigma}$ are identified from the sample counterparts, what we need to determine is whether the model parameters are identified. We say that the elements in $\boldsymbol{\Omega}$ are *identified* if they can be expressed uniquely in terms of the elements of the covariance matrix Σ. If all elements in Θ are identified, we say that the model is identified.

—David W. Kaplan

REFERENCE

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, UK: Cambridge University Press.

COVARIATE

In statistical analysis, the term *covariate* generally refers to the INDEPENDENT VARIABLE. For example, in the analysis of EXPERIMENTS, a covariate is an independent variable not manipulated by the experimenter but still affecting the response variable. In SURVIVAL or EVENT HISTORY ANALYSIS, a covariate is simply an independent variable, continuous or not, that may have an effect on the HAZARD RATE, such as a time-varying (or time-dependent) covariate. However, in LOG-LINEAR MODELING, continuous variables are treated as "cell covariates" by imposing a metric on categorical variables, instead of allowing use of truly continuous variables.

—Tim Futing Liao

COVER STORY

In some laboratory research, the particular hypotheses under investigation are relatively obvious. This transparency of research purpose may prove problematic if participants are inclined to assist the researcher by helping to confirm the apparent hypotheses. For example, suppose participants were asked to complete a questionnaire that asked a series of questions regarding their feelings toward citizens of Uzbekistan. Following this task, participants are shown a video of heroic Uzbekistanis defending their homeland against terrorists. Then, the attitude questionnaire is readministered. In this case, the goal of the research appears obvious. Research on DEMAND CHARACTERISTICS suggests that the apparently obvious aim of the study may produce artifactual results. Some participants will have determined that the purpose of the research is to induce them to indicate positive evaluations of Uzbekistanis and will cooperate, even if the video (the INDEPENDENT VARIABLE) had little real effect on their feelings.

To offset the possibility of distorted results brought on by overly cooperative participants, psychologists have developed more or less elaborate cover stories designed to hide the true purpose of their research. In the example, the pretest might be omitted, and the participants are informed that their task is to judge the technical quality of the video they are to see. Then, after completing this task, they might be administered a posttest that taps their feelings toward the Uzbekistanis. As an addition to the cover story, this task might be introduced in a way that also hides the true nature of the research. These participants' responses could then be compared in a *posttest-only/comparison group design* with those of participants who saw no video or who saw a different video (Crano & Brewer, 2002).

Cover stories are used to disguise the true purpose of research and hence are deceptive. Thus, their ethical use must be considered carefully. We justify cover stories because they are used not principally to deceive but to ensure the validity of research findings at little cost to participants. Typically, they are used in LABORATORY EXPERIMENTS, where the danger of misleading findings due to overly cooperative participants is most severe.

Crano and Brewer (2002) distinguished among three general classes of participants in research: voluntary, involuntary, and nonvoluntary. Nonvoluntary participants, by definition, are unaware they are the subject of scientific investigation, so there is no need for a cover story when using such participants. Voluntary and involuntary participants, however, are well aware of their status. Voluntary participants are content to serve in research, and their positive attitudes may prove dangerous to validity if they motivate attempts to help the investigator confirm research hypotheses. Involuntary respondents also are aware of their role but are unhappy about it, often feeling coerced into participating. These participants' negative attitudes may evolve into attempts to ruin the research; in such cases, a cover story may prove useful because it helps obscure the true intent of the research.

—William D. Crano

REFERENCE

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

COVERT RESEARCH

Covert research is that which proceeds without giving subjects or their legitimate representatives the basic information that they are being researched. All the professional associations relating to social research reflect in their ethical guidelines the principle of INFORMED CONSENT, and they advise uniformly against covert methods.

ETHICAL ISSUES

There are compelling reasons why social researchers should want to avoid covert practice.

First and foremost, there is an issue of autonomy. Subjects are accorded a right to be informed, to control their own observed behavior, and to make their own decisions on the risks or benefits of participation. Moreover, the data they supply have a value, and they may be entitled to put a price on what researchers take from them; researchers who do not inform subjects that they are yielding these data may be thought to be plundering or stealing them.

Furthermore, the use of methods that are not open will get researchers a bad name. At one time, Henle and Hubble (1938) stowed themselves under the beds in student dorms to record conversations; later, Laud Humphreys (1975) posed as a lookout or "watchqueen" to observe homosexual encounters in public toilets. Nowadays, those who research human behavior resist the reputation of eavesdropper or voyeur. The professional associations in the social and behavioral sciences claim the moral high ground and prefer to leave covert research to investigative journalists. They recognize also that behavior may be private, even though the territory in which it is observable may be public.

JUSTIFICATIONS

The consensus of professional associations and codes notwithstanding, there persists a debate concerning the acceptability of covert methods (Bulmer, 1982; Homan, 1992). It is not realistic to insist that all subjects should know that they are being observed and be given the option to withdraw. Some of the most perceptive work in the social sciences has taken the form of retrospective analysis of everyday experience. Erving Goffman interpreted the putting down of towels and newspapers on beaches and trains as territorial behavior: No one would ask that consent be sought or that a notice should be posted saying, "Caution: this beach is being observed by a sociologist." Once possessed of what Wright Mills (1959) called the "sociological imagination," students of human behavior become habitual researchers: The act of research is not the collection of data but the subsequent construction of them, against which there is no law.

The right to know that research is taking place and therefore to decline or avoid it is itself problematic. It is, for example, accorded more to the living than to the dead whose papers turn up in county record offices. The right to refuse research may run contrary to the public interest. For example, it is arguable that social researchers should complement the investigations of journalists and police to expose places where children or other vulnerable groups are being abused or where racist attacks are being sponsored. Indeed, the intellectual freedom and integrity of social researchers are compromised if they are prepared to take "no" for an answer. The pity is that subjects in positions of power are more prone to insulate themselves from open investigators by withholding consent than are the relatively defenseless and disempowered; in consequence, the ethic of openness discriminates against the weak.

Covert methods are, in some circumstances, more likely to be nonreactive. This is true in two senses. First, because subjects will not know that they are on the record, they will not have reason to adjust behavior or data for the sake of a favorable report. Second, the research act will not inhibit or devalue their discourse: Participants in an intimate prayer meeting need to be unselfconscious, and it would therefore be disturbing for them to know that one of their company was only there to observe. One of the ways of overcoming the issue of REACTIVITY with informed subjects is to stay in the field for a sufficient time until one becomes accepted as part of the physical setting; another common strategy is to use a gatekeeper or informed collaborator who can speak for the group while not being a member of it.

COMMENTARY

It is not possible to draw a clear line between covert and open methods. Even those who use the most explicit of methods, such as interviews and questionnaires, deploy skills to redefine the research act and to diminish their subjects' awareness of it. Again, they will ask informants questions about other persons whose consent has not been sought, thus using interviewees as surrogate informants while not physically engaging those who are the effective subjects in the research. Although being ethical in the sense of being in accordance with the professional codes, open methods are not necessarily more moral.

What weighs even more heavily than the argument against covert methods is the professional consensus. The principle of informed consent is that on which the self-respect and reputation of social research rest, and practitioners defy it at their peril.

More than 10 years ago, the standard objections to covert practice were that it constituted an infringement of personal liberty and a betrayal of trust in a world, at a time when these things were highly valued. Recent years have seen the advent of security cameras in public places, computer records and databases that can be hacked into by those with appropriate skills, and cases of dishonesty and sleaze in high places. The milieu of trust and liberty has changed. Social researchers endeavor to convey findings that politicians then spin. It may well be that the credibility of researchers will stand or fall not by the openness of the methods they deploy but by the extent to which they can bypass control and manipulation and offer findings that users can trust.

—Roger Homan

REFERENCES

Bulmer, M. (Ed.). (1982). *Social research ethics.* London: Macmillan.

Henle, M., & Hubble, M. B. (1938). Egocentricity in adult conversation. *Journal of Social Psychology, 9,* 227–234.

Homan, R. (1992). *The ethics of social research.* London: Longman.

Humphreys, L. (1975). *Tearoom trade: Impersonal sex in public places.* Chicago: Aldine.

Mills, C. W. (1959). *The sociological imagination.* New York: Oxford University Press.

CRAMÉR'S *V*. *See* SYMMETRIC MEASURES

CREATIVE ANALYTICAL PRACTICE (CAP) ETHNOGRAPHY

From the 1980s onward, influenced by POSTMODERNISM, feminism, and POSTSTRUCTURALISM, contemporary ETHNOGRAPHERS have challenged standard social science writing as ethically, philosophically, and scientifically problematic. Recognizing that the boundaries between "fact" and "fiction," "subjective and objective," and "author and subject" are blurred, these contemporary ethnographers are writing social science in new ways. At first, this work was referred to as *experimental ethnography* or *alternative ethnography.* More recently, to underscore the processes by which the writing is accomplished and to establish the legitimacy of these ethnographies in their own right, they are increasingly known as *creative analytical practice* or *CAP ethnography.*

CAP ethnography is both "scientific"—in the sense of being true to a world known through empirical work and study—and "literary"—in the sense of expressing what one has learned through evocative writing techniques and forms. Although CAP ethnographers belong to a number of different disciplines, they tend to share some research/writing practices in common. These include using theories that challenge the grounds of disciplinary authority; writing on topics of social and personal importance; privileging nonhierarchical, collaborative, and/or multivocal writing; favoring self-reflexivity; positioning oneself in multiple, often competing discourses; and the signature practice, writing evocatively, for different audiences in a variety of formats.

The varieties of CAP ethnography are many (cf. Richardson 1997). They include AUTOETHNOGRAPHY (cf. Ellis, 2003), FICTION, poetic representation, performance texts, ethnographic drama, story writing, reader's theater, aphorisms, conversations, epistles, mixed media, and layered accounts, as well as hypertexts, museum exhibits, photography, paintings, dance, and other forms of artistic representation. For example, Laurel Richardson created a narrative poem from an open-ended interview with "Louisa May," an unwed mother (see Richardson, 1997). Carolyn Ellis (2003) invented a fictional graduate seminar in which characters, based on composites of actual students, engage in dialogue about autoethnography.

EVALUATION OF CAP ETHNOGRAPHIES

Most CAP ethnographers subscribe to six criteria:

1. *Substantive:* Does the piece make a substantive contribution to our understanding of social life?
2. *Aesthetic:* Does the piece invite interpretive responses? Is it artistically satisfying and complex?
3. REFLEXIVITY: Is the author reflexive about the self as a producer of the text?
4. ETHICS: Does the author hold himself or herself accountable to the standards of knowing and telling of the people studied?
5. *Impact:* Does the text affect the reader? Emotionally? Intellectually? Move others to social actions? Did the text change the writer?
6. *Expresses a reality:* Does the text provide a credible account of a cultural, social, individual, or communal sense of the "real"?

CAP ethnographers recognize that the product cannot be separated from the producer—the problems of subjectivity, authority, authorship, and reflexivity are intricately interwoven into the representational form. Science is one lens, creative arts another. CAP ethnographers try to see through both lenses.

—Laurel Richardson

REFERENCE

Denzin, N., & Lincoln, Y. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Ellis, C. (2003). *The ethnographic "I": A methodological novel on doing autoethnography.* Walnut Creek, CA: AltaMira.

Richardson, L. (1997). *Fields of play: Constructing an academic life.* New Brunswick, NJ: Rutgers University Press.

CRESSIE-READ STATISTIC

The Cressie-Read statistic belongs to the family of power divergence statistics for assessing model fitting and is an alternative to the LIKELIHOOD RATIO STATISTIC, although it does not have a p value associated with the statistic.

—Tim Futing Liao

REFERENCE

Cressie, N., & Read, T. (1984). Multinomial goodness of tests. *Journal of Royal Statistical Society Series B, 46,* 440–464.

CRITERION RELATED VALIDITY

In educational assessment and social science measurement, criterion-related validity is defined within the context of test or measurement validation. Test and measurement validation is the process whereby data are collected to establish the credibility of inferences to be made from test scores or measurements—both inferences about use of scores and measures (as in decision making) and inferences about interpretation (meaning) of scores and measures. Prior to the 1985 *Standards for Educational and Psychological Testing* (American Educational Research Association, 1985), validity was conceptualized, in the behaviorist tradition, as a taxonomy of three major and largely distinct categories (i.e., content validity, criterion-related validity, and CONSTRUCT VALIDITY). These three validation approaches provide accumulative evidence for a test's or a measurement instrument's validity. Criterion-related validity is evidence of a test's or an instrument's association with a criterion. This validation procedure is commonly used for aptitude tests and for other situations in which test scores or measurements are used for prediction. Criterion-related validity is further divided into two subtypes: (a) predictive evidence of criterion-related validity (or predictive validity) and (b) concurrent evidence of criterion-related validity (or concurrent validity). The predictive evidence of criterion-related validity indicates the effectiveness of a test or an instrument in predicting a test taker's future performance in specified activities. For example, to claim that a harp aptitude test has predictive validity, test developers must demonstrate that high scores on this test are associated with superior performance on the harp, and low scores are associated with poor performance. The concurrent evidence of criterion-related validity indicates the test's degree of association with a current criterion. Concurrent validity differs from predictive validity only in that the criterion is concurrent. For example, a test designed to diagnose psychological depression may be validated by administering it to persons whose level of depressive symptomatology is already known. Often, concurrent validity is used as a substitute for predictive validity because there is no delay in the availability of criterion measures. So, for instance, if a scale is constructed to assess students' science aptitude, the concurrent validity may be established by correlating test scores with students' grades in a current science course, instead of with future indicators of science aptitude, such as accomplishments in science, writings on science topics, and so forth. The difference between these two validity evidences lies in the objective of a test. If a test or an instrument is to be used for predicting a criterion trait or behavior, the predictive validity is called for. If a test or an instrument is to be used to identify a criterion trait or behavior, the concurrent validity is used to make the validity claim.

The current thinking about the conceptualization of validity, as reflected in the 1985 and 1999 *Standards for Educational and Psychological Testing,* stresses that it is not the tests that are valid (or not); rather, inferences made about the meaning and the use of test scores are valid (or not). "Validity refers to the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests" (American Educational Research Association, 1999, p. 9). Furthermore, validity is said to be a "unified though faceted concept" (Messick, 1989) for which different types of evidence (not different types of validity) should be gathered to support the inferences made from test scores (or measurements) and proper use of the test (or the instrument). The instances of criterion-related validity evidence discussed above are examples of such types of evidence.

—Chao-Ying Joanne Peng and Daniel J. Mueller

See also CONSTRUCT VALIDITY, VALIDITY

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice, 21*(1), 31–41.

Messick, S. (1989, March). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher,* pp. 5–11.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., pp. 1487–1495). New York: Macmillan.

CRITICAL DISCOURSE ANALYSIS

Critical discourse analysis (CDA) sees language as one element of social events and social practices that is dialectically related to other elements (including social institutions and aspects of the material world). Its objective is to show relationships between language and other elements of social events and practices. It is an approach to language analysis that originates within linguistics but is more socially oriented and interdisciplinary than most linguistics.

Discourse here means language (as well as "body language," visual images, etc.) seen as a moment of the social. The approach is "critical" in the sense that (a) it seeks to illuminate nonobvious connections between language and other elements of social life; (b) it focuses on how language figures in the constitution and reproduction of social relations of power, domination, and exploitation; and (c) it takes a specifically language focus on social emancipation and the enhancement of social justice. In these respects, it can be seen as a branch of critical social science.

CDA is based on the assumption that the language elements of social events (talk, texts) can contribute to change in other social elements—that discourse is socially constructive. A focus of analysis has been on the effects of discourse in constituting, reproducing, and changing ideologies. Both are consistent with a dialectical view of discourse as an element of the social that is different from others while not being discrete—different elements "internalize" each other. One aspect of recent changes in social life is arguably that discourse has become more salient in certain respects in relation to other elements; for instance, the concept of a "knowledge" or "information" society seems to imply that social (e.g., organizational) change is "led" by discourses that may be enacted in new ways of acting, inculcated in new identities, materialized in new plants, and so forth. However, in contrast with a certain discourse idealism, we should take a moderate view of social CONSTRUCTIVISM, recognizing that social constructs acquire intransitive properties that may make them resistant to the internalization of new discourses.

DEVELOPMENT OF THE FIELD

Critical perspectives on the relationship of language to other elements of social life can be traced back to Aristotelian RHETORIC and found across a range of academic disciplines. Although CDA is not original in addressing such issues, it has developed for the first time a relatively systematic body of theory and research. Its main sources include broadly "functionalist" (as opposed to "formalist") traditions within linguistics, including the systemic functional linguistics theory of Michael Halliday; theorizations of "hegemony" and "ideology" within Western Marxism, notably by Gramsci and Althusser; the "critical theory" of the Frankfurt School, including more recently Habermas; the concept of "discourse" in Foucault; and the dialogical language theory of Bakhtin, including its emphasis on "heteroglossia" and "intertextuality" and its theorization of "genre."

The term *critical discourse analysis* was first used around 1985, but CDA can be seen as including a variety of approaches, beginning with *critical linguistics,* an application of Halliday's linguistics to analyzing texts from the perspective of ideology and power. Some work in CDA (especially that of Teun van Dijk) incorporates a cognitive psychology, some includes an emphasis on historical documentation (especially the "discourse-historical" approach of Ruth Wodak), and some (especially Gunther Kress and Theo van Leeuwen) has focused on the "multimodal" character of contemporary texts, particularly the mixture of language and visual image. CDA has progressively become more interdisciplinary, engaging more specifically with social theory and research, and (especially in the case of Fairclough) it is committed to enhancing the capacity of research on the social transformations of the contemporary world (globalization, neoliberalism, new capitalism) to address how language figures in processes of social transformation.

APPLICATION

CDA has recently been drawn on in research within a considerable variety of disciplines and areas of study in the social sciences (e.g., geography and urban studies, health sciences, media studies, educational research, sociology). A major division within this extensive uptake is between studies that incorporate detailed linguistic analysis of texts and studies that work with a generally Foucaultian view of discourse and draw on theoretical perspectives in CDA without such close textual analysis. For some CDA researchers, establishing the value of close textual analysis in social research is a central concern. At the same time, there is a concern to contribute a language focus to social and political debates in the public sphere (Fairclough, 2000).

EXAMPLE

A theme that has attracted considerable international attention and now has its own virtual research network is Language in New Capitalism (www.cddc.vt.edu/host/lnc/), which explores how language figures in the current restructuring and "global" rescaling of capitalism. An example of what CDA can contribute to social research on these transformations is analysis of narratives of global economic change (Fairclough, 2000). Such narratives are pervasive in economic, political, and other types of texts. Economic change is recurrently represented within them as a process that does not involve human agents—so, for instance, new markets "open up" rather than people opening them up, and technologies "migrate" from place to place rather than being transported by companies—which is universal in time and place (it happens in a universal present and everywhere equally). Such representations of economic change are widely worked together with policy injunctions—what "is" unarguably and inevitably the case dictates what we "must" do in policy terms. Analysis of these narratives can put some flesh on the rather skeletal argument advocated by Bourdieu that neoliberal discourse is a powerful part of the armory for bringing into existence a "globalization" (and transformation of capitalism) that is misleadingly represented as already fully achieved.

—Norman Fairclough

REFERENCES

Chouliaraki, L., & Fairclough, N. (1999). *Discourse in late modernity.* Edinburgh, UK: Edinburgh University Press.

Fairclough, N. (2000). *New labour, new language?* London: Routledge Kegan Paul.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. In T. van Dijk (Ed.), *Discourse as social interaction*. Thousand Oaks, CA: Sage.
Wodak, R., & Meyer, M. (2001). *Methods in critical discourse analysis*. Thousand Oaks, CA: Sage.

CRITICAL ETHNOGRAPHY

In recent years, critical ethnography has become an increasingly popular perspective.

However, how scholars from different disciplines have tailored it by using different theoretical or paradigmatic approaches has created a diverse range of applications that result in labeling any form of cultural criticism as *critical ethnography*. Therefore, it may be less important to define critical ethnography than to identify a few fundamental constituent elements.

Reducing critical ethnography simply to social criticism distorts and oversimplifies the critical ethnographic project. At its simplest, critical ethnography is a way of applying a subversive worldview to more conventional narratives of cultural inquiry. It does not necessarily stand in opposition to conventional ETHNOGRAPHY or even to conventional social science. Rather, it offers a more reflective style of thinking about the relationship between knowledge, society, and freedom from unnecessary forms of social domination.

What distinguishes critical ethnography from the other kind is not so much an act of criticism but an act of critique. *Criticism,* a complaint we make when our eggs are undercooked, generally connotes dissatisfaction with a given state of affairs but does not necessarily carry with it an obligation to dig beneath surface appearances to identify fundamental social processes and challenge them.

Critique, by contrast, assesses "how things are" with an added premise that "things could be better" if we examine the underlying sources of conditions, including the values on which our own complaints are based.

Critical ethnography is not a theory because it does not, in itself, generate a set of testable propositions or lawlike statements. Rather, it is a perspective that provides fundamental images, metaphors, and understandings about our social world. It also provides value premises around which to ask questions and suggest ways to change society. It is a methodology that "strives to unmask hegemony and address oppressive forces" (Crotty, 1998, p. 12). Critical researchers begin from the premise that all cultural life is in constant tension between control and resistance. This tension is reflected in behavior, interaction rituals, normative systems, and social structure, all of which are visible in the rules, communication systems, and artifacts that comprise a given culture. Critical ethnography takes seemingly mundane events, even repulsive ones, and reproduces them in a way that exposes broader social processes of control, power imbalance, and the symbolic mechanisms that impose one set of preferred meanings or behaviors over others. Although not all critical ethnographers are equally comfortable with issues such as GENERALIZABILITY, RELIABILITY, VALIDITY, and VERIFICATION, critical ethnography is quite compatible with the conventions of "empirical science" by which the credibility of empirical claims can be evaluated (Maines, 2001).

Two examples illustrate the diversity of critical ethnography. First, Bosworth's (1999) critical ethnography of culture in three English women's prisons identifies a paradox: The conventional idealized images of heterosexual femininity that function as part of the punishment process also become a means for women to resist control and survive their prison experience. She argues that, although an idealized notion of femininity underlies much of the daily routine of women's prisons, it possesses the contradictory and ironic outcome of producing the possibility for resistance (Bosworth, 1999, p. 107). This type of critique helps soften the jagged edges of the gender gap in criminal justice scholarship by reversing the longstanding tradition of looking at women's prison culture through the prism of male prisoners and male scholars.

Ferrell and Hamm (1998) provide a collection of insightful examples that further demonstrate the utility of critical ethnography. By looking at marginalized populations such as pimps, marijuana users, phone sex workers, and domestic terrorists, the contributors challenge the reader to recast conventional, comfortable assumptions and images about crime, deviance, and social control. This, in turn, challenges us to examine more fully the asymmetrical power relations that create and sustain sociocultural power imbalances.

There are, of course, problems with critical ethnography (Thomas, 1992). But science—and critique is part of the scientific project—is self-correcting. One task of those engaged in ethnographic critique lies in balancing the occasionally conflicting tasks of a priori conceptual analysis, interpretive narration, empirical rigor, and theory building, taking care not to reject such tasks as POSITIVIST lest we substitute one intellectual dogma with another. This means that we must continually reassess our own project with the same rigor that we assess our foci of analysis, always bearing in mind that things are never what they seem and that social justice is not simply a goal but a vocation to which all scholars ought strive with verifiable critique.

—Jim Thomas

REFERENCES

Bosworth, M. (1999). *Engendering resistance: Agency and power in women's prisons.* Aldershot, UK: Ashgate/ Dartmouth.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process.* Thousand Oaks, CA: Sage.

Ferrell, J., & Hamm, M. S. (1998). True confessions: Crime, deviance, and field research. In J. Ferrell & M. Hamm (Eds.), *Ethnography at the edge: Crime, deviance and field research* (pp. 2–19). Boston: Northeastern University Press.

Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research.* Beverly Hills, CA: Sage.

Maines, D. R. (2001). *The faultline of consciousness: A view of interactionism in sociology.* New York: Aldine De Gruyter.

Thomas, J. (1992). *Doing critical ethnography.* Newbury Park, CA: Sage.

CRITICAL HERMENEUTICS

Critical hermeneutics grows out of the hermeneutic phenomenology of Paul Ricoeur (1981; Thompson, 1981) and the depth hermeneutics of John Thompson (1990). Following Ricoeur, it decomposes interpretation into a dialectical process involving three interrelated, interdependent, mutually modifying phases or "moments": a moment of social-historical analysis, a moment of formal analysis, and a moment of interpretation-reinterpretation in which the first two moments of analysis are brought together (see Phillips & Brown, 1993, for an empirical example). In interpreting a text, one can and should analyze the social-historical context out of which it has arisen. But one should also analyze the text formally, as a system of signs abstracted from experience using one of the formal approaches to textual analysis such as structural SEMIOTICS or DISCOURSE ANALYSIS. From a critical hermeneutic perspective, neither the analysis of the social-historical context nor the text itself is superior. Each is an indispensable moment in the interpretive process, and methods of interpretation that fail to explicitly include the social-historical context *and* the formal analysis of the text are incomplete. Having completed the social-historical and formal analysis, a researcher must creatively combine these two moments to produce an interpretation of the text and its role in the social system of which it is a part. It is in this moment of interpretation that the researcher develops insights into the role of the text in the ongoing construction of social reality. This decomposition of the process into moments provides a structured framework for the researcher to follow and allows readers to more clearly understand the researcher's methodology and to judge more clearly the value of the final conclusions.

But in addition to being hermeneutic, the methodology is also critical. In addition to being concerned with meaning, critical hermeneutics is concerned with the role of communication in the ongoing maintenance of the asymmetrical power relations that characterize the social world. The examination of the social-historical context of the production, transmission, and reception of a text provides an opportunity for researchers to include a concern for the role of communication in the development of relations of power that lead to enduring asymmetries of opportunities and life chances. It allows for the introduction of an emancipatory interest and focuses attention on the taken-for-granted conventions and rules that underlie a particular social system. Those applying a critical hermeneutic methodology ask how the communication process in a certain social setting functions, for whose benefit it functions, and who is allowed to participate.

By combining an interest in interpretation of texts with the social-historical context of the text, critical hermeneutics makes three important contributions. First, it provides a structured method for examining the role of symbolic phenomena in social life. Second, it provides a structured method for examining the sources of texts and the interests of their producers. Third, it provides a way to integrate formal methods of textual analysis with an interest in the social context in which the texts occur. Critical hermeneutics therefore adds an important dimension to more traditional interpretive methods in social science.

—Nelson Phillips

REFERENCES

Phillips, N., & Brown, J. (1993). Analyzing communication in and around organizations: A critical hermeneutic approach. *Academy of Management Journal, 36*(6), 1547–1576.

Ricoeur, P. (1981). *Hermeneutics and the human sciences: Essays on language, action and interpretation.* New York: Cambridge University Press.

Thompson, J. (1981). *Critical hermeneutics: A study in the thought of Paul Ricoeur and Jurgen Habermas.* Cambridge, UK: Cambridge University Press.

Thompson, J. (1990). *Ideology and modern culture.* Stanford, CA: Stanford University Press.

CRITICAL INCIDENT TECHNIQUE

The critical incident technique (CIT) was first devised and used in a scientific study almost half a century ago (Flanagan, 1954). The method assumed a POSITIVIST approach, which was at the time the dominant paradigm in the social as well as the natural sciences. It was used principally in occupational settings, and its advocates established the VALIDITY and RELIABILITY of the technique (Andersson & Nilsson, 1964; Ronan and Latham, 1974). The approach of Flanagan and the studies he reviewed assumed tangible reality. In contrast, CIT, as developed by Chell (1998), assumes PHENOMENOLOGY. It is intended through the method of, technically, an UNSTRUCTURED INTERVIEW to capture the thought processes, the frames of reference, and the feelings about an incident that have meaning for the RESPONDENT. In the interview, respondents are required to give an ACCOUNT of what the incident means for them in relation to their life situation, as well as their present circumstances, attitudes, and orientation. The approach may be used in an in-depth CASE STUDY, as well as a multisite investigation.

CIT is a QUALITATIVE INTERVIEW procedure, which facilitates the investigation of significant occurrences (events, incidents, processes, or issues) identified by the respondent, the way they are managed, and the outcomes in terms of perceived effects. The objective is to gain an understanding of the incident from the perspective of the individual, taking into account cognitive, affective, and behavioral elements (Chell, 1998, p. 56).

EXAMPLE AND METHOD

Arguably, all research commences with "preliminary design work." This assumes that there is a point to the research; the research questions and ideas are coherent and focused on the aims of the study. Moreover, in empirical work, the researcher needs to consider how to locate the respondent (or subject) and why this particular respondent was chosen. Researchers may use a variety of sources from their own local knowledge, word of mouth, and databases and lists owned by other bodies. This also means that there is likely to be a criterion or criteria for inclusion of the subject that need to be made explicit. An example in which the use of the CIT was helpful in exposing a depth of rich information was a study of personal and business issues on microbusinesses in business services. It showed the importance of personal and family issues relative to other business and economic factors (see Chell, 1998; Chell & Baines, 1998).

Hand in hand with locating subjects is the ability to gain ACCESS to them. This may require skill, patience, and perseverance. It is wise not to take access for granted. Devices such as an initial introductory telephone call or letter of introduction may be used. The latter approach enables the researcher to outline the aims and objectives of the project and offer plausible reasons for seeking to intrude on the subject's time and PRIVACY. Hence, sensitive handling is required from the outset.

Once access has been granted and an appointment made, the researcher will need to "rehearse" the CIT interview. Although the interview is unstructured, there is a sequence of development of the interview that commences with the interviewer introducing himself or herself, his or her research, and the nature of the CIT interview to the subject. Issues such as CONFIDENTIALITY and associated assurances are essential to win the trust of the subject. Once the interviewee is relaxed and comfortable with the procedure, the interview may commence.

The interviewee will then be asked to focus on one or more (usually no more than four) "incidents" that have occurred (over a specified period of time) to the subject or a group or organization of which the subject has intimate knowledge. Different ploys may be used to get the subject to focus on critical events. Respondents vary greatly in the extent to which they can articulate events, and the toughest interviews tend to be those in which there appear to be no events! The interviewer must then attempt to explore lengthy periods of apparent absence of incident to yield useful information. A further difficulty is that of establishing a correct chronology or sequencing of events. This can be facilitated by reference to other events, documents, diaries, and other validating materials. The subject may be encouraged to supply such cross-referencing material.

This helps with TRIANGULATION and VALIDATION of the information.

The respondent may also think of incidents as positive or negative experiences. How subjects frame the incident is entirely under their control, as indeed is the language in which they couch their account. Although it is important that the researcher captures evaluative language, it is also important that he or she listens carefully and PROBES appropriately to grasp the essential details of the incident from the subject's perspective.

Controlling the interview despite its unstructured nature is important. The interview may use generic probes (e.g., What happened next? How did you cope?), which yields information about the context, language, and rapport. When the interviewer does not understand, he or she must interject to get immediate clarification; otherwise, the thread may be lost, and it is often difficult to return to more than a few minor points at the end of an interview. It can also be difficult to remember something that one has not understood! When finally concluding the CIT, the interviewer must take care of all ethical issues and concerns. If ANONYMITY is important, then assurances must be given. If feedback is required in some form, then an arrangement must be made to ensure that it is given. The interviewer must leave the interview feeling that, at whatever time or stage in the future, he or she will be welcome to return.

The final stage in the process is TRANSCRIPTION of the interview and its analysis. The analytical process is likely to be based on GROUNDED THEORY. This means that the researcher abandons preconceptions and, through the process of analysis, builds up an explanatory framework through the CONCEPTUALIZATION of the data. This is THEORY building in its most basic form. It should be used ideally with THEORETICAL SAMPLING techniques so that depth of understanding can be developed from which a coherent body of theory will eventually be formed. The CIT is not easy to conduct well and requires the researcher to be skilled and mature in approach. It can be stressful, but on the whole, it is enjoyable both for the researcher and subject. It is a tool that helps us develop an understanding of dynamic processes within social intercourse. It assumes that real life is messy and that all, including the researcher and the subject, are attempting to make sense of their reality.

—Elizabeth Chell

REFERENCES

Andersson, B. E., & Nilsson, S. G. (1964). Studies in the reliability and validity of the critical incident technique. *Journal of Applied Psychology, 48*(1), 398–403.

Chell, E. (1998). Critical incident technique. In G. Symon & C. Cassell (Eds.), *Qualitative methods and analysis in organizational research: A practical guide* (pp. 51–72). London: Sage.

Chell, E., & Baines, S. (1998). Does gender affect business performance? A study of micro-businesses in business services in the U.K. *International Journal of Entrepreneurship and Regional Development, 10*(4), 117–135.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327–358.

Ronan, W. W., & Latham, G. P. (1974). The reliability and validity of the critical incident technique: A closer look. *Studies in Personnel Psychology, 6*(1), 33–64.

CRITICAL PRAGMATISM

Critical pragmatism is a methodological orientation that believes that social science research should illuminate ideological domination, hegemonic practices, and social injustice; it also advocates an eclectic methodological experimentalism in the pursuit of this illumination.

The critical element in critical pragmatism is derived from CRITICAL THEORY, particularly the work of Herbert Marcuse and Antonio Gramsci. Both these writers urge the need for research into the ways in which societies reproduce themselves and the ways in which people are persuaded to embrace ideas, beliefs, and practices that are patently not in their best interests. Marcuse was interested in the way in which people could learn to challenge the widespread existence of one-dimensional thought—that is, thought that prevented any consideration of the moral basis of society and that dissuaded people from utopian speculation (see Marcuse, 1964). Gramsci was interested in the way people learned to recognize and contest hegemony—that is, the process by which people embrace ways of thinking and acting that are harmful to them and that keep unequal structures intact (see Gramsci, 1971). The pragmatic element comes from the tradition of American PRAGMATISM. Pragmatism has an experimental, anti-foundational tenor that advocates using an eclectic range of methods and approaches to promote democracy. It encourages the constant critical analysis of assumptions and is skeptical of reified, standardized models of practice or inquiry.

The element of FALLIBILISTIC self-criticality endemic to critical pragmatism means avoiding a slavish adherence to a particular methodology, whether this be QUANTITATIVE, QUALITATIVE, POSITIVIST, ETHNOGRAPHIC, statistical, or PHENOMENOLOGICAL. A critically pragmatic stance rejects any premature commitment to one research approach, no matter how liberatory this might appear. It urges a principled methodological eclecticism, a readiness to experiment with any and all approaches in the attempt to illuminate and contest dominant ideology.

Critical pragmatism has as the focus of its research effort a number of questions. How do people learn forms of reasoning that challenge dominant ideology, and how do they learn to question the social, cultural, and political forms that ideology justifies? How do people learn to interpret their experiences in ways that emphasize their connectedness to others and lead them to see the need for solidarity and collective organization? How do people learn to unmask the flow of power in their lives and communities? How do people learn of the existence of hegemony and of their own complicity in its continued existence? Once aware of hegemony, how do they contest its all-pervasive effects? A quantitatively inclined example of critical pragmatism would be John Gaventa's (1980) study of land ownership in Appalachia. A qualitatively inclined example would be Douglas Foley's (1990) ethnography of Mexican American youth in a south Texas town.

—Stephen D. Brookfield

REFERENCES

Foley, D. E. (1990). *Learning capitalist culture: Deep in the heart of Texas.* Philadelphia: University of Pennsylvania Press.

Gaventa, J. (1980). *Power and powerlessness: Quiescence and rebellion in an Appalachian valley.* Chicago: University of Chicago Press.

Gramsci, A. (1971). *Selections from the prison notebooks.* London: Lawrence and Wishart.

Marcuse, H. (1964). *One dimensional man.* Boston: Beacon.

CRITICAL RACE THEORY

Critical race theory is a body of radical critique against the implicit acceptance of White supremacy in prevailing legal paradigms and in contemporary law. The legal literature and civil rights discourse used to be dominated by the writings of White scholars and legal practitioners. Critical race theory emerged in the early 1990s as an intellectual movement shaped substantially, but not exclusively, by progressive scholars of color in the United States (Crenshaw, Gotanda, Peller, & Thomas, 1995). Critical race theory deconstructs the legal, symbolic, and material capital represented by *Whiteness.* Its mission is to critique and, ultimately, transform the law, an institution contributing to the construction and perpetuation of racial domination and subordination. The intertwining of experiential depth, politics, and intellectual rigor makes critical race theory accessible and relevant to social scientists despite its foundations in the discipline of law.

Critical race theory can be seen as part of the radical tradition of oppositional writings (Essed & Goldberg, 2002). These approaches are premised on the idea that science is not "neutral" or "objective." Notions of rights and equality, right and wrong, justice and injustice have different implications for groups who have suffered through history. An important methodological consideration is that race and racism are often understood differently, depending on the positioning of the defining party in the matrix of race relations (Bulmer & Solomos, 2003). Institutions, the court included, tend to privilege the perpetrator's perspective of racism, usually a narrow definition based on proof of intent. The victim, who is exposed to a broader range of experiences of racism, can often understand racism in terms of the societal consequences of exclusion and humiliation.

Critical race theory has been instrumental in mainstreaming *intersectionality* and *multiple identifications* as theoretical frameworks. Race, gender, and other structural categories intersect with and determine legal and social discourses. For example, the disproportionate number of African American women who lose custody of their children through the child welfare system is partly due to cultural bias privileging the White and middle-class ideal of the nuclear family and (full-time) mothering.

Using texts and narratives as data, critical race scholars expose the limitations and distortions of traditional legal analysis. They deconstruct law through contextual analysis of historical and other court cases that have marked the landscape of race relations. The myth of *color blindness* underlying the well-known "separate but equal" case, *Plessy v. Ferguson* (1896), is a case in point. The court treated racial categorizations as unconnected to social status and reconfirmed segregation. Critical race theory offers a framework for understanding *how* and *why* court systems continue to reinforce racial injustice. Collection and analysis of historical and contemporary texts and stories of racial injustice are powerful tools to build bridges of validation and understanding in the pursuit of social transformations.

—Philomena Essed

REFERENCES

Bulmer, M., & Solomos, J. (Eds.). (2003). *Researching race and racism.* London: Routledge Kegan Paul.

Crenshaw, K., Gotanda, N., Peller, G., & Thomas, K. (Eds.). (1995). *Critical race theory.* New York: The New York Press.

Essed, Ph., & Goldberg, T. D. (Eds.). (2002). *Race critical theories: Text and context.* Malden, MA: Blackwell.

Plessy v. Ferguson, 163 U.S. 537 (1896).

CRITICAL REALISM

Critical realism is a way of understanding the nature of both natural and human social sciences. Particularly in the case of the latter, it also advocates some and argues against other ways of practicing them. As a version of REALISM, it is committed to the view that the objects of scientific knowledge both exist and act independently of our beliefs about them. However, the qualification *critical* serves to distinguish this form of realism from others, sometimes called *direct* or *naive* realisms, which assume simple "one-to-one" links between beliefs and reality. In the human social sciences, the qualifier *critical* also often signals a normatively critical orientation to existing social forms.

Critical realists make a clear distinction between the independently existing real beings, relations, processes, and so on (the "intransitive dimension"), which are the objects of scientific knowledge, and the socioculturally produced concepts, knowledge claims, and methods through which we attempt to understand them. The latter (sometimes referred to as the *transitive dimension*) (Bhaskar, 1979/1998) are always provisional and fallible. Critical realists thus fully acknowledge the socially and historically "constructed" character of scientific theoretical frameworks and knowledge claims, but they argue that practices such as experiment, scientific education, and the application of scientific ideas presuppose the independent *existence* of the objects of scientific knowledge. We cannot make sense of these practices other than on the basis that they are *about* something, even though we may doubt the truth of specific assertions about that "something."

In most cases, critical realists reject the relativism often associated with approaches (such as their own), which emphasize the socially constructed character of knowledge. Without commitment to unsustainable notions of "ultimate" or "absolute" truth, it is usually possible to evaluate rival knowledge claims as (provisionally) more or less rationally defensible in light of existing evidence.

Critical realists have provided powerful arguments against both empiricist and relativist accounts of the nature of natural scientific activity. Bhaskar (1975/1997), in particular, used transcendental arguments from practices such as scientific experiment to show the incoherence of empiricist restrictions on scientific ONTOLOGY and accounts of scientific laws. Experimental practice is designed to isolate causal mechanisms, so producing regular event sequences, or "constant conjunctions." But because these constant conjunctions are the outcome of experimental manipulation, they cannot be identical with scientific laws (otherwise, we would have to say that laws of nature were products of human manipulation).

This analysis results in a view of the world as "stratified," distinguishing between the "level" of observed events (the "empirical") and that of the "actual" flow of events (most of which pass unobserved). A third level, misleadingly referred to as the "real" (because all levels are in some sense real), is that of the "causal mechanisms," whose tendencies are captured by statements of scientific laws. Critical realism might reasonably be described as a form of "depth" realism, implying the existence of multiple layers of reality behind or below the flow of sense-experience and subject to discovery by scientific experiment and theoretical analysis. Science is, in this way, seen as a process of discovery of ever deeper levels of reality (genes, molecules, atoms, subatomic particles, fields of force, quasars, etc.) rather than as one of accumulating experiential generalizations. The analysis of practices of applying scientific knowledge in technology yields a further claim about the nature of the world as an object of scientific knowledge. This is its "differentiated" character. Experimental practice involves isolating causal mechanisms from extraneous influences (i.e., the creation of "closed systems"). However, technologies involve application of scientific knowledge in "open systems," involving the interaction of multiple causal mechanisms. In open systems, interaction between mechanisms results in the overriding or modification of their effects, such that regular or predictable flows of events and observations may not occur.

The applicability of the critical realist model of science to the social sciences is contested. For some philosophers of social science, humans and their social relations are so unlike the natural world that it is fundamentally inappropriate to attempt to study them on the basis of any model of science. However, critical realists have generally taken the view that despite important differences between their objects of knowledge and those of the natural sciences, the social studies may still be scientific. However, if we recognize that each science must adopt methods of enquiry and forms of explanation appropriate to its subject matter, it still remains an open question just what it is to study social life in a (critical realist) scientific way.

During the 1970s, several authors addressed this question from realist points of view (e.g., Benton, 1977; Bhaskar, 1979/1998; Keat & Urry, 1975). Bhaskar argued that the social sciences can be scientific—not just *despite* but also *because* of acknowledged differences between nature and society, as well as our relationship to them. Social structures are held to exist only in virtue of the activity of agents, are dependent on agents' beliefs about them, and are only relatively enduring. There are also problems about the partial identity of the observer and what is observed: Social sciences are themselves part of society, and it is not possible to establish experimental closure in the social sciences.

These differences, for Bhaskar (1979/1998), imply only that the social sciences must take account of them in its methodology—for example, by beginning investigations with actors' own understanding of what they are doing. For him, social crises constitute an analog of experimentation in the natural sciences. Collier (1994), however, regards the lack of a social science counterpart to experimentation in the natural sciences as grounds for denying that they (the social "sciences") can be fully scientific: He terms them *epistemoids*. Outhwaite (1987) argues for a close relationship between critical realism and critical theory, whereas Benton (1998) sees in critical realism a basis for a nonreductive realignment of the social and natural sciences, which is important for understanding ecological issues in particular.

Another area of disagreement among critical realists concerns the relationship between explanation and moral or political values. Bhaskar and others argue that it is possible to derive values from scientific explanation on the realist model—hence the possibility of "critical" social science, which is oriented to a vision of human emancipation. Marx's "explanatory critique" of ideology is the best-known example: Other things being equal, social or economic relations that tend to produce false beliefs in their agents are undesirable and ought to be transformed. Against this, Sayer (2000) has argued that existing social forms are only legitimately open to criticism on the basis of accounts of plausible alternatives to them.

In general, for those social scientists who espouse critical realism, there is a strong commitment to the distinction between social structures and human agency. Both are regarded as real, distinct, but interdependent levels. Agents both reproduce and transform social structures through their activities, whereas social structures both enable and constrain social action (Archer, 1995). There is a large and growing body of empirical research in geography, economics, law, sociology, politics, psychology, feminist and environmental studies, and other fields informed by critical realism (for examples of the application of critical realism, see New, 1996; Porter, 1993).

—Ted Benton

REFERENCES

Archer, M. (1995). *Realist social theory: The morphogenetic approach.* Cambridge, UK: Cambridge University Press.

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds.). (1998). *Critical realism: Essential readings.* London: Routledge Kegan Paul.

Benton, T. (1977). *Philosophical foundations of the three sociologies.* London: Routledge Kegan Paul.

Benton, T. (1998). Realism and social science. In M. Archer, R. Bhaskar, A. Collier, T. Lawson, & A. Norrie (Eds.), *Critical realism: Essential readings.* London: Routledge Kegan Paul.

Bhaskar, R. (1997). *A realist theory of science.* London: Verso. (Original work published 1975)

Bhaskar, R. (1998). *The possibility of naturalism.* Hemel Hempstead, UK: Harvester Wheatsheaf. (Original work published 1979)

Collier, A. (1994). *Critical realism.* London: Verso.

Keat, R., & Urry, J. (1975). *Social theory as science.* London: Routledge Kegan Paul.

New, C. (1996). *Agency, health and social survival.* London: Taylor & Francis.

Outhwaite, W. (1987). *New philosophies of social science.* London: Macmillan.
Porter, S. (1993). Critical realist ethnography: The case of racism and professionalism in a medical setting. *Sociology, 27,* 591–609.
Sayer, A. (2000). *Realism and social science.* London: Sage.

CRITICAL THEORY

To the extent that any theorization of societal mechanisms and modes of conducting social relations does not accord with dominant ways of viewing society and social relations, it may be said to be critical. The term *critical theory,* however, is normally reserved for a particular set of ideas associated with what become known as the Frankfurt School of Social Research and its followers, who have modified and extended its original insights, aspirations, and agendas.

Under the auspices of its second director, Max Horkheimer (1895–1973), the Institute for Social Research at Frankfurt centered its interests on the following key areas: EXPLANATIONS for the absence of a unified working-class movement in Europe, an examination of the nature and consequences of capitalist crises, a consideration of the relationship between the political and the economic spheres in modern societies, an account for the rise of fascism and Nazism as political movements, the study of familial socialization, and a sustained critique of the link between POSITIVISM and science. To pursue this study, those figures who stood between Hegel, Marx, and the Frankfurt School—such as Schopenhauer and Nietzsche, who had both questioned the Enlightenment project—required systematic engagement to recover the promise of Marxism.

As noted, the changing climate shaped the work of these scholars. In Nazi Germany, the institute found itself under threat with the result that its leading members, including Theodor Adorno (1903–1969), Max Horkheimer, and Herbert Marcuse (1898–1979), emigrated to the United States (Adorno and Horkheimer returned after World War II). Here they found a self-confident bourgeois liberal-capitalism, with its apparent ability to absorb and neutralize proletarian consciousness. Here was the personification of the ideology of individualism accompanied by ideas of success and meritocracy. Analyzing these trends and their consequences necessitated a consideration of culture that, until this time, had been largely devalued in Marxist circles.

The work of these scholars was interdisciplinary (thus anticipating debates that are more prominent in contemporary times), accompanied by critiques of both positivist and interpretivist approaches to understanding human relations. On one hand, positivism had failed to examine the conditions under which capitalism develops and is sustained. On the other hand, interpretivism was held to be inadequate due to an exclusive concentration on the process of self-understanding and self-consciousness. The result was an uncritical acceptance of dominant forms of consciousness within given societies and a failure to consider the structural determinants of human actions.

Historical investigations were then undertaken into the Enlightenment project. Here we find that

> the individual is wholly devalued in relation to the economic powers, which at the same time press the control of society over nature to hitherto unsuspected heights. . . . The flood of detailed information and candy-floss entertainment simultaneously instructs and stultifies mankind. (Adorno & Horkheimer, 1944/1979, pp. xiv–xv)

On this basis, studies examined the process of ideological incorporation. Herbert Marcuse (1968), in *One Dimensional Man,* thus spoke of the "technical-administrative control" of society.

This meta-critique noted the following question: Having rejected interpretivist social theory, how can the important analysis of mental states be achieved? This is where Sigmund Freud enters as part of the whole history of the relationship between psychoanalysis and social theory. This focus moves away from an understanding of the solitary ego in the social world to the *self-misunderstanding* person who fails to see the causes of his or her own symptoms. However, issues remained. Up to the time of the Frankfurt theorists, Marxists had dismissed psychoanalysis as unworthy of attention and a distraction from the primary purpose of overthrowing an unjust economic system. Psychoanalysis, after all, may be seen as the practice of amelioration, not resolution, let alone revolution. Conceptually speaking, it starts from the individual and tends to play down social conditions and constraints.

Therefore, if Freud's work was to serve the needs of critical theory, it required modification. At this point, Leo Lowenthal and Eric Fromm enter the scene as central to this process in the development of critical theory. Fromm, for example, examined the relationship between the individual and totalitarianism. He also held throughout his life that our capacity for love and freedom is inextricably bound up with socioeconomic conditions. The interest in Freud was also apparent in sociopsychological studies on the structure of modern personality types carried about by Adorno and his associates in 1950, as well as in the writings of Willhelm Reich on the relationship between sexuality and capitalism.

For critical theory, there is a constant interaction between THEORY and facts, and theorists seek to recognize the relationship between the constitution of their propositions and the social context in which they find themselves. REFLEXIVITY concerning the relationship between what is produced, under what circumstances, and with what effects distinguishes traditional from critical theory. In addition, the issue of research results feeding back into social life is not a "problem" for researchers. On the contrary, the adequacy of critical research lies in its value for informing political actions.

Although critical social research aligns itself with the "wishes and struggles of the age," it remains the case that critics, although recognizing the impossibility of separating values from research, regard justifications for conflating the two as untenable. After all, who is to define the values that are of importance? Furthermore, what is the basis of critique? Is it the autonomy of the free individual? The idea of autonomy becomes an abstraction because how can any individual be presumed to be separate from the social relations of whom he or she is a part? The basis of critique then becomes so far removed from reality that it appears aloof and withdrawn. These and other criticisms have been made of research practice informed by critical theory.

Of all contemporary scholars, Jürgen Habermas has continued most prolifically in this tradition and addressed the above issues through extensive modification and critique. This legacy is also evident in the writings of Nancy Fraser (1997) and Axel Honneth (1996), who have contributed to debates on recognition and redistribution, as well as Pierre Bourdieu's (2000) contributions to a "realpolitik of reason." As for Michel Foucault, his own words in "Remarks on Marx" will suffice: "The Frankfurt School set problems that are still being worked on. Among others, the effects of power that are connected to a rationality that has been historically and geographically defined in the West, starting from the sixteenth century on" (Foucault, 1991, p. 117).

—Tim May

REFERENCES

Adorno, T., & Horkheimer, M. (1979). *Dialectic of enlightenment* (J. Cumming, Trans.). London: Verso. (Original work published 1944.)

Bourdieu, P. (2000). *Pascalian meditations* (R. Nice, Trans.). Cambridge, UK: Polity.

Comstock, D. E. (1994). A method for critical research. In M. Martin & L. C. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 625–639). Cambridge: MIT Press.

Foucault, M. (1991). *Remarks on Marx: Conversations with Duccio Trombadori* (R. J. Goldstein & J. Cascaito, Trans.). New York: Semiotext(e).

Fraser, N. (1997). *Justice interruptus: Critical reflections on the 'postsocialist' condition.* London: Routledge Kegan Paul.

Honneth, A. (1996). *The struggle for recognition: The moral grammar of social conflicts* (J. Anderson, Trans.). Cambridge: MIT Press.

Marcuse, H. (1968). *One dimensional man: The ideology of industrial society.* London: Sphere.

Morrow, R. A. (with Brown, D. D.). (1994). *Critical theory and methodology.* London: Sage.

Wiggershaus, R. (1995). *The Frankfurt School: Its history, theories and political significance* (M. Robertson, Trans.). Cambridge, UK: Polity.

CRITICAL VALUES

In a HYPOTHESIS TEST, we decide whether the data are extreme enough to cast doubt on the NULL HYPOTHESIS. If the data are extreme, we reject the null hypothesis; if the data are moderate, we accept it. A threshold between extreme and moderate data, between accepting and rejecting the null hypothesis, is called a *critical value.*

Suppose we flip a coin 10 times. An obvious null hypothesis is that the coin is fair, or equally likely to come up heads or tails. If the null hypothesis is true, we will probably flip close to 5 heads. A sensible test would accept the null hypothesis if the number of heads is close to 5 (moderate) and reject the null hypothesis if the number of heads is far from 5 (extreme). For example, a test might reject the null hypothesis if we flip 9 heads or more or if we flip 1 head or fewer. For this test, 1 and 9 are the critical values.

Critical values are related to SIGNIFICANCE LEVELS. The test above has a significance level of .0215: Under the null hypothesis, there is a .0215 probability that the number of heads will be at least as extreme as the critical values. (This probability is calculated using the BINOMIAL DISTRIBUTION.) We can use the critical values to calculate the significance level, as we did here. It is more common, however, to choose the significance level first and then calculate the corresponding critical values.

Tests using critical values are equivalent to tests using *P* VALUES. That is, rejecting a hypothesis because of extreme data is equivalent to rejecting the hypothesis because of a small p value. For example, suppose 10 coin flips produced 10 heads; 10 is more extreme than the critical value of 9, so we would reject the null hypothesis using the critical value. But we would also reject the null hypothesis if we used the TWO-TAILED p value: The p value for 10 heads is .002, which is less than the significance level of .0215. As this example illustrates, decisions based on p values will always agree with decisions based on critical values.

—Paul T. von Hippel

CRONBACH'S ALPHA

Cronbach's alpha is one type of INTERNAL RELIABILITY estimate used to assess the consistency of responses on a composite measure that contains more than one component (i.e., items, raters). Alpha coefficient, the most widely used reliability estimate, was first introduced by the late Lee J. Cronbach in 1951. According to the Social Sciences Citation Index, between 1951 and 2002, the article was cited 4,912 times by other published articles. The number of published and unpublished articles that report Cronbach's alpha without citing the 1951 article may very likely approach an astronomical figure.

Cronbach's alpha and standardized Cronbach's alpha are formulated as

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s_t^2}\right)$$

and standardized

$$\alpha = \frac{k\bar{r}}{1 + \bar{r}(k-1)},$$

where k is the number of components in a composite measure, s_i^2 is the variance of component i, s_t^2 is the variance of the total score on the measure, and $\bar{r}$ is the average intercomponent correlation. Standardized Cronbach's alpha is used when component standardized scores are computed instead of using raw component scores. The precision of Cronbach's alpha is determined by the standard error of the intercomponent correlations, formulated as

$$\frac{s_r}{\sqrt{[0.5 \times k \times (k-1)] - 1}},$$

where s_r is the standard deviation of the intercomponent correlations. All other things being equal, Cronbach's alpha becomes less precise if the variation among intercomponent correlations increases. It should be noted, however, that the size of Cronbach's alpha is not a function of the size of the standard error. It is possible to have two sets of k components with different amounts of variation among their intercomponent correlations but similar Cronbach's alpha coefficients, which are determined by the average intercomponent correlation.

Conceptually, Cronbach's alpha, rather than standardized Cronbach's alpha, is a lower bound estimate of the *proportion of variance* in a composite measure that is accounted for by a common factor underlying all components. Although this statement suggests that Cronbach's alpha should not be negative and can only vary between 0 and 1, it is not unlikely in practice to observe a negative Cronbach's alpha. As shown in Table 1, a seven-item measure is administered to a group of 10 people. Each item has three response categories, ranging from one to three. The alpha coefficient is negative, and its absolute value is even greater than 1.

Cronbach's alpha can be affected by characteristics of respondents and components. Everything being equal, if respondents are homogeneous (i.e., similar in the characteristic assessed by a measure), s_t^2 tends to decrease. As seen from the above formulas, a small s_t^2 tends to lead to a small alpha coefficient. Second, Cronbach's alpha tends to increase as the number of good-quality components increases. However, this positive feature becomes less salient when the number of components is so large that $\frac{k}{k-1}$ approaches unity. Note that if the components on a measure are heterogeneous, suggesting there is more than one factor or dimension underlying the components, Cronbach's alpha will underestimate the true reliability. In addition, a META-ANALYTIC finding by Peterson (1994) suggests that Cronbach's alpha may underestimate the true reliability when the measure is composed of less

Table 1 A Demonstration of a Negative Cronbach's Alpha

Person	*Item 1*	*Item 2*	*Item 3*	*Item 4*	*Item 5*	*Item 6*	*Item 7*	*Total*
A	1.00	2.00	2.00	2.00	1.00	2.00	2.00	12
B	2.00	1.00	2.00	3.00	2.00	3.00	2.00	15
C	1.00	2.00	1.00	2.00	2.00	2.00	2.00	12
D	2.00	3.00	2.00	1.00	2.00	1.00	2.00	13
F	1.00	2.00	3.00	2.00	1.00	2.00	3.00	14
G	3.00	1.00	2.00	3.00	2.00	1.00	2.00	14
H	1.00	2.00	2.00	2.00	2.00	2.00	1.00	12
I	2.00	1.00	2.00	1.00	2.00	1.00	2.00	11
J	1.00	2.00	1.00	2.00	3.00	2.00	1.00	12
K	1.00	3.00	2.00	1.00	2.00	2.00	2.00	13
Variance	0.45	0.49	0.29	0.49	0.29	0.36	0.29	1.36

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s_t^2}\right) = \left(\frac{7}{6-1}\right)\left(1 - \frac{2.66}{1.36}\right) = -1.12$$

$$\text{Standardized } \alpha = \frac{k\bar{r}}{1+\bar{r}(k-1)} = \frac{7 \times (-.0802)}{1+(-.0802)(7-1)} = -1.08$$

than four items with only two response categories. It should be emphasized that a high Cronbach's alpha (e.g., 0.8) does not necessarily suggest that there is only one factor or dimension underlying the components. High internal consistency is just a necessary but not sufficient condition for unidimensionality. Another related but often-neglected concern is that Cronbach's alpha is not an appropriate reliability estimate for a composite measure when its components are causal indicators of a CONSTRUCT. For instance, a list of behaviors such as watch TV, read book, ski, visit friends, and play chess describes the construct of social activities. However, these activities cannot be considered as parallel components. Therefore, Cronbach's alpha will not provide an adequate reliability estimate for this measure.

—Peter Y. Chen and Autumn D. Krauss

REFERENCES

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98–104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5,* 343–355.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research, 21,* 381–391.

CROSS-CULTURAL RESEARCH

Cross-cultural research can be defined in an anthropological sense to mean any kind of description or comparison of different cultures. It can also be used in the sense of systematic comparisons that explicitly aim to answer questions about the incidence, distribution, and causes of cultural variation (Ember & Ember, 2001). Although it can be useful to identify recurrent themes and patterns in cultures, it can also be risky to generalize about cultural similarities because of the danger of stereotyping. Patterns can provide overall structure to cross-cultural understandings, but they can also hide great individual variations within each culture.

Whether the goal is to describe a unique culture that is different from that of the researcher or to make comparisons across cultures, many issues arise in the conduct of cross-cultural research. Ethnocentrism, or using one's own culture as a lens or standard to judge others, is a barrier to understanding another's cultural experience. Cultural relativism stresses the variability of culture and emphasizes the need to try to understand people in other cultures from within their own cultural context.

Variability within samples in cross-national comparisons, as well as variability on dimensions other than culture, can threaten the VALIDITY of comparative studies (Hujala, 1998). In a study of teachers' thinking between the United States and Finland, the researchers wanted to use the national political borders to define culture as their independent variable. However, the teachers in Finland and the United States differed on other important dimensions. The Finnish teachers had more training compared to the U.S. teachers, and this difference could have explained differences in the teachers' thinking rather than the cultural thinking in a certain society. Gopaul-McNicol and Armour-Thomas (2002) suggest that researchers examine the relationships between culture and class, culture and ethnicity, and culture and language to better understand how to observe meaningful differences within and between cultural groups. They raise issues related to dominant and subdominant cultures within a country and the EPISTEMOLOGICAL assumptions associated with each group. Development of research methods and data collection instruments by the dominant cultural group can result in a relative neglect of individuals' experiences in the nondominant group.

A debate in the cross-cultural research community reflects contrasting positions concerning the extent to which cross-cultural measures can be developed (Evans, 2000). One position is that tests can be transported from one culture to another with appropriate linguistic translation, administration by a native tester, and provision of familiar content. Cultural psychologists adopt an alternative position because they emphasize the importance of symbolic culture, that is, shared value and meaning, knowing, and communication (Greenfield, 1997). Thus, a test can be transported to another culture only if there is universal agreement on the value or merit of particular responses to particular questions, and the same items must mean the same things in different cultures, given a good linguistic translation of the instrument. In addition, the definition of relevant information must be universally the same, and communication with strangers must be acceptable.

To obtain valid data in cross-cultural research, one must determine the local definitions of concepts (Evans, 2000). The measurement of intelligence in Western countries and in rural Kenyan villages serves as one example of the importance of determining local definitions of concepts. Villagers in rural Kenya describe children's intelligence in terms of their ability to identify animals and trees by characteristics, use, and names and by how far they can carry things, such as stools or food, for a specified distance. These are the criteria that villagers use to determine if a child's growth is on track at the age for school entry.

Cultural representation in the research project is important to the conduct of credible cross-cultural research. This principle applies to studies that are conducted within countries where a variety of cultures are present, as well as to cross-national studies. This should entail full participation in the study from the beginning of the process. One example of a cross-cultural study that followed this principle is the ongoing Preprimary Study of the International Association for the Evaluation of Education Achievement (Olmsted & Weikart, 1994). The study is designed to investigate the nature and effects of various kinds of early childhood experiences on the long-term development of young children. The process and instruments for the study were developed by a cross-cultural team of researchers representing each of the countries initially included in the study (i.e., Belgium, China, Finland, Germany, Hong Kong, Italy, Nigeria, Portugal, Spain, Thailand, and the United States). The group decided on the items to include, as well as the examples and wording that would best represent their culture.

In cross-cultural research, validity in measurement can be influenced by the translation of the instrument or by the influence of cultural norms in behaving with a stranger or an elder (Evans, 2000). A linguistically accurate translation may still lead to misunderstanding because of the multiple meanings attached to words. The translation should include a translation not only of the words but also of the meaning. Work with children may especially be influenced by cultural norms if adults do not generally solicit the views of children or if children are not normally expected to speak in the presence of adults. The researcher in such a context might consider a performance assessment that does not involve asking direct questions.

—Donna M. Mertens

REFERENCES

Ember, C. R., & Ember, M. (2001). *Cross-cultural research methods.* Walnut Creek, CA: AltaMira.

Evans, J. L. (2000). *Early childhood counts.* Washington, DC: World Bank.

Gopaul-McNicol, S.-A., & Armour-Thomas, E. (2002). *Assessment and culture: Psychological tests with minority populations.* San Diego: Academic Press.

Greenfield, P. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52*(10), 1115–1124.

Hujala, E. (1998). Problems and challenges in cross-cultural research. *Acta Universitatis Ouluensis, 35,* 19–31.

Olmsted, P., & Weikart, D. P. (Eds.). (1994). *Families speak: Early childhood care and education in 11 countries.* Ypsilanti, MI: High/Scope Press.

CROSS-LAGGED

Cross-lagged models are widely used in the analysis of PANEL DATA, or data collected more than once on the same individuals or units over time, to provide evidence regarding the direction of causality between variables X and Y and to estimate the strength of the CAUSAL effects of each variable on the other. Cross-lagged models involve the estimation and comparison of CORRELATION and/or REGRESSION COEFFICIENTS between each variable measured at one panel wave and the other variable at the *next* panel wave. The underlying logic of such models is that, if X is a cause of Y, then X measured at Time 1 should be related to Y at Time 2; that is, the cross-lagged relationship between X_1 and Y_2 should be nonzero. Furthermore, if X is a more powerful cause of Y than Y is of X, then the cross-lagged relationship between X_1 and Y_2 should be stronger than that between Y_1 and X_2.

A cross-lagged model for a two-wave panel is shown in diagram form in Figure 1. Variables X_2 and Y_2 from Wave 2 are each hypothesized to be determined by their Wave 1 values, the LAGGED value of the other variable, and an ERROR term U; the Wave 1 correlation between variables X and Y is represented by ρX_1Y_1, and the correlation between the equations' error terms is represented by ρU_1U_2. The two equations may be written as follows:

$$Y_2 = \beta_1X_1 + \beta_2Y_1 + U_1, \tag{1a}$$

$$X_2 = \beta_3Y_1 + \beta_4X_1 + U_2, \tag{1b}$$

with the intercepts omitted if the variables are expressed in mean deviation form.

It was originally argued that the causal direction between the two variables could be determined by comparing the magnitude of the two cross-lagged *correlations* in the model, that is, whether ρX_1Y_2 is larger or smaller than ρY_1X_2. Numerous scholars, however, pointed out that the cross-lagged correlations are a function of the causal effects β_1 (or β_3) in Figure 1, as well as the *stability* of each variable over time (i.e., ρX_1X_2 or ρY_1Y_2) and the initial correlation between X and Y (ρX_1Y_1). Consequently, REGRESSION and STRUCTURAL EQUATION MODELS are the more

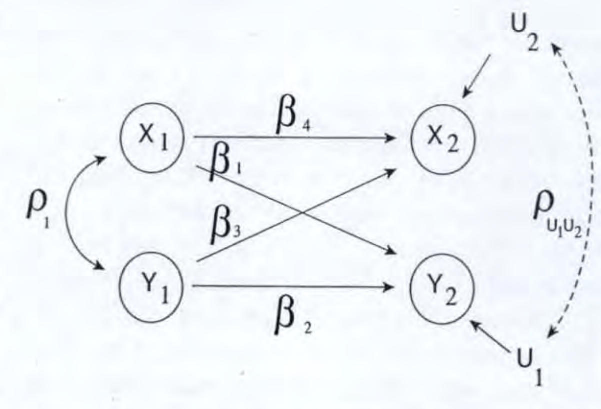

Figure 1 A Two-Wave Panel Model With Cross-Lagged Effects

widely used methods for causal direction and strength, although there are still uses for cross-lagged correlational analysis in testing certain models of spurious association between X and Y (Kenny, 1979).

There are several important threats to causal inference in the cross-lagged panel model. MEASUREMENT ERROR in Y_1 and X_1 will tend to depress the estimates of the stability of Y and X over time, leading to BIAS in the estimates of the cross-lagged effects as well (Kessler & Greenberg, 1981). Second, as if often the case in LONGITUDINAL data, the error terms U_1 or U_2 may be AUTOCORRELATED, leading to a nonzero relationship between Y_1 and U_1 and/or X_1 and U_2. More fundamentally, the cross-lagged model may yield erroneous results because the true causal lag between X and Y is much shorter than the time period between measurements (Plewis, 1985). In such cases, it would be appropriate to specify a "SYNCHRONOUS" or "cotemporal" effects model, in which X_2 and Y_2 are hypothesized to influence each other in the *same* time period. Models with both cross-lagged and synchronous effects may also be specified and estimated either with INSTRUMENTAL VARIABLES or related procedures or, in the multiwave case, by imposing equality or other constraints on the parameters across waves.

—Steven E. Finkel

REFERENCES

Kenny, D. A. (1979). *Correlation and causality.* New York: John Wiley.

Kessler, R., & Greenberg, D. (1981). *Linear panel analysis.* New York: Academic Press.

Plewis, I. (1985). *Analysing change: Measurement and exploration using longitudinal data.* Chichester, UK: Wiley.

CROSS-SECTIONAL DATA

Cross-sectional data come from a study in which all observations for each participant are collected at approximately the same point in time and in which time is not an element in the RESEARCH DESIGN. All observations are considered to have been made concurrently, although literally this is never true. Even within a single session, an individual completes items in a sequence over time, and within a study, all participants do not typically provide data at the same moment. However, the time frame in such a design is typically short, and it is presumed that time of observation during the course of the study had RANDOM effects that produced only error. This is contrasted with LONGITUDINAL designs in which observations are taken at different points in time to see if there were differences in variables due to events occurring over the course of the study.

The term *cross-sectional* is often used to describe QUESTIONNAIRE studies. The simplest is the single-source cross-sectional design in which participants provide all data about themselves, and all items and scales are administered in a single session. The multisource design adds DATA not contained in the participant's questionnaire and might include questionnaires given to other people (e.g., coworker or family member), data provided by observers, or archival data. Another approach could be to divide a long questionnaire into two portions that are administered at different sessions but in which there is no attempt to draw inferences about time.

It is possible to have cross-sectional data without questionnaires and in studies in which the sampling unit is not a person. For example, one might look at relations between different variables in the U.S. census in which all data are considered to have been collected concurrently.

Cross-sectional data can be highly efficient for determining if there are relations between VARIABLES. Their biggest limitation is in not allowing confident CAUSAL conclusions (Shadish, Cook, & Campbell, 2002; Spector, 1994). Without having time built into the design, there is no way to determine the direction of causality between two variables that are related or if there are additional variables that account for the observed relations. For that, one must have either a longitudinal design or an EXPERIMENT.

—Paul E. Spector

REFERENCES

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior, 15,* 385–392.

CROSS-SECTIONAL DESIGN

A cross-sectional design refers to collecting DATA that are assumed to have been collected at one point in time. The objective is to get a "snapshot" or picture of a group. The data hypothetically can represent individuals, groups, institutions, behaviors, or some other unit of analysis, and the data can be collected using a range of procedures. In actuality, the term *cross-sectional design* is often used interchangeably with the term SURVEY and most frequently refers to data that are collected through interviews or questionnaires that are administered to individuals who either represent themselves or who represent households, institutions, or other social entities.

Although not mandatory, cross-sectional designs often assume that the group or population under study is heterogeneous in its characteristics—that is, people of many ages, behaviors, and opinions are represented within the study POPULATION. The objective is to properly "represent" the diversity within the group. With this assumption of heterogeneity come assumptions about the need to appropriately represent or infer to some larger population or group from which the study subjects are drawn when findings are analyzed and conclusions are drawn. As a result, the procedures by which study subjects are selected are of concern. If, within the cross-sectional study, data are not collected from the entire population of interest, then it is important to know how those chosen for study were selected. PROBABILITY or SYSTEMATIC SAMPLING strategies generally are preferred. Because data in cross-sectional designs are collected at one point in time, the emphasis in analysis is to examine the FREQUENCY DISTRIBUTIONS of single

VARIABLES and the ASSOCIATIONS between two or more variables. Researchers sometimes distinguish between descriptive and EXPLANATORY cross-sectional studies, with explanatory studies moving beyond description to examine whether conclusions can be drawn about how certain things influence or result in change (e.g., examining whether exposure to disasters increases psychological distress).

Cross-sectional designs can be contrasted with group comparison or CASE CONTROL designs; LONGITUDINAL, prospective, or PANEL designs; and true EXPERIMENTAL or QUASI-EXPERIMENTAL designs. The objective of longitudinal, experimental, and group comparison designs is to maximize the ability to make PREDICTIONS, investigate CAUSAL linkages, and "explain" why or how something happens, whereas cross-sectional designs are better able to describe relationships between variables. These designs (e.g., experimental) are less concerned with EXTERNAL VALIDITY, or the need to generalize to a larger population, and more concerned with INTERNAL VALIDITY, or the need to accurately measure the relationships between variables or constructs.

Sociologists conducted most of the early research on how to design cross-sectional studies. Paul Lazarsfeld is considered one of the major contributors to early research on cross-sectional studies, and Leslie Kish made major contributions to how participants in cross-sectional studies should be selected or sampled (see Kish, 1965; Lazarsfeld, 1958). Lazarsfeld was particularly instrumental in developing procedures by which causal inferences could be made from cross-sectional data before the advent of computers. His work on the ELABORATION MODEL was extended by Morris Rosenberg (1968).

—Linda B. Bourque

REFERENCES

Aday, L. A. (1996). *Designing and conducting health surveys: A comprehensive guide* (2nd ed.). San Francisco: Jossey-Bass.

Kendall, P. L., & Lazarsfeld, P. F. (1950). Problems of survey analysis. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research: Studies in the scope and method of "the American soldier"* (pp. 160–176). Glencoe, IL: Free Press.

Kish, L. (1965). *Survey sampling*. New York: John Wiley.

Lazarsfeld, P. F. (1958). Evidence and inference in social research. *Daedalus, 87*, 120–121.

Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.

CROSS-TABULATION

A cross-tabulation is a table that displays the RELATIONSHIP between two variables. It is also called a cross-tab, or a CONTINGENCY TABLE. The simplest is a 2 × 2 table, consisting of four CELLS (one for each possible combination of values). Suppose one variable is gender (male = 1, female = 0) and the other variable is vote (yes = 1, no = 0). Gender, the presumed INDEPENDENT VARIABLE, can be placed at the top of the table (as the column variable). Vote, the presumed DEPENDENT VARIABLE, can be placed at the side of the table (as the row variable). The raw frequencies and the percentage frequencies can be placed in each cell; for example, in Cell A, there might be 214 men (55%) who voted. Tables contain more cells as the variables have more categories. Also, the analyst might look at a cross-tab between two variables while controlling on a third variable. For example, one might look at the table for gender and vote within the White subsample and within the non-White subsample.

—Michael S. Lewis-Beck

See also CONTINGENCY TABLE

CUMULATIVE FREQUENCY POLYGON

A cumulative frequency polygon or ogive is a variation on the frequency polygon. Although both are used to describe a relatively large set of quantitative data, the distinction is that cumulative frequency polygons show cumulative frequencies on the y-axis, with frequencies expressed in either absolute (counts) or relative terms (proportions). Cumulative frequencies are useful for knowing the number or the proportion of values that fall above or below a given value.

In the sample cumulative frequency polygon in Figure 1, it is easy to judge the number of males shorter than 68 inches and the number at this height or above. Changing the y-axis values to proportions (200/1,000 = .20,400/1,000 = .40, etc.) would yield the proportion of males shorter than 68 inches and the proportion at this height and above. The same information cannot be obtained so easily in a frequency polygon because the y-axis yields a measure of the

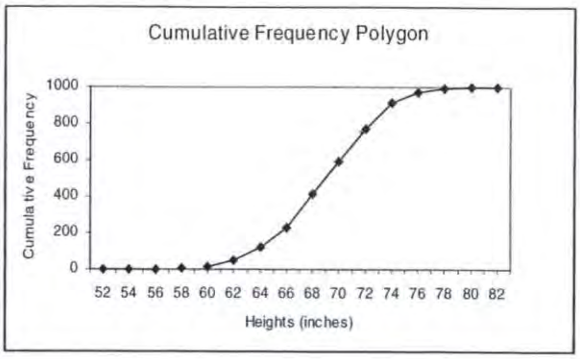

Figure 1 Adult Male Heights Displayed According to the Cumulative Frequency for Each 2-Inch Interval Between 52 and 82 Inches

frequency (absolute or relative) of observations within a single category, not cumulatively across categories.

To create a cumulative frequency polygon, researchers first sort data from high to low and then group them into contiguous intervals. The upper limits to each interval are represented on the x-axis of a graph. The y-axis provides a measure of cumulative frequency. It shows either the number or the proportion of data values falling at or below each upper limit. For absolute frequencies, it is scaled from a minimum of 0 to a maximum that equals the number of measures in the data set. For relative frequencies, it is scaled from 0 to 1.00. Within the field of the figure, straight lines connect points that mark the cumulative frequencies.

Cumulative frequency polygons are used to represent quantitative data when the data intervals do not equal the number of distinct values in the data set. This is always the case for continuous data. When the data are discrete, it is possible to represent every value in the data set on the x-axis. In this case, a cumulative histogram would be used instead of a cumulative frequency polygon. The importance of the distinction is that points are connected with lines in the polygon to allow interpolation of frequencies for x-axis values not at the upper limits (e.g., 67 inches in the figure). When data have no values other than those represented on the x-axis, cumulative frequencies are simply marked by bars (or columns).

Some statisticians reserve the term *cumulative frequency polygon* just for graphs that show the absolute frequencies of values and prefer the term *cumulative relative frequency polygon* for graphs that show proportions. The alternative is to refer, as here, to two types of cumulative frequency polygons—one for counts and one for proportions.

—Kenneth O. McGraw

CURVILINEARITY

Variables in social science research can exhibit a curvilinear relationship to one another, which means that their relationship changes depending on the numerical values each variable takes. Usually, the change is expressed in terms of values of the INDEPENDENT VARIABLE first and then the impact on the DEPENDENT VARIABLE. These changes in SLOPE appear to take the shape of a curve if plotted on a line, as seen on a two-dimensional graph with the independent variable along the x-axis and the dependent variable along the y-axis. Addressing curvilinear relationships is most critical in ORDINARY LEAST SQUARES (OLS) analysis.

An example of this phenomenon from political science is the relationship between economic development (gross national product [GNP] per capita) and democracy across countries. As very poor countries increase their wealth, they also can dramatically increase their levels of democratic performance. Countries at very low levels of wealth can expect to secure for themselves substantial, linearly steep gains in democratic performance. (Naturally, the relationship is not perfect, as no unicausal explanation is satisfactory in accounting for social science phenomena.) However, the initial dramatic gains in democracy become less dramatic as the country becomes much more economically developed. The graphical effect of this curvilinearity is noticeable. The slope of the relationship shifts from steep at low levels of wealth to nearly zero at high levels of wealth, where most countries cluster around the high values of democracy.

For OLS regression to produce estimates that are best linear unbiased estimates (BLUE), the regression assumptions must be met. One of these assumptions is that the independent variables must have linear relationships with the dependent variable. This assumption is clearly violated in the instance of GNP per capita's relationship to democracy. The resulting OLS estimation will produce biased estimates.

The analyst can pursue remedies to address the problem of curvilinearity in OLS regression, all of which are oriented around arithmetically transforming

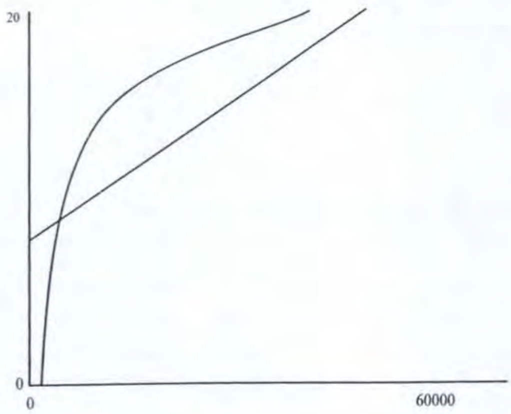

Figure 1 Curvilinear (Versus Linear) Relationship of Economic Development to Democracy

the variable so that the linearity assumption can be met. Social science analysts most commonly employ one of three transformations. The most frequently used is the logarithmic transformation, which can be either to the natural LOGARITHM or to the base 10 logarithm (the effect is the same no matter which one is used). A logarithmic transformation tends to create greater dispersion among tightly clustered data, "linearizing" the scatterplot. The logarithmic transformation linearizes the relationship of economic development to democracy.

Although Figure 1 does not show the actual scatter of points, the curve captures the shape of the distribution of points better than does the straight (linear) line. Statistically expressed, the improvement in fit of the OLS regression model is notable. The R-SQUARED statistic for the linear specification is .27, while the R^2 statistic for the logarithmic specification (X logged) is .30. (The data set has N = 130 countries and excludes two categories of OUTLIER countries: members of OPEC (Organization of Petroleum Exporting Countries) and countries that rely on petroleum exports for more than 50% of their export earnings.)

The other two transformations most commonly employed are the parabolic transformation and the hyperbolic transformation. A parabolic transformation is suggested if the relationship, when graphed, is one of an upside-down U-shaped curve. It involves squaring the independent variable and including it with the untransformed variable in the OLS regression model. The hyperbolic transformation is recommended if there is a threshold for the dependent variable. To take this transformation, researchers should perform the arithmetic operation $1/X$, with X representing the independent variable. These three remedies will assist the researcher in meeting the linearity assumption when the untransformed variables have a curvilinear relationship.

—Ross E. Burkhart

REFERENCES

Burkhart, R. E., and Lewis-Beck, M. S. (1994). Comparative democracy: The economic development thesis. *American Political Science Review, 88,* 903–910.

Draper, N. R. (1998). *Applied regression analysis* (3rd ed.). New York: John Wiley.

Jackman, R. W. (1973). On the relationship of economic development to political performance. *American Journal of Political Science, 17,* 611–621.

Kruskal, J. B. (1968). Transformation of data. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (Vol. 15, pp. 182–192). New York: Macmillan.

Studenmund, A. H. (1992). *Using econometrics: A practical guide* (2nd ed.). New York: HarperCollins.

D

DANGER IN RESEARCH

Researchers potentially face a variety of hazards while conducting fieldwork. These might include accidents, arrest, assault, illness, harassment, and verbal abuse. These risks are probably most common for researchers working in remote locations, in situations of violent social conflict, with some kinds of deviant groups or those employed in dangerous occupations. Female researchers may face risks of sexual assault or sexual harassment. The risks posed by fieldwork shape research agendas by deterring researchers from investigating particular topics or working in particular regions. For example, the dangers involved have probably inhibited research on the violent conflicts occurring in some countries and in some regions. Fieldwork hazard is also strongly implicated in the ethics and politics of field situations, and has an impact on professional practice.

Broadly, dangers are of two kinds—ambient and situational. Ambient danger arises when a researcher is exposed to otherwise avoidable dangers from having to operate in a dangerous setting in order for the research to be carried out. For example, to study the policing of a violent social conflict, a researcher might have to accompany police officers into areas where they are vulnerable to attack. Situational danger arises when the researcher's presence or activities evoke aggression, hostility, or violence from those within the setting. Drug users, for instance, have been known to exhibit aggression toward researchers when under the influence of certain drugs. Researchers working in dangerous settings are probably most vulnerable early on in their field experience, when they still lack situational awareness that makes setting members competent at assessing potential dangers. Research in dangerous situations often depends on the development of an interactional "safety zone," a socially maintained area that extends for a short distance around the field worker and within which the researcher can feel secure and psychologically at ease. Establishment of a safety zone depends in part on the social acceptance of the researcher by those within the setting, as well as on the researcher's ability to size up the possible hazards involved in entering or staying in a particular setting. For example, researchers working with drug abusers might need to be alert to the possibility of fights or disputes erupting around them. They might also need to exercise caution with respect to drive-by shootings and police raids.

Personal style and demeanor can be important in some kinds of potentially dangerous situations. Researchers might find it appropriate to avoid dressing or acting in ways that single them out for undue attention and may also need to adopt a self-confident and determined demeanor to avoid denoting potential victim status to those in the setting. Failure to do so has sometimes resulted in researchers being attacked. Being socially "sponsored" by someone well-known and trusted in the setting can help to avoid danger. In potentially hazardous settings, such individuals can warn the researcher of imminent danger or draw attention to volatile or threatening situations. Some researchers, such as anthropologists who work in remote areas, may face particular hazards. There can be a risk of contracting infectious and parasitic diseases. Falls and vehicle accidents have been identified as common causes of death and injury among anthropologists.

It should not be assumed that research in dangerous settings is impossible. Indeed, the risks involved can sometimes be exaggerated. Fieldwork dangers can be negotiated, but they need to be approached with foresight and planning. It is important to evaluate the possibility of danger, its potential sources, and how it might be managed or exacerbated by the researcher's actions. Traditionally treated rather lightly, often as a source of "war stories," the risks involved in fieldwork are now being discussed more often. Concerns about insurance liability and legal responsibilities toward employees and students have encouraged a more explicit acknowledgment of the risks involved in fieldwork. Institutional responses to fieldwork dangers now increasingly take the form of risk assessments, precautionary guidelines, policy frameworks that specify good practice, managerial responsibilities, and training provision. It is important that researchers are aware of, and contribute to, the development of such responses.

—Raymond M. Lee

See also ETHICAL PRINCIPLES, FIELD RELATIONS, PARTICIPANT OBSERVATION, POLITICS OF RESEARCH

REFERENCE

Lee, R. M. (1995). *Dangerous fieldwork.* London: Sage.

DATA

Data are the recorded empirical observations on the CASES under study, such as the annual percentage changes in the gross domestic product, the number of children in each family, scores on an attitude scale, fieldnotes from observing street gang behavior, or interview responses from old-age pensioners. Social scientists gather data from a wide variety of sources, including experiments, surveys, public records, historical documents, statistical yearbooks, and direct field observation. The recorded empirical observations can be qualitative as well as quantitative data. With quantitative data, the empirical observations are numbers that have intrinsic meaning, such as income in dollars. Qualitative data may be kept in text form, or assigned numerical codes, which may represent an efficient way to record and summarize them. For example, in an anthropological study, a researcher may record whether the market vendor brought produce to the market by cart, animal, motor vehicle, or hand. The researcher may go on to code these observations numerically, as follows: cart = 1, animal = 2, motor vehicle = 3, hand = 4. This coding facilitates storage of the information by computer. Increasingly, computer-assisted qualitative data analysis software is being used. Finally, it should be noted that the word *data* is plural.

—Michael S. Lewis-Beck

DATA ARCHIVES

Data archives are resource centers that acquire, store, and disseminate digital DATA for secondary analysis for both research and teaching. Their prime function is to ensure long-term preservation and future usability of the data they hold.

Data archiving is a method of conserving expensive resources and ensuring that their research potential is fully exploited. Unless preserved and documented for further research, data that have often been collected at significant expense, with substantial expertise and involving respondents' contributions, may later exist only in a small number of reports that analyze only a fraction of the research potential of the data. Within a very short space of time, the data files are likely to be lost or become obsolete as technology evolves.

HISTORY

The data-archiving movement began in the 1960s within a number of key social science departments in the United States that stored original data of survey interviews. The movement spread across Europe, and in 1967, the UK Data Archive (UKDA) was established by the National Research Council for the Social Sciences. In the late 1970s, many national archives joined wider professional organizations: the Council of European Social Science Data Archives (CESSDA) and the International Federation of Data Organizations (IFDO). These were established to promote networks of data services for the social sciences and to foster cooperation on key archival strategies, procedures, and technologies.

The first data archives collected data of specific interest to QUANTITATIVE RESEARCHers in the social sciences. In the 1960s, these were largely opinion poll or election data, but as the trend for large-scale surveys grew, by the late 1970s, the UKDA began to acquire major government surveys and censuses. Because of their large sample sizes and the richness of the information collected, these surveys represent major research resources. Examples of major British government series include the General Household Survey, the Labor Force Survey, and the Family Expenditure Survey. In the United States, key government survey series include the Survey of Income and Program Participation, the Current Population Survey, and the National Health Interview Survey.

By the 1990s, the UKDA collection had grown to thousands of data sets spanning a wide range of data sources. Well-known U.K.-based academic repeated large-scale surveys include the British Social Attitudes Survey, the British Household Panel Survey, and the National Child Development Study. In the United States, the General Social Survey, the Panel Study of Income Dynamics, and the National Longitudinal Survey of 1972 are examples of long-running survey series. Major cross-national series include the World Values and European Values Surveys, the Eurobarometer Surveys, and the International Social Survey Program Series.

In the 1990s, the research community recognized the needs of qualitative researchers by funding a Qualitative Data Archive (Qualidata) at the University of Essex in 1994, and in 1995, a History Data Service (HDS) was established devoted to the archiving and dissemination of a broad range of historical data.

DATA HOLDINGS

Social science data archives acquire a significant range of data relating to society, both historical and contemporary, from sources including surveys, CENSUSes, registers, and aggregate statistics.

Data are created in a wide variety of formats. Data archives typically collect numeric data, which can then be analyzed with the use of statistical software. Numeric data result when textual information (such as answers to survey questions) has been coded, or they may represent individual or aggregated quantities, such as earned income.

The demand for access to digital texts, images, and audiovisual material has meant that material from in-depth interviews, FIELDNOTES to recorded interviews, and open-ended survey questions are now available for computer analysis.

Finally, data sets from across the world are available through national data archives' reciprocal arrangements with other national data archives.

ACQUIRING DATA

A key concern for a data archive is to ensure that the materials it acquires are suitable for informed use and meet demand. All materials deposited are selected and evaluated, and they must meet certain criteria, such as being documented to a minimum standard. Furthermore, acquisitions policies must be flexible and responsive to changes in both data and information needs of the research communities, as well as in the rapidly changing climate of technology.

LONG-TERM PRESERVATION

It is the responsibility of data archives to keep up with technological advances by monitoring hardware and software developments and migrating their collection accordingly. When technology changes, the data in their holdings are technically transformed to remain readable in the new environment. Computer programs are maintained to allow data to be easily transformed from an in-house standard to the various formats required by users.

PREPARING DATA FOR ARCHIVING

On acquisition, data archivists undertake data-processing activities. First, the data set is checked and validated, for example, by examining numeric data values and ensuring that data are anonymous so that the risk of identifying individuals is minimal. Second, "metadata" (data about data) are produced with the aim of producing high-quality finding aids and providing good user documentation. Metadata cover information describing the study and the data and include enhancements added to a numeric data file (such as creating variable and value labels that enable users to view variable information and frequencies). A systematic catalogue record is always created for studies, detailing an overview of the study, the size and content of the data set, its availability, and terms and conditions of access. User guides further contain information on how the data were collected, the original QUESTIONNAIRES, and how to use the data.

PROVIDING ACCESS TO DATA

Data supplied by data archives can be used for reporting of statistics, such as basic frequency counts, and for more in-depth secondary analysis. Secondary analysis strengthens scientific inquiry, avoids duplication of data collection, and opens up methods of data collection and measurement. Reusing archived data enables new users to ask new questions of old data; undertake COMPARATIVE RESEARCH, REPLICATION, or restudy; inform RESEARCH DESIGN; and promote methodological advancement. Finally, data can provide significant resources for training in research and substantive learning.

Users typically request data in a particular format, such as a statistical or word-processing package. These days, data can be accessed via instant Web download facilities or can be dispatched on portable media such as CD-ROM. The 21st century has seen a move toward sophisticated online analysis tools, where users can search and analyze subset data via a Web browser, such as NESSTAR.

Data supplied by data archives are normally anonymous to ensure that the risk of identifying individuals is minimal. Users are typically required to sign an agreement to the effect that they will not attempt to identify individuals when carrying out analyses.

—Louise Corti

REFERENCES

ICPSR. (2002). *Guide to social science data preparation and archiving* [online]. Available: http://www.ifdo.org/archiving_distribution/datprep_archiving_bfr.htm.

IFDO. (n.d.). [online]. Available: http://www.ifdo.org/.

Ryssevik, J., & Musgrave, S. (2001). The social science dream machine: Resource discovery, analysis, and delivery on the Web. *Social Science Computing Review, 19*(2), 163–174.

UK Data Archive (n.d.). [Online]. Available: http://www.data-archive.ac.uk/.

DATA MANAGEMENT

We need to manage research data in a way that preserves its utility over time and across users. This means we must pay attention to data documentation and storage so that, at a minimum, others can access our data; replicate our analyses; and, preferably, extend them. How we achieve these goals depends upon whether we produce the data (primary data collection) or simply reuse it (secondary data analysis; see SECONDARY DATA, SECONDARY ANALYSIS OF SURVEY DATA), as well as what kind of data we are dealing with.

As we discuss managing social science data, we distinguish between the two most common types of data: SURVEY data and records data. Social scientists deal with other types, such as data from EXPERIMENTS and audiovisual recordings. Audio or video recordings create their own data management problems, and people in the broadcast industry likely have better insights to management and archival issues, so we will not attempt to give advice in this area. Increasingly, economists collect experimental data from laboratory settings to learn how economic agents react to controlled economic opportunities. These experiments involve customized software that will, itself, generate the data for analysis. In general, many of the same principles that apply to survey data apply to experimental data, but the applications are so specialized that we will say little more about experimental data.

The primary data collector bears the most onerous and important data management responsibilities; in order for others to exploit the data, the collector must successfully execute the primary data management task. The data management task begins with a plan that encompasses the entire project from data collection through analysis. Naturally, the sort of data collected dictates the plan for data management. For secondary data, the management task begins downstream from collection, and the scope of the data management problem crucially depends on the sorts of data resources that the collector has passed along to the user. It is likely that the best strategy for the data manager of secondary data is to start with the data system used by the primary data collector.

For a variety of data management applications, the relational database technology dominates other strategies. Most social scientists are not familiar with how these systems work and therefore tend to manage their data within software systems with which they are familiar, such as SAS, SPSS, or Stata. For small projects, this may be the only feasible approach; it does not make sense to learn a new technology that might be used for only a few months, unless one foresees a career requiring strong data management skills.

We will begin our discussion of data management strategies with records data. These data are especially well suited to relational databases, and applications that exemplify the technique are easier to understand in the context of records data. Although survey data are perhaps more widely used, the reader will probably better understand how relational database methods can be used with survey data after working through an example of research data derived from administrative records.

ADMINISTRATIVE RECORDS DATA

Records data come from many sources and almost always originate with someone other than the researcher. The appropriate data management structure with records data depends upon the complexity of the data. Typical records data have a simple structure and can be maintained within the statistical package used for analysis. For example, national income account and other macroeconomic data have a fixed set of variables that repeats for various time periods. Gross national product data are quarterly, unemployment rates are monthly, and so forth. Data on stock transactions are available in a variety of temporal aggregations and even disaggregated as the full series of transactions.

However, some records data can be very complex. For example, states have massive databases on earnings reported for unemployment insurance purposes, tax returns, and people receiving benefits from transfer programs such as TANF (the former cash assistance for poor families with children, a program that has a variety of state-specific names following the reform of the welfare system in the late 1990s), Medicaid, and food stamps. These data cover both individuals and families with records from applications, payments, and administrative actions.

Let us consider the example of transfer program administrative data to describe how a relational database strategy allows the researcher to organize the data for analysis. We will simplify the example for exposition, so the reader should not infer that any particular state organizes its data in the manner we describe.

Relational databases consist of a set of tables, each of which contains a fixed set of variables. A relational database does not require the tables to be related in any particular way, nor does it require the data to be combined and extracted in any particular way. If a table contains data on a person (or company, transaction, etc.), we will use the term *row* to refer to the set of variables for that table that applies to that particular person (or company, transaction, etc.). This flexibility makes relational databases excellent candidates for archiving and retrieving complex data from a wide range of applications. Because household size and income determine the size of welfare payments, one likely table in a database for welfare data would describe the person receiving payments under a particular program. Let us use the term *assistance group* to describe the set of people eligible for benefits under a particular program. In a single household, a person may be part of one assistance group for TANF purposes and another group for food stamp purposes. Our hypothetical state will, as a result, have data describing each assistance group for a type of benefit, with each assistance group having a unique identifier. A TANF assistance group data set might have a table that identifies the name; address; demographic characteristics (date of birth, ethnicity, sex, citizenship, etc.); Social Security number; and possibly a state identification number for the person receiving benefits on behalf of each assistance group. Another set of tables would document the checks issued to the assistance groups, with one record containing the assistance group identifier, the date a payment was made, the size of the payment, and indicators for any special administrative details relating to the payment. Similarly, there might be a table that stores the outcome of eligibility reviews or records from visits by a caseworker.

Another table in the database might relate to individuals. For example, for each month that a person generates a payment to an assistance group, there might be a row in the table that contains the person's name, date of birth, sex, ethnicity, disability status, the program under which a payment was made, and the identifier of the assistance group to which the payment was made for the benefit.

States also maintain the database showing quarterly earnings covered by the unemployment insurance program. A table in this database would contain the Social Security number of the worker, the Employer Identification number for the firm, the quarter for earnings, and the amount of the worker's earnings.

From this latter set of records, we could construct a quarterly history of the earnings for any person in the state. Similarly, we could construct a month-by-month history for any person in the state that showed in which months the person generated food stamp payments or TANF payments. It is also possible to create an event history of any person's membership in an assistance group and how that relates to earned income. We could also use these data to determine whether the child becomes the beneficiary of benefits as a part of a new assistance group when a child's parent loses welfare benefits. Data from different tables can be *joined* together by matching rows that share a common value of some identifier, such as a Social Security number. *The advantage of a relational database is not that it is in the right format for any particular analysis, but that it allows the researcher to combine data elements together into the right format for virtually any analysis.* This flexibility explains its dominance for large commercial applications that range from e-commerce Web sites to accounting and inventory systems. Frequently, companies integrate *all* their record systems within a collection of tables of a large relational database, providing extraordinary power to track activity within the enterprise. In a well-designed relational database, all redundant data items are eliminated. This reduces both maintenance costs and retrieval errors.

When researchers know exactly what data they need and in what format it must be presented, they are likely to extract from records data only the variables they think they need and to move the data into a form convenient for their analysis strategy (e.g., readable by their statistical package). Such researchers are unlikely to use a relational database, and if the project develops as expected, this will be the right decision. However, whatever work such researchers have done with the primary data may be of little value to others using the data who have a different vision of which data are needed and how they should be organized. This violates the basic goal we set out for data management: Preserve the utility of the data across time and across users. Relational databases may not be the ideal data management system for any particular application, but they will be very good systems for almost any application.

SURVEY DATA

Survey data often possess a complex structure. Moreover, survey data are often collected with secondary analysis in mind, and for the most influential surveys, secondary analysis is the primary objective. Surveys are natural applications for relational database techniques because they are frequently made up of sets of tables of variables that are related to one another in a variety of ways.

The samples themselves most often come from multistage STRATIFIED random SAMPLES rather than the SIMPLE RANDOM SAMPLING we see in textbooks. In addition, we often interview multiple people from a unit chosen from the SAMPLING FRAME. It is essential that the data file preserve the relationships between the data and the sampling structure and respondent relationships. For example, in area PROBABILITY SAMPLING, this means retaining for each observation the identification of the primary sampling unit and substratum, including census tract and block group identifiers, along with the probabilities of selection at each step in the sampling process. These geographic attributes define the way in which various observational units (respondents) are connected to one another. The recent explosion in the use of Geographic Information Systems (GIS) data argues strongly that the data collector capture the longitude and latitude for dwelling units either using a global positioning satellite receiver or by geocoding the street address to latitude and longitude using a package such as Maptitude or ArcView. Latitude and longitude data are the basis for organizing most GIS data, so these should be the primitives of location, not block group, census tract, or some other indexing system.

Survey data frequently utilize complex questionnaires, and this creates serious data management problems. Indeed, as computer-assisted techniques dominate the industry, instruments that heretofore would have been impossible for interviewers to administer have become the rule. We encourage the reader interested in survey data to also read the entry on COMPUTER-ASSISTED PERSONAL INTERVIEWING, which deals with many of the same issues and, together with this entry, provides a holistic view of the survey data management process from design through data dissemination.

Instruments often collect lists, or rosters, of people, employers, insurance plans, medical providers, and so on and then cycle through these lists asking a set of questions about each person, employer, insurance plan, or medical provider in the roster. These sets of related answers to survey questions constitute some of the tables of a larger relational database where the connections among the tables are defined by the design of the questionnaire.

One can think of each question in a survey as a row within a table with a variety of attributes that are linked in a flexible manner with other tables. The attributes (or "columns") within a question table would be the following:

- The question identifier
- Descriptors that characterize the content of the question (alcohol use, income, etc.)

- The question text
- A set of questions or check items that leads into the question
- A set of allowable responses to the question and data specifications for these allowable responses (whether the answer is a date, time, integer, dollar value, textual response, or a numerical value assigned to a categorical response, such as 1 = yes, 0 = no)
- Routing instructions to the next question or check item that are contingent on the response to the current question
- Instructions to assist the interviewer and respondent in completing the question
- Alternate language versions for question attributes that are in text form
- Comments about the accuracy or interpretation of the item or its source

We often refer to these attributes of questions as "metadata." In complex surveys, these pieces of information that describe a question are connected to that question. For example, metadata include which questions lead into a particular question, and questions to which that question branches. These linkages define the flow of control or skip pattern in a questionnaire. With a sophisticated set of table definitions that describes a questionnaire, one can "join" tables and rapidly create reports that are CODEBOOKS, questionnaires, and other traditional pieces of survey documentation.

If we break down the survey into a sequence of discrete transactions (questions, check items, looping instructions, data storage commands, etc.) and construct a relational database, with each transaction being a row in a database table and the table having a set of attributes as defined in the relational database, we can efficiently manage both survey content and survey data.

Relational database software is a major software industry segment, with vendors such as Oracle, Sybase, IBM, and Microsoft offering competitive products. Many commercial applications use relational database systems (inventory control; accounting systems; Web-based retailing; administrative records systems in hospitals, welfare agencies, and so forth, to mention a few), so social scientists can piggyback on a mature software market. Seen in the context of relational databases, some of the suggested standards for codebooks and for documenting survey data, such as the data documentation initiative, are similar to relational database designs but fail to use these existing professional tool sets and their standard programming conventions.

The primary data collector has several data management choices: (a) Design the entire data collection strategy around a relational database, (b) input the post-field data and instrument information into a relational database, or (c) do something else. For simple surveys, the third alternative likely translates into documenting the survey data using a package like SAS, SPSS, or Stata. These are excellent statistical packages, and one can move data among them with a package like Stata's Stat/Transfer. Statistical packages are, themselves, starting to incorporate relational database features. For example, SAS supports SQL (standard query language) queries to relational databases, and it also connects to relational databases. The trend for many years has been toward relational databases to manage databases, starting with the largest, most complex projects. When setting up large projects, social scientists would do well to build their data management strategies and staff around relational database management systems.

—Randall J. Olsen

See also QUALITATIVE DATA MANAGEMENT

REFERENCES

Elmasri, R. A., & Navathe, S. B. (2001). *Fundamentals of database systems.* New York: Addison-Wesley.

Gray, J., & Reuter, A. (1992). *Transaction processing: Concepts and techniques.* San Francisco: Morgan Kaufmann.

Kroenke, D. M. (2001). *Database processing: Fundamentals, design and implementation.* Upper Saddle River, NJ: Prentice Hall.

Stern, J., Stackowiack, R., & Greenwald, R. (2001). *Oracle essentials: Oracle9i, Oracle8i and Oracle 8.* Sebastopol, CA: O'Reilly.

DEBRIEFING

Originally a military term, *debriefing* means questioning or instructing at the end of a mission or period of service. As used in human research, debriefing refers to a conversation between investigator and subject that occurs after the research session. Debriefing is the post-session counterpart of INFORMED CONSENT and should be conducted in a way that benefits and respects the subject.

Debriefing may have several purposes. Generally, it is an opportunity for the subject to ask questions and for the investigator to thank the subject for participating, more fully explain the research, and discuss the subjects' perception of the research experience. Research participation can have educational or therapeutic value for participants, and debriefing is an appropriate opportunity to consolidate these values through appropriate conversation and handouts. Debriefing also provides an important opportunity for the researcher to learn how subjects perceived the research and why they behaved as they did. The perceptive researcher may learn as much from the debriefing as from the research itself. For these values of debriefing to be realized, it is essential that the researcher prepare appropriate information to be conveyed, provide a respectful opportunity for the subject to ask questions and express reactions, and allow an unhurried period in which the interested subject and appreciative investigator can interact. The debriefing should be planned and scheduled as an integral part of the research, and it should be appropriate to the circumstances. For example, if the research was about a private or sensitive matter, the debriefing should occur privately. In the case of research on children or on adults of limited autonomy, the subject's parent or guardian, as well as the subject, should be debriefed appropriately.

Sometimes, investigators promise to send subjects the results of the study rather than discussing the study with them after their session. There are many problems with this arrangement. Such promises are often broken. Completion of the research typically occurs so much later that it may be impossible to re-contact the subjects; if re-contact occurs, the information arrives out of context and after the experience has been forgotten. Besides, the results of a single study are often meager and of limited interest to subjects. Of greater value is the knowledge that the researcher gained from the literature search that (should have) preceded the study and that can easily be summarized in layman's language during the debriefing.

When the research involves concealment or DECEPTION (any procedure that causes the subject to believe something that is not so), the debriefing should include explaining the nature of the deception and why it was employed (*dehoaxing*). After a deceptive or stressful session, the debriefing should also include *desensitizing*, an effort to return the subject to a positive emotional state. Deception is easy to debrief when honest, straightforward induction involves having subjects

- Consent to participate in one of various specified conditions (as in placebo research)
- Consent to deception
- Waive their right to be informed (see Sieber, 1992, for a discussion of these methods)

Such subjects expect to have the real situation revealed in a debriefing. In contrast, subjects who were falsely informed may react negatively to the revelation that they have been fooled, and they may wonder whether they can believe the dehoaxing. Desensitizing, under such conditions of mistrust, may be impossible. Similarly, subjects who do not know they have participated in research may cling to the belief that they had engaged in a naturally occurring (uncontrived) situation, and may be difficult to debrief. Whatever device was used to deceive (e.g., a confederate, misrepresentation of an object or procedure), a convincing demonstration of the deception is usually easy to arrange. For example, if false feedback was given on a test performance, the dehoaxing could include giving subjects their own tests, still in sealed envelopes as subjects had submitted them.

WHEN NOT TO DEHOAX

Dehoaxing may involve explanation of the research design and procedures, including the hypotheses and the comparisons that were being made. However, debriefing should be appropriate to the culture and level of sophistication of the subjects and should not involve blaming, demeaning, or confusing subjects. Hence, sophisticated details may be inappropriate, as in situations such as the following in which subjects most likely would be unable to understand or benefit from a detailed dehoaxing:

- A study of children's ability to resist temptation to take (steal) items that do not belong to them, as a function of stage of moral development
- A study of child-rearing practices and child aggression that compares populations of parents who espouse corporal punishment with parents who espouse other approaches
- A study of explanations of illness by members of cultures that tend to invoke explanations involving evil spirits or witchcraft

It might be appropriate to provide a simple, nonjudgmental description of the research, such as, "We wanted to learn the various methods parents use to discipline their children. We appreciate your letting us observe how you interact with your child." If subjects express a desire to know more, additional nonjudgmental information could be offered, such as, "Here are some of the kinds of methods we observed."

Designing appropriate debriefing requires sensitivity to the needs, culture, and feelings of subjects. The main purposes of debriefing are to thank subjects for participating; answer their questions; learn their reactions; and, in a respectful way, convey honest and perhaps educationally beneficial information to subjects.

—Joan E. Sieber

REFERENCES

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47,* 1597–1611.

Holmes, D. (1976). Debriefing after psychological experiments: Effectiveness of post-experimental desensitizing. *American Psychologist, 32,* 868–875.

Sieber, J. E. (1992). *Planning ethically responsible research.* Newbury Park, CA: Sage.

DECEPTION

Deception takes the form of a lapse or calculated misrepresentation in the process of informing participants that research is taking place and of its nature, purpose, and consequences.

Some measure of deception is widely practiced in social research, although it is not always identified as such. The story given to participants to inform their consent is often partial; investigators fear that they will lose willing respondents if they reveal the whole truth. They also use misleading descriptions of their interests, for example, talking of recreation and health when they mean sex and drugs.

In the 1960s and 1970s, however, before social researchers had developed a collective conscience, deception was rather more blatant; it often took the form of the adoption of a role, disguise, or false identity by investigators entering a research field. For example, Rosenhan (see Bulmer, 1982) and his collaborators simulated the symptoms of insanity in order to gain admission to mental institutions and test screening procedures. Howard Griffin took a medication to change the color of his skin in order to conduct a pseudo-participant study of being black in the American Deep South; contrary to expectation, this procedure proved to be irreversible. In order to secure the cooperation of a pastor of an Afro-Caribbean congregation in Bristol, Ken Pryce signified beliefs he did not hold and allowed himself to be baptized (Homan, 1992).

Stanley Milgram's infamous studies of obedience and authority involved deception of a different kind. Milgram misled those he hired into thinking they were his assistants; they were then inveigled into believing that they were inflicting pain on other participants, who were, in reality, trained actors. Subjects suffered various adverse aftereffects as they came to terms with their willingness to cause pain and distress.

Milgram's case and others have prompted concerns about the effect deception has upon those who practice it and upon the reputation of research professionals. At the individual level, deceit and betrayal may become habitual, and distrust endemic. For the community of researchers, there is a risk that they gain a reputation for deviousness and fraudulence. The maintenance of an unfamiliar role puts considerable strain upon those who are trained in social science rather than in drama school. A change of color, residence in a mental hospital, or the company of a criminal gang are not undertakings one would wish upon oneself, let alone prescribe for one's students.

But whether deception in social research is always morally wrong is an open question. We sympathize with participants in the obedience experiments who suffered nervous aftereffects. We turn on those who conducted these projects and reckon that a general taboo on deception would have forestalled these consequences. But when deception is used to bypass the barriers of powerful or socially disapproved groups like the Ku Klux Klan, the outrage is rather muted. In the instance of groups that can look after themselves and resist conventional forms of inquiry, there may be an argument for deceptive strategies. Researchers must consider the conflicting responsibilities they have to their subjects and to the publication of truth.

—Roger Homan

REFERENCES

Bulmer, M. (Ed.). (1982). *Social research ethics.* New York: Holmes and Meier.

Homan, R. (1992). *The ethics of social research.* London: Longman.

Lauder, M. (2003). Covert participant observation of a deviant community: Justifying the use of deception. *Journal of Contemporary Religion, 18*(2), 185–196.

Milgram, S. (1974). *Obedience to authority.* London: Tavistock.

DECILE RANGE

The decile range is calculated in a similar way, and has a similar role, to the INTERQUARTILE RANGE. A frequency distribution is arrayed from low to high and divided into 10 equal groups in terms of the number of cases in each group. The first and last decile are removed, and the decile range is the difference between the lowest and the highest values that are left. As with the interquartile range, the decile range is a way of dealing with the problem of OUTLIERS that affects the RANGE. The term is sometimes used, though incorrectly and inaccurately, as a synonym of "decile."

—Alan Bryman

DECONSTRUCTIONISM

Deconstructionism is concerned with identifying and problematizing the categories and strategies that produce particular versions of reality. Texts are taken apart to observe how they are constructed in such a way as to exhibit particular images of people, places, and objects. French social theorist Jacques Derrida (1976) first introduced deconstructionism to philosophy in the late 1960s as part of the poststructuralist movement. Although closely linked to literary theory, deconstructionism has also found resonance in the study of architecture, art, culture, law, media, politics, and psychology.

Rooted in close readings of a text, deconstructionism enables us to identify how the construction of the text is dependent on contradictions, confusions, and unstated absences. Derrida objects to the logic of fixed categories that have dominated Western thought and philosophy, especially the either/or logic of binary oppositions such as mind/body, masculine/feminine, health/illness, and individual/society. Instead of the either/or logic of binary oppositions, Derrida argues that we should accept the logic of both/and. Phenomena are best understood when we consider what they appear to both include and exclude. For example, the individual/society opposition can be alternatively understood as conjoined features of a system, where each component requires the other to exist and to make sense (Burr, 1995). Oppositions serve as devices to highlight ethical, ideological, and political processes at work in a text.

Ian Parker (1988) recommends three steps or stages to deconstructionist analysis. Step 1 involves identifying BINARY oppositions or polarities in the text and illustrating the way in which one particular pole is privileged. Step 2 shifts the subordinate pole of the opposition into the dominant position. This serves to demonstrate how the privileged pole depends on and could not exist, and would not function, without the other opposition. Finally, in Step 3, the opposition is reconsidered, and new concepts and practices are developed. Parker has used these stages of deconstructionism when analyzing the discourse of a British radio soap opera, *The Archers*. Through an analysis of transcripts of this soap opera, Parker contests the distinctions between reality and fiction, and everyday explanations and theoretical explanations that are used when discussing gender relations.

Deconstructionism is controversial. Derrida's work has been praised for making one of the most significant contributions to contemporary thinking while simultaneously being denounced as corrupting widely accepted ideas about texts, meanings, and identities. Proponents of deconstructionism argue that it provides a novel and stimulating approach to the study of language and meaning, and that it can disrupt theories and serve to open up debate and conflicts. Equally, deconstructionism can serve to change power relations and can encourage subversion of dominant ideas. Opponents, however, claim that deconstructionism is concerned with verifying that which cannot be verified or proven. Verification through deconstructionism is immensely difficult given that its own object of inquiry is unstable. It is impossible to locate the meaning of a text because it consists of an endless chain of signification. Therefore, deconstructionist analysis is limited to merely exemplifying meaning in a text. Deconstructionism has also been criticized for prioritizing the role of texts when understanding meaning making and power struggles, to the detriment of considering the role of the audience.

—Sharon Lockyer

REFERENCES

Burr, V. (1995). *An introduction to social constructionism*. London: Routledge.

Derrida, J. (1976). *Of grammatology* (G. Spivak, Trans.). Baltimore, MD: Johns Hopkins University Press.

Parker, I. (1988). Deconstructing accounts. In C. Antaki (Ed.), *Analysing everyday explanation: A casebook of methods* (pp. 184–198). London: Sage.

DEDUCTION

In logic, deduction is a process used to derive particular statements from general statements. The logic is used in the social sciences to test THEORY. A HYPOTHESIS is deduced from a theory and is tested by comparison with relevant DATA. If the data do not agree with the hypothesis, the theory is rejected as false. If the data agree with the hypothesis, the theory is accepted provisionally; it is corroborated. Testing such a hypothesis is assumed to be a test of the theory itself.

Notions of deduction go back to antiquity, to Euclidean geometry and Aristotelian logic. Deduction is an essential aspect of the HYPOTHETICO-DEDUCTIVE METHOD, which is associated with Karl Popper's FALSIFICATIONISM. Popper advocated the logic of deduction to overcome the deficiencies of INDUCTION. The core of his argument is that because observations, or data, do not provide a reliable foundation for scientific theories, and because inductive logic is flawed, a different logic is needed for developing theories. His solution was to accept that all data collection is selective and involves interpretation by the observer, and then to use a logic that is the reverse of induction.

Instead of looking for confirming evidence to support an emerging generalization, as occurs in induction, Popper argued that the aim of science is to try to refute the tentative theories that have been proposed. The search for truth is elusive because we have no way of knowing when we have arrived at it. All that can be done is to eliminate false theories by showing that data do not fit with them. Theories that survive the testing process are not proven to be true because it is possible that further testing will find disconfirming evidence. It may be necessary to discard a theory, or at least to modify it and then conduct further research on it. Therefore, science is a process of conjecture and refutation, of trial and error, of putting up a tentative theory and then endeavoring to show that the theory is false.

An example of what a deductive theory might look like comes from a reconstruction of Durkheim's theory of egoistic suicide. It consists of five propositions using three main concepts: suicide rate (the number of suicides per thousand of a population); individualism (the tendency of people to think for themselves and to act independently, rather than to conform to the beliefs and norms of some group); and Protestantism (a collection of Christian religious groups formed following the Reformation and the subsequent fragmentation of Roman Catholicism). The five propositions are as follows:

1. In any social grouping, the suicide rate varies directly with the degree of individualism (egoism).
2. The degree of individualism varies directly with the incidence of Protestantism.
3. Therefore, the suicide rate varies with the incidence of Protestantism.
4. The incidence of Protestantism in Spain is low.
5. Therefore, the suicide rate in Spain is low (Homans, 1964, p. 951).

The conclusion to the argument becomes the hypothesis to be tested. If the hypothesis is refuted, the theory is assumed to be false. By changing the country, for example, predictions can be made and the hypothesis tested in other contexts.

Like inductive logic, deductive logic has been severely criticized. Some major criticisms are the following: If observations are interpretations, and we can never observe reality directly, regularities cannot be established confidently nor theories refuted conclusively; the tentative acceptance of a yet-unrefuted theory requires some inductive support; paying too much attention to logic can stifle scientific creativity; and the process of accepting or rejecting theories involves social and psychological processes, not just logical ones (see Blaikie, 1993, and Chalmers, 1982, for critical reviews).

—Norman Blaikie

REFERENCES

Blaikie, N. (1993). *Approaches to social enquiry.* Cambridge, UK: Polity.

Chalmers, A. F. (1982). *What is this thing called science?* St. Lucia: University of Queensland Press.

Homans, G. C. (1964). Contemporary theory in sociology. In R. E. L. Faris (Ed.), *Handbook of modern sociology* (pp. 951–977). Chicago: Rand McNally.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

DEGREES OF FREEDOM

The number of independent observations available for PARAMETER ESTIMATION indicates the degrees of freedom, *df*. Generally speaking, a test statistic has degrees of freedom, determined by mathematical constraints on the quantities to be estimated. As examples, degrees of freedom must be calculated with the T TEST, the F RATIO, the CHI-SQUARE, the STANDARD DEVIATION, or REGRESSION analysis. In the case of regression analysis, that number may be calculated by subtracting from the sample size the number of coefficients to be estimated, including the constant.

Consider the simple case of estimating the STANDARD DEVIATION of a population variable, from a random sample of N observations. The formula is as follows:

$$SD = \sqrt{\frac{\sum(U - \bar{U})^2}{N - 1}},$$

where SD = the standard deviation estimated in the sample $\sqrt{\ }$ = square root, $\sum$ = sum, U = sample observations on the population variable, $\bar{U}$ = the average score of the sample observations on U, and N = the sample size. Note that the numerator is divided by $N - 1$, not N. Substantively, this adjustment may appear small. Nevertheless, it is theoretically important. The correction is necessary to avoid an exaggerated, BIASED estimate of the population standard deviation. In calculating this 1 population parameter, we have exhausted 1 degree of freedom. This is so because of the characteristics of the sample mean, used in the formula. There are N number of $(U - \bar{U})$ scores. However, just $N-1$ of those are independent because once they are computed, the last, the Nth value, is necessarily fixed. This comes from the fact that always $\sum(U - \bar{U}) = 0$. For instance, suppose $N = 3$, and $(U_1 - \bar{U}) = 6$, $(U_2 - \bar{U}) = 12$. Then, it must be that $(U_3 - \bar{U}) = -18$. The restriction on the deviations from the mean, that they sum to zero, dictates that the last deviation is not "free" but fixed, so 1 degree of freedom is lost. In the example, only one population parameter, the standard deviation, is estimated, leaving degrees of freedom = $N - 1$.

In the more complicated case of regression analysis, the $df = N - K$, where N = the sample size and K = the number of parameters to be estimated. Suppose the MULTIPLE REGRESSION model of $Y = a + bX + cZ + e$, where $N = 45$. There are three parameters to be estimated—the intercept a and the two slopes, b and c. Hence, $df = N - K = 45 - 3 = 42$. With regression analysis, it is important to have a sufficient number of degrees of freedom. Take the extreme, when the SIMPLE CORRELATION (REGRESSION) model $Y = a + bX + e$ is fitted to a sample with $N = 2$. Here the R-SQUARED = 1.0, because the straight line of the ORDINARY LEAST SQUARES fit connects the two data points without error. This is not a perfect explanation but mere mathematical necessity, for the degrees of freedom have been exhausted, that is, $N - K = 2 - 2 = 0$. Generally, in regression analysis, one wants many more independent observations than independent variables. When K begins to approach N, analysts pay special attention to the ADJUSTED R-SQUARED, which corrects for lost degrees of freedom.

—Michael S. Lewis-Beck

REFERENCES

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction.* Thousand Oaks, CA: Sage.

DELETION

Deletion is a method for dealing with MISSING DATA, which are usually eliminated (or transformed) before analysis begins. The most common type of missing data deletion by computer is called LISTWISE deletion (also called casewise deletion). With listwise, a case is eliminated if it has a missing value on any of the variables under analysis. An alternative to listwise deletion is PAIRWISE deletion, where a case is eliminated only if it has missing value(s) on a pair of variables being studied. Analysis might then proceed, for example, from the estimated CORRELATION matrix. The pairwise strategy has the advantage of keeping up SAMPLE size, but has the disadvantage that it can produce statistical inconsistencies in the results.

—Michael S. Lewis-Beck

DELPHI TECHNIQUE

This is a qualitative method for obtaining consensus among a group of experts. Named after the Delphic oracle, skillful at forecasting the future in ancient times, it was invented by the RAND Corporation in the 1960s for forecasting the probability, desirability, and impact of future events. Its use has now been extended to a wide range of situations where convergence of opinions is desirable. There are now three distinct forms: *exploratory Delphi,* as developed by RAND; *focus Delphi,* seeking views of disparate groups likely to be affected by some policy; and *normative Delphi,* gathering experts' opinions on defined issues to achieve consensus (e.g., to set goals and objectives). The essential element in the Delphi process is anonymity of participants when giving their opinion, which alleviates problems that could be caused by domination of the group by a few powerful individuals. Quantitative estimates can also be obtained (e.g., spending priorities and relative allocations of money).

The Delphi process proceeds in an iterative way as follows:

Round 1. Participants are invited to comment on a subject or a set of opinions provided by the facilitating group. Opinions should be based on participants' personal knowledge and experience. The facilitators then analyze the results and provide feedback, such as graphical summaries, rankings, or lists under a limited number of headings. The results are then fed back to participants as input into a second-round questionnaire.

Round 2. In view of the feedback based on Round 1, participants are asked to comment on the summarized views of others and also reconsider their opinions. As before, the facilitators analyze the opinions received, and these are again fed back to participants as input into the questionnaire for the next round. This process of repeated rounds continues until "adequate" convergence is obtained. Usually, three rounds are sufficient. Often, in addition to scoring their agreement with results, participants are asked to rate the degree of confidence they have in their opinions—this can be used in weighting responses to produce summary results, provided the more expert participants do not turn out to be the most modest. With the advent of the Internet, and, in the future, The Grid, there is now the opportunity to automate much of the facilitators' analyses and speed up the process.

It is important that participants be, in some sense, "experts" for valid results to be reached, and there should be sufficient participants that the extreme views of a few do not influence the results for the group disproportionately. Users of a service can be considered "experts" in what they want from a service, but professionals may be better informed as to potential risks and benefits. Both can be accommodated in the same study, with each group being asked to take into account the other's views (e.g., see Charlton, Patrick, Matthews, & West, 1981). Consideration also needs to be given at the start as to what is fed back to participants after each round, and how "agreement" is defined, including the treatment of outliers. The validity and acceptability of the results need to be assessed at the end of the study. The Delphi method may be most appropriate when opinions are being sought, rather than in situations where evidence can be synthesized using conventional statistical methods. Delphi may be seen more as a method for structuring group communication than providing definitive answers.

—John R. H. Charlton

REFERENCES

Charlton, J., Patrick, D. L., Matthews, G., & West, P. A. (1981). Spending priorities in Kent: A Delphi study. *Journal of Epidemiology and Community Health, 35,* 288–292.

Linstone, H. A. (1978). The Delphi technique In R. B. Fowles (Ed.), *Handbook of futures research.* Westport, CT: Greenwood.

Jones, J. M. G., & Hunter, D. (1995). Consensus methods for medical and health services research. *British Medical Journal, 311,* 376–380.

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F. B., Ashkam, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development [entire issue]. *Health Technology Assessment, 2*(3).

Pope, C., & Mayes, N. (1995). Qualitative research: Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *British Medical Journal, 311,* 42–45.

DEMAND CHARACTERISTICS

In some research, typically laboratory-based experimental research, the expectations of the researcher are transparent, the hypotheses under investigation are discernable, and the hoped-for response is obvious. Such research contexts are said to be characterized by high degrees of experimental demand characteristics. Demand characteristics are cues provided participants (by the experimenter, the research context, etc.), often unintentionally and inadvertently, regarding the appropriate or desirable behavior. A smile by the researcher when the participant has made the sought-for response, a slight nod of the head, a minor change in vocal inflection—all or any of these actions might bias the outcome of an investigation. Demand characteristics seriously undermine the INTERNAL and EXTERNAL VALIDITY of research because they can cause outcomes that are misattributed to the experimental treatment. When variations in results are the result of demand effects, the interpretation that implicates the experimental treatment as the cause of observed results is in error.

Robert Rosenthal (1976) was most persistent in calling social scientists' attention to the distorting effects of experimental demand, and in a series of studies, he showed how experimenters' expectations could be transmitted to research participants. In one of his most controversial studies, he suggested that teachers' expectations regarding the future achievement of their pupils affected students' actual achievement gains (Rosenthal & Jacobson, 1968), and later meta-analytic research has supported this interpretation. As a result of heightened attention to the potentially biasing effects of demand characteristics, social researchers today are careful not to make obvious the hypotheses under investigation, often resorting to elaborate COVER STORIES to hide the true purpose of their research. In addition, to shield against inadvertent cueing of the sought-for responses, studies often are conducted in such a manner that neither participant nor researcher knows the particular treatment condition to which any given individual has been (randomly) assigned. Such DOUBLE-BLIND techniques are now common in social research and represent a powerful defense against results that otherwise might be tainted by demand characteristics.

Demand characteristics are related to the particular mental set adopted by the participants, as well as their self-consciousness as subjects of research. Crano and Brewer (2002) grouped participants into three general categories: voluntary participants, who are aware of, and satisfied with, their role as subjects of study; involuntary participants, who feel they have been coerced into serving in a study; and nonvoluntary participants, who are unaware that they are under investigation. Of these, voluntary participants are most likely to be affected by demand characteristics because they have made the conscious decision to participate; do not feel unduly coerced; and, thus, may be prone to try to "help" the researcher by acting so as to confirm the hypotheses that are apparently under study. Involuntary participants are more likely to attempt to sabotage the research, in that they feel they were coerced into serving as participants. Nonvoluntary participants are not likely to attempt to help or hurt the research because they are, by definition, unaware that they are under investigation.

—William D. Crano

REFERENCE

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Rosenthal, R. (1976). *Experimenter effects in behavioral research.* New York: Irvington.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart and Winston.

DEMOCRATIC RESEARCH AND EVALUATION

Deliberative democratic research and evaluation aspires to arrive at unbiased conclusions by considering all relevant interests, values, and perspectives; by engaging in extended dialogue with major stakeholders; and by promoting extensive deliberation about the study's conclusions.

—Ernest R. House

REFERENCE

House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research.* Thousand Oaks, CA: Sage.

DEMOGRAPHIC METHODS

Demographic methods include a wide range of measures and techniques used to describe POPULATIONS. Although usually applied to human populations, they are increasingly used to describe nonhuman populations, ranging from animals and insects to manufactured products.

The most basic demographic techniques, such as the calculation of proportions and ratios, describe the composition of a population, such as its age-sex structure, racial composition, or geographic distribution. Measures such as rates and probabilities describe population dynamics. Some demographic methods focus on the separate processes producing change in a population. Single and multiple decrement LIFE TABLES, for example, focus on exit processes, such as mortality. More complex techniques consider both entrances (such as fertility) and exits in combination to estimate measures of reproductivity, intrinsic rates of growth or decline, and the parameters of stationary and stable populations. Techniques such as direct standardization or the calculation of shares or indexes of dissimilarity allow comparisons across two or more populations.

DATA SOURCES

The use of demographic methods to describe human populations is usually based on secondary data generated by national censuses; surveys; vital statistics registration systems; and, less frequently, population registers. The goal of a national CENSUS is to elicit a complete count of all people on a specific date, either all people present in the country (a de facto count), or all people residing in the country (a de jure count). In practice, census counts are marred by some degree of underenumeration or undercounting of some individuals (often assessed through CAPTURE-RECAPTURE models) and overcounting of other individuals. Most national censuses also include information such as the age, sex, relationship to household head, educational attainment, and marital status of each individual counted in the census.

Vital statistics registration systems describe the incidence of births, deaths, and sometimes other demographic events, such as marriages and divorces, as they occur over time. Vital statistics systems also often include additional information, such as age, sex, and marital status of the person experiencing the demographic event. Population registers combine the benefits of regular censuses and vital registration systems by keeping continuous track of the entire population and vital events, but they are difficult and expensive to maintain and thus much rarer.

BASIC DEMOGRAPHIC MEASURES

Proportions (or percentages) and ratios can be used to describe the structure of a population or the distribution of its elements along a variable of interest. Because many of the demographic and social processes producing change in a population are strongly age-graded, more complex demographic techniques often control for, or take into account, the age structure of the population. The complete age structure of a population can be described by tabulating the absolute numbers of people or by calculating the proportions (or percentages) of people in each age interval. The results are often portrayed graphically in a POPULATION PYRAMID. It is also common to calculate summary measures describing a particular feature of the age-sex composition of a population. One important summary measure is the age-dependency ratio, defined as the sum of the number of people aged less than 18 and the number of people aged 65 and over (the "dependent population") divided by the number of people aged 18–64 (the "working population"). Another summary measure is the sex ratio, defined as the number of males divided by the number of females.

MEASURES OF CHANGE

Almost all demographic methods and techniques describing change in a population build upon the basic demographic model in which the size of population at time t is a function of the size of the population at an earlier point in time, $t - n$, and the numbers of entrances and exists between time t and time $t - n$. A national population, for example, changes between time t and time $t + n$ through the quartet of classic demographic events: births (B), deaths (D), arrivals of in-migrants (I), and departures of out-migrants (O).

The most basic demographic measure describing change is the rate. A rate is the ratio of the number of demographic events occurring within a specified time frame (typically 1 year) to the number of people at risk of experiencing that event. The crude death rate for a population, for example, can be calculated as the total number of deaths during a specified year divided by the estimate of the total population midway through the year. Rates are often multiplied by a constant, k, such as 1,000 or 100,000. The crude death rate (CDR) of the United States in the year 2000, for example, was 8.5 deaths per 1,000 people:

$$\text{CDR} = \frac{D}{N} \times k = \frac{2{,}403{,}000}{281{,}422{,}000} \times 1{,}000 = 8.5.$$

The crude birth rate (CBR) of the United States in the year 2000, defined in parallel fashion to the CDR above (but with the number of births substituted for the number of deaths), was 14.4 births per 1,000 people.

$$\text{CDR} = \frac{B}{N} \times k = \frac{4{,}065{,}000}{281{,}422{,}000} \times 1{,}000 = 14.4.$$

The difference between the CBR and the CDR is the crude rate of natural increase (CRNI). In the case

Table 1 Measures of Reproductivity for the U.S. Population, 2000

Age Interval	*Age-Specific Fertility Rate Per 1,000 Women*	*Age-Specific Fertility Rate × Proportion of Births That Are Female*	*Proportion of Daughters Surviving to the Age of Their Mother*	*Product*
15–19	49.4	24.1	0.9902	23.9
20–24	112.3	54.8	0.9879	54.1
25–29	121.4	59.2	0.9854	58.4
30–34	94.1	45.9	0.9822	45.1
35–39	40.4	19.7	0.9776	19.3
40–44	8.4	4.1	0.9707	4.0
Sum	426.0	207.9		204.8
Sum × 5	2,130.0	1,039.5		1,023.8

SOURCES: Martin, Hamilton, Ventura, Menacker, and Park (2002); National Center for Health Statistics (2003).

NOTE: Age-specific fertility rates for the youngest and oldest age intervals include births to women younger than 15 and births to women older than 44, respectively. The age-specific fertility rates referring to female babies were estimated using a sex ratio at birth of 1,049 boys per 1,000 girls.

of the United States in 2000, the annual rate of natural increase was 5.9 people per 1,000.

$$\text{CRNI} = \text{CBR} - \text{CDR} = 14.4 - 8.5 = 5.9.$$

The crude in-migration rate or the crude out-migration rate can be calculated in a manner directly analogous to the crude birth rate or crude death rate. Many countries do not, however, collect comprehensive data on out-migration, and data on in-migration are often flawed. Thus, migration rates often refer to net migration ($I - O$).

DEATH RATES

More refined rates refer to subpopulations defined through age, sex, or a social characteristic, such as race or ethnicity. Because the force of mortality varies so strongly by age, death rates are often presented for specific age groups. The formula written below for an age-specific death rate uses a common style of notation in which the capital letters M, D, and N refer to the death (or mortality) rate, number of deaths, and number of people, respectively, and the right-hand and left-hand subscripts identify the lower bound and the width of the age interval, respectively. The death rate for people aged 15–24 in the United States in the year 2000 was 79.9 deaths per 100,000 young adults aged 15 to 24:

$${}_{10}M_{15} = \frac{{}_{10}D_{15}}{{}_{10}N_{15}} \times k = \frac{31,300}{39,183,900} \times 100,000 = 79.9.$$

If the focus is on infants, the infant mortality rate (IMR) is often substituted for the age-specific death rate for the population aged 0 to 1 because of common errors in the census enumeration of infants. The infant mortality rate is calculated as the number of deaths to infants divided by the number of births. In the United States in 2000, preliminary data on infant deaths and births yield an infant mortality rate of 6.9 infant deaths per 1,000 births.

FERTILITY RATES

Fertility rates typically refer to women in the reproductive ages because of the difficulties gathering accurate data on the ages (or other attributes) of male parents. The formula below for an age-specific fertility rate uses the capital letters F, B, and N to refer to the fertility rate, the number of births born to women in the age interval, and the number of women in the age interval, with the subscripts referring, as before, to the lower bound and width of the age interval. The fertility rate for women aged 15–19 in the United States in the year 2000 was 48.5 births per 1,000 women:

$${}_{5}F_{15} = \frac{{}_{5}B_{15}}{{}_{5}N_{15}} \times k = \frac{469,000}{9,670,000} \times 1,000 = 48.5.$$

Summing the age-specific fertility rates within the reproductive age span of women (typically 15–44 or 15–49) and multiplying by the width of the age categories yields the total fertility rate (TFR). The sum of the age-specific fertility rates presented in column 1 of Table 1 shows that the total fertility rate for the United States in 2000 was 2,130 births per 1,000 women. The TFR can be interpreted as the total number of births to a hypothetical cohort of women experiencing the age-specific fertility rates observed in 2000 as they live through their reproductive life span.

MEASURES OF REPRODUCTIVITY

Classic measures of reproductivity focus on one sex, typically the female population. The calculation and interpretation of the gross reproduction rate (GRR) parallel the calculation and interpretation of the total fertility rate with one exception—the summed age-specific fertility rates consider only female births rather than all births and are sometimes called "maternity rates" rather than fertility rates. If age-specific maternity rates are unavailable, they are often estimated by multiplying age-specific fertility rates by the proportion of all births that are female. In the formula below, the right-hand superscript refers to the sex of the births:

$$\begin{aligned} \text{GRR} &= n\sum\nolimits_n F_x^f = n\sum\nolimits_n F_x \times \frac{B^f}{B} \\ &= 5(49.4 + 112.3 + \cdots + 8.4) \\ &\quad \times \frac{1{,}000}{2{,}049} = 1039.5. \end{aligned}$$

The net reproduction rate (NRR) takes mortality into consideration by multiplying the age-specific maternity rates by the proportion of newly born daughters surviving to the age of their mother, ${}_nL_x/nl_0$, which is derived from a LIFE TABLE for females. The results, displayed in detail in Table 1, show an NRR of 1,023.8 for the United States in the year 2000:

$$\begin{aligned} \text{NRR} &= n\sum\nolimits_n F_x^f \times \frac{{}_nL_x^f}{nl_0} \\ &= 5[(24.1 \times 0.9902) + (54.8 \times 0.9879) + \ldots] \\ &= 1023.8. \end{aligned}$$

Thus, a cohort of 1,000 U.S. women subject to the age-specific fertility rates observed in the year 2000 could expect to have 2,130.0 births by the end of their reproductive life span. Of these, 1,039.5 would be daughters, and 1,023.8 would survive to the age of their mother.

Every population can be described at a moment in time by a schedule of age-specific fertility and mortality rates. If these rates were to continue indefinitely, the result would be a stable population marked by a birth rate, a death rate, a growth rate, and an age structure that are "intrinsic" to the schedules of age-specific vital rates and that do not depend on the age structure of the original population.

STANDARDIZATION

Because fertility and mortality rates differ greatly across age intervals, it can be misleading to compare crude fertility and mortality rates across populations if the age structures of the populations differ. The techniques of direct and indirect standardization allow more appropriate comparisons of rates across populations, although the former approach is preferable to the latter (see Shryock, Siegel, & Associates, 1973).

A crude rate for a population can be rewritten as the sum of the product of age-specific rates and the age-specific proportions of people in each age interval. For example, the crude death rate for population i can be written as

$$\text{CDR}_i = \frac{{}_\omega D_{0i}}{{}_\omega N_{0i}} = \sum \left(\frac{{}_nD_{xi}}{{}_nN_{xi}}\right)\left(\frac{{}_nN_{xi}}{{}_\omega N_{0i}}\right).$$

A directly standardized rate (DSDR_j) for population j is calculated by multiplying the age-specific rates of population j by the age-specific proportions of the standard or reference population, population i:

$$\text{DSDR}_j = \sum \left(\frac{{}_nD_{xj}}{{}_nN_{xj}}\right)\left(\frac{{}_nN_{xi}}{{}_\omega N_{0i}}\right).$$

The directly standardized death rate can then be compared to the crude death rate of the standard population because they were calculated using the same age structure. When comparing the rates of two populations, the choice of the "standard" population is largely arbitrary, although it is generally better to use the one with the more regular age structure. However, conventions sometimes govern the choice of the standard population. The U.S. Census Bureau, for example, recently declared the 2000 U.S. population to be the standard for historical comparisons involving the American population.

The difference between any crude rate and a directly standardized rate ($\text{CDR}_i - \text{DSDR}_j$) is the sum of the differences in the age-specific rates weighted by the age composition of the standard population. It is also possible to calculate the corresponding directly standardized rate for population j. The raw difference between the crude death rates of the two populations ($\text{CDR}_i - \text{CDR}_j$) can then be decomposed into three parts: a component attributable to differences in rates ($\text{CDR}_i - \text{DSDR}_j$); a component attributable to differences in age structure ($\text{CDR}_i - \text{DSDR}_i$); and an interaction term ($\text{DSDR}_j + \text{DSDR}_i - \text{CDR}_i - \text{CDR}_j$). Some scholars divide the interaction term between the rate and age components (see Preston, Hueveline, & Guillot, 2001).

Direct standardization and its corollary, the decomposition of the difference between two rates (or means), is a technique with wide applicability outside of the classic demographic concerns of fertility or mortality. It can be used whenever a rate (or mean) can be conceptualized as the sum of the product of i-specific rates and the relative proportion of the population along dimension i. It is also possible to standardize rates (and to decompose differences between rates) along two or more dimensions of interest (see RATE STANDARDIZATION).

MEASURES OF INEQUALITY AND SEGREGATION

Some demographic measures and techniques assess and compare the distribution of counts across units of analysis. For example, income (a count of number of dollars earned) may be distributed unequally across households (the unit of analysis) in a population. The distribution of such variables is often described through "shares." Shares are calculated by first ranking the units of analysis by the values of the count variable, from lowest to highest, and dividing the units of analysis into groupings of equal size, typically quintiles (fifths) or deciles (tenths). In 2000, for example, the poorest fifth of all U.S. households earned 3.6% of the grand total of all income earned by U.S. households, whereas the top quintile earned 49.7%.

Alternatively, a population (counts of people) can be distributed unequally across the categories of a variable (e.g., geographic units). The index of dissimilarity (D) is often used to compare the distributions of two or more populations across a categorical variable. The index of dissimilarity is calculated using the formula

$$D = \frac{1}{2}\sum_i \left(\left| \frac{y_i}{\sum_i y_i} - \frac{x_i}{\sum_i x_i} \right| \right).$$

For example, the categorical variable could refer to the four major regions in the United States (Northeast, Midwest, South, and West); y to African Americans; and x to Whites. The numbers (in 1,000s) of African Americans living in each of the four regions in the year 2000 are 6,100; 6,500; 18,982; and 3,077; the corresponding numbers for Whites are 41,534; 53,834; 72,819; and 43,274. A value of 0 for the index of dissimilarity would suggest that African Americans and white Americans are distributed similarly across the four regions, whereas a value of 1 would suggest that the two populations live in entirely different regions. The value of the index of dissimilarity is .20, suggesting that about a fifth of either African Americans or Whites would have to shift to another region for the two distributions to agree exactly. Although the index of dissimilarity is easy to calculate and interpret, it has several shortcomings. It is, for example, sensitive to the number of categories. Other indexes, such as the GINI COEFFICIENT (or Gini ratio) or the entropy index (H), can be calculated in its stead, although they, too, have shortcomings. See Siegel (2002) for a discussion of these alternatives.

—Gillian Stevens

REFERENCES

Martin, J. A., Hamilton, B. E., Ventura, S. J., Menacker, F., & Park, M. M. (2002, February 12). *Births: Final data for 2000* (National Vital Statistics Report, Vol. 50, No. 5). Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics. (2003). *United States life tables,* 2000. Retrieved from www.cdc.gov/nchs/data/lt2000.pdf.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes.* Oxford, UK: Basil Blackwell.

Shryock, H. S., Siegel, J. S., & Associates. (1973). *The methods and materials of demography.* New York: Academic Press. (Condensed edition by Edward G. Stockwell.)

Siegel, J. S. (2002). *Applied demography.* San Diego, CA: Academic Press.

DEPENDENT INTERVIEWING

Dependent INTERVIEWING is a technique used in some LONGITUDINAL RESEARCH such as PANEL surveys, where the same individuals are reinterviewed at subsequent points in time. Depending on the survey design, interviews may take place quarterly, annually, biannually, or over some other specified period. Dependent interviewing can be defined as a process in which a question asked of a respondent within the current survey wave is informed by data reported by a respondent in a previous wave (Mathiowetz & McGonagle, 2000). For example, if, at one survey interview, a respondent reported being employed as a bus driver, dependent interviewing would use that response in the question asked about the respondent's job at the following interview. Instead of simply asking the respondent "What is your current job?" at both interview points, a dependent interviewing approach might ask, "Last year you told us you were working as a bus driver. Is that still the case?" The respondent can then either confirm the previous response or give a different answer if some change has occurred.

Longitudinal surveys are subject to the types of error normally found in survey data, including recall error, interviewer effects, and "noise" in the coding process. Longitudinal data, where continuous records of activities such as labor market histories or continuous income records are collected, suffer from an additional measurement problem known as the *seam effect* or the *seam problem* (Doyle, Martin, & Moore, 2000; Lemaitre, 1992). The seam effect occurs when an artificially high level of observed change in activity spells at the seam between two survey periods is reported by respondents. Respondents tend to report events that they have already reported in the previous interview and pull them into the current interview period, known as *forward telescoping,* with the result that events become bunched around the seam between the two survey periods. The use of dependent interviewing is designed to reduce these effects and produce more consistent data over time. Dependent interviewing has become more common with the advent of computer-assisted interviewing technologies for data collection, which have made it technically feasible to feed forward previously reported information from one interview into the next interview. Dependent interviewing has benefits in terms of data quality, reducing respondent and interviewer burden, as well as the cost of data collection and processing. Despite this, there are also concerns. Depending on how it is implemented, dependent interviewing could result in an underestimate of the true level of change within the population because it is easier for the respondent to simply agree with his or her previous report.

—Heather Laurie and Mike C. Merrett

REFERENCES

Doyle, P., Martin, B., & Moore, J. (2000). *Improving income measurement.* The Survey of Income Program Participation (SIPP) Methods Panel. Washington, DC: U.S. Bureau of the Census.

Lemaitre, G. (1992). *Dealing with the seam problem* (Survey of Labour and Income Dynamics Research Papers 92–05). Ottawa: Statistics Canada.

Mathiowetz, N., & McGonagle, K. (2000). An assessment of the current state of dependent interviewing in household surveys. *Journal of Official Statistics, 16*(4), 401–418.

DEPENDENT OBSERVATIONS

Dependent observations are observations that are somehow linked or clustered; they are observations that have a systematic relationship to each other. Clustering typically results from research and sampling design strategies, and the clustering can occur within space, across time, or both.

Single and multistage cluster samples produce samples that are clustered in space, typically geographical space. Identifying and selecting known and defined clusters, and then selecting observations from the chosen clusters, produce observations that are not independent of each other. Rather, the observations are connected by the cluster, whether it is a geographic entity (e.g., city), organization (e.g., school), or household.

Longitudinal research designs produce observations that are linked across time. For example, even if individuals are selected randomly, a researcher may obtain an observation (e.g., attitude_{it}) on each individual i at various points in time, t. The resultant observations (e.g., attitude_{i1} and attitude_{i2}) are not independent because they are both attached to person i.

The occurrence of dependent observations violates a key assumption of the general linear model: that the error terms are independent and identically distributed. Observations that are linked in space or across time are typically more homogeneous than independent observations. Because linked observations do not provide as much "new" information as independently selected observations might, the resultant sample characteristics are less heterogeneous. In QUALITATIVE RESEARCH, this means that researchers should make inferences cautiously (King, Keohane, & Verba, 1994). In quantitative research, this means that standard errors of estimates will be deflated (Kish, 1965).

To correct for dependent observations resulting from cluster sampling techniques, sampling weights should be used if they are a function of the dependent variable, and thus the error term. In addition, Winship and Radbill (1994) recommend using the White heteroskedastic consistent estimator for standard errors instead of trying to explicitly model the structure of the error variances. This estimator can also be used to correct for multiplicity in the sample and is available in standard statistical packages.

To correct for temporally dependent observations, researchers typically assess and model the order of the temporal autocorrelation and introduce lagged endogenous and/or exogenous variables. Such techniques have been studied extensively in the time-series literature.

Instead of simply correcting for dependence, researchers are increasingly interested in modeling and explaining the nature of the dependence. Links between individuals or organizations, and clusters of such units, are of primary interest to social science researchers. For decades, network theorists and analysts have focused their efforts on describing and explaining the nature of dependence observed within a cluster or clusters; recent advancements extend these methods to include affiliation networks (Skvoretz & Faust, 1999) and methods for studying networks over time (Snijders, 2001). More recently, Bryk and Raudenbush (1992) have developed a framework for modeling clustered, or what they call "nested," data. For such HIERARCHICAL (NON)LINEAR MODELS (also known as mixed models, because they typically contain both fixed and random effects), cross-level interactions in MULTILEVEL ANALYSIS permit an investigation of contextual effects: how higher-level units (schools, organizations) can influence lower-level relationships.

—Erin Leahey

REFERENCES

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

King, G., Keohane, R. D., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research.* Princeton, NJ: Princeton University Press.

Kish, L. (1965). *Survey sampling.* New York: Wiley.

Skvoretz, J., & Faust, K. (1999). Logit models for affiliation networks. *Sociological Methodology, 29,* 253–280.

Snijders, T. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology, 31,* 361–395.

Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research, 23*(2), 230–257.

DEPENDENT VARIABLE

A dependent variable is any variable that is held to be causally affected by another variable, referred to as the INDEPENDENT VARIABLE. The notion of what can be taken to be a dependent variable varies between experimental and nonexperimental research [see entries on DEPENDENT VARIABLE (IN EXPERIMENTAL RESEARCH) and DEPENDENT VARIABLE (IN NONEXPERIMENTAL RESEARCH)].

—Alan Bryman

DEPENDENT VARIABLE (IN EXPERIMENTAL RESEARCH)

A dependent variable (DV) is a measurement. In well-designed experimental research, the DV is the measure of the effect of the INDEPENDENT VARIABLE (IV) on participants' responses. The measure is termed dependent because it is expected to be influenced by (or is dependent on) systematic variations in the independent variable. With appropriate experimental design, observed variations in the DV are attributable to the effects of variation of the independent variable, and hence are fundamental to the attribution of cause.

Strictly speaking, in nonexperimental research, all variables are considered dependent, insofar as they are not systematically manipulated. Thus, in correlational studies, all variables are dependent, because such research uses no manipulation. In multiple regression prediction research, dependent measures are used as predictors and sometimes are termed independent (or predictor) variables. In some research traditions, predictors are sometimes interpreted causally if they are clearly unlikely to serve as outcomes. For example, a series of naturalistic studies of ambient temperature and aggression suggests a moderately strong positive relationship. It is extremely unlikely that people's aggression affects the temperature, so the correlation is interpreted as heat causes people to become more aggressive. Relationships of this sort, based on common sense or theory, often are interpreted causally, but the inferences made are tentative and not as clear as those drawn from experimental contexts, because extraneous variables that may co-occur with the critical variable may prove to be the true causal factor. For instance, in the heat-aggression example, it may be that the tendency of people to go out on the streets in hot weather to escape the heat of their homes is the true instigating factor for aggression. Crowds (of unhappy people) may be the necessary ingredient in the equation.

The DV may take on many forms. The most highly self-conscious form of DV is the self-report in which participants express an attitude, belief, or feeling; indicate their knowledge about a phenomenon; make a causal attribution about their, or another's, actions; and so on. Measures of this type must be interpreted with caution for at least two reasons. First, people generally are reluctant to report socially undesirable actions or attitudes. A DV used to assess offensive acts or beliefs might evoke socially desirable, but untrue, responses. A more pernicious problem arises from the tendency of participants to attempt to assist the investigator by answering in such a way as to help confirm the research hypotheses. Overcooperative responding to the perceived needs of the experimenter (see DEMAND CHARACTERISTICS) can prove misleading and result in a confirmation of an incorrect hypothesis.

At the opposite end of the self-conscious response continuum are DVs collected on respondents who are unaware they are being observed. Such unobtrusive or nonreactive measures are common in research conducted outside formal laboratory settings (see LABORATORY EXPERIMENT) and offer the advantage of truthful, or at least non-self-conscious, response on the part of unwary participants. However, such measures can prove insensitive, and often are less reliable than more straightforward self-report measures (Webb, Campbell, Schwartz, Sechrest, & Grove, 1981). In addition, the ethical justifiability of such measures has come under intense debate. As such, they should be used only after very careful consideration.

In addition to the self-conscious/non-self-conscious responding dimension, DVs also differ in terms of structure. Close-ended DVs are highly structured; they may involve true/false or multiple-choice response formats. These forms of DV are most appropriate in hypothesis-testing research, where the issues under investigation are precisely specified and the requisite information clearly specified on the basis of theory. In less formal, hypothesis-generating research, such restrictive measures may prove unsatisfactory, and more open-ended measures are recommended. Such measures may involve the researcher's field notes of an observational setting, notes on an unstructured conversation, and so on. In these instances, the format of the observation is open. Although open-ended DVs require coding and theory-based interpretation to support the research process, they may supply useful insights that may be developed to support more focused research efforts.

—William D. Crano

REFERENCE

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. (1981). *Nonreactive measures in the social sciences*. Boston: Houghton-Mifflin.

DEPENDENT VARIABLE (IN NONEXPERIMENTAL RESEARCH)

A dependent variable is held to be causally affected by an INDEPENDENT VARIABLE. Strictly speaking, a dependent variable is randomly distributed, although this randomness can be influenced by a number of structural factors and in the most rigorous research setting by an EXPERIMENT in which the independent variable is manipulated. However, in nonexperimental research, variables are not manipulated, so that it is sometimes deemed to be unclear which among the various variables studied can be regarded as DEPENDENT VARIABLES. Sometimes, researchers use the term loosely to refer to any variables they include on the left-hand side of a REGRESSION equation where causal inference is often implied. But it is necessary to discuss what qualifies as a dependent variable in different RESEARCH DESIGNS. When data derive from a CROSS-SECTIONAL DESIGN, information on all variables is collected coterminously and no variables are manipulated. This means that the issue of which variables have causal impacts on other variables may not be immediately obvious. Although it is difficult to establish causality, especially when using cross-sectional data, researchers use a few rules of thumb as to which variable to regard as dependent. Five such rules, which are not necessarily mutually exclusive, may serve as a guide for use in both cross-sectional and longitudinal research.

1. One variable precedes another in time. Even in cross-sectional surveys, certain variables record events that occurred earlier than others, such as the educational attainment of the respondent's parent versus the educational attainment of the respondent, which can be regarded as dependent.

With LONGITUDINAL RESEARCH, such as a PANEL design, the issue is somewhat more complex, because although no independent variables are manipulated, the fact that data can be collected at different points in time allows some empirical leverage on the issue of the temporal and, hence, causal priority of variables. Strict experimentalists might still argue that the lack of manipulation casts an element of doubt over the issue of causal priority. To deal with this difficulty, analysts of panel data resort to the so-called cross-lagged panel analysis for assessing causal priority among a pair of variables measured at two or more points in time. The line between dependent and independent variables in such analysis is, in a sense, blurred because the variable considered causally precedent is treated as an independent variable in the first regression equation but as a dependent variable in the second. Conversely, the variable considered causally succedent is treated as dependent in the first equation and as independent in the second. Judging from analyses like this, it appears that the term "dependent variable" can be quite arbitrary; what matters is clear thinking and careful analysis in nonexperimental research.

2. Variables are from a well-known sequence. For example, events in one's LIFE COURSE, such as first sexual intercourse and first birth, have a clear causal order.

3. One event freezes before another starts. The amount of donation to one's alma mater may depend on the type of bachelor's degree one receives. The act of donation does not begin until the degree is completed.

4. One variable is more changeable than another. To take a simple example: Imagine that we conduct a survey that shows that gender and a scale of political conservatism are related. One's gender hardly changes, but one's political orientation may. It seems reasonable to assume that political conservatism is the dependent variable. Of course, there may be INTERVENING VARIABLES or MODERATING VARIABLES at work that influence the causal path between gender and conservatism, but the basic point remains that it is still possible and sensible to make causal inferences about such variables.

5. One variable is more fertile than another. The more fertile (i.e., more likely to produce consequences) one, say, one's religious affiliation, could be used as an independent variable, whereas the less fertile one, say, the number of children one has, would be the dependent variable.

Researchers employing nonexperimental research designs (such as a SURVEY design) must engage in *causal inference* to tease out which are the dependent variables. Essentially, this process entails a mixture of commonsense inferences from our understanding about the nature of the social world, including the five rules discussed earlier, and from existing theory and research, as well as analytical methods in the area that is the focus of attention. It is this process of causal inference that lies behind such approaches as CAUSAL MODELING and PATH ANALYSIS to which the cross-lagged panel analysis is related.

Making causal inferences becomes more difficult when causal priority is less obvious. For example, if we find, as a result of our investigations in a large firm, that employees' levels of routine in their jobs and their organizational citizenship behavior are negatively correlated, it may be difficult to establish direction of causality, although existing findings and theory relating to these variables might provide us with helpful clues. It could be that employees in more routine jobs are less inclined to put themselves out for the firm, but it could also be the case that employees who exhibit organizational citizenship behavior are less likely to be assigned to routine jobs. A panel study is often recommended in such circumstances as a means of disentangling cause and effect.

—Alan Bryman and Tim Futing Liao

REFERENCES

Davis, J. A. (1985). *The logic of causal order* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–055). Beverly Hills, CA: Sage.

Lieberson, S. (1987). *Making it count: The improvement of social research and theory.* Berkeley: University of California Press.

Rosenberg, M. (1968). *The logic of survey analysis.* New York: Basic Books.

DESCRIPTIVE STATISTICS. *See* UNIVARIATE ANALYSIS

DESIGN EFFECTS

The design effect indicates the impact of the sample design on the VARIANCE of an estimate. The impact is measured relative to the variance of the equivalent estimate obtained from a SIMPLE RANDOM SAMPLING of the same size. Thus, the design effect for an estimate of a POPULATION parameter y can be written

$$\mathrm{DEFF}(\hat{y}_{\mathrm{Com}}) = \frac{V(\hat{y}_{\mathrm{Com}})}{V(\hat{y}_{\mathrm{SRS}})}, \tag{1}$$

where $\hat{y}_{\text{Com}}$ is the estimate of y under the complex sample design in question and $\hat{y}_{\text{SRS}}$ is the estimate from a simple random sample of the same size.

The design effect is specific to both the estimate and the sample design. Consequently, estimates of different quantities from the same survey may have different design effects, as may estimates of the same quantity from different SURVEYs. Three important influences on the design effect are sample clustering, sample stratification, and sample inclusion probabilities.

Multistage clustered sample designs (see MULTISTAGE SAMPLING) tend to result in design effects greater than 1. This is because sample units within clustering units tend to be more similar to one another than are sample units in the population as a whole (i.e., clusters are relatively homogeneous). This may be true, for example, of households living within small areas, pupils attending the same school, employees of the same employer, and so on. The design effect due to clustering will depend on the extent of the homogeneity within clusters and the (mean) cluster sample size:

$$\text{DEFF}(\hat{y})_{\text{Clus}} = 1 + (b-1)\rho, \tag{2}$$

where b is the (mean) cluster sample size and ρ is the intracluster CORRELATION coefficient.

The design effect due to proportionate stratification (see STRATIFIED SAMPLE) will tend to be less than 1. Again, this is because strata are relatively homogeneous. But unlike clustering, which limits the sample to a subset of clusters, stratification ensures that the sample represents all strata. The higher the correlation between stratum and target variable, the smaller (more beneficial) the design effect due to stratification.

The design effect due to variable inclusion probabilities (encompassing both selection probabilities and response probabilities) will tend to be greater than 1 and will tend to be greater the greater the variation in the probabilities. There are notable exceptions, such as the use of optimal allocation to strata in order to increase precision of estimation in the situation where stratum population variances are known or can be estimated. The design effect due to variable inclusion probabilities can be written as follows:

$$\text{DEFF}(\hat{y})_{\text{Inc}} = \frac{n}{N^2 S^2}\sum_{h=1}^{H}\frac{N_h^2 S_h^2}{n_h}, \tag{3}$$

where n_h, N_h, and S_h^2 are, respectively, the sample size, population size, and population variance in class h, the population being divided into H mutually exclusive and comprehensive classes, whereas n, N, and S^2 are the overall sample size, population size, and population variance.

Many survey technical reports publish tables of design effects for a range of estimates. These can be useful in informing the design of future surveys as well as the interpretation of results from the current survey.

A closely related concept is the *design factor,* which indicates the impact of the sample design on the STANDARD ERROR of the estimate, and hence on the width of the CONFIDENCE INTERVAL:

$$\begin{aligned}\text{DEFT}(\hat{y}_{Com}) &= \frac{s.e.(\hat{y}_{\text{Com}})}{s.e.(\hat{y}_{\text{SRS}})} \\ &= \sqrt{\text{DEFF}(\hat{y}_{\text{Com}})}.\end{aligned} \tag{4}$$

—Peter Lynn

REFERENCES

Kish, L. (1965). *Survey sampling.* New York: Wiley.

Kish, L. (1995). Methods for design effects. *Journal of Official Statistics, 11*(1), 55–77.

DETERMINANT

The determinant of a square MATRIX is a unique scalar associated with that matrix. The determinant of a matrix $\mathbf{A}$ is most commonly denoted as det $\mathbf{A}$ or $|\mathbf{A}|$. The notation for determinants substitutes vertical lines for the brackets of matrices. Thus, for

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad |\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Texts may sometimes refer to

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

as the determinate function and the resulting scalar as the value of the determinant.

Calculation of the determinant is straightforward for matrices of order 1 and 2. If $\mathbf{A} = [a_{11}]$, then $|\mathbf{A}| = |a_{11}| = a_{11}$. If

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

then

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Note that the determinant for the second-order example is the product of the elements on the main diagonal minus the product of the elements on the off-diagonal. A technique known as row (or column) expansion allows for the systematic calculation of determinants for matrices of order 3 and above. In brief, for matrix **A** of order n, row expansion produces the equation

$$|\mathbf{A}| = \sum_{j=1}^{n} (-1)^{i+j}\, a_{ij} |\mathbf{A}_{ij}|$$

for any one row $i = 1, 2, \ldots, n$. For column expansion, the equation is

$$|\mathbf{A}| = \sum_{i=1}^{n} (-1)^{i+j}\, a_{ij} |\mathbf{A}_{ij}|$$

for any one column $j = 1, 2, \ldots, n$. In these equations, $|\mathbf{A}_{ij}|$ is the determinant of the submatrix formed by eliminating the ith row and jth column of **A**. This is often referred to as the minor of a_{ij} from **A**. Thus, expanding by the first row for matrix **A** of order 3,

$$|\mathbf{A}| = \sum_{j=1}^{3} (-1)^{i+j}\, a_{ij} |\mathbf{A}_{ij}| = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Several theorems allow the manipulation of the rows and columns of a determinant to facilitate the calculations. Among them are the following:

- The determinant of a matrix **A** is equal to the determinant of the transpose of **A** (i.e., $|\mathbf{A}| = |\mathbf{A}^{\mathrm{T}}|$).
- Exchanging any two rows of a determinant changes its sign.
- If every element of a row is multiplied by a factor k, then the value of the determinant is multiplied by k.
- The value of a determinant is unchanged if a scalar multiple of one row is added to another row.

If the determinant of a matrix is nonzero, the matrix is said to be nonsingular and will have an inverse. Determinants are used to calculate the inverses of matrices, both of which may then be used to solve systems of equations such as those found in REGRESSION analysis. Determinants are also used with Cramer's rule, a method of solving SIMULTANEOUS linear EQUATIONS, and in determining whether the extreme points of FUNCTIONS are maxima or minima.

If the determinant of a matrix equals zero, the matrix is said to be singular. This indicates that two or more of the rows or columns are LINEARLY DEPENDENT. Algebraically, this means that an inverse of the matrix cannot be calculated. As a practical matter, it means that two or more of the variables in the model represented by the matrix are linear TRANSFORMATIONS of each other.

—Timothy M. Hagle

REFERENCES

Cullen, C. G. (1990). *Matrices and linear transformations* (2nd ed.). New York: Dover.

Pettofrezzo, A. J. (1978). *Matrices and transformations.* New York: Dover.

DETERMINATION, COEFFICIENT OF.

See *R*-SQUARED

DETERMINISM

Determinism is a term usually used pejoratively to describe an argument or methodology that simplistically reduces causality to a single set of factors acting more or less directly to produce outcomes. So, for example, economic determinism attributes social, political, and cultural effects to economic relations as the fundamental causal factors; technological determinism does the same with technological innovation; textual determinism attributes the impact of mediated messages on individuals and audiences to the textual attributes of the messages or the medium, and so on. A related but different term is reductionism, where an analysis is accused of reducing a complex set of circumstances to a limited set of supposedly salient characteristics.

The methodologies most often accused of determinism are those that favor a totalizing approach to explanation, such as Marxism, Freudianism, informationalism, textualism, systems theory, and so on; those that want to privilege certain natural, social, or physical characteristics as fundamental to social activity, such as climate, class, gender, race, or geography; or those that single out particular spheres of human activity as overwhelmingly powerful in their impact on other spheres, such as technology, sex, or work.

Of course, there is poor intellectual work that manifests determinist and/or reductionist flaws. But what often underlies debates about determinism are the more fundamental issues of where a researcher might look for adequate causal explanation in human phenomena, whether such a search is worth undertaking at all, whether fields of human activity in time and space can be thought of as totalities (or sets of overlapping totalities), and whether there are systematic hierarchies of factors in social causation. In short, arguments about determinism can be proxy arguments about causality.

Thinkers have approached causality in a number of different ways. One interesting outcome of sociometric modeling has been questions about quantifiable predictability as a corollary of causality: The debate is between deterministic, mathematical models and stochastic, statistical approaches (the latter incorporating chance outcomes). At the metatheoretical level, some argue that the structure of a whole determines the nature and function of its parts; systems theory approaches are typical of this approach. Others argue that the whole is the sum of its parts, and that only the parts are knowable in any worthwhile sense; pragmatic approaches often take this position.

A third approach looks to neither the parts nor the whole but to the relations between forces in constant conflict that structure the particular parts in relation to any whole at any specific juncture of time and space. This conflict produces outcomes in particular periods and places that determine the nature of relationships between the parts and so structure the whole. It is an approach that developed in the second half of the 20th century, in different disciplines and from different totalizing perspectives (e.g., the educational psychologist Jean Piaget called it "operational structuralism," and for the Marxist geographer David Harvey, it is "dialectical, historical-geographical materialist theory").

It is an approach that views causation as a complex and continuing process of conflict among abstract forces, manifesting themselves in specific material ways in specific historical and geographical contexts, producing outcomes that themselves then feed into the process of further causation. It rejects notions of direct, linear causality between factors, but does seek to identify hierarchies of power or dominance in the complex relationships that determine outcomes.

David Harvey's work is preoccupied with these methodological questions with respect to urbanization and presents a thorough exploration of the spatial and temporal dimensions of causality. *Social Justice and the City* (1973) is an early, detailed exposition of the methodological issues; *The Condition of Postmodernity* (1989) is his best known work.

—Chris Nash

REFERENCES

Harvey, D. (1973). *Social justice and the city.* Baltimore: Johns Hopkins University Press.

Harvey, D. (1989). *The condition of postmodernity.* Oxford, UK: Basil Blackwell.

DETERMINISTIC MODEL

A deterministic model is one that contains no random elements.

Deterministic models began to be widely used to study physical processes in the early 18th century with the development of differential equations by mathematicians such as Jakob Bernoulli, Johann Bernoulli, and Leonhard Euler. Differential equations, which give the value of a variable as a function of its value at an infinitesimally earlier time point, remain the most common form of deterministic MODEL. Differential equations proved applicable to a wide variety of astronomical and mechanical phenomena, and have produced a rich mathematical literature. Finite difference equations—where the value of a variable is a function of its value at a time period lagged by a finite value (typically 1)—are another form of deterministic model, albeit a form that has proven to be less mathematically elegant than differential equations. Difference equations have been commonly used to model biological and social processes where change occurs more slowly, for example, across generations in a biological population or across budget years in an economic process.

The values of a deterministic model are completely predictable provided one knows the functional form of the model, the values of its coefficients, and the initial value of the process. During the 18th and 19th centuries, the physical universe was assumed by many natural scientists to operate according to purely deterministic laws. This meant, in theory at least, that it was entirely predictable—in the widely quoted formulation of the French mathematician Pierre Simon de Laplace in his work, *Analytic Theory of Probability* (1812):

> Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective positions of the beings which compose it . . . it would embrace in the same formula both the movements of the largest bodies in the universe and those of the lightest atom; to it nothing would be uncertain, and the future as the past would be present to its eyes. (Mackay, 1977, p. 92)

As Laplace notes, a curious feature of deterministic models is that not only can the future be predicted from the model, but one can also deduce behavior in the past. Thus, for example, astronomers modeling the orbit of a newly discovered comet with deterministic equations can figure out not only where the comet is going, but where it came from. Similarly, a deterministic model of the path of a ballistic missile can be used to determine its approximate launch point, even if the launch was not observed.

The 20th century saw three challenges to the concept of a deterministic universe. First, early in the century, the theory of quantum mechanics developed and argued that at an atomic level, physical processes had intrinsically random elements, in contrast to the purely deterministic world envisioned earlier. Second, throughout the 20th century, the mathematical fields of probability and statistics—including the study of STOCHASTIC processes, which incorporate randomness—developed to a level of sophistication comparable to the 19th-century mathematics of deterministic processes. Subsequent work in the social sciences has consequently focused primarily on models with explicitly specified random elements.

Finally, near the end of the 20th century, chaotic processes began to receive considerable attention (see Kiel & Elliott, 1996). Chaotic models are deterministic—they are typically nonlinear difference equations—but generate sequences of values that appear to be random, and those values cannot be predicted over long periods unless the initial values can be assessed with infinite precision, which is not possible in practice.

Despite these challenges, deterministic models remain useful in a variety of contexts. Notwithstanding quantum mechanics and chaos theory, deterministic equations can provide extremely accurate predictions for many physical processes that occur above the atomic level. Deterministic equations are also useful as approximations for social processes, even when these are known to contain random components. Examples include the growth of populations, budgets, and international arms races.

—Philip A. Schrodt

REFERENCES

Kiel, L. D., & Elliott, E. (1996). *Chaos theory in the social sciences.* Ann Arbor: University of Michigan Press.

Mackay, A. L. (1977). *Scientific quotations.* New York: Crane, Russak.

DETRENDING

Often, TIME-SERIES data may exhibit an upward or downward trend over time. A trend may be present in the form of a change in the mean value and/or VARIANCE over time. A plot of a time series may be linear, demonstrating a trending mean over time. Regressing one such variable on another will often yield a high R-SQUARED, yet the estimated parameters may not reflect the relationship between the two variables. Rather, the estimated parameters will be BIASED upwards because of the trend, resulting in a SPURIOUS RELATIONSHIP between the dependent and independent variables, reflecting the common trend present in both of them. When this is the case, observations may be detrended before fitting a model in order to obtain true estimates of the relationship between the dependent and independent variables. Figure 1 shows a series that is trending over time, and that same series after the trend has been removed.

A simple solution to this problem is to include a time, or trend, variable as an INDEPENDENT VARIABLE in a REGRESSION analysis. The trend variable need not be linear; if the series were to exhibit a cubic trend, then the trend variable could be X^3, where X is time. Furthermore, some time series exhibit seasonal

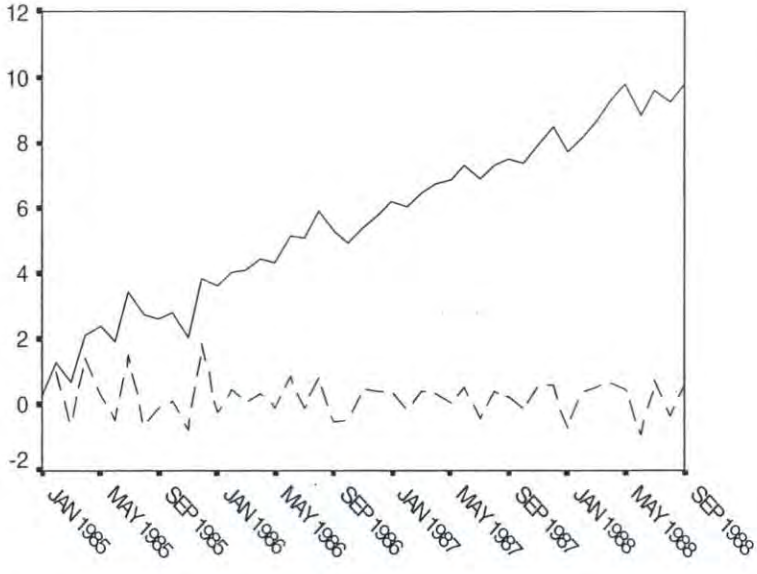

Figure 1 Example of a Trending Series and Same Series After Detrending

trends, and it is appropriate in these instances to include seasonal trend variables as well. Introducing a trend variable on the right-hand side of the equation will yield for its coefficient an estimate of the true relationship between it and the dependent variable.

If the dependent and independent variables exhibit different trends, say, the dependent a linear trend and the independent a quadratic, then it may be useful to employ a more complex detrending process. First, regress the dependent variable on time to obtain the RESIDUALS, $\hat{u}_{1t}$. Then, regress the independent variable on time to obtain the residuals, $\hat{u}_{2t}$. Finally, regress $\hat{u}_{1t}$ on $\hat{u}_{2t}$. With the time trend removed from both variables, the coefficient on $\hat{u}_{2t}$ will be the true relationship between the dependent and independent variable. This method for detrending is additionally useful if you wish to report not merely the relationship between the dependent variable and the independent variable, but their unique trends over time as well.

Detrending in this manner is only applicable for the series in question if the trend is deterministic. If the trend changes over time, it is said to be STOCHASTIC, and in this instance, ARIMA modeling may be an appropriate method to use to discover the underlying relationship. A UNIT ROOT test may be appropriate to determine whether the series is stationary. Trends are removed in ARIMA modeling by differencing series, or by subtracting previous values of the variable from the value at time t. Simple linear trends, such as those easily removed by including a trend variable inregression, may also be removed in ARIMA models with a first-differenced equation, or an equation where the dependent variable is change in y from time, t, to previous time, $t - 1$.

—Andrew J. Civettini

REFERENCES

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). St. Louis, MO: McGraw-Hill.

Harvey, A. C. (1990). *The econometric analysis of time series* (2nd ed.). New York: Philip Allan.

DEVIANT CASE ANALYSIS

Deviant case analysis is a procedure used in CONVERSATION ANALYSIS (CA) for checking the validity and generality of proposed phenomena of conversational interaction. It bears comparison to hypothesis testing and falsification as general methodological principles. However, it has special features that make it an essential ingredient in CA.

The main aim of CA is to discover robust rules of conversational sequencing that are not merely statistically observed regularities, but ones that are demonstrably used and oriented to, as normatively expectable, by conversational participants. Through inductive analysis of a corpus of conversational data, potential rules of conversational order may be proposed, such as in greetings exchanges, or in the ways in which invitations may be either accepted or rejected. Deviant cases are those that do not appear to conform to the proposed rule or pattern.

The importance of these cases, beyond their status as disconfirmations, is that they are often the examples that turn out to best demonstrate participants' normative orientations, somewhat echoing the "breaching experiments" that were characteristic of early ETHNOMETHODOLOGY. Unlike those experiments, however, CA's deviant cases are ones found naturally occurring in a corpus of recorded talk.

Stephen Clayman and Douglas Maynard (1995) define three uses of deviant cases in CA: (a) Deviant cases may turn out, upon analysis, to instantiate the same normative orientations as are proposed for the standard cases; (b) a new analysis may be proposed that now takes full account of both the (apparently) deviant cases, and also the main set; and (c) a deviant case may turn out to exhibit a different kind of order in its own right, such that it properly belongs in a different collection.

The classic example of the second type, where a revised analysis is proposed, is an early CA study by Emanuel Schegloff (1968) on the ways in which telephone conversations routinely begin. Using a corpus of 500 examples, Schegloff had proposed a "distribution rule for first utterances," to the effect that the person answering the call is, expectably, first to speak. One case was found to deviate from the rule. This led to a questioning and radical revision of what the rule proposed. Rather than proposing a rule in which call answerers make first turns, which might be considered first parts of a greeting exchange (e.g., "hello," answered by "hello"), a revised rule could be proposed that accounted for all 500 cases. This was the summons-answer sequence, in which the first utterance of "hello" (or whatever) was now understood as the second part of a different kind of pair, the answer to the summons, which was the phone ringing.

—Derek Edwards

REFERENCES

Clayman, S. E., & Maynard, D. (1995). Ethnomethodology and conversation analysis. In P. ten Have & G. Psathas (Eds.), *Situated order: Studies in the social organization of talk and embodied activities* (pp. 1–30). Washington, DC: University Press of America.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist, 70,* 1075–1095.

DEVIATION

Deviation is a very general term referring to some type of discrepancy. A simple example is the deviation or difference between an observation and the POPULATION mean, μ, the average of all individuals if they could be measured. The average (expected) squared deviation between an observation and μ is the population VARIANCE. A related measure is the squared deviation between observations made and the sample MEAN, which is represented by the SAMPLE variance. That is, deviations are defined in terms of the sample mean rather than population mean.

In a more general context, deviation refers to the discrepancy between a sample of observations and some model that supposedly represents the process by which all data are generated. Perhaps the simplest and best-known example is determining whether observations were sampled from some specified DISTRIBUTION, with a NORMAL DISTRIBUTION typically being the distribution of interest. The most commonly used measure of deviation between the distribution of the observed values, versus the hypothesized distribution, is called a Kolmogorov distance (e.g., see Conover, 1980). The method has been extended to measuring the overall deviation between two distributions (Doksum & Sievers, 1976), which can result in an interesting perspective on how two independent groups compare. (For details and appropriate software, see Wilcox, 2003.)

The notion of deviation arises in several other situations. In regression, for example, a common model is $Y = \beta_0 + \beta_1 X + \varepsilon$, where ε is usually taken to be a variable having a mean of zero. So, the model is that if we are told that $X = 10$, for example, the mean of Y is equal to $\beta_0 + \beta_1 X = \beta_0 + \beta_1 10$, where, typically, β_0 and β_1 are unknown parameters that are estimated based on the sample of observations available to us. In essence, the model assumes a linear association between the mean of Y and X. Here, deviation might refer to any discrepancy between the assumed model and the data collected in a particular study. So, for example, if Y represents cognitive functioning of children and X represents level of aggression in the home, a common strategy is to assume that there is a linear association between these two measures. A fundamental issue is measuring the deviation between this assumed linear association and the true association in an attempt to assess the adequacy of the model used to characterize the data. That is, is it reasonable to assume that for some choice of β_0 and β_1, the mean of Y is equal to $\beta_0 + \beta_1 X$? Measures of how much the data deviate from this model have been devised, one of which is essentially a Kolmogorov distance, and they can be used to provide empirical checks on the assumption of a linear model (e.g., Wilcox, 2003).

Yet another context where deviation arises is categorical data. As a simple example, consider a study where adult couples are asked to rate the effectiveness of some political leader on a 5-point scale. A possible issue is whether a woman's response is independent of her partner's response. Now, the strategy is to measure the deviation from the pattern of responses that is expected under independence. More generally, deviation plays a role in assessing goodness of fit when examining the plausibility of some proposed model (e.g., see Goodman, 1978).

—Rand R. Wilcox

REFERENCES

Conover, W. J. (1980). *Practical nonparametric statistics.* New York: Wiley.

Doksum, K. A., & Sievers, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika, 63,* 421–434.

Goodman, L. A. (1978). *Analyzing qualitative/categorical data.* Cambridge, MA: Abt.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques.* San Diego, CA: Academic Press.

DIACHRONIC

Diachronic features describe phenomena that change over time, implying the researcher employs a chronological, historical, or longitudinal perspective (cf. SYNCHRONIC).

—Tim Futing Liao

DIARY

Diaries are a natural method of data collection for obtaining personal data on a regular (usually day-to-day) time basis (Corti, 1993). They are well-adapted to getting contextualized, detailed, sequential information, and, compared to QUESTIONNAIRES, data from diaries are usually more reliable because they minimize problems of bias in retrospective recall and (being self-administered) can ensure ANONYMITY and freedom from INTERVIEWER EFFECTS. However, there can be problems associated with selection or volunteer bias. Consequently, incentives are often paid to motivate diarists to continue for the full term of their diary-keeping. Diaries are also used in conjunction with follow-up interviews to expand on episodic behavior (Zimmerman & Wieder, 1977).

FORMS OF DIARY

The diary method has a range of structured forms:

- *Unstructured diaries* have little or no format specified. They are often unsolicited productions, and the diarist decides on what is significant to report. Free-format diaries are autobiographical, in natural language (or in code, like Pepys' diary), and more a qualitative resource than a mode of systematic data collection.
- *Semistructured diaries* have a specific form or schema imposed by the researcher, and they are usually solicited. This format allows diaries to be directly compared, but also allows the diarists to state the events in their own words and expand on them. The resulting data often integrate qualitative and quantitative modes.
- *Structured diaries* are more akin to the life history in the sense that the units of time are fixed, and two crucial components of the data are (a) what the specific activity is (relevant activities are usually prespecified and precoded), and (b) how long the given event or activity lasted. This generates directly comparable quantitative data. The most common example is the time-use diary (Juster & Stafford, 1985).

The length of time covered by diaries varies considerably. Time-use studies tend to last only a few days, but other forms are usually timed to cover significant events or provide stable estimates of behavior and tend to be between 1 week and 1 month in length. The diary method has been revolutionized in recent years by the Internet. The online diary has become very popular and covers the range of diary forms outlined above.

APPLICATIONS OF DIARY METHODS

Household surveys often include a diary component, and market researchers and economists use diaries for estimating patterns of consumer expenditure. They have also been used extensively in health and illness studies (Verbrugge, 1980), in diet and nutrition, in researching alcohol and recreational drug use, and in compliance with drug regimes. Sexual behavior has been studied by individual diary-keeping methods at least since the original Kinsey studies in the 1930s and, more recently, in studies of risk behavior among men who have sex with men (Coxon, 1996).

—Anthony P. M. Coxon

REFERENCES

Corti, L. (1993). Using diaries in social research (*Social Research Update,* Iss. 2). Guildford, UK: University of Surrey, Department of Sociology.

Coxon, A. P. M. (1996). *Between the sheets: Sexual diaries and gay men's sex in the era of AIDS.* London: Cassell.

Juster, F. T., & Stafford, F. P. (1985). *Time,goods and well-being.* Ann Arbor, MI: Institute for Social Research.

Verbrugge, L. (1980). Health diaries. *Medical Care, 18,* 73–95.

Zimmerman, D. H., & Wieder, D. (1977). The diary-interview method. *Urban Life, 5,* 479–498.

DICHOTOMOUS VARIABLES

A dichotomous VARIABLE is one that takes on one of only two possible values when observed or measured. The value is most often a representation for a measured variable (e.g., age: under 65/65 and over) or an attribute (e.g., gender: male/female).

If the dichotomous variable represents another variable, that underlying variable may be either observed or unobserved. For example, a dichotomous variable may be used to indicate whether a piece of legislation passed. The dichotomous variable (pass/fail) is a representation of the actual, and observable, vote on the legislation. In contrast, each legislator's vote is the result of an individual, and unobserved, PROBABILITY distribution between voting either for or against the legislation.

Dichotomous variables are most commonly measured using 1 and 0 as the two possible values. The use of 1 and 0 usually has no specific meaning relating to the variable itself. One could just as easily choose 2 and 0, or 1 and −1 as the two values. Even so, the 1, 0 coding does have the advantage of showing, when cases are summed and averaged, the proportion of cases with a score of 1. The choice and assignment of 1 and 0 is usually based on ease of interpretation and the nature of the specific research question.

Dichotomous variables may be used as either DEPENDENT (ENDOGENOUS) or INDEPENDENT (EXOGENOUS) VARIABLES. Many research questions in the social sciences use dichotomous dependent variables. Such questions often involve individual binary choices or outcomes (e.g., to purchase or not, to vote for or against, to affirm or reverse). Estimating models with dichotomous dependent variables often requires the use of nonlinear techniques because linear models (such as ORDINARY LEAST SQUARES) may be unrealistic on a theoretical level and produce inefficient or inconsistent results.

When used as independent variables, dichotomous variables are often referred to as "controlling" or "DUMMY" VARIABLES. Unfortunately, referring to dichotomous independent variables in this way may lead to inadequately justifying the measurement of the variable or its inclusion in the model. As with any other variable, a dichotomous independent variable should have a sound theoretical basis and be properly operationalized.

As a cautionary note, one must not be overly eager to "dichotomize" variables (i.e., convert a measurable underlying variable to dichotomous form) for three reasons. First, such conversions may not be theoretically justifiable. The popularity of estimation techniques such as PROBIT ANALYSIS and LOGIT caused some to unnecessarily convert continuous dependent variables to dichotomous form. Second, even when theoretically justified, such conversions necessarily result in a loss of information. It is one thing, for example, to know that the Supreme Court affirmed a decision, but quite another to know whether it did so by a 9–0 or 5–4 vote. Third, forcing a variable into just two categories may not always be easy or appropriate if there is a nontrivial residual category. For example, if one measures handedness dichotomously (left/right), the number of ambidextrous individuals or those with arm injuries that required converting to their other hand may pose coding or analytical problems.

—Timothy M. Hagle

REFERENCES

Hardy, M. A. (1993). *Regression with dummy variables.* Newbury Park, CA: Sage.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics.* Cambridge, UK: Cambridge University Press.

DIFFERENCE OF MEANS

Often, social scientists are interested in finding out whether the means of two social groups are identical, for example, whether the income levels are the same between two social classes or between women and men. This requires SIGNIFICANCE TESTING, and the NULL HYPOTHESIS is that there exists no difference between the two means μ_1 and μ_2. When only two group means are involved, a simple *T*-TEST is the most often applied procedure, which is also applied to testing DIFFERENCE OF PROPORTIONS, whereas ANALYSIS OF VARIANCE is a natural choice when multiple groups are involved to detect BETWEEN-GROUP DIFFERENCES. A closely related procedure is MULTIPLE CLASSIFICATION ANALYSIS. There are also various MULTIPLE COMPARISON procedures available for dealing with different situations, such as the BONFERRONI TECHNIQUE for related or dependent significance tests. In sum, testing *difference of means* is just one of the many ways of making STATISTICAL COMPARISON.

—Tim Futing Liao

DIFFERENCE OF PROPORTIONS

When people are divided into groups, different proportions of each group will display certain traits, attitudes, or outcomes. For example, different proportions of men and women vote for Democrats, different proportions of Americans and Japanese suffer from breast cancer, and different proportions of mainline and fundamentalist Protestants approve of abortion for unmarried women.

The difference between the abortion attitudes of mainline and fundamentalist Protestants may be estimated using data from the U.S. portion of the 1990 World Values Survey, summarized in the following table:

	Mainline	*Fundamentalist*	*Totals*
Approve	172 (n_{11})	23 (n_{12})	195 ($n_{1.}$)
Disapprove	404 (n_{21})	125 (n_{22})	529($n_{2.}$)
Totals	576 ($n_{.1}$)	148 ($n_{.2}$)	724 ($n_{..}$)

NOTE: To simplify the example, we treat the World Values Survey like SIMPLE RANDOM SAMPLING. The actual survey design was more complicated.

Of $n_{.1} = 576$ mainline Protestants, $n_{11} = 172$ approved of abortion for unmarried women, a proportion of $p_1 = 172/576 = .2986$. Of $n_{.2} = 148$ fundamentalist Protestants, $n_{12} = 23$ approved of abortion for unmarried women, a proportion of $p_2 = 23/148 = .1554$. Among the surveyed Protestants, then, the *difference of proportions* was $p_1 - p_2 = .1432$. The surveyed mainline Protestants were 14.32% more likely to approve of abortion for unmarried women.

If we wish to generalize from this sample to the U.S. Protestant population, we need to know the SAMPLING DISTRIBUTION for the difference of proportions. In large samples, the difference of proportions $p_1 - p_2$ approximates a NORMAL DISTRIBUTION, whose STANDARD DEVIATION is known as the STANDARD ERROR *for the difference of proportions.*

Different standard errors are used for CONFIDENCE INTERVALS and HYPOTHESIS tests. For confidence intervals, the most common standard error formula is

$$s_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_{.1}} + \frac{p_2(1-p_2)}{n_{.2}}},$$

where

$$p_1 = \frac{n_{11}}{n_{.1}} \quad \text{and} \quad p_2 = \frac{n_{12}}{n_{.2}},$$

and the corresponding $(1 - \alpha) \times 100\%$ confidence interval is

$$p_1 - p_2 \pm z_{1-\alpha/2} s_{p_1-p_2},$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2) \times 100$th PERCENTILE of the standard NORMAL DISTRIBUTION.

Although popular, this confidence interval has poor coverage, especially when the table counts n_{ij} are low. Better coverage, especially for low counts, is obtained by adding 1 to each n_{ij} and 2 to each $n_{.j}$ (Agresti & Cato, 2000). The improved standard error is

$$\tilde{s}_{p_1-p_2} = \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_{.1}+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_{.2}+2}},$$

where

$$\tilde{p}_1 = \frac{n_{11}+1}{n_{.1}+2} \quad \text{and} \quad \tilde{p}_2 = \frac{n_{12}+1}{n_{.2}+2}$$

and the improved $(1-\alpha) \times 100\%$ confidence interval is

$$\tilde{p}_1 - \tilde{p}_2 \pm z_{1-\alpha/2} \tilde{s}_{p_1-p_2}.$$

For the Protestant data, $\tilde{p}_1 - \tilde{p}_2 = 0.1393$, $\tilde{s}_{p_1-p_2} = 0.0355$, and $z_{.975} = 1.96$, so a 95% confidence interval runs from 0.0698 to 0.2088. In other words, the population of mainline Protestants is probably between 6.98% and 20.88% more likely to approve of abortion for unmarried women. (Sometimes, the confidence limits calculated from this formula exceed the lower bound of $-p_1$ or the upper bound of $1 - p_1$. On such occasions, it is customary to adjust the confidence limits inward.)

If we wish to test the NULL HYPOTHESIS of equal proportions, we use a slightly different standard error. Under the null hypothesis, we regard the surveyed Protestants as a single sample of $n_{..} = 724$ people, of whom $n_{1.} = 195$ approve of abortion for unmarried women, a proportion of $p = 195/724 = 0.2693$. In this setting, an appropriate standard error formula is

$$\hat{s}_{p_1-p_2} = \sqrt{p(1-p)\left(\frac{1}{n_{.1}} + \frac{1}{n_{.2}}\right)},$$

where

$$p = \frac{n_{1.}}{n_{..}}.$$

For the Protestant data, $\hat{s}_{p_1-p_2} = 0.0409$. An appropriate test statistic is

$$z = \frac{p_1 - p_2}{\hat{s}_{p_1-p_2}},$$

which, under the null hypothesis, follows an approximate standard normal distribution. For the Protestant data, $z = 0.1432/0.0409 = 3.503$, which has a two-tailed P VALUE of 0.00046. So we reject the null hypothesis that mainline and fundamentalist Protestants are equally likely to approve of abortion for unmarried women.

When squared, this z statistic becomes the CHI-SQUARE (χ^2) TEST statistic used to test for association in a 2×2 contingency table. For the Protestant data $\chi^2 = z^2 = 12.27$, which, like the z statistic, has a p value of 0.00046.

The z and χ^2 tests are considered reasonable approximations when all table counts n_{ij} are at least 5 or so. For tables with lower counts, z and χ^2 should be avoided in favor of Fisher's exact test (Hollander & Wolfe, 1999).

—Paul T. von Hippel

REFERENCES

Agresti, A., & Cato, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician, 54*(4), 280–288.

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences.* Upper Saddle River, NJ: Prentice Hall.

Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods* (2nd ed.). New York: Wiley.

DIFFERENCE SCORES

A difference score is the difference of two values, usually two values on the same variable measured at different points in time. For example, suppose variable X is an attitude scale on Aggressiveness and is administered to subjects at Time 1, and then again 6 months later at Time 2. Then, $X_1 - X_2$ equals the difference score, or change, in Aggressiveness from one time to the next. Some variables measured repeatedly over time, such as number of homicides in a nation, tend to exhibit an upward drift. Such variables may be converted to difference scores in order to flatten out that trend.

—Michael S. Lewis-Beck

See also CHANGE SCORES

DIMENSION

A term used by Paul Lazarsfeld (1958) to refer to the different aspects that a CONCEPT may assume. Many concepts are regarded as multidimensional in the sense that it may be possible to specify on a priori grounds different components of those concepts. *Dimension* also refers to the directional spread in a data MATRIX or a CONTINGENCY TABLE, such as a row (vector or variable), a column (vector or variable), etc.

—Alan Bryman

REFERENCE

Lazarsfeld, P. (1958). Evidence and inference in social research. *Daedalus, 87,* 99–130.

DISAGGREGATION

Disaggregation is the breaking down of aggregates into smaller and smaller units. Data on geographic units, say, the American states, may be disaggregated to congressional districts, or to the still-smaller unit of the county. Aggregated data are often aggregations from a smaller to a larger geographic unit, as in the foregoing example. However, it may refer to human populations. For example, behavior observed at the group level, such as cooperation in a work organization, might be disaggregated to the individual level, such as the cooperativeness of the individual worker. Or, as another example, U.S. presidential voting behavior might be studied at the disaggregated level of the individual voter, or at the highly aggregated level of the presidential election outcome in the nation as a whole.

—Michael S. Lewis-Beck

DISCOURSE ANALYSIS

There is a wide range of theory and method that uses the term *discourse analysis* (henceforth DA), largely because the varieties of DA derive from different academic disciplines. All of them are concerned with the structures and functions of discourse, or talk and text. However, most varieties of DA have little obvious connection with each other. We focus here on the kind of DA that is closely tied to the analysis of social actions and interaction, and that has developed primarily within the social sciences, and social psychology in particular. This excludes the cognitive psychology of "discourse processing," as well as developments within linguistics that extend grammatical and pragmatic analysis beyond the boundaries of single sentences. An impression of the broad range of kinds of DA can be gleaned from Teun Van Dijk's (1985) four-volume collection, as well as from other entries in this encyclopedia, particularly those on CRITICAL DISCOURSE ANALYSIS, which is grounded in linguistics; CONVERSATION ANALYSIS, which is based in sociology; and NARRATIVE ANALYSIS, which is derived from linguistics and literary criticism.

In their influential work *Discourse and Social Psychology,* Jonathan Potter and Margaret Wetherell (1987) defined a theoretical and methodological approach to discourse that focuses on everyday descriptions and evaluations, or "versions," of things. Versions have three major features: *construction, function,* and *variability.*

1. *Construction* has two senses. Discourse is both constructed (built from ready-made linguistic resources) and constructive (offering particular versions of the world, as distinct from alternative versions).
2. Discourse is always performative, or *functional.* That is, things said or written in everyday talk and text are invariably produced with regard to some context of interaction or argument, where they perform actions such as evaluating, criticizing, requesting, confessing, claiming, defending, refusing, and so on.
3. Versions are *variable,* which is to say that we should not expect people to be consistent. It is this latter property, variability, that generated a profound critique of traditional theory and methods concerning the nature and measurement of attitudes, and also the conduct and interpretation of interviews.

The three features are mutually implicated. Discourse is *constructive* of whatever it describes, in that an indefinitely extensive range of alternative descriptions is always possible, without their becoming simply or objectively false. This permits the choice of any particular description to be *functional,* or action-performative, not only in clear cases, such as requests and questions, but also in the case of ostensibly simple, straightforward, factual descriptions. Indeed, DA's major analytic concern has been with the constructive and functional work done by factual descriptions. Furthermore, given the functional work that versions perform, *variability* across versions, even within the talk of the same speaker, can be expected on the basis that different versions may be doing different things on different occasions. Again, this principle has generated far-reaching critiques of other methods, including traditional uses of INTERVIEWS and QUESTIONNAIRES, in which variability is systematically avoided or removed in favor of defining a person's consistent attitude or understanding.

HISTORICAL DEVELOPMENT

DA's sources include linguistic philosophy, POSTSTRUCTURALISM, RHETORIC, ETHNOMETHODOLOGY, and conversation analysis. These coalesced in an approach to discourse that focuses on its place in social practices and its role in defining, legitimating, and undermining factual versions of the world. Its immediate origin was in the sociology of science. Nigel Gilbert and Michael Mulkay (1984) had a problem finding consistency across the things scientists said in interviews, in their informal talk, and in the accounts given in scientific papers of the grounds on which knowledge claims are made and refuted. Two crucial analytic moves were necessary. First, rather than using current scientific consensus as their own criterion of truth and error, they decided to treat all notions of truth and error as participants' constructions, and to focus on analyzing how scientists produced them. Second, rather than worrying about how to eliminate inconsistency, they chose to make inconsistency, or variability, the central phenomenon. Two contrasting INTERPRETATIVE REPERTOIRES were identified in scientific accounts: an *empiricist repertoire* that accounted for scientific truth, and a *contingent repertoire* that called upon personal and social factors and foibles in accounting for error. The study emphasized how both kinds of accounts, rather than just the standard empiricist repertoire, were essential elements in scientific explanation.

Potter and Wetherell's contribution was to extend these developments in the sociology of science into a broadly applicable theory and method for analyzing discourse in general. They first extended the principles of construction, function, and variability, and of interpretative repertoires, to the study and critique of attitude theory and measurement in social psychology. Rather than looking for coherent and consistent psychological states such as "attitudes," which might be assumed to be expressed in things people say, Potter and Wetherell, along with other discourse-based critics such as Michael Billig, focused on the ways in which descriptions and evaluations are produced in talk. Again, variability and the functional, action-performative nature of discourse came to the fore, and criticisms were produced of the older established methods by which such things are usually hidden from view in pursuit of consistency, reliability, or measures of central tendency.

Recent work in DA has taken two main directions, reflected in different methodological emphases and research topics. One direction has been to develop the notion of "interpretative repertoire" into a concept capable of linking the details of discourse practices, generally made available through transcripts of interview talk, to larger historical and ideological themes. Repertoires are considered to provide the resources on which people draw when constructing versions of people and events in the world. This strand of DA (e.g., Wetherell & Potter, 1992) focuses on controversial topics with clear ideological import, such as race, gender, and discriminatory practices. It is influenced by Michel Foucault's concept of discourses as historically evolved concepts and frameworks of explanation related to the growth of institutions such as the law, education, medicine, and psychiatry, tracing their infusion into commonsense language and ideology. In contrast to Foucault's discourses, interpretative repertoires are intended to be more analytically tractable, being closer to discourse *practices,* or the actions and interactional work performed by actual instances of talk and text. The second direction in DA's development is called *discursive psychology,* which, taking its major analytic impetus from conversation analysis, focuses on the close analysis of recordings of everyday, naturally occurring talk and, to a lesser extent, textual materials such as newspaper reports.

DISCURSIVE PSYCHOLOGY

Discursive psychology (henceforth DP) is the application of discourse analytic principles to psychological themes and concepts (Edwards, 1997; Edwards & Potter, 1992). In spite of the focus on psychology, however, its range of relevance and application is very broad. This is because psychological themes are pervasive in the workings of commonsense accountability. In everyday talk, and across a range of institutional settings, there is a prime concern with what people know, think, feel, understand, want, intend, and have in mind, as well as their personal dispositions, and so forth. Indeed, such themes are often defining features of the normative business of institutional settings, such as in education, in courtrooms, and in psychotherapy and counseling, where what people think, understand, and remember are central concerns.

There are three major strands of research in DP:

1. Respecification and critique of psychological topics and explanations, which reworks notions such as "memory" and "emotion" as discourse practices
2. Investigations of how everyday psychological terms (the "psychological thesaurus") are used

3. Studies of how psychological themes are handled and managed in talk and text, without having to be overtly labeled as such

The use of the term *psychological* here does not imply a treatment of these themes and topics as rooted in psychological processes and explanations. Rather, DP avoids psychological theorizing in favor of the pragmatics of social actions. All three strands tend to be relevant in actual studies, where the emphasis is on empirical analysis of everyday and institutional talk and text.

ANALYTICAL PRINCIPLES

There is a range of analytical principles that guides empirical work in DP. The following is a representative rather than definitive or exclusive list.

1. There is a preference for *naturally occurring* data, or discourse produced as part of everyday and institutional life, rather than data obtained through, say, research interviews. This is not crucial, but it reflects the principle that discourse performs social actions; a lot of social science's "qualitative" data are interview talk produced aside from, and somewhat reflective on, the lives discussed.

2. Ask not what state of mind the talk/text expresses, nor what state of the world it reflects, but what *action* is being done by saying things that way.

3. Look for *participants' concerns*; their categories, concepts, the things they are dealing with. Look for how they use psychological concepts, or orient to psychological concerns. DP is not, at any stage, looking for evidence of underlying psychological processes.

4. For any issue that the analyst might bring to the data, try *topicalizing* it instead. That is, try seeing to what extent it is something that the participants themselves handle or deal with in some way.

5. Focus on *subject-object* relations (or "mind-world" relations). Look for how descriptions of people and their mental states are tied to, or implied by, descriptions of actions, events, and objects in the external world, and vice versa.

6. Examine how the *current speaker/writer* attends reflexively to his or her own subject-object issues: his or her grounds for knowing and telling things; how he or she deals with the possibility of not being believed, or of being considered biased or emotionally involved.

7. For any content of talk, ask *how, not why,* it is said. Ask "What does it do, and how does it do it?" "Why" questions often depend on normative assumptions about mind, language, and social settings, and they are often best translated into "how" questions. So, instead of asking "Why did X say that?" we can ask, "Did X say that in some way that attends to her possible motives or reasons for saying it?"

8. Analyze *rhetorically.* Ask "What is being denied, countered, forestalled, etc., in talking that way?"

9. Analyze *semiotically.* In the case of DA (see also SEMIOTICS), this means asking, "What is *not* being said here that could have been said by using closely similar words or expressions?" The principle is that language is a system of differences, such that all words, all details, have meanings because there are alternatives. The selection of a particular word or expression is crucial, and we can get to it analytically by imagining plausible alternatives and by looking to see what alternative descriptions may actually be in play. This is a method for seeing the specificity of what is said, not a claim that any such alternatives are actually entertained by speakers.

10. Analyze *sequentially.* For any stretch of discourse, look at the immediately prior and subsequent discourse, or turns at talk, to see what the content of the current turn or section is dealing with and making relevant. This is the major principle of conversation analysis, enshrined in the PROOF PROCEDURE, and it is an essential ingredient in DP for all the same reasons. What we are analyzing is not a collection of speakers' thoughts being put into words, like quotations, but a sequence of actions being performed in sequentially relevant ways.

11. When recurrent patterns are found, look for *deviant cases,* which are examples that do not seem to fit the developing analysis, and see if the analysis needs to be changed or the phenomenon redefined. Again, DEVIANT CASE ANALYSIS is a principle derived from conversation analysis.

AN ILLUSTRATION

Discourse analysis deals with bulky and varied data, not coded and reduced to categories, as in CONTENT ANALYSIS, but retaining its full textual, recorded, or transcribed form, as in conversation analysis. One consequence of that is that all of its principles cannot easily be demonstrated with brief examples. Nevertheless, we can illustrate the kind of data and analysis involved.

The following brief extract comes from a relationship counseling session in which husband Jeff and wife Mary are talking with a counselor about their marital difficulties. Relevantly to the extract, Mary has had an extramarital affair and is talking about Jeff's reactions to being told about it (see Edwards, 1997, for more details).

1. *Mary:* U::m (1.0) and then::, (.) obviously
2. you went through your a:ngry stage,
3. didn't you?
4. (.)
5. Ve:ry upset obviously, hh an:d uh, (0.6)
6. we: started ar:guing a lot, an:d (0.6)
7. just drifted awa:y.

Mary describes Jeff's reactions as "angry" (line 2) and "very upset" (line 5). Drawing on the *constructive, rhetorical,* and *semiotic* principles, note that these categories characterize Jeff's reactions as angry rather than, say, as having arrived at a condemnatory judgment of his wife's actions and character. They define his reaction as emotional, and as a specific kind of emotion. "Angry" sets up various possibilities that a cognitive "judgment," for example, would not imply. Emotion discourse often trades on its commonsense irrationality, for example, which can be used to lead us away from making inferences about Mary and focus more on Jeff's state of mind. Similarly, a view, opinion, or judgment can (normatively) be more enduring than an emotional reaction, such as anger, and less expected to change. These are conceptual and rhetorical possibilities at this point in the analysis. Their significance will depend on what Mary and Jeff go on to say, because it is their significance *for them* that we are trying to find.

The description "your angry *stage*" starts to exploit a notion of anger as a temporary state that has its proper occasions and durations. It is a description that implies a kind of script (a routine, expectable sequence of events) for emotional reactions. For example, while acknowledging that Jeff's anger is normatively expectable, the idea of an angry *stage* projects the implication that it should not endure for an unreasonably long time. Mary makes rhetorical room here for something that she actually does go on to develop after this extract, which is the notion that Jeff's reactions are starting to become the problem that they have in their relationship. His enduring emotional reaction can start to become the topic for the counseling to focus on—his inappropriate feelings, rather than her unfaithful behavior and possible proclivities.

Therefore, the expression "your angry stage" is constructive, rhetorical, and performative. It is not just descriptive of Jeff (as an object in the world), nor merely an expression of Mary's mental understanding of him (her thoughts put into words). Rather, it works to *define* the problem they have as one residing in Jeff, while at the same time avoiding overt blame of him by Mary. Indeed, the term "your" angry stage, along with the normative notion "stage" and the expression "obviously" (line 1), present his reaction as one to which he is normatively entitled, at least for a limited period.

Note also how the description fits into a narrative sequence. The next thing that Mary recounts (and, by implication, what not only follows but *follows from* Jeff's reactions; see the entry on narrative analysis), is how "we started arguing a lot, and just drifted away" (lines 6–7). Their problems become joint ones, arguments, and a kind of nonagentive, nonblaming, "just" drifting apart. Jeff's "stage" is already past tense ("you *went* through your angry stage, *didn't* you?"), which again implies that his anger ought now to be over. The general direction of Mary's account, produced through its specific details, is to shift our understanding of their relationship troubles from a focus on her extramarital affair toward her husband's persisting emotional difficulties in dealing with it.

—Derek Edwards

REFERENCES

Edwards, D. (1997). *Discourse and cognition.* London: Sage.
Edwards, D., & Potter, J. (1992). *Discursive psychology.* London: Sage.
Gilbert, G. N., & Mulkay, M. (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse.* Cambridge, UK: Cambridge University Press.
Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour.* London: Sage.
Van Dijk, T. A. (Ed.). (1985). *Handbook of discourse analysis: Vols. 1–4.* London: Academic Press.
Wetherell, M., & Potter, J. (1992). *Mapping the language of racism: Discourse and the legitimation of exploitation.* Brighton, UK: Harvester Wheatsheaf.

DISCRETE

The term refers to a variable such as a CATEGORICAL variable, which groups units into discrete clusters. It is also often used to refer to instances where the units that make up a variable are discrete. Thus, number of children as a variable can only be in discrete units—you cannot have 3.67 children.

—Alan Bryman

See also Attribute, Categorical, Nominal Variable

DISCRIMINANT ANALYSIS

The basic idea of discriminant analysis is clearly illustrated by a study from sociologist Fred Provoost (1979). He conducted a scientific investigation into the discrimination between the rich and poor neighborhoods in the Dutch-speaking part of Belgium and the district of Brussels. Six discriminating variables were distinguished (mean income, educational attainment, occupation, housing situation, an index of social and cultural participation, and the presence of services), and these six variables, taken together, were examined for their capability of discriminating significantly between the two kinds of neighborhoods.

Research involving discrimination between social groups consists of three steps. First, and as reliably as possible, we make an a priori classification of groups. Because this classification is the subject of explanation, it represents the (dichotomous) dependent variable Y with categories such as those of rich and poor neighborhoods. Second, statistical data are gathered for a great number of characteristics, their discriminating capacity is determined by means of statistical analysis, and a weighted sum is calculated. These characteristics are the independent variables $X_1, X_2, \ldots$ (discriminating variables); their weights are $k_1, k_2, \ldots,$ respectively (discriminant weights); and the weighted sum is $f = k_1X_1 + k_2X_2 + \cdots$ (discriminant function). Third, if the analysis of the second step appears to be successful (significant discriminating capacities of the weighted sum and of each of the characteristics separately), then we have a good instrument for classifying additional units into one of the groups. This last step, classification, offers the opportunity for a well-founded determination of whether the neighborhoods not included in the analysis are rich or poor.

It follows from the foregoing that the structure of discriminant analysis is the same as that of multiple REGRESSION analysis except for one point: The measurement level of the dependent variable Y is now categorical (here even dichotomous) instead of interval. The causal diagram, which illustrates the research problem, is shown in Figure 1.

The units of analysis are districts or neighborhoods. The discriminating variables, mean income (X_1), educational attainment (X_2), ... up to the level of services (X_6), are measured at the interval or ratio level. As in multiple regression analysis, they can be considered as causal factors in a multicausal model. The dependent variable Y represents the groups. It can be seen as a categorical variable, here a dichotomy with the categories poor and rich. Anticipating the dummy approach, we can use the codes 0 and 1, respectively. We will no longer speak of the prediction of Y scores,

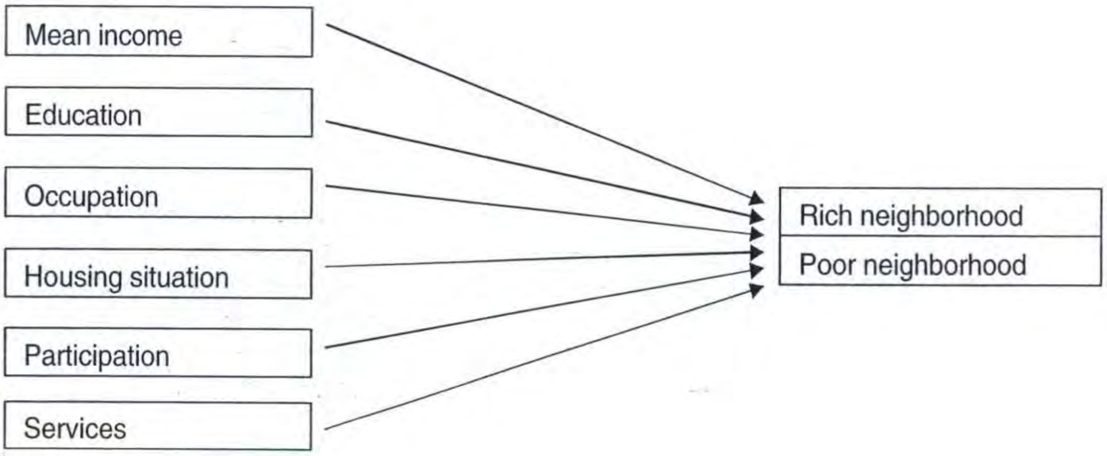

Figure 1

but rather of the classification into one of the two groups. As in multiple regression analysis, the model is also additive. This means that there are no effects of interaction in principle, that is, a weighted sum of X variables is sought, and product terms like X_1X_2 are not considered.

THE MODEL OF DISCRIMINANT ANALYSIS

In discriminant analysis, we want to examine whether a set of variables ($X_1, X_2, \ldots$) is capable of distinguishing (i.e., discriminating in the neutral sense) between a number of groups. Therefore, we search for a linear combination of the discriminating variables ($X_1, X_2, \ldots$) in such a way that the groups (poor and rich neighborhoods) are maximally distinguished. Such a linear combination is called a discriminant function and generally has the following form:

$$F - \bar{F} = k_1(X_1 - \bar{X}_1) + k_2(X_2 - \bar{X}_2) + \cdots + k_p(X_p - \bar{X}_p),$$

or

$$f = k_1x_1 + k_2x_2 + \cdots + k_px_p.$$

In this formula, f and x_i are expressed as deviations of mean (small letters). The coefficients k_i are called discriminant weights. The variables x_1 to x_p are the discriminating variables, where p is the total number of X variables.

This can be written much simpler in matrix notation as $\mathbf{f} = \mathbf{Xk}$, in which $\mathbf{f}$ is an $n \times 1$ column vector of discriminant scores, $\mathbf{X}$ is an $n \times p$ matrix of scores of the discriminating variables, and $\mathbf{k}$ is a $p \times 1$ column vector of discriminant weights.

In the case of two groups, there is only one discriminant function. If, on the other hand, several groups were to be compared, then the number of discriminant functions would maximally be equal to $\min(g-1, p)$. For example, if $g = 4$ groups are analyzed by means of $p = 5$ discriminating variables, then we have $g - 1 = 3$ and $p = 5$, so that the maximum number of discriminant functions is equal to 3.

OBJECTIVES OF THE TECHNIQUE

The purposes of discriminant analysis are threefold. Keeping the example of rich and poor neighborhoods (two groups) in mind and restricting ourselves to two discriminating variables (x_1 and x_2), these purposes can be summarized as follows:

1. We search for a linear function $f = k_1x_1 + k_2x_2$, called a discriminant function, in such a way that the scores of the neighborhoods on this function f (discriminant scores) exhibit the property of maximizing the ratio of between-group and within-group variability. This comes down to the calculation of the discriminant weights k_1 and k_2 and, next, to the calculation of the n discriminant scores.

2. We examine whether the discrimination established by the function f can be generalized to the population. Is there a significant difference between the two group centroids, that is, between the means of the two groups for variables X_1 and X_2 taken together? This comes down to the performance of Hotelling's T^2 test, which is an extension of the simple Student's t-test. In the more general case of three or more groups, this will be Wilks's Λ test.

3. Making use of the discriminant function f, we also want to determine for each neighborhood, and eventually for new neighborhoods, to which of the two groups they belong. Given our knowledge of a neighborhood's score of income and level of services, can we predict whether it is rich or poor? This comes down to determining a cutoff point (in the middle between the two group centroids) and verifying whether the discriminant scores are situated to the left or to the right of this point. If a discriminant score is equal to the cutoff point, then the corresponding neighborhood can be randomly assigned to one of the two groups.

This process of classification of units into groups is mostly considered in the broader context of statistical decision theory and Bayesian statistics.

CALCULATION OF THE DISCRIMINANT FUNCTION f

Function f is a linear combination of x variables: $f = k_1x_1 + k_2x_2 + \cdots + k_px_p$. This function f will be calculated in such a way that it shows the highest possible discrimination between the two groups, that is, that the ratio of the between- and within-DISPERSION is maximal. For that purpose, the variation of f is split into a between and within part, and the discriminant weights are calculated in such a way that the ratio between/within is as high as possible.

To solve this problem of maximization, the eigenstructure of a typical matrix will be examined. In discriminant analysis, this typical matrix is $\mathbf{W}^{-1}\mathbf{B}$, which represents the between/within ratio of variations and covariations. Most of the references at the end of this entry give the details of this eigenstructure examination.

CLASSIFICATION AND PREDICTION

After having found the discriminant function, the expected f value (= discriminant score) of each of the individual neighborhoods can be calculated by substituting the scores of the X variables into the function f. The same operation can also be performed for the group centroids in order to obtain an f score, $\bar{f}_1$, for the centroid of Group 1 and $\bar{f}_2$ for the centroid of Group 2. The point in the middle of these two projections of the group centroids on the f-axis can be chosen as the cutoff point. If the two groups are of equal size, then this cutoff point would lay in the origin, i.e., $f_c = 0$. It will lie to the left of the origin when Group 1 contains fewer neighborhoods than Group 2.

Looking at the discriminant scores of the neighborhoods, we can classify each neighborhood in one of the two groups. Those scoring to the right of f_c ($f_i > f_c$) are assigned to Group 2 and those to the left of f_c ($f_i < f_c$) are classified as Group 1. We can then indicate the predicted group membership. Comparing this with the original Y scores, we can see which percentage of neighborhoods (e.g., 80%) are correctly classified. The discriminant function is not capable of a perfect classification when there is some overlap between the two groups.

SIGNIFICANCE TESTING

We want to examine whether there is a significant difference between the two group centroids, that is, between the means of the two groups for the X variables taken together. This is a test of the global model. In addition, we want to perform separate significance tests, called *univariate tests,* because the X variables are considered one at a time.

For the multivariate test, we make use of Wilks' lambda, Λ, a measure, just like Student's t and Fisher's F, that confronts the ratio of the between-dispersion and the within-dispersion from the sample with this ratio under the null hypothesis. Wilks' Λ is the multivariate analogy of this ratio in its most general form, so that t and F are special cases of Λ. This measure is therefore used for more complex multivariate analysis techniques, such as MULTIVARIATE ANALYSIS OF VARIANCE AND COVARIANCE, CANONICAL CORRELATION ANALYSIS, and multiple discriminant analysis.

When only two groups are involved, as in our example, then Hotelling's T^2 test can be performed, which is a special case of Wilks' lambda test. T^2 has the nice property that a special matrix is included, $d'C_w^{-1}d$, which is known as Mahalanobis' distance, a generalized distance in which the unequal dispersions of the discriminating variables and their mutual associations are taken into account.

EXTENSIONS

Discriminant weights are commonly computed from the eigenstructure of the data. Other formulations are possible, for example, starting from the total-sample covariance or correlation matrix ($\mathbf{C}_T$ or $\mathbf{R}_T$) or from a dummy regression approach obtained by regressing the dummy variable Y on the discriminating variables and leading to the same discriminant weights up to a proportionality transformation. In short, quoting Paul Green, there is an "embarrassment of riches" insofar as computational alternatives for discriminant weights are concerned.

In the foregoing discussion, it was taken for granted that the variables were all measured at the interval or ratio level. It was also assumed that the relationships were all linear. These two requirements can, however, be relaxed. In addition to SPSS, the computer program CRIMINALS, from Leiden University in The Netherlands, performs discriminant analysis, which is nonmetric as well as nonlinear. The interested reader is referred to Gifi (1980).

Linear discriminant analysis is also a technique for which a lot of additional requirements have to be fulfilled; next to linearity, also normality and homogeneity of covariance matrices. In cases in which the covariance matrices do differ significantly, a quadratic discriminant function might be more appropriate. Literature on this topic can be found in Lachenbruch (1975). In case of serious violations of requirements, LOGISTIC REGRESSION is always a good alternative for discriminant analysis.

—Jacques Tacq

REFERENCES

Cooley, W., & Lohnes, P. (1971). *Multivariate data analysis.* New York: Wiley.

Gifi, A. (1980). *Nonlinear multivariate analysis.* Leiden, The Netherlands: Leiden University Press.
Green, P. (1978). *Analyzing multivariate data.* Hinsdale, IL: Dryden.
Lachenbruch, P. A. (1975). *Discriminant analysis.* New York: Hafner.
Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis.* London: Sage.
Van de Geer, J. (1971). *Introduction to multivariate analysis for the social sciences.* San Francisco: Freeman.

DISCRIMINANT VALIDITY

Discriminant (also referred to as divergent) validity is evidence that a measure is not unduly related to other similar, yet distinct, constructs (Messick, 1989). CORRELATION coefficients between measures of a construct and measures of conceptually different constructs are usually given as evidence of discriminant validity. If the correlation coefficients are high, this shows a lack of discriminant validity or weak discriminant validity, depending on the theoretical relationship and the magnitude of the coefficient. On the other hand, if the correlations are low to moderate, this demonstrates that the measure has discriminant validity. Discriminant validity is especially relevant to the validation of tests in which irrelevant variables may affect scores; for example, it would be important to show that a measure of passionate love is unrelated to social desirability.

Other methods of demonstrating discriminant validity are multitrait-multimethod (MTMM) (see MIXED METHODS RESEARCH) and FACTOR ANALYSIS. The MTMM is useful for examining both convergent and discriminant validity. This procedure involves measuring more than one construct by more than one method. For example, a new measure of job satisfaction could be correlated with existing, validated measures of life satisfaction and relationship satisfaction. Self-report data could be obtained, as well as supervisor and partner ratings on each of the three constructs (i.e., three different methods). Theoretically, job, life, and relationship satisfaction are conceptually different constructs; thus, the correlations between these measures would need to be low in order to indicate discriminant validity. Similarly, the correlations between ratings from different observers about the different constructs would be low if discriminant validity is present.

Factor analysis can be used to show that similar but conceptually different constructs are distinct; for example, Pierce, Gardner, Cummings, and Dunham (1989) assessed the discriminant validity of their newly developed organization-based self-esteem measure by factor analyzing it with several affective measures (e.g., organizational commitment). Their findings showed that organization-based self-esteem could be distinguished from affective measures because it loaded on a different factor compared to the other measures. Conversely, factor analysis can also be used to demonstrate lack of discriminant validity. For example, Carless (1998) used confirmatory factor analysis to show that the subscales of a measure of transformational leadership were highly correlated and formed a higher-order construct of transformational leadership. Factor analysis is an essential technique when studying variables that have conceptual overlap and when data are collected from only one source (e.g., an individual respondent). Depression and anxiety, for example, are theoretically distinct constructs, but there is also considerable empirical overlap between them; depressed individuals are also often highly anxious. In order to validate scales, it is important not only to examine the magnitude of the correlation between these variables, but also to demonstrate discriminant validity by factor analysis.

—Sally A. Carless

REFERENCES

Carless, S. A. (1998). Assessing the discriminant validity of the transformational leadership behaviour as measured by the MLQ. *Journal of Occupational and Organizational Psychology, 71,* 353–358.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education and Macmillan.
Pierce, J. L., Gardner, D. G., Cummings, L. L., & Dunham, R. B. (1989). Organization-based self-esteem: Construct definition, measurement, and validation. *Academy of Management Journal, 32,* 622–648.

DISORDINALITY

Disordinality is an INTERACTION where the RELATIONSHIP between X_1 and Y is different and opposing depending on the value of X_2. In the usual interaction, the effect of X_1 on Y is merely different, depending on the value of X_2. Suppose the question is how Education (X_1, measured in years) relates to Income (Y, measured in dollars) for men and women (X_2, scored $1 = \text{male}$, $0 = \text{female}$) in a sample from the American labor force. Now regress Y on X_1, first for the male subsample and then for the female subsample. If the finding is that the REGRESSION slope is significantly steeper for men, we may conclude that there is an interaction, with the impact of education on income dependent on gender. This result would be an ordinal interaction. But suppose instead that our finding was that the regression slope for men was positive, but the slope for women was not only significantly different but negative. That would be a disordinal interaction, where X_1 has a positive effect within one group and a negative effect within the other. It is useful to imagine that, in this instance, the two regression lines, rather than being nonparallel, as in the usual ordinal interaction, actually cross each other.

—Michael S. Lewis-Beck

See also ORDINAL INTERACTION

DISPERSION

Dispersion roughly refers to the degree of scatter or variability in a collection of observations. For example, individuals might differ about whether a political leader is doing a good job; children might respond differently to a method aimed at enhancing reading; and even in the physical sciences, measurements might differ from one occasion to the next because of the imprecision of the instruments used. In a very real sense, it is dispersion that motivates interest in statistical techniques.

A basic issue is deciding how dispersion should be measured when trying to characterize a POPULATION of individuals or things. That is, if all individuals of interest could be measured, how should the variation among these individuals be characterized? Such measures are population measures of dispersion. A related issue is deciding how to estimate a population measure of dispersion based on a SAMPLE of individuals.

Choosing a measure of dispersion is a complex issue that has seen many advances during the past 30 years. The choice depends in part on the goal of the investigator, with the optimal choice often changing drastically depending on what an investigator wants to know or do. More than 150 measures of dispersion have been proposed, comparisons of which were made by Lax (1985) based on some fundamental criteria that are relevant to a range of practical problems. Although most of these measures seem to have little practical value, at least five or six play an important and useful role.

Certainly the best known measure of dispersion is the population VARIANCE, which is typically written as σ^2. It is the average (or expected) value of the squared difference between an observation and the population MEAN, μ. That is, if all individuals could be measured, the average of their responses is called the population mean, μ, and if, for every observation, the squared difference between it and μ were computed, the average of these squared values is σ^2. In more formal terms, $\sigma^2 = E(X - \mu)^2$, where X is any observation we might make and E stands for expected value. The (positive) square root of σ^2, σ is called the (population) STANDARD DEVIATION. Based on a sample of n individuals, if we observe the values $X_1, \ldots, X_n$, the usual estimate of σ^2 is the sample variance:

$$s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2,$$

where $\bar{X} = \sum X_i/n$ is the sample mean.

For some purposes, the use of the standard deviation, σ, stems from the fundamental result that the probability of an observation being within some specified distance from the mean, as measured by σ, is completely determined under normality. For example, the probability that an observation is within 1 standard deviation of the mean is 0.68, and the probability of being within 2 standard deviations is 0.954. These properties have led to a commonly used measure of EFFECT SIZE (a measure intended to characterize the extent to which two groups differ) as well as a frequently employed rule for detecting OUTLIERS (unusually large or small values). Shortly after a seminal paper by Tukey (1960), it was realized that even very small departures from normality can alter these properties substantially, resulting in practical problems that commonly occur.

Consider, for example, the two distributions shown in Figure 1. One is a standard normal, meaning it has mean 0 and variance 1. The other is not a normal distribution, but rather something called a mixed

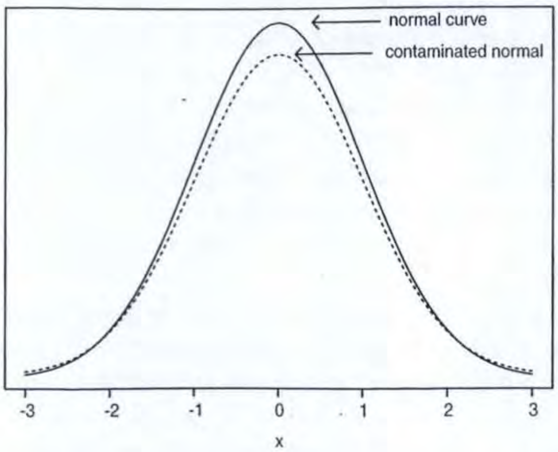

Figure 1

or contaminated normal. That is, two NORMAL DISTRIBUTIONS, both having the same mean but different variances, are mixed together. (A precise description of this particular mixed normal can be found in Wilcox, 2003, but the details are not important here.) The important point is that, despite the obvious similarity between the two distributions, their variances differ substantially: The normal has variance 1 but the mixed normal has variance 10.9. This illustrates the general principle that small changes in the tails of a distribution can substantially affect σ^2.

Now consider the property that under normality, the probability that an observation is within 1 standard deviation of the means is 0.68. Will this property hold under nonnormality? In some cases, it is approximately true, but in other situations, this is no longer the case, even under small departures from normality (as measured by any of several commonly used metrics by statisticians). For the mixed normal in Figure 1, this probability exceeds 0.999.

The sample variance reflects a similar property: It is very sensitive to OUTLIERS. In some situations, this can be beneficial, but for a general class of problems, it is detrimental. For example, a commonly used rule is to declare the value X an outlier if it is more than 2 (sample) standard deviations from the mean, the idea being that under normality, such a value is unusual from a probabilistic point of view. That is, declare X an outlier if

$$\frac{X - \bar{X}}{s} > 2.$$

This rule suffers from masking, however, meaning that even obvious outliers are not detected because of the sensitivity of s to outliers. Consider, for example, the values 2, 2, 3, 3, 3, 4, 4, 4, 100,000, 100,000. Then, $\bar{X} = 20002.5$, $s = 42162.38$, and $(100{,}000 - 20{,}002.5)/s = 1.9$. So, the value 100,000 would not be declared an outlier based on the rule just given, yet it is clearly unusual relative to the other values.

For detecting outliers, two alternative measures of dispersion are typically used. The first is called the median absolute deviation (MAD) statistic. It is the median of the n values

$$|X_1 - M|, \ldots, |X_n - M|,$$

where M is the usual median of $X_1, \ldots, X_n$. Now X is declared an outlier if

$$\frac{X - M}{\text{MAD}/0.6745} > 2.24.$$

The other alternative measure is the INTERQUARTILE RANGE, which is just the difference between the upper and lower quartiles. (This latter measure of dispersion is used by boxplot rules for detecting outliers.) An important feature of both of these measures of dispersion is that they are insensitive to extreme values; they reflect the variation of the centrally located observations, which is a desirable property when detecting outliers with the goal of avoiding masking.

Another approach when selecting a measure of dispersion is to search for an ESTIMATOR that has a relatively low STANDARD ERROR over a reasonably broad range of situations. In particular, it should compete well with s under normality, but it is desirable to maintain a low standard error under nonnormality as well. So, for example, if it is desired to compare two groups in terms of dispersion, the goal is to maintain high power regardless of whether sampling is from a normal distribution or from the mixed normal in Figure 1. This is roughly the issue of interest in the paper by Lax (1985). Among the many methods he considered, two estimators stand out. The first is called a percentage bend midvariance, and the other is a biweight midvariance. The tedious details regarding how to compute these quantities are not given here, but they can be found in Wilcox (2003) along with easy-to-use software. In terms of achieving a relatively low standard error, the sample standard deviation, s, competes well with these two alternative measures under normality, but for nonnormal distributions, s can perform rather poorly.

EFFECT SIZE

The variance has played a role in a variety of other situations, stemming in part from properties enjoyed by the normal distribution. One of these is a commonly used measure of effect size for characterizing how two groups differ:

$$\Delta = \frac{\mu_1 - \mu_2}{\sigma_p},$$

where μ_1 and μ_2 are the population means, and where, by assumption, the two groups have equal standard deviations, σ_p. The idea is that if $\Delta = 1$, for example, the difference between the means is 1 standard deviation, which provides perspective on how groups differ under normality. But concerns have been raised about this particular approach, because under nonnormality, it can mask a relatively large difference (e.g., Wilcox, 2003).

STANDARD ERRORS

The notion of variation extends to SAMPLING DISTRIBUTIONS in the following manner. Imagine a study based on n observations resulting in a sample mean, say, $\bar{X}$. Now imagine the study is repeated many times (in theory, infinitely many times), yielding the sample means $\bar{X}_1, \bar{X}_2, \ldots,$ with each sample mean again based on n observations. The variance of these sample means is called the squared standard error of the sample mean, which is known to be σ^2/n under random sampling. That is, it is the variance of these sample means: $E(\bar{X} - \mu)^2 = \sigma^2/n$. Certainly, the most useful and important role played by the variance is making INFERENCES about the population mean based on the sample mean. The reason is that the standard error of the sample mean, $\sigma/\sqrt{n}$, suggests how inferences about μ should be made assuming normality, but the details go beyond the scope of this entry.

WINSORIZED VARIANCE

Another measure of dispersion that has taken on increasing importance in recent years is the Winsorized variance; it plays a central role when comparing groups based on trimmed means. It is computed as follows. Consider any γ such that $0 \leq \gamma < 0.5$. Let $g = \gamma n$, rounded down to the nearest integer. So if $\gamma n = 9.8$, say, $g = 9$. Computing a γ-trimmed mean refers to removing the g smallest and g largest observations and averaging those that remain (see TRIMMING). Winsorizing n observations refers to setting the g smallest values to the smallest value not trimmed, and simultaneously setting the g largest values equal to the largest value not trimmed. The sample variance, based on the Winsorized values, is called the Winsorized variance. Both theory and simulation studies indicate that trimming can reduce problems associated with means due to nonnormality, particularly when sample sizes are small. Although not intuitive, theory indicates that the squared standard error of a trimmed mean is related to the Winsorized variance (e.g., Staudte & Sheather, 1990), and so the Winsorized variance has played a role when testing hypotheses based on trimmed means. Also, it has been found to be useful when searching for robust analogs of PEARSON'S CORRELATION (e.g., Wilcox, 2003).

Numerous multivariate measures of dispersion have been proposed as well, several of which currently have considerable practical value. Some of these measures are described in Rousseeuw and Leroy (1987) and Wilcox (2003).

—Rand R. Wilcox

REFERENCES

Lax, D. A. (1985). Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association, 80,* 736–741.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression & outlier detection.* New York: Wiley.

Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing.* New York: Wiley.

Tukey, J. W. (1960). A survey of sampling from contaminated normal distributions. In I. Olkin et al. (Eds.), *Contributions to probability and statistics.* Stanford, CA: Stanford University Press.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques.* San Diego, CA: Academic Press.

DISSIMILARITY

Dissimilarity is a generic term used to describe measures of similarity or dissimilarity (also called "proximity" measures, especially when used in conjunction with distance models of MULTIDIMENSIONAL SCALING and CLUSTER ANALYSIS). Such measures may be (a) direct judgments or ratings of similarity made by subjects, or (b) aggregate measures derived from an array of variables. Direct judgments are made between a pair of objects, using a rating scale graded between "totally similar" and "totally dissimilar." The judgments can then be treated as "quasi-distances" and directly scaled. Aggregate measures differ by LEVEL OF MEASUREMENT and are usually noncorrelational, because for distance model representation, a dissimilarity measure should be nonnegative and symmetric, and it should obey the triangle inequality.

—Anthony P. M. Coxon

REFERENCE

Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification, 5*, 5–48.

DISTRIBUTION

Distributions are ubiquitous in social science. They appear in the frameworks for analyzing particular topical domains, where they provide a diversity of shapes to represent the range of possible variability in fundamental quantities; in theories, where they provide basic tools for deriving predictions; and in empirical work, where underlying distributional forms are estimated and distributions serve to characterize operation of the unobservables.

CLASSIFICATION OF DISTRIBUTIONS

Distributions may be classified along several dimensions—continuous versus discrete, univariate versus multivariate, modeling versus sampling—or according to mathematical characteristics of their associated functions. For introduction and comprehensive exposition, see the volumes in Johnson and Kotz's series, *Distributions in Statistics,* and their revisions (e.g., Johnson, Kotz, & Balakrishnan, 1994), as well as the little handbook by Evans, Hastings, and Peacock (2000).

ASSOCIATED FUNCTIONS

Mathematically specified distributions have associated with them a variety of functions. The most basic is the distribution function (also known as the cumulative distribution function), which is defined as the probability α that the variate X assumes a value less than or equal to x and is usually denoted $F_X(x)$, or simply $F(x)$. Probably the best known of the associated functions is the PROBABILITY DENSITY FUNCTION, denoted $f(x)$, which, in continuous distributions, is the first derivative of the distribution function with respect to x (and which, in discrete distributions, is sometimes called the probability mass function). For example, in the normal family, the BELL-SHAPED CURVE depicting the probability density function is a more common representation than the ogive depicting the distribution function. Two of the most useful for social science are the quantile function, which, among other things, provides the foundation for whole-distribution measures of inequality, such as Pen's Parade, and the hazard function.

Of course, all the associated functions are related to each other in specified ways. For example, as already noted, among continuous distributions, the probability density function is the first derivative of the distribution function with respect to x. The quantile function, variously denoted $G(\alpha)$ or $Q(\alpha)$ or $F^{-1}(\alpha)$, is the inverse of the distribution function, providing a mapping from the probability α to the QUANTILE x. Important relations between the two include the following (Eubank, 1988; Evans, Hastings, & Peacock, 2000):

$$F[Q(\alpha)] = \alpha, \text{ for } F \text{ continuous}$$

$$Q[F(x)] = x, \text{ for } F \text{ continuous and strictly increasing.}$$

DISTRIBUTIONAL PARAMETERS

Distributions also have associated with them a number of parameters. Of these, three are regarded as basic—the location, scale, and shape parameters. The location parameter is a point in the variate's domain, and the scale and shape parameters govern the scale of measurement and the shape, respectively. It is sometimes convenient to adopt a uniform notation across distributional families, such as that in Evans, Hastings, and Peacock (2000), which uses lowercase, italicized English letters a, b, and c to denote the location, scale, and shape parameters, respectively. Variates differ in the number and kind of basic parameters. For example, the normal distribution has two parameters, a location parameter (the mean) and a scale parameter (the standard deviation).

SUBDISTRIBUTION STRUCTURE

Distributions are useful for representing many phenomena of special interest in social science. These include, besides inequality, the subgroup structure of a population. In general, two kinds of subgroup structure are of interest, and they may be represented by versions of the distributional operations of CENSORING AND TRUNCATION. In the spirit of Kotz, Johnson, and Read (1982, p. 396) and Gibbons (1988, p. 355), let censoring refer to selection of units by their ranks or percentage (or probability) points; and let truncation refer to selection of units by values of the variate. Thus, the truncation point is the value x separating the subdistributions; the censoring point is the percentage point α separating the subdistributions. For example, the subgroups with incomes less than $20,000 or greater than $80,000 each form a truncated subdistribution; the top 5% and the bottom 10% of the population each form a censored subdistribution.

There is a special link between these two kinds of subgroup structure and Blau's (1974) pioneering observation that much of human behavior can be traced to the differential operation of quantitative and qualitative characteristics. Quantitative characteristics—both cardinal characteristics, such as wealth, and ordinal characteristics, such as intelligence or beauty—generate both truncated and censored subdistribution structures. For example, the subgroups "rich" and "poor" may be generated by reference to an amount of income or by reference to percentages of the population. However, qualitative characteristics—such as race, ethnicity, language, or religion—may be so tightly related to quantitative characteristics that the subgroups corresponding to the categories of the qualitative characteristic are nonoverlapping and thus provide the basis for generating censored subdistribution structures. For example, in caste, slavery, or segmented societies, the subdistribution structure of interest may be a censored structure in which the percentages pertain to the subsets formed by a qualitative characteristic—such as "slave" and "free."

DISTRIBUTIONS IN THEORETICAL WORK

Often, theoretical analysis of a process or derivation of predictions from a theory makes use of modeling distributions. This approach makes it possible to investigate the possible variability of a process or, relatedly, the impact of the shape of underlying distributions, such as the income distribution. The protocol involves selecting appropriate modeling distributions and preparing them for use in the particular theoretical problem.

To illustrate, we consider three theoretical problems. The first, based on a theory of choice attributed to St. Anselm, investigates the effect of inequality on own income. The second, based on a theory of status integrating elements pioneered by Berger and his associates (Berger, Fisek, Norman, & Zelditch, 1977), Goode (1978), Sørensen (1979), and Ridgeway (2001), investigates the effect on status of rank on valued quantitative characteristics (Jasso, 2001b). The third, based on a theory of justice and comparison processes, investigates the effects of valuing cardinal versus ordinal goods and of the distributional shapes of cardinal goods and their inequality on the distribution of such outcomes as happiness and the sense of being justly or unjustly rewarded (Jasso, 2001a).

The second and third problems include attentiveness to ranks (always in status theory, sometimes in justice-comparison theory). Ranks, as is well known, are directly represented by the rectangular distribution (the discrete rectangular in problems involving small groups and the continuous rectangular in problems involving large societies).

The first and third problems include distributions of cardinal goods, and these require selection of modeling distributions from among distributions with a positive support. One approach is to consider behavior at the extremes, that is, whether there is a minimum income or a maximum income. This approach leads to four combinations: (a) a variate with nonzero infimum and a supremum, (b) a variate with nonzero infimum and no supremum, (c) a variate with zero infimum and a supremum, and (d) a variate with zero infimum and no supremum. To illustrate, the wealth distribution of a primitive society in good times may be modeled by the first type and in bad times (such as famine) by the third type; the wealth distributions of advanced societies near the end of a war may be modeled by the second type among the winners and by the fourth type among the losers. Table 1 provides example(s) of each of the four combinations of properties. As shown, the Pareto variate, as well as the exponential variate, exemplify the combination of nonzero infimum and no supremum. The lognormal exemplifies the combination of

Table 1 Prototypical Distributions of Material Goods

	Supremum	
Infimum > 0	*Yes*	*No*
Yes	Quadratic	Exponential Pareto
No	Power-Function	Lognormal

NOTE: In both the power-function and the quadratic variates, the supremum is also the maximum. The lognormal and the power function have an infimum at zero; the power-function of $c > 1$ also has a minimum of zero.

zero infimum and no supremum. The power-function variate has a supremum but no infimum. Finally, the quadratic, a symmetric variate, has both an infimum and a supremum.

In using distributions to model cardinal quantities in the first and third problems, it is useful to have at hand formulas for the variates' associated functions expressed in terms of two substantively meaningful parameters, the mean and inequality. Accordingly, the first step is to adapt the selected variates so that the versions have two parameters, and these two parameters are the arithmetic mean and a general inequality parameter of which all measures of relative inequality are monotonic functions. Table 2 provides the formulas for the cumulative distribution function, the probability density function, and the quantile function for the five variates highlighted in Table 1.

In the Anselmian problem, the protocol involves setting up the quantile function for each modeling distribution and finding the first partial derivative of own income with respect to inequality. Results include the prediction that the proportion of the population for whom own income increases as income inequality decreases varies greatly—for example, 37% when income is power-function distributed, 50% when income is lognormally distributed, and 63% when income is Pareto distributed. In Anselmian theory, these individuals include the poorest members of the society, and they cannot be characterized as being egoistic or altruistic, for the same choice that increases their own income also decreases income inequality.

In the status and justice-comparison problems, an early task is to find the distributions of status and of the justice-comparison outcomes, respectively. This is done by using the formulas for status and for justice-comparison (provided by the ASSUMPTIONS of the respective theories) and the formulas for the modeling distributions, and applying change-of-variable, convolution, and other techniques (described in standard sources). In the case of status, it is easy to find the distribution in the case of one-good status because the status formula, already expressed in terms of ranks, is seen by inspection to be that for the quantile function of the positive exponential (Sørensen, 1979). Other status distributions, in cases of two goods simultaneously conferring status, include the Erlang and an as-yet unnamed family (Jasso, 2001b, p. 122). Justice-comparison distributions that have been obtained include the negative exponential, the positive exponential, the normal, the logistic, and symmetric and asymmetric Laplace (Jasso, 2001a, pp. 683–685).

Further analysis of the status and justice-comparison distributions yields a large variety of testable predictions, including, for example, the predictions that the rate of out-migration, the public benefit of religious institutions, and conflict severity among warring subgroups all increase as inequality in the distribution of valued material goods increases.

Of course, an important challenge is to obtain distribution-independent results. For example, use of tools from the study of probability distributions yields the result that two singly sufficient but not necessary conditions for the justice-comparison distribution to be symmetric about zero are that (a) the actual reward and the just reward be identically and independently distributed, and (b) the actual reward and the just reward be identically distributed and perfectly negatively related (Jasso, 2001a, pp. 683–685).

DISTRIBUTIONS IN EMPIRICAL WORK

Social science researchers are constantly concerned about the distributions of their DATA, especially the distribution of the outcome or DEPENDENT VARIABLE. Although there exist various reasons for studying distributions of variables, two major purposes stand out—to gain knowledge through analyzing distributions and to approximate underlying distributions.

1. *Gaining further knowledge.* Often, the basic information of distributional parameters is not sufficient, and researchers resort to graphical means to assist their understanding and possibly further

Table 2 Modeling Distributions for Cardinal Goods and Associated Basic Functions

Variate Family	*Cumulative Distribution Function*	*Probability Density Function*	*Quantile Function*
Exponential $x > \frac{\mu}{c}$, $c > 1$	$1 - \exp\left[-\frac{cx-\mu}{\mu(c-1)}\right]$	$\frac{c}{\mu(c-1)}\exp\left[-\frac{cx-\mu}{\mu(c-1)}\right]$	$\mu\left[\frac{1}{c} + \frac{c-1}{c}\left(\ln\frac{1}{1-\alpha}\right)\right]$
Lognormal $x > 0$, $c > 0$	$F_N\left\{\frac{[\ln(\frac{x}{\mu}) + \frac{c^2}{2}]}{c}\right\}$	$\frac{1}{xc\sqrt{2\pi}}\exp\left\{-\frac{(\frac{c^2}{2} + \ln\frac{x}{\mu})^2}{2c^2}\right\}$	$\mu\exp\left[cQ_N(\alpha) - \frac{c^2}{2}\right]$
Pareto $x > \frac{\mu(c-1)}{c}$, $c > 1$	$1 - \left[\frac{\mu(c-1)}{cx}\right]^c$	$\left[\frac{\mu(c-1)}{c}\right]^c cx^{-c-1}$	$\frac{\mu(c-1)}{c(1-\alpha)^{1/c}}$
Power-Function $0 < x < \frac{\mu(c+1)}{c}$, $c > 0$	$\left[\frac{xc}{\mu(c+1)}\right]^c$	$\left[\frac{c}{\mu(c+1)}\right]^c cx^{c-1}$	$\frac{\mu(c+1)\alpha^{1/c}}{c}$
Quadratic $\frac{\mu(2+c)}{2} < x < \frac{\mu(2-c)}{2}$, $0 < c < 2$	$\frac{1}{2}\left\{1 - \cos\left[3\arccos\left(\frac{x-\mu}{\mu c}\right) - 4\pi\right]\right\}$	$\left(-\frac{6}{\mu^3c^3}\right)x^2 + \left(-\frac{12}{\mu^2c^3}\right)x + \frac{3(c^2-4)}{2\mu c^3}$	$\mu\left\{1 + c\sin\left[\frac{\arcsin(2\alpha-1)}{3}\right]\right\}$

NOTE: For all variates, $x > 0$; other restrictions as indicated. The terms $F_N(\cdot)$ and $Q_N(\cdot)$ denote the cumulative distribution function and the quantile function, respectively, of the unit normal variate. Inequality is a decreasing function of c in the Pareto and the power-function variates and an increasing function of c in the exponential, lognormal, and quadratic variates.

operations on the data. These can be a HISTOGRAM before applying a particular CODING scheme, a STEM-AND-LEAF DISPLAY or a BOXPLOT for catching OUTLIERS and potential INFLUENTIAL CASES, a Lorenz curve for illustrating distributional inequality, or a SCATTERPLOT for displaying the distribution of one variable against another. Contemporary statistics offers many useful tools for exploring and examining distributions, especially when STATISTICAL COMPARISON is the aim. For example, the RELATIVE DISTRIBUTION METHOD provides researchers with an invaluable way to compare two income distributions.

2. *Approximating underlying distributions.* It is common practice in the social sciences to apply some form of REGRESSION model to our outcome variables, ranging from a simple LINEAR REGRESSION to a nonlinear random coefficient model. By doing so, we assume we know the underlying mechanism that generates the dependent variable. However, too often, convention governs our modeling attempts, without much due attention to distributional properties. A little exploration would go a long way in better capturing the underlying distribution. Figure 1 gives the distributions of a sample of women from the 1991–2000 British Household Panel Survey, with a histogram representing the empirical distribution of the data and the lines representing three theoretical distributions. The variable measures depression and records the number of items on which a woman reported "feeling worse" than before.

Clearly, the normal distribution does not represent the data well. The Poisson distribution is better, but among the three candidates considered, the negative binomial distribution outshines the others by a close approximation of the data, hence of the underlying distribution.

FURTHER READING

The volumes in Johnson and Kotz's series, *Distributions in Statistics,* and their revisions represent an

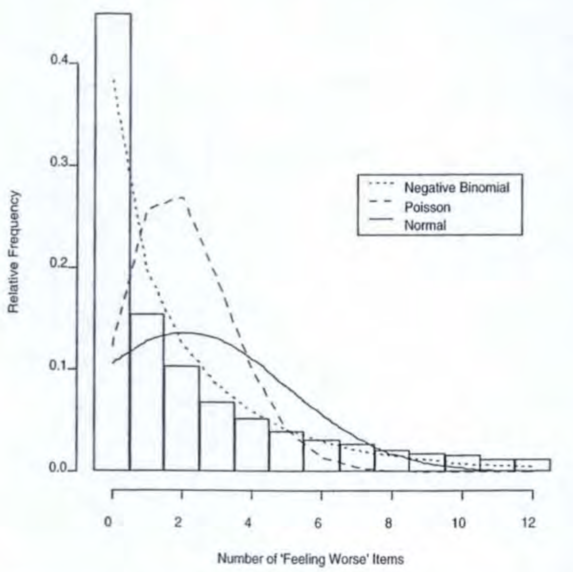

Figure 1 Distributions of GHQ12, 1991–2000 BHPS ($N = 7{,}855$)

invaluable resource for both theoretical and empirical work in the social sciences.

—Guillermina Jasso and Tim Futing Liao

REFERENCES

Berger, J., Fisek, H., Norman, R., & Zelditch, M. (1977). *Status characteristics and social interaction: An expectation states approach.* New York: Elsevier.

Blau, P. M. (1974). Parameters of social structure. *American Sociological Review, 39,* 615–635.

Eubank, R. L. (1988). Quantiles. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 7, pp. 424–432). New York: Wiley.

Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions* (3rd ed.). New York: Wiley.

Gibbons, J. D. (1988). Truncated data. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 9, p. 355). New York: Wiley.

Goode, W. J. (1978). *The celebration of heroes: Prestige as a control system.* Berkeley: University of California Press.

Jasso, G. (2001a). Comparison theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 669–698). New York: Kluwer Academic/Plenum.

Jasso, G. (2001b). Studying status: An integrated framework. *American Sociological Review, 66,* 96–124.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions, Vol. 1.* New York: Wiley.

Kotz, S., Johnson, N. L., & Read, C. B. (1982). Censoring. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 1, p. 396). New York: Wiley.

Ridgeway, C. L. (2001). Inequality, status, and the construction of status beliefs. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 323–340). New York: Kluwer Academic/Plenum.

Sørensen, A. B. (1979). A model and a metric for the analysis of the intragenerational status attainment process. *American Journal of Sociology, 85,* 361–384.

DISTRIBUTION-FREE STATISTICS

In the strictest sense of the phrase, a distribution-free statistic is one whose SAMPLING DISTRIBUTION does not depend on the POPULATION from which a random sample was drawn. That is, the statistic has the same sampling distribution irrespective of the form of the parent population, whether it be the normal, exponential, or any other DISTRIBUTION.

The simplest example of a distribution-free statistic is the SIGN TEST statistic, defined as the number of plus signs among, say, n differences within a set of n paired sample observations that have no ties. The number of plus signs among these differences follows the BINOMIAL DISTRIBUTION with PARAMETER p representing the PROBABILITY of a plus sign and n the sample size, regardless of the form of the distribution from which the random sample of pairs was drawn. Another distribution-free statistic, or at least approximately so, is the standardized mean (or Z-SCORE) of a random sample of at least 30 observations drawn from any continuous population with known VARIANCE. According to the CENTRAL LIMIT THEOREM, the sampling distribution of such a standardized sample mean is approximately the normal distribution.

In practice, any distribution-free statistic permits a researcher to carry out a distribution-free test, that is, one that does not require any assumptions about the distribution from which the sample was drawn in order to be regarded as a valid test. A valid test is defined as a test for which the exact probability of a TYPE I ERROR is the stated α value, or one for which the P VALUE is exact. The term *distribution-free test* has now been essentially superseded by the term nonparametric test, which includes not only distribution-free tests but also any statistical test whose NULL

HYPOTHESIS does not include the value of a parameter. This latter type of test includes tests of goodness-of-fit, tests for trend, tests for AUTOCORRELATION, and tests for randomness, among others. Nonparametric statistics include a very large number of methods of statistical analysis that are especially useful for small sample sizes.

—Jean D. Gibbons

DISTURBANCE TERM

Disturbance term refers to the ERROR term in a statistical model. If the disturbance is truly STOCHASTIC, then it is WHITE NOISE.

—Tim Futing Liao

DOCUMENTS OF LIFE

Documents of life are all those documents in which people reveal their social and personal characteristics in ways that are accessible for research. The documents could include diaries; letters; photographs; life stories; inscriptions on tombstones; furnishings; videos; and, these days, personal Web sites. In an early description of the method, humanistic anthropologist Robert Redfield described them as "personal documents," seeing them as one in which the human and personal characteristics of somebody who is, in some sense, the author of the document find expression, so that through its means the reader of the document comes to know the author and his or her views of events with which the document is concerned (Redfield, in Gottschalk, Kluckhohn, & Angell, 1945, p. vii). They are not the second-order constructs of reality made by social scientists, but first-order accounts that attempt to enter the subjective world of informants, taking them seriously on their own terms.

—Ken Plummer

See also AUTOBIOGRAPHY, AUTOETHNOGRAPHY, DIARIES, LIFE COURSE RESEARCH

REFERENCES

Allport, G. (1942). *The use of personal documents in psychological science.* New York: Social Science Research Council.

Gottschalk, L., Kluckhohn, C., & Angell, R. (1945). *The use of personal documents in history, anthropology, and sociology.* New York: Social Science Research Council.

Hoskins, J. (1998). *Biographical objects: How things tell the stories of people's lives.* London: Routledge.

Plummer, K. (2001). *Documents of life-2: An invitation to a critical humanism.* London: Sage.

DOCUMENTS, TYPES OF

A document is any kind of physically embodied text, and a handwritten or printed text on paper, such as a letter or government report, is the archetypal document. However, the range of documents is much wider than this. We may classify documents as *written* (where the text is visible words), *audio* (where the text is sound that can be accessed through aural channels), and *visual* (where the text is a designed or pictorial representation). A text may be produced with a pen, a pencil, a paint brush, a printing machine, a tape recorder, or a computer, and it may be inscribed on clay, stone, parchment, paper, magnetic disks, or electronic display screens. The full range of documents includes newspapers, diaries, stamps, directories, handbills, maps, photographs, paintings, gramophone records, tapes, and computer files.

Documents are sometimes seen as the particular source material for the historian, but they have a wide range of uses across the social sciences. The classical sociologists made far greater use of documentary sources than they did of survey or observational methods. Karl Marx based *Capital* on his heavy use of the publications of the factory inspectors, Max Weber used religious tracts and pamphlets in *The Protestant Ethic and the Spirit of Capitalism,* and Emile Durkheim drew on a variety of official statistics to produce *Suicide.*

Documents can be classified most usefully in terms of their authorship and the conditions for access to them.

Authorship can be subdivided according to whether documents originate in the personal sphere or the public, official sphere, and official documents can be subdivided according to their origin in state or private bureaucracies. This differentiation rarely can be drawn with any precision, particularly in the pre-modern period. In medieval Europe, for example, a distinction between personal family papers, church papers, and state papers would not make a great deal of sense because of the highly personalized nature of patrimonial authority.

Access concerns the conditions under which documents may be available to people other than their authors. Documents may be "closed" if they are available only to a very limited group of people, such as those who produce them and their bureaucratic superiors. If documents are "restricted," they are accessible to a wider group, but under limited conditions determined by their producers or their guardians. The most liberal conditions are "open" access through archiving or publication. Archival access exists where documents have been stored and made available to a wide public: Documents may be available with only minimal bureaucratic restrictions, such as the need to register or to acquire a reader's ticket (see DATA ARCHIVES, ARCHIVING QUALITATIVE DATA). Where documents are published, access is almost completely open because they are accessible to all who are able to afford them.

These two dimensions can be combined as shown in Table 1 to produce 12 different kinds of documents. Closed personal documents (Type 1) include letters, diaries, household account books, and many other domestic items. These kinds of documents are normally available only to the individual who owns them or to the immediate household that produced them. Personal documents may move from one category of access to another. Some may become available more widely if they are deposited and stored in public records offices (Type 3), and some may be published (Type 4). DIARIES, for example, begin as purely closed documents, but are often produced with the intention of making them available to a wider readership. The records of many landed and wealthy families have been deposited in public archives and so become more easily accessible. Many such documents, however, remain in private hands—if they survive for any period at all—and are closed to public access unless specific permission is granted by the owner (Type 2).

Official documents in the private sphere are those produced by organizations such as businesses, schools, hospitals, and churches. Confidential organizational documents (Type 5) include medical records, records of educational attainment, and company personnel records, all of which are normally available only to those with an administrative or professional responsibility for the people to whom they relate. Restricted organizational documents (Type 6), such as share registers or lists of mortgages, are normally made available to researchers, if at all, when they are no

Table 1 A Typology of Documents

		Authorship		
			Official	
		Personal	*Private*	*State*
Access	Closed	1	5	9
	Restricted	2	6	10
	Open-archival	3	7	11
	Open-published	4	8	12

longer of practical business relevance. Specific restrictions may be applied to the consultation and use of the documents. For some, however, there is a legal requirement to deposit them in a public archive (Type 7): Share registers in England and Wales usually must be sent each year to the companies' registration office. Organizational documents that are published (Type 8) may also be published as a legal or other official requirement, as is the case with the published reports and accounts of companies, but many are published as commercial ventures. Examples include timetables and directories, reference books, newspapers, films (see FILM AND VIDEO IN RESEARCH), and other mass media products.

Government documents, whether local or national, involve the same categories of access and are, perhaps, the single largest type of document available to social researchers. Those with closed access (Type 9) include criminal records and security reports, local authority housing records, and current taxation records. Many are covered by official secrecy legislation that prevents unauthorized disclosure outside the immediate department responsible. Under such legislation, many documents remain closed on a permanent basis, although some papers may be opened up for restricted access (Type 10), and others may become fully available in public archives (Type 11). Those state documents that enter public archives often do so only after a specified period, when they are no longer regarded as confidential. Many state documents are produced for publication (Type 12), and this includes acts of Parliament, reports of royal commissions, statistical reports (see OFFICIAL STATISTICS), and research reports.

The general principles involved in using documents are similar to those in any other area of social research. Nevertheless, the particular features of documentary sources do imply specific judgments and specific techniques. The effective use of documents depends on them being appraised in terms of four criteria: authenticity, credibility, representativeness, and meaning.

The criterion of *authenticity* requires that documents be assessed for their soundness and authorship. An assessment of soundness means discovering whether the document is an original or a copy and, if a copy, whether it is a copy of an original or a copy of a copy. Copying, whether by handwriting, by re-typesetting, or by photocopying, can result in missing or unreadable text, but even originals may contain typographical errors or be incomplete. An "unsound" document is one that is not close enough to an idealized original because it is corrupted in some way. A researcher may need to reconstruct what a sound document would look like before making use of the unsound one available. The more corrupted a document is, the more difficult such a task will be, and, in extreme cases, it may not be possible to reconstruct a sound version of the original document.

Once soundness has been assessed, authorship can be considered. It is important to know, for example, whether diaries attributed to particular individuals were actually written by the individuals concerned. Issues of forgery or fraudulent authorship can arise, even in apparently straightforward cases. Fraudulent documents may be very important sources of evidence, but the fact that they are fraudulent must be known. Authorship can be assessed through the use of both internal and external evidence. Internal evidence of vocabulary and literary style can be combined with external evidence from chemical tests on paper and ink, and from estimations of plausibility, to arrive at a judgment about the attribution of authorship. In many cases, however, the question of authorship is far from straightforward. Official documents are produced by an administrative machine in which any named person may play a minimal role. Similarly, books and newspapers are the result of a complex division of labor in which the work of named writers is processed by copy editors, subeditors, and editors.

The question of *credibility* concerns the sincerity and accuracy of a document. All documents are selective, because it is impossible to construct accounts that are independent of particular standpoints. Nevertheless, they can be more or less credible as accounts, depending on how sincerely the choice of a point of view is made and whether the account gives an accurate report from that standpoint. The sincerity of authors reflects their motivation and the degree of choice that they exercise. People write for self-justification, to propagandize or to deceive, for financial gain, or simply to report with as much objectivity as possible, and in each case, it is important to assess the extent to which the document is shaped by these reasons. Official documents, for example, may be presented in the form of "information," but may actually involve an intent to persuade people of a particular position. Similarly, newspapers are produced by journalists and others who are paid to write marketable material and who may be subject to political pressure from a proprietor.

Even a document that has been sincerely produced may be inaccurate. To assess the accuracy of a report, it is necessary to look at the conditions under which it was compiled and, in particular, how close the author was to the events reported. Accuracy is generally held to be greatest in primary sources—first-hand accounts—because these are felt to minimize any loss of accuracy due to lapses of memory, inadequate records, or ignorance. But even first-hand observers face the problem of recording their observations in a form that they can use at a later time to construct accurate reports. It is generally very difficult, for example, to record conversations verbatim. Any system of recording, especially a mechanical one, is likely to be intrusive and so may be avoided. Observers must often rely on memory, and thus, accuracy of recall is a problem even in primary sources.

Assessing *representativeness* depends upon a knowledge of documents' survival and their availability. It is important to know whether the documents consulted are representative of all the relevant documents, and this depends upon what proportion of the relevant documents have actually survived and whether they are available for researchers to use. If they are to survive, documents must be stored in some way. This storage can range from deposition in a public archive to dumping in a cardboard box. Not all documents get to be stored. Many public and private documents are destroyed soon after their production, whereas others are stored for a period only to be destroyed at a later date. The increasing use of computers to produce and store documents means that, in many cases, no historical record is produced: Documents are edited and updated in real time, so only the latest revision is ever available. Where computer files are archived in backup form and survive for later research use, it is difficult to know how their contents relate to the original continuous database. Many personal documents are not retained, and those that survive are likely to do so because their owners have judged them to be out of the ordinary in some way. Only a small proportion of the vast number of documents produced in modern bureaucratic organizations can be retained. Official documents are stored while in current use and may then be weeded out before being transferred to an archive. The number of documents that survive may decrease over time through deterioration and decay or through periodic "clear-outs." Such losses may sometimes occur at random, but they may also involve unknown systematic processes.

The availability of official documents that have survived is limited by considerations of confidentiality and official secrecy. State documents often enter the public sphere after a particular period of time has lapsed, but some documents may be permanently restricted. Problems may be even greater in the personal and private spheres, and especially with family and household documents, where there are generally no legal requirements about access at all. Personal documents also tend to lack the kind of index or calendar that allows the representativeness of official documents to be assessed. Unless they are stored in family archives—which is unusual except for very wealthy families—private household materials are neither archived nor catalogued.

Although documents must be assessed for their authenticity, credibility, and representativeness, it is not necessary that they actually be fully authentic, credible, or representative. What is important is that the researcher knows that he or she is dealing with documents that are inauthentic, incredible, or unrepresentative to some degree. It is then possible to move on to the question of their meaning.

The *meaning* of documents is an issue that arises at two levels: the literal and the interpretive. A literal understanding of a document involves its physical readability, whether it is in a language that can be read, and such issues as dating. Reading faded or decayed documents may require technologies such as microscopes and infrared light, and the reading of computer files will always require specific hardware and software. The reading of audio and visual materials will routinely require access to particular technologies. Issues of literal meaning are, in principle, resolvable on a practical basis, leaving the researcher with a meaningful text to analyze.

It is at this point that the more fundamental question of interpretive meaning arises, and this is a matter that raises complex theoretical issues. Interpretation is a hermeneutic task that involves an appreciation of the social and cultural context and forms of discourse that structure a text. It is necessary to grasp the underlying point of view, from which the individual concepts in a text acquire their meaning. The assessment of authenticity, credibility, and representativeness discloses the context from which the document emerged and within which it must be interpreted.

Methods of textual analysis cannot be considered in any detail in this entry, but the two principal approaches are quantitative CONTENT ANALYSIS and qualitative semiotic approaches (see QUALITATIVE CONTENT ANALYSIS). Whereas the former involves the use of methods of counting and statistical analysis, the latter involves methods of decoding and deconstruction. Ultimately, however, all interpretation is a qualitative task, and even a quantitative content analysis rests upon some assessment of significance through an often-implicit semiotic process.

The four criteria of authenticity, credibility, representativeness, and meaning should not be seen as separate and distinct stages in the assessment of documents. They are interdependent, and a researcher cannot adequately consider one without simultaneously considering the others. In this sense, the appraisal of documents is a never-ending process. The interpretive meaning that a researcher aims to produce is a tentative and provisional judgment that must be constantly revised as new discoveries and new problems force a reappraisal of the evidence. An understanding of this is essential if a researcher is to make sensible use of documents.

—John Scott

REFERENCE

Scott, J. (1990). *A matter of record: Documentary sources in social research.* Cambridge, UK: Polity.

DOUBLE-BLIND PROCEDURE

Blind experimental procedures are designed to eliminate biases produced by expectations of study participants, experimenters, and data analysts. Investigators commonly employ any of three levels of blindness in their experimental designs. In a simple blind EXPERIMENT, participants are unaware of their assignment to the treatment or to the control group. In a double-blind experiment, both the participant and the experimenter are unaware of the group assignment. Finally, in a triple-blind experiment, the participants, experimenter, and those responsible for data entry and analysis are all unaware of the treatment versus control condition of each study trial.

Blind procedures are necessary because participants respond not only to the experimental stimulus but also to their impression of the experiment's HYPOTHESIS. If the manipulation of the experiment is readily apparent, participants may alter their response in accordance with what they perceive to be the desired response. In this case, the measured effect of the experiment represents an unknown combination of intended treatment effects and the demand characteristics of the experimental setting.

A more subtle source of BIAS lurks in the interaction between the experimenter and the participant. When an experimenter expects a certain response from participants in one condition, he or she may artificially generate or enhance this response by influencing these participants with (often unintentional) nonverbal cues, extra encouragement, or heightened attention. The bias that results from these behaviors is called the experimenter expectancy effect. Because participant and experimenter biases are systematically related to expectations of the experimental manipulation, they have the potential to confound the interpretation of experimental results. The same type of expectancy effects may lead the people handling the data to consciously or unconsciously skew the results in an expected direction.

The double-blind procedure, which is also referred to as a "masked condition" or an "experimentally naïve" procedure, is often used in medical research. For example, in a drug trial, a sugar pill identical in appearance to the experimental pill is administered to the control group. The experimental and sugar pill treatments are randomly assigned by an investigator who does not interact with participants. The identity of the pills keeps both the participants and the experimenters who actually administer the drugs blind to the nature of each experimental trial. Double blind procedure has the added benefit of preventing those administering the treatment from succumbing to the temptation to give needier patients the experimental drug rather than the placebo, thereby distorting the results.

Many social science investigations involve treatments that, unlike a pill, cannot be engineered to look the same to treatment and control subjects. Consider a study comparing the effects of telephone versus e-mail messages on attitude change; here, both participants and experimenters are fully aware of whether the study involves a phone call or an e-mail. In such instances, double blindness is approximated by keeping both participants and experimenters unaware of the hypotheses, or by minimizing contact between participants and the experimenter who knows the hypothesis. For example, a third party can administer the treatment before the experimenter collects data on the DEPENDENT VARIABLE. In sum, a "blind" PROTOCOL helps the investigator to minimize systematic human bias that is introduced when participants, experimenters, and data analysts are aware of the experimental treatment. Double-blind procedures afford the investigator more confidence when drawing causal relationships between the treatment and the outcome variable.

—Donald P. Green and Elizabeth Levy Paluck

REFERENCES

Aronson, E., Carlsmith, M., & Ellsworth, P. C. (1990). *Methods of research in social psychology.* New York: McGraw-Hill.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

DUAL SCALING

This is a method for multidimensional quantification of categorical data in a two-way table (e.g., a CONTINGENCY TABLE, an examinees-by-questions table of multiple-choice responses). The word "dual" reflects its symmetric scaling of rows and columns of a table. Dual scaling determines weights for options of multiple-choice items in such a way that scores of the subjects, given by weighted sums of weights of chosen options, attain greatest discriminability, hence, maximal internal consistency; in turn, those option weights are expressed as weighted sums of scores of the examinees who chose the options. This property is often referred to as Louis Guttman's principle of internal consistency. Frederic M. Lord verified that such a weighting scheme as this maximizes the internal consistency reliability of the derived scores.

HISTORICAL DEVELOPMENT

Early work on multidimensional quantification of categorical data can be traced back to Marion Richardson and G. Frederick Kuder in 1933, using what Paul Horst in 1935 called the method of reciprocal averages. H. O. Hirschfeld in 1935 presented a mathematically equivalent method, simultaneous linear regressions, as termed by James C. Lingoes in 1964. Ronald A. Fisher in 1940 proposed DISCRIMINANT ANALYSIS of categorical data, and K. Maung in 1941 showed that Fisher's quantification method could be equivalently formulated by four distinct objective functions. Whereas Maung's paper was for the contingency table, Guttman in 1941 presented three mathematically equivalent approaches to the quantification of multiple-choice data, also extending it to paired comparison and rank order data in 1946. The foundation for quantification was firmly laid by 1946. In 1950, Chikio Hayashi launched a project on Hayashi's quantification theory in Japan, Type III of which corresponds to dual scaling of the contingency table. In the early 1960s, Jean-Paul Benzécri started research in France, which was developed into CORRESPONDENCE ANALYSIS for the contingency table and multiple correspondence analysis for multiple-choice data. Both Hayashi and Benzécri had many followers, as referred to by the Hayashi School and the Benzécri School of data analysis. Similar consorted efforts led to the Leiden Group (headed by Jan de Leeuw) in the Netherlands and the Toronto Group (headed by Shizuhiko Nishisato) in Canada in the late 1960s.

A plethora of names have been proposed for the method, such as Guttman scaling (Guttman); the method of reciprocal averages (Richardson, Kuder, Horst); simultaneous linear regressions (Hirschfeld); Fisher's appropriate scoring (Fisher); Hayashi's theory of quantification Type III (Hayashi); optimal scaling (R. Darrell Bock); correspondence/multiple-correspondence analysis (Benzécri, Brigitte Escofier); biplot (Ruben Gabriel, John Gower, David Hand); homogeneity analysis (de Leeuw, Willem Heiser, Jacqueline Meulman); and dual scaling (Nishisato). The name *dual scaling* was proposed in 1976 in response to the criticism that the popular name *optimal scaling* was not specific enough to describe the method. According to Meulman in 1998, dual scaling is a "comprehensive framework for multidimensional analysis of categorical data," (p. 291) because it handles a wider range of categorical data than such methods as correspondence analysis and optimal scaling, of which applications are mainly to incidence data.

BASIC FORMULAS

The method employs singular value decomposition for categorical data. Consider a two-way contingency table with the typical element being the frequency f_{ij} in row i and column j. The singular value decomposition of this element can be expressed as

$$f_{ij} = \frac{f_{i.}f_{.j}}{f_{..}}[1 + \rho_1 y_{i1}x_{j1} + \rho_2 y_{i2}x_{j2} + \cdots + \rho_k y_{ik}x_{jk} + \rho_K y_{iK}x_{jK}], \quad (1)$$

where $f_{i.}$ is the marginal frequency of row i, $f_{.j}$ is that of column j, ρ_k is the kth singular value ($\rho_1 \geq \rho_2 \geq \cdots \geq \rho_k$), y_{ik} is the optimal weight of row i of the kth solution, and x_{jk} is that of column j. The solution, corresponding to 1 inside the bracket, is called a trivial solution, typically deleted from analysis, and is the frequency expected when the rows and the columns are statistically independent. The others are called proper solutions. For any proper solution, say, solution k, there exist dual relations (Nishisato, 1980) or transition formulas (Benzécri et al. in 1973): For an $m \times n$ data matrix,

$$y_{ik} = \frac{1}{\rho_k}\frac{\sum_{j=1}^{n} f_{ij}x_{jk}}{f_{i.}} \quad \text{and} \quad x_{jk} = \frac{1}{\rho_k}\frac{\sum_{i=1}^{m} f_{ij}y_{ik}}{f_{.j}}. \quad (2)$$

The equations are nothing but an expression of simultaneous linear regressions of rows on columns and columns on rows. It also indicates the basis for the method of reciprocal averages: Start with arbitrary row weights to calculate weighted column means, which are, in turn, used as weights to calculate weighted row means, and continue the process. This reciprocal averaging scheme always converges (Nishisato, 1980) to the optimal sets of row weights and column weights, the proportionality constant being the singular value. Nishisato (1980) calls weights y and x *normed weights* and ρy and ρx *projected weights,* also referred to as *standard coordinates* and *principal coordinates,* respectively (see Greenacre, 1984). Geometrically, y and ρx span the same space, and so do ρy and x. The angle of the discrepancy between row space and column space is given by the cosine of the singular value (Nishisato & Clavel, 2003).

Table 1

Party	Plan	Ranking (1)	(2)	(3)	(4)	(5)	(6)	Dominance Matrix (1)	(2)	(3)	(4)	(5)	(6)
Subject	1	6	1	5	4	3	2	−5	5	−3	−1	1	3
	2	6	1	5	2	4	3	−5	5	−3	3	−1	1
	3	3	5	2	4	1	6	1	−3	3	−1	5	−5
	4	3	4	2	6	1	5	1	−1	3	−5	5	−3
	5	5	3	1	4	6	2	−3	1	5	−1	−5	3
	6	2	6	3	5	4	1	3	−5	1	−3	−1	5
	7	1	2	4	5	3	6	5	3	−1	−3	1	−5
	8	4	3	2	6	5	1	−1	1	3	−5	−3	5
	9	2	1	4	5	3	6	3	5	−1	−3	1	−5
	10	6	1	4	3	5	2	−5	5	−1	1	−3	3

Contingency tables and multiple-choice data, expressed as response patterns, are examples of *incidence data,* which are decomposed into the expression of equation (1). In contrast, *dominance data* include rank-order data and paired comparison data, for which the observations are first coded as follows:

$$ {}_{i}f_{jk} = \begin{cases} 1 & \text{if Subject } i \text{ prefers Object } j \text{ to } k, \\ 0 & \text{if Subject } i \text{ makes a tied judgment,} \\ -1 & \text{if Subject } i \text{ prefers Object } k \text{ to } j. \end{cases} \tag{3} $$

These are transformed to dominance numbers by

$$ e_{ij} = \sum_{j=1}^{n} \sum_{k=1\ (j\neq k)}^{n} {}_{i}f_{jk}. \tag{4} $$

The matrix of dominance numbers has the property that row elements sum to zero. Thus, to arrive at weights for subjects and weights for objects, the method of reciprocal averages, for example, is carried out with the understanding that each cell of the dominance matrix is based on $n-1$ responses.

INCIDENCE DATA

In 1993, Nishisato classified categorical data into two types, incidence data and dominance data. Incidence data consist of 1, 0, or frequencies; a trivial solution exists; the chi-square metric is used; and even when all variables are perfectly correlated, more than one dimension is generally needed for a complete description of data. Incidence data are the data type for which most quantification methods mentioned earlier are developed. See Gifi (1990), Greenacre (1984), and Nishisato (1980, 1994) for analytical examples of such incidence data as contingency tables, multiple-choice data, and sorting data.

DOMINANCE DATA

Dominance data consist of responses *greater, equal,* or *smaller*; no trivial solution is involved; the Euclidean metric is used; and when variables are perfectly correlated, a single dimension is sufficient to describe data. Because dual scaling also handles dominance data, let us look at one example. Ten subjects ranked the following six Christmas party plans according to the order of their preference: (1) potluck in the group during the daytime, (2) pub-restaurant crawl after work, (3) reasonably priced lunch in an area restaurant, (4) evening banquet at a hotel, (5) potluck at someone's home after work, (6) ritzy lunch at a good restaurant. The data from *Psychometrika* (Nishisato, 1996) and the corresponding dominance matrix are in Table 1. Relative variances of five proper solutions are 0.45, 0.27, 0.14, 0.08, and 0.06, respectively. Look at the first two proper solutions, and a joint plot of the *normed* weights y of subjects and the *projected* weights ρx of the party plans, the practice that leads to a solution to Clyde H. Coombs's problem of multidimensional unfolding (Nishisato, 1994) (see Figure 1). From this graph, calculate the distances between subjects and party plans, and rank the distances within each subject, to obtain the rank-2 approximation to the input ranks (see Table 2). Look at Subject 2: The ranking of the distances reproduces exactly one's ranking of the plans.

Table 2

	Squared Distances						Rank-2 Approximation					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
S1	1.94	0.73	1.72	1.06	1.67	1.37	6	1	5	2	4	3
S2	2.20	0.98	2.02	1.36	1.91	1.71	6	1	5	2	4	3
S3	0.67	1.75	1.21	1.53	0.81	1.98	1	5	3	4	2	6
S4	0.49	1.63	1.01	1.38	0.69	1.79	1	5	3	4	2	6
S5	1.72	1.58	1.18	1.27	1.77	0.54	5	4	2	3	6	1
S6	1.70	2.38	1.37	1.95	1.96	1.51	3	6	1	4	5	2
S7	1.10	1.42	1.55	1.45	0.89	2.14	2	3	5	4	1	6
S8	1.61	1.70	1.07	1.33	1.71	0.76	4	5	2	3	6	1
S9	1.40	1.28	1.74	1.45	1.11	2.18	3	2	5	4	1	6
S10	2.13	1.02	1.79	1.21	1.91	1.25	6	1	4	2	5	3

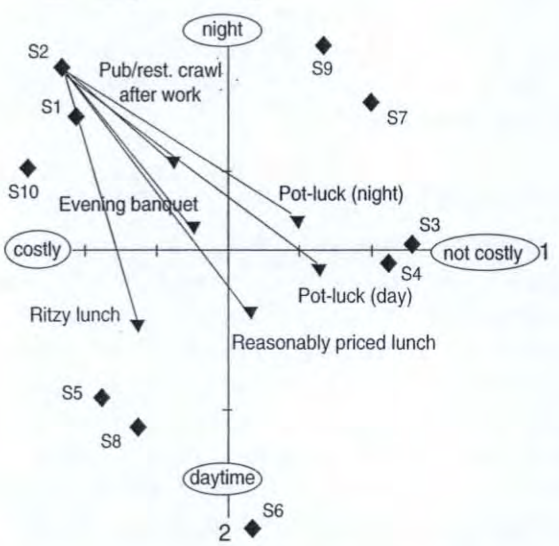

Figure 1 Joint Plot of Respondents and Party Plans

Other cases require more dimensions than two. Overall, the rank-2 approximation is close to the rank-5 data. The graph reveals individual differences in ranking as affected by the party cost and the party times. Observe their closeness to particular plans.

—Shizuhiko Nishisato

REFERENCES

Benzécri, J. P., et al. (1973). *L'analyse des données: II. L'analyse des correspondances.* Paris: Dunod.

Gifi, A. (1990). *Nonlinear multivariate analysis.* New York: Wiley.

Greenacre, M. J. (1984). *Theory and method of correspondence analysis.* London: Academic Press.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis.* New York: Wiley.

Meulman, J. J. (1998). Book review of W. J. Krzanowski, & F. H. C. Marriott, Multivariate analysis. Part I. Distributions, ordinations, and inference, *Journal of Classification, 15,* 287–293.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Nishisato, S. (1994). *Elements of dual scaling: An introduction to practical data analysis.* Hillsdale, NJ: Lawrence Erlbaum.

Nishisato, S. (1996). Gleaning in the field of dual scaling. *Psychometrika, 61,* 559–599.

Nishisato, S., & Clavel, J.G. (2003). A note on between-set distances in dual scaling and correspondence analysis. *Behaviormetrika, 30,* 87–98.

DUMMY VARIABLE

A set of dummy variables allows us to specify categorical explanatory variables as independent variables in statistical analyses. Some of the information we wish to incorporate into statistical models is organized as classification schemes that represent a set of nominal categories. Marital status, gender, race/ethnicity, political party, or industrial sector can all be represented as a set of exhaustive and mutually exclusive categories, with each observation being assigned to a single category. In the simplest case, we have two categories (e.g., men and women), which we can represent with a single dummy variable coded "1" if a man, "0" if not a man. This type of CODING is called binary coding and indicates the presence of an attribute or membership in a particular category. For classifications with more than two categories, we can fully represent the categorical information and produce coefficients that provide interpretable information if we strategically code $j - 1$ variables (Hardy, 1993; Hardy & Reynolds, 2004).

When specified on the right-hand side of an equation (as independent variables), researchers may want to know how differences in marital status, for example, are related to variation in a dependent variable, such as depressive symptoms. At issue is how to represent the information contained in a j-category classification scheme without imposing unrealistic distributional assumptions. If, for example, marital status was reported for a sample of adult men and women in the United States, we could specify categories such as (1) currently married, (2) currently divorced, (3) currently separated, (4) currently widowed, and (5) never married. If we simply included the variable "marital status" in a statistical model, without modifying in any way the coding of 1 through 5 assigned to the different categories, we would be assuming interval-level measurement of marital status, that is, that the numerical difference of one unit between "currently married" and "currently divorced" is equivalent to the unit difference between "currently widowed" and "never married," but only half the difference between "currently widowed" and "currently divorced," and that these categories were appropriately ordered, with "currently married" being scored at the lowest value in the range and "never married" at the highest value. This approach is clearly nonsensical; the numbers assigned to categories as well as the order of the categories are arbitrary. We require an approach consistent with nominal distributional properties, which, in this case, suggests that we need a numerical way to isolate the distinctive categories so that we can rely on counts (i.e., frequencies) and sample partitioning to provide appropriate estimates.

Although coding possibilities are many, three approaches are most frequently used: BINARY CODING, CONTRAST CODING, and EFFECTS CODING. To return to the example of marital status, we require four dummy variables to fully represent the information in the five-category classification. Using binary coding and choosing "currently married" as our reference group (the group to which remaining categories will be compared), we can define four dummy variables named "divorced," "separated," "widowed," and "never." Respondents whose marital status matches the variable name are coded "1" on that dummy variable. For example, currently divorced respondents are coded "1" on "divorced" and coded "0" on the remaining three variables; respondents who were never married are coded "1" on "never," "0" otherwise, and so on. That four dummy variables fully capture the information of a five-category nominal variable is demonstrated by the fact that any respondent coded "0" on all four variables must be married.

Regardless of which coding scheme is used, the appropriate *inference* test to determine whether the classification variable—in this example, marital status—improves the fit of the model is the F-test comparing model fit with and without the set of dummy variables. Dummy variables expand the flexibility of the GENERAL LINEAR MODEL, allowing us to test a broader range of hypotheses, including direct and indirect effects, INTERACTIONS, and different functional forms.

Binary-coded dummy variables also can be specified as dependent variables in statistical models. A set of estimation techniques for models with limited dependent variables includes binary probit and binary LOGIT models (Long & Cheng, 2004). Discrete time EVENT HISTORY models also use binary coded dependent variables to indicate whether or not an event has occurred (Allison, 2004). Binary-coded dependent variables are also used in various selection models, such as the Heckman two-stage procedure and endogenous switching models (Fu, Mare, & Winship, 2004).

—Melissa A. Hardy

REFERENCES

Allison, P. D. (2004). Event history analysis. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis*. London: Sage.

Fu, V., Mare, R., & Winship, C. (2004). Sample selection bias models. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis*. London: Sage.

Hardy, M. A. (1993). *Regression with dummy variables*. Newbury Park, CA: Sage.

Hardy, M. A., & Reynolds, J. (2004). Incorporating categorical information into regression models: the utility of dummy variables. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis*. London: Sage.

Long, J. S., & Cheng, S. (2004). Regression models for categorical outcomes. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis*. London: Sage.

DURATION ANALYSIS. *See* EVENT HISTORY ANALYSIS

DURBIN-WATSON STATISTIC

The Durbin-Watson (1950) statistic (*DW* or *d*) is a commonly used and routinely reported diagnostic test for the presence of first-order auto or SERIAL CORRELATION in the error of a TIME-SERIES REGRESSION model. The statistic is calculated using the RESIDUALS from the regression we care about, $\hat{\mu}_t$ as follows:

$$DW = \frac{\sum_2^n (\hat{\mu}_t - \hat{\mu}_{t-1})^2}{\sum_1^n \hat{\mu}_t^2}.$$

The NULL HYPOTHESIS for the test is that there is no first-order autocorrelation in the regression ERRORS. The test has two alternatives: positive and negative first-order autocorrelation.

It is illustrative to express the test statistic directly in terms of the first-order autocorrelation ρ:

$$DW \approx 2(1 - \rho).$$

Written in this form, it is easy to see the relationship between the value of the test statistic and the degree of autocorrelation in the model residuals. Specifically, if there is no autocorrelation ($\rho = 0$), *DW* is approximately 2; if there is perfect positive autocorrelation ($\rho = 1$), *DW* is approximately 4; and if there is perfect negative autocorrelation ($\rho = -1$), *DW* is approximately 0.

The CRITICAL VALUES for the test vary as a function of SAMPLE size and the DEGREES OF FREEDOM associated with the regression of interest. Typically, the *DW* test is computed for the alternative hypothesis of positive first-order autocorrelation so that we are looking for a value of *DW* that is significantly smaller than 2. Given difficulties in deriving the distribution of the *DW* statistic under the null hypothesis, two critical values are needed for each alternative hypothesis.

The test is appropriate only if there are no LAGGED dependent variables on the right-hand side of the regression we care about. If the regression contains a lagged dependent variable, alternative tests include Durbin's alternative, Durbin's H, or Lagrange multiplier tests. Similarly, *DW* will not detect higher orders of autocorrelation. To test for second and higher orders of autocorrelation in the errors, one can use Lagrange multiplier tests.

Finding a significant *DW* statistic indicates autocorrelation, which will invalidate HYPOTHESIS tests on the regression PARAMETERS we care about. A significant *DW* statistic is generally evidence of MODEL misspecification, but may indicate that the errors are themselves serially correlated (Wooldridge, 2000). The appropriate remedial response varies based on the cause of the problem. The model may require that the right-hand side of the equation include a lagged dependent variable or other dynamic REGRESSORS. Alternatively, we may need to estimate the regression with feasible GENERALIZED LEAST SQUARES (GLS) by transforming the variables, reestimating the original regression to remove the autocorrelation, and estimating the resulting regression.

To illustrate, suppose we wish to model the voter turnout rate of women in the United States, (v_t), from 1921 to the present, as a function of the mean level of education of women (e_t) and percent of women employed outside the home (w_t). To test for first-order serial correlation, we would first save the residuals from the following regression model:

$$v_t = \beta_0 + \beta_1 e_t + \beta_2 w_t + \mu_t.$$

The residuals would then be used to compute the *DW* statistic as above. For a regression with a sample of size $t = 45$ and 3 degrees of freedom (two regressors plus a constant), the 5% critical values for the alternative of positive autocorrelation are 1.45 (lower) and 1.62 (upper). Assuming an estimated $DW = 1.8$, we could not reject the null hypothesis that there is no positive autocorrelation, and we would infer that there is probably no first-order autocorrelation. If the estimated $DW = 1.4$, we would reject the null hypothesis of no positive autocorrelation. For all values between 1.45 and 1.62, the test is inconclusive, and we would infer that there probably is first-order autocorrelation.

—Suzanna De Boef

REFERENCES

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regressions I. *Biometrika, 37,* 409–428.

Wooldridge, J. M. (2000). *Introductory econometrics: A modern approach.* Cincinnati, OH: South-Western Thomson Learning.

DYADIC ANALYSIS

Often in social research, pairs of people (e.g., married couples, roommates, friends, and coworkers) are measured. The statistical analysis of data from pairs is called *dyadic analysis*. The analysis of dyadic data is much more complicated than the analysis of data from individuals (Kashy & Kenny, 2000). Particular consideration needs to be given to the interdependence of the data and to the UNIT OF ANALYSIS.

Typically, the scores of the two people are correlated, resulting in what is called *interdependence*. The correlation of scores invalidates the use of SIGNIFICANCE TESTING from an analysis that uses person as the unit of analysis. The p values from testing may be larger or smaller than they appear to be. This interdependence of scores can even be negative, such that the two members of the dyad are more different from one another than two randomly chosen people. The degree of interdependence can be measured by a Pearson correlation coefficient when members can be distinguished by some variable. For example, heterosexual married couples can be distinguished by their gender. When members cannot be distinguished (e.g., coworkers), the INTRACLASS CORRELATION should be used to test for measuring and testing nonindependence.

To solve the problems created by nonindependence, one typically needs to make dyad, not person, the unit of analysis. In some cases, one can conduct analyses using the sum and/or the difference between the two members of the dyad as the variables in the analysis.

Although nonindependence can be viewed as a problem, it represents an opportunity to study various dyadic phenomena. For instance, reciprocity, compensation, synchrony, and similarity imply nonindependence in the responses of the members of the dyad. Sources of nonindependence such as these can be due to nonrandom pairing (e.g., assortative mating), common fate, and mutual influence.

Occasionally in dyadic analysis, an index is created for each dyad. For example, the similarity of the members' leisure preferences is indexed by a correlation coefficient. Although dyad indexes can be informative, there are several pitfalls in their interpretation (Cronbach, 1955).

It can be useful in dyadic analysis to study how it is that the characteristics of one person influence the behavior of the person and the person's partner. So, for each person, there is an outcome that is regressed on the person's own characteristics and the characteristics of the person's partner. As an example, it could be examined how the level of depression of the person and the person's roommate affects the person's satisfaction with the roommate. Kashy and Kenny (2000) refer to this model as the Actor-Partner Interdependence Model. Alternatively, it is possible to estimate models of mutual influence using feedback loops and instrumental variables (Duncan, Haller, & Portes, 1975).

Sometimes, with dyadic data, each person is paired with several others. For instance, each person might rate everyone else in her sorority. Such round-robin data requires the use of the SOCIAL RELATIONS MODEL.

—David A. Kenny

REFERENCES

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52,* 177–193.

Duncan, O. D., Haller, A., & Portes, A. (1968). Peer influence on aspiration: A reinterpretation. *American Journal of Sociology, 75,* 119–137.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 451–477). New York: Cambridge University Press.

DYNAMIC MODELING. *See* SIMULATION

E

ECOLOGICAL FALLACY

In 19th-century Europe, suicide rates were higher in countries that were more heavily Protestant; Emile Durkheim inferred that suicide was promoted by the social conditions of Protestantism (*Le suicide,* 1897). This is an "ecological inference"—a conclusion about individual behavior drawn from data about aggregate behavior.

The Protestant countries, of course, were different from the Catholic countries in many ways besides religion (the problem of "confounding"). Moreover, Durkheim's data did not tie individual suicides to any particular religious faith: He had rates for countries, not for religious denominations within geographic areas.

The problem of confounding must be dealt with in any observational study, but the second problem is specific to ecological studies: Putative causes and effects are measured for groups rather than for individuals. Moreover, the analytic interest is in one kind of grouping—for Durkheim, religion—whereas data are available for an entirely different sort of grouping (geography).

If there is no confounding, the expected difference between effects for groups and for individuals is "aggregation bias"; in general, the difference is attributable partly to confounding and partly to aggregation bias. The "ecological fallacy" consists of thinking that relationships observed for groups necessarily hold for individuals: If countries with more Protestants tend to have higher suicide rates, then Protestants must be more likely to commit suicide.

Ecological studies can provide useful clues, but conclusions about individuals are in general only weakly supported by data on groups. The source of the problem is confounding and aggregation bias. Indeed, as shown by Robinson (1950), individual-level relationships can be reversed by aggregation. This is "Simpson's Paradox" for the correlation coefficient.

Statistical procedures have been proposed for disentangling individual-level from group-level behavior, including "ecological regression" and "cross-level" or "hierarchical" regression models. However, each method makes its own rather strong behavioral assumptions, which may seem implausible when stated explicitly. For instance, ecological regression makes the "constancy assumption." According to this assumption, with an application like Durkheim's, individual behavior depends on religious affiliation but not geographical location. Protestants all over Europe must have similar propensities to commit suicide, and Catholics are just as homogeneous.

Generally, the identifying assumptions in the models cannot be validated by the data. Moreover, in test situations where individual-level data are available—so that estimates can be compared to reality—the track record of the models is mixed at best. For additional discussion from various perspectives, see the references below.

—David A. Freedman

REFERENCES

Firebaugh, G. (2001). Ecological fallacy. *International encyclopedia for the social and behavioral sciences* (Vol. 6, pp. 4023–4026). Oxford, UK: Pergamon.

Freedman, D. A. (2001). Ecological inference and the ecological fallacy. *International encyclopedia for the social and behavioral sciences* (Vol. 6, pp. 4027–4030). Oxford, UK: Pergamon.

Freedman, D. A., Klein, S. P., Ostland, M., & Roberts, M. R. (1998). Review of "A solution to the ecological inference problem." *Journal of the American Statistical Association, 93*, 1518–1522. (Discussion appears in Vol. 94, pp. 352–357.)

Goodman, L. (1953). Ecological regression and the behavior of individuals. *American Sociological Review, 18*, 663–664.

Grofman, B., & Davidson, C. (1992). *Controversies in minority voting: The voting rights act in perspective.* Washington, DC: Brookings Institution.

King, G. (1997). *A solution to the ecological inference problem.* Princeton, NJ: Princeton University Press.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.

ECOLOGICAL VALIDITY

Ecological validity refers to the extent to which behavior indicative of behavior studied in one environment (often, reference is to a laboratory setting) can be taken as characteristic of (or generalizable to) an individual's cognitive processes in a range of other environments (often glossed as "everyday" or "natural").

Discussions of the problem of ecological validity first came to prominence in psychological research in the United States owing to the work of Egon Brunswik (1943) and Kurt Lewin (1943), two German scholars who emigrated to the United States in the 1930s. Following in this tradition, developmental psychologist Urie Bronfenbrenner argued that there are three conditions that ecologically valid research must fulfill: It must (a) maintain the integrity of the real-life situations it is designed to investigate, (b) be faithful to the larger social and cultural contexts from which the subjects come, and (c) be able to demonstrate that the experimental manipulations and outcomes are "perceived by the participants in a manner consistent with the conceptual definitions explicit and implicit in the research design" (Bronfenbrenner, 1979, p. 35). In similar terms, Neisser (1976) noted marked discontinuities between the spatial, the temporal, and the intermodal relationships between real objects and events, on one hand, and the objects and events characteristic of laboratory-based research, on the other, as a fundamental shortcoming of cognitive psychology.

In many discussions of ecological validity, it is assumed that it is possible to discover and directly observe the ways that laboratory tasks occur (or don't occur) in nonlaboratory settings (e.g., Simon, 1976). A variety of findings in contemporary research, however, indicate that the requirements for establishing ecological validity place an enormous analytical burden on social scientists (see Cole, 1996, chaps. 8–9 for an extended treatment of the associated issues). Once we move beyond the laboratory in search of real-world analogues, the ability to identify tasks is markedly weakened. Failure to define the parameters of the analyst's task or failure to ensure that the task-as-discovered is also the subject's task vitiate the enterprise.

The problem arises, as Valsiner and Benigni (1986) point out, because standardized cognitive experimental procedures are meant to embody closed analytic systems. Consequently, attempting to establish task equivalence in order to generalize beyond the experimental circumstances amounts to imposing a closed system upon a more open behavioral system. Insofar as the social scientist's closed system does not capture veridically the elements of the open system it is presumed to model, experimental results systematically misrepresent the life process from which they are derived.

—Michael Cole

REFERENCES

Bronfenbrenner, U. (1979). *The ecology of human development.* Cambridge, MA: Harvard University Press.

Brunswik, E. (1943). Organismic achievement and environmental probability. *The Psychological Review, 50*, 255–272.

Cole, M. (1996). *Cultural psychology: A once and future discipline.* Cambridge, MA: Harvard University Press.

Lewin, K. (1943). Defining the "field at a given time." *Psychological Review, 50*, 292–310.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco: W. H. Freeman.

Simon, H. A. (1976). Discussion: Cognition and social behavior. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, NJ: Erlbaum.

Valsiner, J., & Benigni, L. (1986). Naturalistic research and ecological thinking in the study of child development. *Developmental Review, 6*(3), 203–223.

ECONOMETRICS

Literally interpreted, *econometrics* means "economic measurement." Although measurement is an important part of econometrics, the scope of econometrics is much broader, as can be seen from the following quotations.

> Econometrics, the result of a certain outlook on the role of economics, consists of the application of mathematical statistics to economic data to lend empirical support to the models constructed by mathematical economics and to obtain numerical results. (Tintner, 1968, p. 74)

> Econometrics may be defined as the social science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena. (Goldberger, 1964, p. 1)

> The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. (Haavelmo, 1944, p. iii)

As the preceding definitions suggest, econometrics is an amalgam of economic theory, mathematical economics, economic statistics, and mathematical statistics. Instead of being an adjunct to one of these disciplines, econometrics is now taught as a separate subject in most undergraduate and graduate departments of business and economics, for reasons explained below.

Economic theory makes statements or hypotheses that are mostly qualitative in nature. For example, microeconomic theory states that, other things remaining the same (the famous *ceteris paribus* clause), a reduction in the price of a commodity is expected to increase the quantity demanded of that commodity. Thus, economic theory postulates a negative or inverse relationship between the price and the quantity demanded of a commodity. The theory itself, however, does not provide any numerical measure of the relationship between the two. That is, it does not tell by how much the quantity demanded will go up or down as a result of a certain change in the price of that commodity. It is the job of the econometrician to provide such numerical estimates. Put differently, econometrics gives empirical content to economic theory. Sometimes econometrics can refine an existing theory by explicitly considering variables that were subsumed in the *ceteris paribus* clause. Qualitative variables such as sex, religion, and ethnicity have been added to labor market studies to explain labor force participation (i.e., the decision to work or not). Traditionally, variables such as hourly earnings and level of education were considered the primary determinants of labor force participation.

The main concern of mathematical economics is to express economic theory in mathematical form (equations) without regard to measurability or empirical verification of the theory. The econometrician uses the mathematical equations provided by the mathematical economist but puts these equations in such a form that they lend themselves to empirical testing. This conversion of mathematical equations into econometric equations (i.e., STOCHASTIC equations) requires a great deal of ingenuity and practical skill.

Economic statistics is concerned primarily with collecting, processing, and presenting economic data in the form of charts and tables. These are the jobs of the economic statistician. It is he or she who is primarily responsible for collecting data on such variables as gross domestic product (GDP), employment, unemployment, and prices. The data thus collected constitute the raw data for econometric work. The economic statistician does not go any further, not being concerned with using the collected data to test economic theories. Someone who does that becomes an econometrician.

Although mathematical statistics provides many tools used in the trade, the econometrician needs special methods in view of the unique nature of most economic data, namely, that the data are not generated as the result of controlled experiments. The econometrician, like the meteorologist, generally depends on data that cannot be controlled directly. As Spanos (1999) observed:

> In econometrics the modeler is often faced with *observational* as opposed to *experimental* data. This has two important implications for empirical modeling in econometrics. First, the modeler is required to master very different skills than those needed to analyze experimental data. . . . Second, the separation of the data collector and data analyst requires the modeler to familiarize himself/herself thoroughly with the nature and structure of data in question. (p. 21)

METHODOLOGY OF ECONOMETRICS

How do econometricians proceed in their analysis of an economic problem? That is, what is their methodology? Broadly speaking, there are three approaches to econometric methodology: classical, Leamer's (1978), and Hendry's (1995).

A discussion of all these approaches to econometric methodology would go far afield. The main features of the classical methodology are presented here because of its historical importance and the fact that this methodology still dominates empirical work in many areas. After discussion of this methodology, the other two approaches will be discussed.

Classical Methodology

Classical methodology proceeds as follows:

Economic theory → Mathematical model of theory → Econometric model of theory → Data → Estimation of econometric model → Hypothesis testing → Forecasting or prediction → Using the model for control or policy purposes.

Each arrow indicates a sequential or stepwise process.

To best illustrate the classical methodology, we consider the well-known Keynesian theory of consumption.

Statement of Theory or Hypothesis

Keynes (1936) stated that

> The fundamental psychological law ... is that men [and women] are disposed, as a rule and on average, to increase their consumption as their income increases, but not as much as the increase in their income. (p. 36)

In short, Keynes postulated that the *marginal propensity to consume* (MPC), the rate of change of consumption for a unit (say, a dollar) change in income, is greater than zero but less than 1.

Specification of the Mathematical Model of Consumption

Although Keynes postulated a positive relationship between consumption and income, he did not specify the precise form of the functional relationship between the two. For simplicity, a mathematical economist might suggest the following form of the Keynesian consumption function:

$$Y = \beta_1 + \beta_2 X; \quad 0 < \beta_2 < 1 \tag{1}$$

where Y is consumption expenditure and X is income. β_1 and β_2, the parameters of the model, are, respectively, the intercept and slope coefficients, with the slope coefficient representing the MPC.

Specification of the Econometric Model of Consumption

The mathematical model of the consumption function given in equation (1) is of limited use to the econometrician because it assumes that there is an *exact* or *deterministic* relationship between consumption expenditure and income. But relationships between economic variables generally are inexact. To allow for the inexact relationship, the econometrician would modify the mathematical model of the consumption function by writing it as

$$Y = \beta_1 + \beta_2 X + u \tag{2}$$

where u, known as the *disturbance,* or ERROR term, is a RANDOM (or stochastic) VARIABLE that has well-defined properties. The error term u may represent all those factors that affect consumption but are not explicitly accounted for in equation (2). Equation (2) is an example of an *econometric model.* More technically, it is an example of a LINEAR REGRESSION MODEL.

Obtaining Data

To estimate equation (2)—that is, to obtain the numerical estimates of β_1 and β_2, which in the present context are known as regression coefficients—we need data. For illustrative purposes, we obtained data on Y (personal consumption expenditure) and X (GDP) for the United States for the period 1982–1996, the data being measured in 1992 in billions of dollars.

Estimation of Econometric Model

Using the method of ORDINARY LEAST SQUARES (OLS), and based on the U.S. data, the following regression results were obtained:

$$\begin{aligned} \hat{Y}_t &= -184.0780 + 0.7064X_t \\ SE: &\quad (46.2619) \quad (0.0078) \\ R^2 = &\quad 0.9984 \end{aligned} \tag{3}$$

where the hat over Y_t means that it is the estimated value of mean, or average, consumption expenditure, and where figures in the parentheses are the estimated standard errors. The subscript t indicates time. [The classical linear model makes certain assumptions about the error term u, namely, that it has a mean of zero, that it has a constant variance, and that the errors are uncorrelated. For hypothesis testing, it is often assumed that the error term is normally distributed. If the normality assumption is invoked, we can use the METHOD OF MAXIMUM LIKELIHOOD ESTIMATION (ML) to estimate the parameters of the regression model. The ML estimators of β_1 and β_2 are the same as those obtained by OLS.]

Hypothesis Testing

Keynes asserted that the marginal propensity to consume is positive but less than 1. In the estimated regression model given in equation (3), the MPC is 0.7064. Is it statistically different from 1? In other words, we wish to test the null hypothesis that the true MPC is 1. How does one test this hypothesis?

Given the assumptions of the CLASSICAL LINEAR REGRESSION model, one can test the hypothesis by the T-TEST, which in the present case gives

$$t = \frac{0.7064 - 1}{0.0078} = -3.7641. \tag{4}$$

Because there are 15 observations in the data, this t value has 13 DEGREES OF FREEDOM. Therefore, this t value is significant at the 1% level of significance (the actual p value is much lower). We can thus reject the null hypothesis that the true MPC is 1, perhaps substantiating Keynes' hypothesis.

Forecasting or Prediction

Assuming that the econometric model is a reasonably good approximation of reality, one can use it to forecast the future value of Y on the basis of a known or expected future value of the predictor variable X. Suppose we were to predict the mean consumption expenditure for 1997. Suppose that the value of GDP for 1997 was $7,269.8 billion (which was the actual value). Using this value of GDP in equation (3), the reader can verify that the predicted mean consumption expenditure is about $4,951.31 billion.

Use of Models for Control or Policy Purposes

Suppose that the government believes that the consumption expenditure of about $4,900 billion will keep unemployment at its current level. What level of GDP will guarantee the target level of consumption expenditure? Assuming that the estimated consumption expenditure regression in equation (3) is reasonably good, all one has to do is to put $4,900 on the left-hand side of that equation and solve for X. Simple arithmetic will show that an expenditure level of $7,200 billion will maintain the current level of unemployment. In this exercise, Y is the *target variable* and X is the *control variable.*

The preceding eight-step procedure is, in a nutshell, the classical methodology. There are several critics of this methodology. Two of them are considered below.

Hendry's Top-Down Methodology

According to Hendry (1995), the classical methodology is too static. That is, it does not take into account changes over time. In other words, he would develop econometric models in a dynamic setting. He would make his econometric model as comprehensive as possible and then whittle it down to a smaller model after considerable diagnostic testing. Because the bulk of Hendry's initial work was done at the London School of Economics (LSE), his methodology is also known as the LSE methodology.

The LSE starting point is that economic theory postulates a long-run, or equilibrium, relationship between economic variables, say Y ("permanent" consumption) and X ("permanent" income), the term "permanent" indicating the level that households regard as likely to persist in the future. This relationship can be summarized as

$$Y_t = \alpha X_t. \tag{5}$$

Hendry and his colleagues at LSE index the observations with the time subscript t because their methodology was developed to deal mainly with TIME-SERIES economic data.

Of course, the long-term relationship postulated in equation (5) takes time to achieve; therefore, the LSE methodology proposes the following type of dynamic procedure in order to reach equation (5):

$$\begin{aligned} Y_t = {} & \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_m X_{t-m} \\ & + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \cdots + \delta_l Y_{t-l} + u_t. \end{aligned} \tag{6}$$

In this model, Y depends not only on current and lagged X values but also on its own lagged values, the lags being represented by the lagged subscript notation. Of course, more X variables and their lags can be introduced.

As equation (6) suggests, we regress Y at time t on the values of X at times $t, (t-1), (t-2), \ldots,$ and $(t-m)$ as well as the lagged values of Y at times $(t-1), (t-2), \ldots, (t-l)$, the choice of the lagged terms m and l being an empirical question.

The model in equation (6) is an example of an *autoregressive distributed lag* (ADL) model. It is autoregressive because we are regressing the current Y value on its previous lagged values, and it is distributed because the effect of the X is spread over time. Such ADL models are also known as *dynamic models* because they explicitly consider the behavior of the Y variable over time. Once equation (6) is estimated, one can get back to the long-run relationship with appropriate algebraic manipulations.

Leamer's Methodology

Leamer is a strong critic of the classical econometric methodology, which he calls *average economic regression* (AER). In his view, in nonexperimentally collected data, the AER strategy is questionable. To him and other critics, once a model is specified, estimating its parameters and engaging in hypothesis testing is trivial, but the task of determining what the appropriate model is to begin with is very demanding. This task is the subject of *specimetrics.* According to Leamer (1978):

> Specimetrics describes the process by which a researcher is led to choose one specification of the model rather than another; furthermore, it attempts to identify the inferences that may be properly drawn from a data set when the data generating mechanism is ambiguous. (p. v)

In short, Leamer and his followers strongly believe that one must pay very careful attention to specimetrics, that is, the choice of the appropriate model. Once that is done, one may then follow the classical or AER methodology.

Besides this fundamental difference between Leamer's and the AER methodologies, Leamer is a strong proponent of BAYESIAN METHODS. The classical and Hendry's approaches to econometric modeling are based on classical statistics. In a Bayesian approach to, say, regression models, "parameters" such as β_1 and β_2 in equation (1) are regarded as "random," with some probability distribution for each (called the *priors* in Bayesian language). Priors typically are chosen subjectively based on such factors as previous research and one's understanding of a phenomenon. Given the data at hand, we estimate β_1 and β_2 and then, using the sample estimates and the priors, we obtain what are called *posterior* (probability) distributions of β_1 and β_2. Thus, in the Bayesian approach, β_1 and β_2 are estimated partially from the data and partially by the prior information about these coefficients. This gives us the posterior (probability) distributions of these coefficients. In the next round of analysis, these posterior distributions become the prior distributions. Thus, the Bayesian approach is a "learning by doing" procedure. As more data become available, researchers revise prior knowledge about the coefficients.

In the classical regression procedure, on the other hand, β_1 and β_2 are regarded as fixed numbers, although their values are unknown. Given the data, one can use OLS or maximum likelihood (ML) to estimate the parameters and engage in hypothesis testing. If we have a new set of data, we start fresh, without incorporating the results from the previous sample. Thus, there is a fundamental difference between the two approaches.

—Damodar N. Gujarati

REFERENCES

Goldberger, A. S. (1964). *Econometric theory.* New York: John Wiley & Sons.

Gujarati, D. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica, 12* (Suppl.), preface, p. iii.

Hendry, D. F. (1995). *Dynamic econometrics.* New York: Oxford University Press.

Keynes, J. M. (1936). *The general theory of employment, interest, and money.* New York: Harcourt Brace Jovanovich.

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data.* New York: John Wiley & Sons.

Mittelhammer, R. C., Judge, G. G., & Miller, D. J. (2000). *Econometric foundations.* New York: Cambridge University Press.

Spanos, A. (1999). *Probability and statistical inference: Econometric modeling with observational data.* Cambridge, UK: Cambridge University Press.

Tintner, G. (1968). *Methodology and mathematical economics and econometrics.* Chicago: University of Chicago Press.

EFFECT SIZE

A standardized numerical INDEX of the magnitude of an effect or relationship is called an *effect size.* The terms "magnitude" and "standardized" are particularly important in the above definition and deserve further explanation. The magnitude of an effect is critical for the proper interpretation of social scientific findings. The alternative to consideration of effect size is to base conclusions solely on the results of statistical SIGNIFICANCE TESTING. Because statistical significance is so heavily influenced by SAMPLE size, it is possible for a very weak and, perhaps, trivial effect to be statistically significant thanks to a large sample. On the other hand, a strong effect might go undetected by a test of statistical significance if the sample is very small. Thus, in addition to using tests of statistical significance, one does well to base one's conclusions on effect size, which is designed to be independent of sample size.

The fact that effect sizes are presented in standardized, rather than raw score, units also is important, especially in the social sciences. Many of our measures have arbitrary scales of measurement (e.g., LIKERT SCALES). Because the scales are arbitrary, they have no inherent meaning and may vary across studies. Suppose that the job satisfaction scores of employees after a given workplace intervention were 0.5 units higher, on average, than before the intervention. The recipient of this information can make no sense out of it, if for no other reason than the fact that job satisfaction could be measured on an infinite number of scales. The value of 0.5 might be an enormous change or nearly no change at all. If, on the other hand, the results were reported in a form such as "Postintervention job satisfaction was 0.5 STANDARD DEVIATIONS higher than it was before the intervention," it would be much easier to draw conclusions. Whatever the standard deviation (i.e., whatever the scale), satisfaction was half of a standard deviation higher postintervention. A study attempting to replicate this finding might use a measure of job satisfaction that uses a different scale, but if the results are reported in standard deviation units, then they will be directly comparable to those of any other such study. It is for this reason that effect sizes usually are standardized.

Although there exist many different types of effect size index, a few in particular are worth mentioning here. For correlational data, the most common effect size is the square of the CORRELATION coefficient. The coefficient itself might be a simple correlation, a multiple correlation, or a partial correlation, among other types. The square of a correlation is particularly handy because it reflects the percentage of VARIANCE in one VARIABLE accounted for by another.

For procedures for finding out the DIFFERENCE OF MEANS, such as those associated with ANALYSIS OF VARIANCE designs, there are various available indices. The most common is Cohen's d statistic, which is defined in a variety of ways, the most common being the difference between two means divided by the pooled standard deviation, s_{pooled}, where s_{pooled} is a function of the within-group variances and sample sizes.

$$s_{\text{pooled}} = \left\{\left[(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2\right] / (N_1 + N_2 - 2)\right\}^{.5}.$$

The resulting d value is simply the mean difference if the dependent variable were scaled to have unit variance. Another way of interpreting d is that it is the proportion of scores in one group that are less than the average score of the other group.

It should be noted that both of the examples of effect size indices given here reflect magnitude of effect and are standardized.

—Jose M. Cortina

EFFECTS CODING

Effects coding is a technique for coding $j - 1$ DUMMY VARIABLES to fully represent the information contained in a j-category classification. Coefficients for effects-coded dummy variables allow the researcher to compare predicted values for designated categories to an "average" value for the overall sample. This "average value" is the unweighted mean of group means, calculated as $[\sum_{j=1}^{J} \bar{X}_j] \div J$. Only when subgroup sizes are equal does the unweighted mean of group means equal the grand mean (Hardy, 1993). Regression coefficients therefore quantify the difference between the mean of the specified subgroup and the unweighted mean of all subgroup means.

To illustrate, consider the relationship between marital status (married, separated, divorced, widowed, never married) and hourly wages (in US$). In this example, marital status consists of five (J) subgroups, represented by four ($J - 1$) dummy variables. Effects coding for the dummy variables is illustrated in Table 1. We select married as the omitted category (see Hardy, 1993), which is coded –1 on all four dummy variables. Each dummy variable codes observations in one of the remaining four subgroups "1" and observations in the other three subgroups "0." For example, EFFNM (the effects-coded dummy variable indicating those who never married) codes the never married "1," those currently married "–1," and remaining groups (separated, widowed, and divorced) 0.

Table 1

	Example of Effects-Coded Dummy Variables					
Subgroups	*EFFNM*	*EFFS*	*EFFW*	*EFFD*	*Means (US$)*	*N of cases*
Married	−1	−1	−1	−1	15.33	1,661
Never married	1	0	0	0	12.95	613
Separated	0	1	0	0	11.64	146
Widowed	0	0	1	0	12.94	390
Divorced	0	0	0	1	11.93	14
All cases					14.26	2,773

Table 2

Model	*B*	*Std. Error*	*t*	*Sig.*
(Constant)	12.96	0.471	27.527	0.000
EFFNM	−0.005	0.536	−0.010	0.992
EFFS	−1.318	0.704	−1.871	0.061
EFFD	−0.022	0.569	−0.038	0.970
EFFW	−1.027	1.755	−0.585	0.559

Consider descriptive statistics and *zero-order correlations* for effects-coded dummy variables. The mean for each effects-coded dummy variable reports the discrepancy in subgroup size between the reference group (coded −1) and the subgroup coded 1, as in $(n_j - n_{\text{ref}})/N$. A negative sign indicates that the reference group has more observations than subgroup j, whereas a positive mean indicates that subgroup j has more observations than the reference group, and the magnitude of the mean indicates the discrepancy in the group sizes relative to the total sample size, or the proportion of excess cases.

When effects-coded dummy variables are used as independent variables in ordinary least squares regression, the *constant* reports the unweighted mean of the subgroups' means. The *coefficients* report each subgroup mean's deviation from this reference point. Sample regression results are reported in Table 2 for a regression of hourly wages on effects-coded dummy variables for marital status. Regression coefficients in Table 2 are unstandardized.

The grand mean for the full set of observations ($N =$ 2,773) is 14.26. Because subgroups are unequal in their numbers of observations, however, the unweighted mean of the subgroup means is 12.96, found by adding the five subgroup means and dividing by 5. This value is reported by the constant. When we evaluate the t-tests, we see that the subgroup means for those widowed, divorced, separated, and never married do not significantly differ from this value. For example, separated's (EFFS) coefficient of −1.318 indicates that the mean hourly wage for separated people is \$1.32 less than the unweighted mean of \$12.96, or \$11.64. Although a difference of more than \$1.00 per hour may seem sizable, the standard error indicates that this wage gap is not reliably measured (e.g., the 95% confidence interval is −1.70 to 1.60). However, the F test indicates that the null hypothesis (which asserts that hourly wage does not differ by marital status) can be rejected at better than the 0.001 alpha level because those currently married have an average hourly wage of \$15.33 $[-1(-0.005 - 1.318 - 0.022 - 1.027) + 12.960 = 15.33]$, which is significantly different from \$12.96 as well as significantly different from the average hourly wage of all other subgroups.

—Melissa A. Hardy and Chardie L. Baird

See also CODING

REFERENCES

Cohen, J., & Cohen, P. (1983). *Applied multiple regression* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Hardy, M. A. (1993). *Regression with dummy variables.* Newbury Park, CA: Sage.

EFFECTS COEFFICIENT

The total effect, direct and indirect, of a variable in a path model is termed the effects coefficient. Suppose the following RECURSIVE PATH model:

$$Y = a + bX + e,$$
$$Z = c + dX + fY + u.$$

The variable X has a direct effect on Z, estimated by d, and an indirect effect on Z via the influence on Y, estimated by $b \times f$, or bf. The effects coefficient of X

on Z therefore is $d + bf$. Effects coefficients enable the analyst to take into account indirect as well as direct effects in a CAUSAL system, arriving at a summary of the impact of a variable as it works its way through a causal system. However, if that variable is not truly exogenous, calculation of its effects coefficient can be problematic.

—Michael S. Lewis-Beck

See also PATH ANALYSIS

EFFICIENCY

An efficient estimator has a PROBABILITY distribution with the smallest VARIANCE. For example, W and Z are ESTIMATORS of the parameter and their variances are $\text{var}(W)$ and $\text{var}(Z)$, respectively. If $\text{var}(W) < \text{var}(Z)$, then W is a more efficient estimator of θ than is Z. Often, this is referred to as relative efficiency because efficiency requires making a comparison between two or more estimators. Efficiency is a desirable property for estimators because the estimator with smaller variance has a greater chance on any given try of producing the true value θ, given that the estimator is UNBIASED.

To visualize efficiency, think of two people playing darts. The goal is to hit as many bull's eyes as possible. Each person represents a different estimator, and the bull's eye is the mean value of the probability distribution for each estimator. Player 1 throws 100 darts that all fall within a 5-inch circle of the bull's eye. Player 2 also throws 100 darts, but they land within a 12-inch circle of the bull's eye. Player 1 is more efficient than Player 2 because the darts of Player 1 lie closer to the mean (bull's eye). Player 1 also has a higher probability of hitting the bull's eye than does Player 2 because of the smaller variance in the throwing ability of Player 1, given that both players' throws are unbiased (that is, they are as likely to be on one side of the bull's eye as the other, and by the same distance).

An efficient estimator will not always produce an estimate equal to the true value of θ because efficiency is a property of the probability distribution from which the estimates are drawn. This means that in repeated SAMPLING even an efficient estimator will fail to produce an estimate equal to the true θ. However, by using an efficient estimator, one has a higher level of certainty that the estimate is close to the true value for θ. Additionally, an efficient estimator will fail to produce an estimate of θ equal to its true value if the estimator is biased.

There are so many different estimators that tend to put restrictions on the categories of estimates that we compare to determine their relative efficiency (Kennedy, 1993). First, we look at only unbiased estimates. Unbiased estimators on average produce estimates of θ equal to the true value of the parameter θ. By limiting our comparisons to only unbiased estimators, we compare distributions with the same mean and that produce parameter estimates on average equal to the true value of the parameter. Many times, it is difficult to determine which estimator is the most efficient among several unbiased estimators. An additional restriction can be added, limiting comparisons to only linear estimators. Linear, unbiased estimators with the smallest variance (efficient) represent a special class of estimators called BEST LINEAR UNBIASED ESTIMATOR (BLUE).

—Heather L. Ondercin

REFERENCES

Kennedy, P. (1993). *A guide to econometrics* (3rd ed.). Cambridge, MA: MIT Press.

EIGENVALUES

Eigenvalues, also known as characteristic roots or latent roots, are a special set of scalars related to MATRIX equations and are an important mathematical concept in a number of statistical methods including FACTOR ANALYSIS and PRINCIPAL COMPONENTS ANALYSIS.

Let $\mathbf{A}$ be a $k \times k$ square matrix of complex numbers; a scalar λ that belongs to the set of complex numbers is said to be an eigenvalue of $\mathbf{A}$ if there exists a nonzero $k \times 1$ column vector $\mathbf{X}$ such that

$$\mathbf{AX} = \lambda\mathbf{X}.$$

This column vector $\mathbf{X}$ is known as the eigenvector of $\mathbf{A}$. Because it is positioned to the right of $\mathbf{A}$, it is called a "right eigenvector." A row eigenvector that is positioned to the left of $\mathbf{A}$ is called a "left eigenvector." Each eigenvalue is associated with a pair of eigenvectors—a left and a right eigenvector. The decomposition of $\mathbf{A}$ into eigenvalues and eigenvectors is known as eigen decomposition. The equation above also can be expressed more compactly as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{X} = 0,$$

where **I** is the identity matrix. When an eigenvalue is distinct from all other eigenvalues, its eigenvector is unique. As an example of the application of eigenvalues, an eigenvalue in a dimension from a principal components analysis measures the goodness of fit and gives the proportion of variance in the original variables accounted for by the principal component.

—Tim Futing Liao

See also FACTOR ANALYSIS

EIGENVECTOR. *See* EIGENVALUES; FACTOR ANALYSIS

ELABORATION (LAZARSFELD'S METHOD OF)

In its most elementary form, application of the elaboration method implies that one starts with a two-dimensional table showing the relationship between two variables and then one introduces a third variable for elaboration and elucidation. Table 1 [inspired by the basic text by Lazarsfeld (1955) and by Hyman (1955)] serves as an example dealing with the (fictitious) relationship between the dichotomous variables X-Sex (1 = Woman, 2 = Man) and Y-Car accident last year (1 = Yes, 2 = No). The table is presented in percentages, with the absolute numbers between parentheses.

Table 1

	X-Sex		
Y-Car Accident	*1 = Man*	*2 = Woman*	*Total*
1 = Yes	38%	19%	28%
	(100)	(50)	(150)
2 = No	62%	81%	72%
	(160)	(220)	(380)
Total	100%	100%	100%
	(260)	(270)	(530)

It appears that women are better drivers than men: Whereas 38% of the men were involved in a car accident last year, this is true for only 19% of the women. What is the cause of this difference? Are women more careful, considerate drivers? A simple explanation might be that men, in the past, drove many more miles than women. It stands to reason that the more a person drives, the greater that person's chances are of getting involved in a road accident. The logic underlying the elaboration method is that if this latter interpretation were true, comparing men and women who drive the same number of miles per year would result in no difference in the number of accidents between men and women. We introduce a third variable, T-Mileage (1 = High, 2 = Low) and investigate the relationship between X and Y within the categories of the third variable (we "control for T" or "we hold T constant"). Introduction of the third variable is shown in Table 2.

Within the subtable for high mileage ($T = 1$), people have many more accidents than in the subtable for low mileage ($T = 2$), but there is no longer a difference between men and women within these subtables (i.e., when comparing men and women who drive the same amount of miles). Lazarsfeld called this type of outcome "interpretation." Representing direct causal effects by an arrow, the causal scheme of this outcome is $X \rightarrow T \rightarrow Y$, meaning X is a cause of Y but indirectly through T (as a result of all kinds of social circumstances, women drive less than men and *therefore* are involved in fewer accidents). If there were other differences left between men and women that caused women to be less involved in accidents than men, then even in the subtables controlling for T, there would have been differences in percentages between men and women.

Another type of outcome is SPURIOUSNESS. For this type, the original ASSOCIATION also disappears, as above, but now the causal order of X and T is changed: X is seen as a consequence of T rather than as a cause. The causal scheme is $X \leftarrow T \rightarrow Y$. Let us assume that for a table with X-Education and Y-Conservatism we find that less-educated people are more conservative than people with more education. At the same time, it may be true that (for T-Age) the older one gets, the more conservative one becomes; furthermore, younger people may on average be better educated. If age is responsible in this way for the original relationship between education and conservatism, the original relationship will disappear when controlling for age in the manner described above. From the interpretation outcome we learn that X and Y are indirectly causally related to each other; spuriousness unmasks an original CORRELATION and shows that the relationship is noncausal: Age (T) causes X (Education), and Age also causes Conservatism (Y), but X in no way causes Y.

Table 2

	T-Mileage (1 = High)			*T-Mileage (2 = Low)*		
	X-Sex			*X-Sex*		
Y-Car accident	*1 = Man*	*2 = Woman*	*Total*	*1 = Man*	*2 = Woman*	*Total*
1 = Yes	50%	50%	50%	7%	8%	7%
	(95)	(35)	(130)	(5)	(15)	(20)
2 = No	50%	50%	50%	93%	92%	93%
	(95)	(35)	(130)	(65)	(185)	(270)
Total	100%	100%	100%	100%	100%	100%
	(190)	(70)	(260)	(70)	(200)	(270)

The third type of elaboration is specification or INTERACTION. In this type of elaboration, the relationships between X and Y are different in the subtables. For example, in the first subtable above (for high mileage), it might occur that men cause more accidents than women, but when both drive less (low mileage), there may be no difference between men and women. Interaction specifies how the (causal) relationship between two variables differs within the categories of a third variable.

The elaboration method fulfills, for categorical variables, much the same role as MULTIPLE REGRESSION ANALYSIS fulfills for CONTINUOUS VARIABLES. Its basic principles can be extended to more than three variables and to more complicated situations, although a more formal approach is then required. Such a formal approach has been developed within the context of the LOG-LINEAR MODEL. Lazarsfeld's ultimate aim was to develop a sound approach toward CAUSAL MODELING using "sociological" data, meaning (for him) CATEGORICAL data obtained by means of nonexperimental survey designs. Many of his insights into the problems and possibilities of causal analysis can be found in modern treatments of causality, such as Pearl (2000).

—Jacques A. Hagenaars

REFERENCES

Hagenaars, J. A. (1998). Categorical causal modeling: Latent class analysis and directed log-linear models with latent variables. *Sociological Methods and Research, 26*, 436–486.

Hyman, H. (1955). *Survey design and analysis*. New York: Free Press.

Lazarsfeld, P. F. (1955). Interpretation of statistical relations as a research operation. In P. F. Lazarsfeld & M. Rosenberg (Eds.), *The language of social research* (pp. 111–125). Glencoe, IL: Free Press.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

ELASTICITY

Social science analysts often are interested in knowing the effect that a change in one variable produces on another. One useful way of capturing the direction and magnitude of this effect is the concept of *elasticity*, which is defined as follows: The elasticity of a function is the proportionate change in the dependent variable divided by the proportionate change in the independent variable at given values of both variables. That is,

$$E = \frac{\%\text{ change in } y}{\%\text{ change in } x} = \frac{(\Delta y/y) \times 100}{(\Delta x/x) \times 100} = \frac{\Delta y}{\Delta x} \times \frac{x}{y}, \qquad (1)$$

where the symbol Δ before a variable means a discrete change in the value taken by the variable and y and x are the DEPENDENT VARIABLE and an INDEPENDENT VARIABLE, respectively.

As an example, suppose that we are interested in measuring the effect of the consumption of calories on a person's weight under a particular diet. Let us assume that y is a person's weight in kilograms and x is the consumption of calories measured in kilocalories per day. Then, if we use absolute changes, the answer to the previous question would be arbitrarily affected by the choice of units. That is, if the per-day consumption of calories increases from 1,000 kilocalories to 1,100 kilocalories and people, as a result, increase their weight from 60 to 63 kilograms, we get the impression that the effect is quite small. After all, an increase of 100 kilocalories produces a weight gain of only 3 kilograms. However, by changing the unit of measure for weight from kilograms to grams, we find that an increase of 100 kilocalories causes a weight gain of 3,000 grams. We now get the impression that the effect is large. Using the ratio between percentage changes avoids this problem. The given weight gain is 5% whether measured in terms of kilograms [(3k/60k) × 100] or in terms of grams [(3,000k/60,000k) × 100].

It is important to note that if the variation in the value taken by the variables is not small, it is common to use an average between the values taken by the variables before and after the variation is produced. In this case, the elasticity takes the form of the following expression:

$$E = \frac{\Delta y}{\Delta x} \times \frac{\frac{(x_1+x_2)}{2}}{\frac{(y_1+y_2)}{2}} = \frac{\Delta y}{\Delta x} \times \frac{x_1+x_2}{y_1+y_2}. \qquad (2)$$

The subscripts 1 and 2 indicate the values taken by the variable before and after the variation is produced, respectively. Notice that if the variation in the value taken by the variables is not small, we will have the problem that the elasticity in equation (1) will be different under an increase in the consumption of calories than under a decrease. That is, when the consumption of calories rises from 1,000 to 1,100 kilocalories, the elasticity would be

$$\frac{3}{100} \times \frac{1{,}000}{60} = 0.5.$$

When the consumption of calories falls from 1,100 to 1,000 kilocalories, it would be

$$\frac{3}{100} \times \frac{1{,}100}{63} = 0.524.$$

By using expression (2), however, we get an estimate of the elasticity equal to 0.512 whether there is an increase or a decrease.

Furthermore, when Δ is very small (i.e., an infinitesimal increase), the previous formulas converge to one in which derivatives are used. That is,

$$E = \frac{\partial y}{\partial x} \times \frac{x}{y}. \qquad (3)$$

The concept of elasticity is also very important in REGRESSION analysis. For instance, when the regression is a linear one, like $y = \alpha + \beta x + \varepsilon$, the use of equation (3) gives an elasticity equal to $\beta(y/x)$. However, in the case that the regression is a logarithmic or a semilogarithmic one, we have the following cases:

$$\ln y = \alpha + \beta_1 \ln x_1 + \beta_2 x_2 + \varepsilon$$

$$\text{use} \quad \begin{cases} E_{y,x_1} = \frac{\partial \ln y}{\partial \ln x} = \beta_1 & (4a) \\ E_{y,x_2} = \frac{\partial \ln y}{\partial x} \cdot x = \beta_2 x_2 & (4b) \end{cases}$$

and

$$y = \alpha + \beta \ln x + \varepsilon \quad \text{use}$$

$$E_{y,x} = \frac{\partial y}{\partial \ln x} \cdot \frac{1}{y} = \beta \frac{1}{y}. \qquad (5)$$

Let's try to see how it works by using an example taken from Lindahl's (2002, p. 7) study of intergenerational income mobility in Sweden between 1962 and 1973. The purpose of Lindahl's study was to obtain a measure of the degree of equality of opportunities, as represented by intergenerational income mobility. Here we interpret only the case of families with only one male son. The regression results are as follow:

$$y = 8.29660 + 0.227x_1 + 0.05712x_2 - 0.00067955x_3, \qquad (6)$$

where y is the long-run log income of the sons, x_1 is long-run log income of the fathers, x_2 is the fathers' age, and x_3 is the square of the fathers' age. The reason for the inclusion of the fathers' age in the regression is to correct for the fact that the long-run income of a person probably is higher at the age of 50 than at the age of 25. Conversely, it was not necessary to correct for the sons' age because every son in the study was of the same age.

We can apply formula (4a) in the previous example to estimate the intergenerational income elasticity, which is equal to 0.227. This elasticity implies that, controlling for fathers' age effects, a 1% rise in the long-run income of the fathers produces a 0.227% rise in the long-run income of the sons.

For further readings on elasticity, see Eatwell et al. (1987).

—Osiris Jorge Parcero

REFERENCES

Eatwell, J., Milgate, M., & Newman, P. (1987). *The new Palgrave: A dictionary of economics*. London: Macmillan.

Lindahl, L. (2002). *Do birth order and family size matter for intergenerational income mobility? Evidence from Sweden* (Working Paper No. 5). Stockholm: Swedish Institute for Social Research.

EMIC/ETIC DISTINCTION

Linguist Kenneth L. Pike, in 1954, coined the terms *emic* and *etic* from *phonemic* and *phonetic*. Pike used emic to refer to the intrinsic cultural distinctions meaningful to the members of a cultural group and etic to refer to the extrinsic ideas and categories meaningful for researchers.

For example, modern medical science defines "diseases" in precise, culture-free (etic) ways for any patient, whereas traditional peoples define "illnesses" differently based on their particular cultural contexts (emically). Western botanists categorize flora and fauna etically, based on the Linnaean taxonomy, whereas indigenous peoples categorize them emically, based on their particular folk-science worldview. Some cultures in Papua, New Guinea (PNG) classify bats with "birds" rather than "mammals" because they fly like most birds. Optical physicists divide the color spectrum according to the wavelength of light etically, whereas local peoples divide it in different ways into emically meaningful divisions based on their languages and cultures. For example, the Selepet people of PNG use a single color term that includes the English *yellow, orange*, and *brown*.

Anthropologist Marvin Harris borrowed the terms from Pike in 1964 and applied them to his own CULTURAL MATERIALISM method of studying cultures. Misunderstanding Pike, he used the terms very differently. Pike had argued for an emic approach to studying a culture if one wanted to understand it correctly. In contrast, Harris defined the emic method as referring to the ideal reasons the natives give for their customs, and he used etics to refer to the subconscious real reasons for customs. For example, the Yanomami Indians say they raid enemy villages to kidnap young women (an emic explanation, in Harris's definition); but the real etic reason, which the Indians don't recognize, is that the enemy villages are encroaching on the scarce wild game resources needed by the raiding group.

COGNITIVE ANTHROPOLOGISTS follow Pike in using the emic/etic concept as a research method, whereas cultural materialists use the method as Harris remodeled it. This resulted in some confusion among anthropologists and linguists by the 1970s as to what emics and etics were supposed to mean. By the 1980s, the terms were appearing in many of the writings of social scientists. The meanings of the terms had multiplied in many more directions, in part because the emic/etic concept was now being used in other academic disciplines (including psychology, sociology, medicine, psychiatry, economics, and religion). These new meanings are described by Headland, Pike, and Harris (1990, pp. 15–23).

One commonality in both Pike's and Harris's models of emics and etics was that they both agreed that an emic description of a culture is not necessarily a model of which the natives are consciously aware. Even though many definitions refer to the emic model as the insider's view, that is not always so. Whereas emic constructs may make sense to those who hold them, those people often are not aware of them unless an outside ethnographer explains their own emic constructs to them.

—Thomas N. Headland and
Kenneth A. McElhanon

See also COMPONENTIAL ANALYSIS

REFERENCES

Harris, M. (1976). History and significance of the emic/etic distinction. *Annual Review of Anthropology, 5*, 329–350.

Headland, T. N., Pike, K. L., & Harris, M. (1990). *Emics and etics: The insider/outsider debate*. Newbury Park, CA: Sage.

Jahoda, G. (1977). In pursuit of the emic-etic distinction: Can we ever capture it? In Y. H. Poortinga (Ed.), *Basic problems in cross-cultural research* (pp. 55–63). Amsterdam: Swets & Zeitlinger.

Lett, J. (1987). The importance of the emic/etic distinction. In *The human enterprise: A critical introduction to anthropological theory*. Boulder, CO: Westview.

Pike, K. L. (1967). *Language in relation to a unified theory of the structure of human behavior* (2nd ed.). The Hague: Mouton. (Original work published 1954.)

EMOTIONS RESEARCH

Emotions research focuses on understanding the social, psychological, and physiological processes involved in emotional experience; the social construction of emotions (including historically and cross-culturally); the role of emotions in social life; the management of emotions in social interaction; and the role of emotions in the research process itself.

Researchers in this subfield collect data on emotions via methods that may closely resemble those used in other branches of sociology and social psychology, among them SURVEYS, EXPERIMENTS, FIELD RESEARCH, and IN-DEPTH INTERVIEWS. Their analyses entertain a variety of qualitative and quantitative techniques. A few approaches bear special mention.

In research based on affect control theory (ACT) (Heise, 1987), emotions are defined as arising from the actions and the confirmed or disconfirmed identities of people involved in interactions. Social actors may respond to the emotional cues that result from disconfirmed identities by redefining situations or other people's identities. Data collection for ACT frequently involves asking participants to rate concepts on three dimensions—evaluation (good vs. bad), potency (powerful vs. weak), and activity (lively vs. inactive)—using a SEMANTIC DIFFERENTIAL SCALE. The majority of analyses using ACT are quantitative.

In qualitative, field- and interview-based studies of emotions, researchers note social actors' emotional displays as they occur in context and record participants' descriptions of felt emotions that could not be inferred from observation alone. Researchers also note the consequences that unfold from emotion-related interactions. Those who apply CONVERSATION ANALYSIS techniques to study emotions in face-to-face communication may examine highly specific exchanges. Field researchers and in-depth interviewers, by contrast, explore a range of situations and conditions in order to understand how emotions work in conjunction with other social concepts and processes such as identity, perception, social relationships, social status, group dynamics, or social movements.

Arising from a critical reaction to the "affective neutrality" that social scientists aspired to maintain in the past, many field researchers now include themselves as research participants rather than acting solely as detached, unemotional observers. They treat their emotional experiences during fieldwork as valuable data, and they incorporate reflexive observations and insights in their FIELDNOTES and qualitative analyses (Kleinman & Copp, 1993). Researchers who include emotions as data can gain deeper sociological insights about the participants and the settings they study, the social constraints under which participants operate, and the social implications of their research.

To study the lived experience of emotions and to better understand how people experience emotions as complex processes, field researchers may employ the sociological technique of *systematic introspection* (Ellis, 1991). This method involves recording one's emotional reactions to people, objects, and events and then analyzing what those feelings mean and how they inform ongoing social situations and cultural practices. Systematic introspection may be conducted by trained participants or practiced by researchers as an AUTOETHNOGRAPHIC research tool.

—Martha Copp

See also REFLEXIVITY

REFERENCES

Ellis, C. S. (1991). Sociological introspection and emotional experience. *Symbolic Interaction, 14*, 23–50.

Heise, D. R. (1987). Affect control theory: Concepts and model. *Journal of Mathematical Sociology, 13*, 1–33.

Kleinman, S., & Copp, M. A. (1993). *Emotions and fieldwork.* Thousand Oaks, CA: Sage.

EMPIRICISM

Like "POSITIVISM," the term "empiricism" has come to be used by social scientists in ways that often indicate little more than negative evaluation. Indeed, these two words are frequently treated as synonyms. Both are now closely associated with the discredited idea that science is the only genuine source of knowledge because its conclusions are logically derived from empirical data.

In EPISTEMOLOGY, "empiricism" refers to a distinctive approach which claims that all knowledge comes from the senses, by contrast with the RATIONALIST argument that there is innate knowledge (see Hamlyn, 1967). In large part, empiricism arose out of the scientific revolution of the 17th century, with its emphasis on empirical (especially experimental) investigation of the physical world as the key to scientific knowledge. This was a departure from earlier ideas about the process of inquiry, in which reliance on ancient texts, religious faith, and/or systematic philosophizing had been influential.

It is perhaps worth emphasizing the radical and progressive character of 17th-century empiricism. This arose from the fact that it involved a critique both of previous claims to knowledge and of the means by which these had been justified. Moreover, it demanded that any criticism be answered in terms of an appeal to evidence of a kind that was, in principle, available to anyone. In this way, all forms of authority were opened up to potential challenge.

The meaning and validity of empiricist theories of knowledge have been matters of recurrent philosophical dispute. Against them, it has been argued that no data are simply available to researchers or observers without judgment or interpretation, and there are, therefore, no data whose validity can be taken as indubitable. Equally, it is claimed that science does not, and could not, operate simply by seeking logically to infer laws from observational data. Instead, it requires development of theories, which are then tested against data; a process that can only indicate likely truth or falsity, rather than establishing VALIDITY beyond all doubt.

In the context of the social sciences, *empiricism* generally has been used to refer to approaches that are felt to place too much emphasis on empirical data, discouraging theory construction. An example is C. Wright Mills's critique of "abstracted empiricism" (Mills, 1959). Empiricism also has been seen as involving an excessively narrow conception of evidence, one that limits it to what is observable or quantifiable. As this suggests, "empiricist" is a term of criticism most commonly applied to QUANTITATIVE RESEARCH, though there are examples of its application to QUALITATIVE RESEARCH. For example, some kinds of ETHNOGRAPHIC research have been challenged, notably by CRITICAL THEORISTS, for being preoccupied with appearances, that is, with documenting commonsense interpretations of the world, thereby neglecting the underlying structures that generate those ideological appearances. On the other side, although few would seek to defend empiricism in any narrow form, it can be argued that criticism of it has become exaggerated today, encouraging the presentation of speculative theorizing as fact.

—Martyn Hammersley

REFERENCES

Bryant, C. G. A. (1985). *Positivism in social theory and research*. London: Macmillan.

Hamlyn, D. W. (1967). Empiricism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (pp. 499–505). New York: Macmillan.

Mills, C. W. (1959). *The sociological imagination*. New York: Oxford University Press.

EMPTY CELL

An empty cell occurs when there are no cases in the CELL of a CONTINGENCY TABLE. Empty cells create problems of interpretation, as do cells containing very small numbers of cases.

—Alan Bryman

See also SPARSE TABLES

ENCODING/DECODING MODEL

The encoding/decoding model of media research, originally formulated by Stuart Hall (1974) at the University of Birmingham's Centre for Contemporary Cultural Studies in the 1970s, has been influential in cultural studies and the sociology of mass communications. It recalls the tripartite structure of the most elementary of communications models—sender, message, and receiver—but emphasizes the relatively autonomous moments of encoding and decoding and the textual polysemy (multiple meanings) of television programs in particular. The meaning of any message is not determined, in the end, by the intention of the communicator. It is not to be assumed, however, that audiences simply misunderstand intended meaning. As Umberto Eco (1965/1972) argued, "aberrant decoding" of mass communications is normal because modern media address such large and socially heterogeneous audiences that inevitably will bring their different frameworks of understanding to bear on the interpretation of messages.

Aberrant decoding of media messages is not just a matter of individual differences, according to Hall, but is also socially motivated. In a capitalist society, class differences are especially important and are represented by contrasting value systems and discourses. Hall appropriated Frank Parkin's (1973) distinction between dominant, subordinate, and radical meaning/value systems in order to examine what he termed dominant, negotiated, and oppositional decodings. It is reasonable to suppose that the mass media typically will present the legitimacy of dominant—in effect, "bourgeois"—values and ways of making sense of the world. Accordingly, preferred meanings and readings are routinely encoded into media texts, usually representing the ideological interests of the powerful. The relatively powerless may either accept or refuse the dominant ideological framework encoded in, say, television news programs. For instance, regarding the coverage of industrial conflict from the point of view of management, a worker of left-wing persuasion may interpret it as biased. In between taking the preferred reading straight (dominant decoding) or questioning it fundamentally (oppositional decoding), various kinds of negotiated decoding might be made more commonly by audience members.

Hall's encoding/decoding model was a key feature of the paradigm of hegemony theory in cultural and media studies. This was a significant revision of the classical Marxist dominant ideology thesis. Hall wanted to retain, to some extent, the argument that the dominant ideas in society were those of the ruling class while investigating the sheer SEMIOTIC complexity of ideological process through channels of communication. It was evidently the case that subordinate groups in society are not just indoctrinated by the dominant ideology but also are likely to question its validity according to their own interpretive frameworks, experiences, and interests. Ideological power, then, was a matter of discursive negotiation in the struggle for social leadership rather than simply imposed from above.

The encoding/decoding model was a heuristic device for putting hegemony theory into operation in critical media research. A student of Hall, David Morley (1980), applied it to study how, in educational settings, groups from different class backgrounds interpreted a newsmagazine program, *Nationwide,* on British television during the late 1970s. He reached two important conclusions that were influential in the further development of research on encoding and decoding. First, class was neither the only nor necessarily the most important factor in differential decoding. Ethnicity, gender, and generation were key sociological determinants as well. In subsequent work by Morley and his followers, recognition of finely differentiated decoding of media genres proliferated to such an extent that the very notion of a preferred meaning and the ideological critique implicit in it was undermined. Second, Morley (1986) himself became dissatisfied with the artificiality of FOCUS GROUP research, which he had employed in his *nationwide* study, and more interested in the ETHNOGRAPHY of media use, especially gender relations surrounding television in the home and the idea of the active audience. As the encoding/decoding model developed, then, it became less concerned with questioning the ideological power of mass media and more focused on domesticated micropolitics and, in effect, the vicissitudes of sovereign consumption (see McGuigan, 1992, for a fuller critical assessment).

—Jim McGuigan

REFERENCES

Eco, U. (1972). Towards a semiotic inquiry into the television message. In *Working papers in cultural studies* (Vol. 3, pp. 103-121). Birmingham: Centre for Contemporary Cultural Studies. (Original work published 1965)

Hall, S. (1974). The television discourse—Encoding and decoding. In *Education and culture* (Vol. 35, pp. 8-14). Paris: United Nations Educational, Social, and Cultural Organization.

McGuigan, J. (1992). *Cultural populism.* London: Routledge.

Morley, D. (1980). *The "Nationwide" audience.* London: British Film Institute.

Morley, D. (1986). *Family television—Cultural power and domestic leisure.* London: British Film Institute.

Parkin, F. (1973). *Class inequality and political order.* London: Paladin.

ENDOGENOUS VARIABLE

An endogenous variable is a factor in a CAUSAL MODEL or a causal system whose value is determined by the states of other variables in the system, in contrast with an *exogenous variable*, the value of which is determined outside the system. Related but non-equivalent distinctions are those between DEPENDENT and INDEPENDENT VARIABLES and between explanandum and explanans. A factor can be classified as endogenous or exogenous only relative to a SPECIFICATION of a MODEL representing the causal relationships producing the outcome $\mathbf{y}$ among a set of causal factors $\mathbf{X}(x_1, x_2, \ldots, x_k)$ where $\mathbf{y} = \mathbf{M}(\mathbf{X})$. A variable x_j is said to be endogenous within the causal model $\mathbf{M}$ if its value is determined or influenced by one or more of the independent variables $\mathbf{X}$ (excluding itself). A purely endogenous variable is a factor that is entirely determined by the states of other variables in the system. (If a factor is purely endogenous, then in theory we could replace the occurrence of this factor with the functional form representing the composition of x_j

as a function of X.) In real causal systems, however, there can be a range of endogeneity. Some factors are causally influenced by factors within the system but also by factors not included in the model. A given factor therefore may be partially endogenous and partially exogenous—partially but not wholly determined by the values of other variables in the model.

Consider a simple causal system—farming. The outcome we are interested in explaining (the dependent variable or the explanandum) is crop output. Many factors (independent variables, explanans) influence crop output: labor, farmer skill, availability of seed varieties, availability of credit, climate, weather, soil quality and type, irrigation, pests, temperature, pesticides and fertilizers, animal practices, and availability of traction, among others. These variables are all causally relevant to crop yield, in a specifiable sense: If we alter the levels of these variables over a series of tests, the level of crop yield will vary as well (up or down). These factors have real causal influence on crop yield, and it is a reasonable scientific problem to attempt to assess the nature and weight of the various factors. We can also notice, however, that there are causal relations among some of but not all these factors. For example, the level of pest infestation is influenced by rainfall and fertilizer (positively) and pesticide, labor, and skill (negatively). Pest infestation therefore is partially endogenous within this system and partially exogenous, in that it is also influenced by factors that are external to this system (e.g., average temperature, presence of pest vectors, decline of predators).

The concept of endogeneity is particularly relevant in the context of TIME SERIES analysis of causal processes. It is common for some factors within a causal system to be dependent for their value in period n on the values of other factors in the causal system in period $n - 1$. Suppose that the level of pest infestation is independent of all other factors within a given period but is influenced by the level of rainfall and fertilizer in the preceding period. In this instance, it would be correct to say that infestation is exogenous within the period but endogenous over time.

—Daniel Little

See also EXOGENOUS VARIABLE

REFERENCES

Hendry, D. F. (1995). *Dynamic econometrics*. Oxford, UK: Oxford University Press.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

EPISTEMOLOGY

In philosophy, epistemology refers to a THEORY of knowledge, a theory of how human beings come to have knowledge of the world around them—of how we know what we know. Epistemology provides a philosophical grounding for establishing what kinds of knowledge are possible and for deciding how knowledge can be judged as being both adequate and legitimate. In the social sciences, the term is used in the context of deciding which scientific procedures produce reliable social scientific knowledge.

Two theories of knowledge have predominated in philosophical discourse since the scientific revolution in the 17th century: RATIONALISM (represented by René Descartes, Gottfried Leibniz, and Benedict de Spinoza) and EMPIRICISM (represented by John Locke, George Berkeley, and David Hume). The concern was to find a secure foundation for scientific knowledge and to distinguish this from belief and prejudice. Rationalism is based on the idea that reliable knowledge is derived from the use of "pure" reason, from establishing indisputable axioms and then using formal logic to arrive at conclusions. From this point of view, mathematics produces such knowledge. Empiricism, on the other hand, relies on the use of the human senses to produce reliable knowledge. This means that knowledge of the world can be obtained only through direct sense-experience.

In the context of the social sciences, these philosophical positions can be further elaborated in terms of two dominant epistemological positions and their associated ontological positions (see ONTOLOGY). In the first epistemological position, known as nominalism, the concepts that are used in description and explanation are simply regarded as convenient, collective names that are invented as summaries of the general categories of things that have been observed, such as "social actors" or "social groups." These collectivities should neither be confused with reality itself nor attributed with the capacity to act; reality is made up of events, and only individuals can act. In the second epistemological position, known as REALISM, scientific concepts are viewed as revealing something about social reality that is not necessarily observable. Such concepts are designed to penetrate beyond observable events to a reality that underlies and explains them (see also CRITICAL REALISM).

When nominalism and realism are combined with the two major alternative ontological positions, materialism and idealism, a four-way classification scheme is generated. Empiricism combines a materialist ontology (see IDEALISM) with a nominalist epistemology, substantialism combines a materialist ontology with a realist epistemology, subjectivism combines an idealist ontology with a nominalist epistemology, and rationalism combines an idealist ontology with a realist epistemology (Johnson, Dandeker, & Ashworth, 1984).

In empiricism, reality is viewed as being constituted of material things that can be observed by the use of the human senses. Concepts and generalizations are shorthand summaries based on many observations. Substantialism also adopts a materialist view of reality but accepts that people in different times and places can interpret reality differently. Nevertheless, the material world is seen to constrain human actions and social relations. Because subjectivism rejects the notion of a material world and views reality as being socially constructed and interpreted, knowledge of this reality is available only from the accounts that social actors can give of it. Finally, rationalism views reality as both real and general; it exists independently of people, their consciousness, and their circumstances. Because this reality is made up of ideas, knowledge of it can be obtained only by examining thought process, the innate ideas shared by human beings—in short, the structure of mind itself. These four positions must be regarded as ideal types, between which there are inherent tensions (Johnson et al., 1984). They are associated with the major PHILOSOPHIES OF SOCIAL SCIENCE. Empiricism is associated with POSITIVISM and FALSIFICATIONISM, substantialism is associated with CRITICAL REALISM, and subjectivism is associated with INTERPRETIVISM. Rationalism can be found in Emile Durkheim's work on suicide but is now uncommon in the social sciences.

—Norman Blaikie

REFERENCES

Blaikie, N. (1993). *Approaches to social enquiry.* Cambridge, UK: Polity.

Crotty, M. (1998). *The foundations of social research.* London: Sage.

Johnson, T., Dandeker, C., & Ashworth, C. (1984). *The structure of social theory.* London: Macmillan.

EQS

EQS is a statistical software package for performing analysis of STRUCTURAL EQUATION MODELS. It is known for its ease in handling non-normal and noncontinuous data. For further information, see the Web site of the software: http://www.mvsoft.com/

—Tim Futing Liao

ERROR

In everyday usage, "error" means "mistake." In arithmetic, "error" means the degree of inaccuracy in a calculation. In statistics, the term "error" may not have a negative connotation. It generally represents the difference or discrepancy between the value of an entity observed (e.g., the sample mean) and its "true" or expected value (e.g., the population mean). Very often, this difference is due to the result of chance elements. That is why it is often called a stochastic error or random error term. (The word "stochastic" comes from the Greek word *stokhos,* meaning the "bull's eye." The outcomes of throwing darts onto a dart board is a stochastic process, that is, a process fraught with misses. In statistics, the word implies the presence of a random variable—a variable whose outcome is determined by a chance experiment.)

There are several types of errors that one may encounter in practice. These include WHITE NOISE error, TYPE I ERROR, TYPE II ERROR, errors in MEASUREMENT, errors in variables, errors in equations, model specification error, errors resulting from missing observations, and FORECASTING errors.

WHITE NOISE ERROR

Let u_t be a random error term. If it has a zero mean, constant variance, and no autocorrelation (or SERIAL CORRELATION), it is called a white noise error term, denoted by $u_t \sim WN(0, \sigma)^2$. If in addition to being serially uncorrelated, u_t is serially independent, then u_t is independent white noise, denoted by $u_t \sim iid(0, \sigma)^2$, where *iid* means independently and identically distributed. If, in addition, u_t is normally distributed, then it is said to be normal white noise or Gaussian white noise, denoted by $u_t \sim iidN(0, \sigma)^2$. It may be noted that zero correlation means independence only in the normal case.

The white noise error term plays a crucial role in TIME SERIES modeling (see FORECASTING) as well as in the classical linear regression model.

ERRORS OF OBSERVATION OR MEASUREMENT

In collecting data, errors often arise because of faulty measurement instruments or because of human factors. A common error is *rounding error*, or *error of approximation*. In this type of error, instead of recording observations, say, to the fourth decimal point, one rounds them to three decimal points. In survey-type data, the questioner may not obtain information on all the questions because of nonresponse, or not all questionnaires may be returned. Sometimes information on variables such as income and wealth is difficult to come by because respondents are reluctant to disclose the true information for fear of creating problems with tax authorities.

ERRORS IN VARIABLES

Sometimes information on variables is inherently difficult to obtain. For instance, data on expected, or anticipated, inflation is hard to come by because one person's expectation of inflation may not coincide with another person's. Economists have devised various schemes, or *proxy variables*, to measure the anticipated or expected inflation. One measure of expected inflation is obtained by surveying leading economists about their views on inflation expected in the near future. The views of these economists may be based on mathematical models or may be purely personal. To take another example from economics, the variable "natural rate of unemployment" is very difficult to measure. Some economists take some kind of weighted average of the actual rate of unemployment in the past several months or quarters and use that as a proxy for the natural rate of unemployment. All these proxy measures are likely to contain errors.

ERRORS IN TESTING HYPOTHESES

In testing statistical hypotheses, one may commit a TYPE I ERROR or a TYPE II ERROR. A Type I error consists of rejecting a true hypothesis, whereas a Type II error consists of not rejecting a hypothesis even though it may be false. Both these types of errors are fundamental to the Neyman-Pearson, or classical, theory of hypothesis testing. Ideally, one would like to minimize both these types of errors; however, for a *given* sample size, it is not possible to minimize both types of errors simultaneously. Classical hypothesis testing usually fixes the probability of a Type I error at some arbitrary LEVEL OF SIGNIFICANCE, such as 1%, 5% or 10%, and does not worry much about a Type II error; In some respects, a Type I error is regarded as more "serious" than a Type II error, in the sense that convicting an innocent person is more serious than letting a guilty person go free.

EQUATION ERROR

Equation error typically is associated with REGRESSION analysis. In regression analysis, the objective is to explain the behavior of one variable, Y (the DEPENDENT VARIABLE), in terms of one or more X variables (the EXPLANATORY VARIABLES). No matter how many X variables are considered, one can never explain the behavior of Y completely in terms of the chosen X variables. The difference between the actual values of Y and those estimated from the regression model represents the equation error term. Symbolically, the regression model may be written as

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_k + u_i.$$

What this model states is that regardless of the number of explanatory variables (the Xs) considered in the model, there is bound to be some variation in Y that cannot be explained by all the X variables. There is bound to be some intrinsic randomness in Y that cannot be explained, and this intrinsic randomness is represented by the STOCHASTIC error term, u_i. This error term may be due to several factors, such as vagueness of the theory underlying the regression model, unavailability of data on some explanatory variables, or use of a poor proxy variable (e.g., using the actual inflation rate instead of the expected inflation rate in models of unemployment). Sometimes, a researcher may know all the explanatory variables in a model but, following the principle of parsimony (or Occam's razor), may decide to include only a few core variables in the analysis.

The stochastic error term u_i plays an extremely critical role in regression analysis because the properties of the estimated regression coefficients—the β coefficients—depend critically on the assumptions about the stochastic error term. In the classical regression model, it is assumed that the values of the X, or explanatory, variables are fixed in repeated sampling, and the error term u_i is assumed to be Gaussian white noise. As a result, the estimated β coefficients also follow the normal distribution. This makes the task of hypothesis testing about the true, or population, β coefficients much simpler. Also, under the normality assumption about the error term u, the MAXIMUM LIKELIHOOD ESTIMATION (MLE) and ORDINARY LEAST SQUARES (OLS) estimators of the β coefficients are the same.

It should be noted that in practice one does not observe the error term u_i. Instead, what one observes is the RESIDUAL term $\hat{u}_i$, which is the difference between the actual value of Y_i and its value estimated from the regression model. In other words, $\hat{u}_i$ is a proxy for the true u_i. It can be shown that as the sample size increases indefinitely, the residuals tend to converge to their true values.

MODEL SPECIFICATION ERRORS

A theory may suggest the variables that may be included in a model, but the theory often is unable to suggest the specific functional form in which the variables should enter the model. Consider, for instance, the demand for a product (Y) in relation to its price (P) and the income of the consumer (I). Now consider two possible (regression) models:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 P_i + \beta_3 I_i + u_i, \\ \ln Y_i &= \alpha_1 + \alpha_2 \ln P_i + \alpha_3 \ln I_i + u_i, \end{aligned}$$

where u_i is the stochastic error term and "ln" stands for the natural logarithm.

The first model is a linear regression model, whereas the second is a log-linear regression model. Although both models contain the same variables, the way they enter the model makes a big difference in the interpretation of the model. In the linear model, for instance, β_2 (the partial slope coefficient) measures the rate of change in the mean value of Y with changes in the value of P, whereas in the log-linear model α_2 measures the *elasticity* of Y with respect to P; that is, it measures the percentage change in Y for a (small) percentage change in P. Obviously, the two measures are not the same.

ERRORS DUE TO MISSING OBSERVATIONS

Researchers often encounter the problem of MISSING DATA or missing observations. This is quite common in survey-type data and occurs when some respondents do not answer all the survey questions. In time-series data, one can encounter the problem of missing data in several ways. For example, data on some time series are available on a monthly basis (e.g., the unemployment rate), but for some other series the data are available only on a quarterly basis (e.g., those on the Gross Domestic Product, or GDP). Sometimes there are gaps in the data because of special circumstances. For example, during World War II, some data could not be collected.

If we ignore the data on the missing observations and do statistical analysis on whatever data are available, we are likely to obtain estimates of the parameters of interest that are subject to BIAS or are inconsistent. Consider, for example, the following simple regression model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i,$$

where Y is expenditure on leisure travel, X is personal income, and u is the stochastic error term. Suppose data are obtained on 1,000 individuals from a survey that asked them information about travel expenditure in the last 6 months. Suppose further that 100 individuals did not incur any travel expenditure. If one decides to do regression analysis on the basis of the data on 900 individuals only (a method referred to as using CENSORED DATA), it can be shown that the OLS estimators of the β coefficients are biased as well as inconsistent. The problem gets complicated if the income data on the 100 individuals is also unavailable (this is the case of truncation; see CENSORING AND TRUNCATION).

FORECASTING ERRORS

In regression models involving TIME SERIES data, one of the objectives is to use the estimated model to forecast the *future* value(s) of the dependent variables, given the values of the explanatory variables.

Suppose one has developed a regression model to explain the sales of cars in the United States using data, say, for the past 40 quarters. Assume that the model is a good one as judged on the usual statistical criteria. Suppose one wants to use this model to forecast car sales for the next 5 quarters. Assume that the data on the explanatory variables are available for the 5 quarters. The values of the car sales predicted for the next 5 quarters can be compared with the actual sales data, as information becomes available. No matter how good the model, there is no guarantee that the forecast values and the actual values of car sales will be identical; the difference is the *forecast error.*

A variety of forecasting techniques have been developed, all aiming to reduce the forecast error as far as possible. These range from the simple MOVING AVERAGE type of models to very sophisticated models such as ARIMA (autoregressive integrated moving average) and VECTOR AUTOREGRESSION (VAR). In developing these models, the white noise error term plays a crucial role. To evaluate the performance of the various forecasting models, researchers often use the Akaike information criterion (AIC) or the Schwarz information criterion (SIC). The lower the value of these criteria, the better is the model.

—Damodar N. Gujarati

REFERENCES

Darnell, A. C. (1995). *A dictionary of econometrics.* Cheltenham, UK: Edward Elgar.

Diebold, F. X. (2001). *Elements of forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Gujarati, D. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

ERROR CORRECTION MODELS

The error correction model (ECM) is a TIME SERIES regression model that is based on the behavioral assumption that two or more time series exhibit an equilibrium relationship that determines both short-run and long-run behavior. The ECM was first popularized in economics by James Davidson, David F. Hendry, Frank Srba, and Stephen Yeo in 1978. In 1987, Robert F. Engle and Clive W. J. Granger demonstrated that cointegrated time series are well represented by an error correction model. Since that time, the ECM has become associated with cointegrated time series. It is, however, important to note that the ECM is a general transformation of an autoregressive distributed lag (ADL) model. Specifically, the ADL model can be rewritten as an ECM such that there are no restrictions on the ADL parameters in the ECM representation. The ECM may thus be used to model equilibrium relationships involving stationary time series as well as cointegrated time series.

The ECM model is given by

$$\Delta y_t = \lambda_0 + \lambda_1 \Delta x_t - \gamma(y_{t-1} - \beta x_{t-1}) + \varepsilon_t,$$

where $\Delta y_t = y_t - y_{t-1}$, $\Delta x_t = x_t - x_{t-1}$, and γ, the error correction rate, gives the rate at which disequilibrium—$(y_{t-1} - \beta x_{t-1})$—is corrected. The term in parentheses may include additional independent variables as well. (See Banerjee, Dolado, Galbraith, and Hendry [1993] for a general discussion of ECMs.)

Assume we believe that consumers' sentiment is tied to public views of the president's ability to manage the economy: As people feel more positive about the president's management skills, they also feel more positive about the economy. In other words, we believe that there is a long-run equilibrium relationship between economic approval and consumer sentiment. If news coverage becomes increasingly critical of the president's economic policy such that the public reassesses its view of the president's managerial skill (downward), sentiment may be too high for current economic approval. In this case, the two series may be said to be out of equilibrium, and we would expect the public's economic sentiment to drop. We can capture this relationship in an ADL model or as an ECM. However, if we wish to know how long this adjustment would take—how quickly or slowly economic sentiment would react—we can estimate this effect directly only with the ECM. As parameterized above, the rate of error correction is given by γ. A γ of 0.2 indicates that 20% of the disequilibrium is corrected in the following time period, an additional 20% in the next term period, and so on, until equilibrium is restored. A γ of 0.8 implies a much quicker readjustment. The ECM thus directly estimates the error correction rate. In addition, because the ECM includes both long-run (levels) and short-run (changes) variables on the right-hand side of the model, we can capture the responsiveness of consumer sentiment to short-run changes in evaluations as well as long-run levels of evaluations. The ability to account for how high or low a time series is (how positive or negative evaluations are), as well as the direction it is moving, is an especially attractive feature of the ECM.

In practice, many models of equilibrium relationships are premised on permanent memory and cannot be estimated using a distributed lag model. In this case, the ECM typically is the model of choice. In addition to the single-equation formulation above, often analysts will rely on the Engle-Granger two-step method for estimating the ECM in the context of cointegration (Engle & Granger, 1987). In this case, the analyst estimates the equilibrium in a first-step regression and enters the first-stage model residuals (the disequilibrium) in the second-stage error correction model.

—Suzanna De Boef

REFERENCES

Banerjee, A., Dolado, J. J., Galbraith, J., & Hendry, D. F. (1993). *Cointegration, error correction and the econometric analysis of nonstationary series.* New York: Oxford University Press.

Davidson, J., Hendry, D. F., Srba, F., & Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal, 88,* 661–692.

Engle, R., & Granger, C. W. J. (1987). Co-integration and error-correction: Representation, estimation, and testing. *Econometrica, 55,* 251–276.

ESSENTIALISM

The term *essentialism* is used in philosophy and social science to refer to someone else's allegedly mistaken belief that he or she has accurately described the essential nature of something or someone, or at least that this should be the aim of philosophy or science. The traditional philosophical distinction between the essential natures of things and their appearances feeds into the beginnings of modern social science, with Karl Marx, for example, stressing that science would be unnecessary if things really were as they appeared to be, and that in the study of society the "power of abstraction" replaces procedures such as chemical analysis. Although some commentators commend Marx for essentialism, drawing parallels with Aristotelian philosophy, most see it as an error, whether or not they think he committed it. Those who take the latter view tend to prefer to call his position REALISM.

In recent decades, discussion of essentialism has mostly followed Karl Popper's critique, first advanced in *The Open Society and Its Enemies*, of what he called methodological essentialism: "the view . . . that it is the task of pure knowledge or 'science' to discover and to describe the true nature of things, i.e. their hidden reality or essence" (Popper, 1945/1966, p. 31). To this view, Popper opposes his "fallibilist" position that science is better understood in terms of the refutation of error, and in terms of "verisimilitude" rather than truth.

Popper also upheld a version of realism, but modern realists have mostly argued that his realism is too weak and his critique of essentialism mistaken. Rom Harré (1986, pp. 103–105) and Roy Bhaskar (1978), for example, have defended essences understood in a modest sense, along with notions of natural kinds and natural necessity, in a model of science seen as intrinsically open to refutation and further advance. Other realists, such as Andrew Sayer (1997), have defended realism against the accusation of essentialism and also argued in defense of moderate varieties of essentialism. Technical philosophical discussions of "dispositional essentialism" (Ellis & Lierse, 1994), the claim that a property such as being positively charged is *essentially* linked to a disposition to attract something with a negative charge, are echoed in social science, in Sayer's defense of realist abstraction against "associational thinking" based on regularities that may be merely accidental (Sayer 2000, 2001; see also Holmwood, 2001).

Recent feminist and antiracist critiques of essentialism ("all women/black people are [intrinsically] *xyz*") have sometimes drawn on deconstructionist or "poststructuralist" positions developed in the critique of essentialism in metaphysics and phenomenology. Other feminists, such as Fuss (1989), however, have criticized the dichotomy between essentialism and social constructionism. These substantive debates therefore intersect with methodological ones; Delanty (1997), for instance, has argued for a synthesis of realist and constructivist or relativist positions in the philosophy of social science.

—R. William Outhwaite

REFERENCES

Bhaskar, R. (1978). *A realist theory of science*. Hassocks, UK: Harvester Press.

Delanty, G. (1997). *Social science: Beyond constructivism and realism.* Buckingham, UK: Open University Press.

Ellis, B. D., & Lierse, C. (1994). Dispositional essentialism. *Australasian Journal of Philosophy, 72*(1), 27–45.

Fuss, D. (1989). *Essentially speaking.* New York: Routledge.

Harré, R. (1986). *Varieties of realism*. Oxford, UK: Blackwell.

Holmwood, J. (2001). Gender and critical realism: A critique of Sayer. *Sociology, 35,* 947–965.

Popper, K. (1966). *The open society and its enemies* (5th ed., Vol. 1). London: Routledge. (Original work published 1945.)

Sayer, A. (1997). Essentialism, social constructionism and beyond. *Sociological Review, 45*, 453–487.

Sayer, A. (2000). System, lifeworld and gender: Associational versus counterfactual thinking. *Sociology, 34*, 707–725.

Sayer, A. (2001). Reply to Holmwood. *Sociology, 35*, 967–984.

ESTIMATION

One major purpose of statistics is to make an inference about a POPULATION using information on a SAMPLE. Suppose that we are interested in studying the annual income of the adult population in a small American city of 100,000 inhabitants. Then the population would be formed by all the economically active adults in the city at the time of study. If we draw 100 adults randomly from the city, these 100 adults will be our sample. The annual incomes of the adult population in the city can be described by a variety of numerical measures called PARAMETERS. Examples of the parameters include the MEAN annual income of the adult population, the VARIANCE (a measure of variability) of the income, and the ASSOCIATION between the person's education and his or her income. The typical objective of statistical analysis is to make an inference about one or more population parameters. In this article, we consider estimation of population parameters. After a brief discussion of point and interval estimates, we will turn to the ordinary least squares ESTIMATOR and the maximum likelihood estimator. Throughout the discussion, we will use a simple numerical example to illustrate the basic ideas behind these estimators. Almost all textbooks on the subject provide a more general treatment.

POINT AND INTERVAL ESTIMATES

Statistical investigation of the annual income of the adult population in the city could give two types of estimate. The sample mean of \$45,500, for instance, could be given. The sample mean of $\bar{Y} = [\sum_{i=1}^{n}(Y_i)]/n$ is a single-number or point estimate. We hope that this estimate is not far from the population mean μ. Alternatively, the investigation could present a two-number estimate, for instance, of (\$42,000, \$49,000), meaning that the average annual income of the city population will be likely to fall between \$42,000 and \$49,000. This alternative form of estimate is called an interval estimate and is designed to enclose the population parameter under study. An interval estimate typically has a higher probability of including the population parameter than the point estimator. Both point and interval estimates have to be produced by an estimator, which is a procedure or method for calculating an estimate from a sample.

THE ORDINARY LEAST SQUARES ESTIMATOR

The ORDINARY LEAST SQUARES (OLS) estimator is often used to estimate the parameters in a LINEAR REGRESSION that is concerned with how one continuous dependent variable Y_i is related to independent variables $x_{1i}, x_{2i}, \ldots, x_{ki}$. Such an estimator could take the form

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + e_i. \quad (1)$$

When a linear regression has only one independent variable, it is a simple linear regression of the form

$$Y_i = \beta_0 + \beta_1 x_i + e_i. \quad (2)$$

Such a regression could be used to study how annual income is related to years of education in the city population. Figure 1 is a hypothetical graph that plots annual income against years of education for a sample from the city population. The plot shows a positive relationship between annual income and education. On average, those who have had more years of education tend to have higher incomes. Although many different mathematical functions can be used to model the association between annual income and education, statistical practice usually uses the linear model, such as equation (2). In this particular example, β_0 is the amount of annual income for those who have not received any formal education and β_1 is the amount of annual income associated with one additional year of education. Equation (2) then describes the relationship between annual income and education in the population. A statistical investigation makes use of a sample of the population to calculate $\hat{\beta}_0$ and $\hat{\beta}_1$, which are estimates of the population parameters β_0 and β_1.

Estimating $\hat{\beta}_0$ and $\hat{\beta}_1$ from the sample amounts to fitting a straight line through the set of data points in Figure 1. We know that no possible line can predict all data points perfectly because education does not predict income perfectly. We want the line to be positioned in such a way that it best represents the set of data points as a whole. We could try to fit the line by eye, but this works only approximately for a simple linear regression or a two-variable problem. More systematically, we resort to the OLS estimator that yields the fitted line

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i. \quad (3)$$

The vertical distance between the observed data point y_i and the fitted line represented by $\hat{y}_i$ (or $y_i - \hat{y}_i$) is referred to as deviation or prediction error. The basic idea of the OLS estimator is to select $\hat{\beta}_0$ and $\hat{\beta}_1$ so that we minimize the sum of all the deviations, $\sum_{i=1}^{n}(y_i - \hat{y}_i)$. However, some data points lie above the fitted line and some lie below the line. The positive errors above the line tend to cancel the negative errors below the line, leaving the sum of all the deviations much smaller than it should be. The standard solution to this problem is to square each error and then minimize the sum of these squared errors (SSE):

$$\mathrm{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2.$$

The values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that correspond to the minimum of SSE can be obtained by solving the equations $\partial\mathrm{SSE}/\partial\hat{\beta}_0 = 0$ and $\partial\mathrm{SSE}/\partial\hat{\beta}_1 = 0$. Taking the partial derivatives of SSE with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$, and setting the results to zero, we obtain

$$\partial\mathrm{SSE}/\partial\hat{\beta}_0 = -2\left(\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_i\right) = 0$$

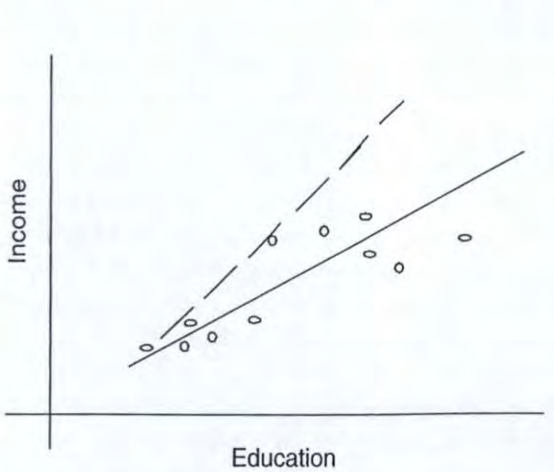

Figure 1 Scatterplot of Income Against Education and the Fitted Line

Table 3 Calculations for the Ordinary Least Squares Fitted Line

Individual i	y_i	x_i	$x_i y_i$	x_i^2
1	35	12	420	144
2	57	18	1026	324
3	26	11	286	121
4	45	12	540	144
5	58	14	812	196
6	40	12	480	144
7	55	16	880	256
8	48	15	720	225
Total	$\sum y_i = 364$	$\sum x_i = 110$	$\sum x_i y_i = 5164$	$\sum x_i^2 = 1554$

and

$$\partial\mathrm{SSE}/\partial\hat{\beta}_1 = -2\left(\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2\right) = 0.$$

These two equations are the OLS equations and can be solved simultaneously for $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\,\bar{x}.$$

Suppose that our sample consists of eight individuals with annual income 35, 57, 26, 45, 58, 40, 55, and 48 (in thousands of dollars) and years of education 12, 18, 11, 12, 14, 12, 16, and 15, respectively. Table 1 illustrates the OLS calculations for the parameters $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\hat{y}_i = -7.18 + 3.83x_i. \quad (3)$$

The OLS estimate of the slope thus shows that each year of education is associated with \$3,830 additional annual income.

THE MAXIMUM LIKELIHOOD ESTIMATOR

MAXIMUM LIKELIHOOD ESTIMATOR (MLE) is behind most of the statistical models used in social sciences, such as logistic regression, probit regression, multinomial regression, Poisson regression, negative binomial regression, a variety of event history models, growth curve models, and multilevel linear models. The EM algorithm, which gained popularity over the past 15 years, is essentially a computing method for maximum likelihood estimation. Even the parameter estimates of the OLS regression discussed earlier can be considered the estimates of the MLE under the assumption that the data have a normal distribution.

Suppose that we want to know the mean annual income of the adult population in our small city. A feasible way of estimating this is to draw and work with a random sample. Suppose that our sample consists of eight individuals with incomes as specified earlier (the MLE should really apply to large samples; the small sample of eight is meant to simplify the illustration). To proceed with the MLE, we must specify the DISTRIBUTION of the data. By far the most assumed distribution for linear data is the NORMAL DISTRIBUTION. Thus, we assume that each income of the eight individuals comes from a normal distribution: $Y_1 \sim N(\mu, \sigma^2), \ldots, Y_8 \sim N(\mu, \sigma^2)$ with density

$$f(y_i) = 1/\sqrt{2\pi\,\sigma^2}\, e^{-[(y_i-\mu)^2/2\sigma^2]}.$$

The parameter μ is not simply the average annual income. Rather, it is the mean of the normal distribution representing the annual income in the population.

The next step in the MLE is to write down the likelihood function of the data. The likelihood function for individual i is the density for the individual viewed as a function of the unknown parameters:

$$L_i(\mu; y_i) = 1/\sqrt{2\pi\,\sigma^2}\, e^{-[(y_i-\mu)^2/2\sigma^2]}.$$

Assuming that the individual observations in the sample are independent, we can write down the joint likelihood function for the whole sample as

$$L(\mu; y_1, y_2, \ldots, y_8) = \prod_{i=1}^{8} L_i = \left(1/\sqrt{2\pi\sigma^2}\right)^8 e^{-\left[\sum_{i=1}^{8}(y_i-\mu)^2/2\sigma^2\right]}.$$

The MLE is that value of the μ that maximizes the joint likelihood function $L(\mu, y_1, y_2, \ldots, y_8)$. In practice, however, it is much easier to maximize Ln $L(\mu)$, the natural logarithm of $L(\mu)$, than to maximize $L(\mu)$. The quantities associated with the likelihood function $L(\mu)$ often become too small or large for computing. The justification for this is that whatever maximizes the log likelihood must also maximize the likelihood. With the logarithm, $L(\mu)$ is simplified to

$$\text{Ln } L(\mu) = 8\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{\sum_{i=1}^{8}(y_i-\mu)^2}{2\sigma^2}.$$

Finding the MLE value requires taking the partial derivative of the log likelihood function with respect to μ, setting the derivative equal to zero, and then solving the equation for the unknown parameter:

$$\frac{\partial \text{ Ln } L(\mu)}{\partial \mu} = \frac{\sum_{i=1}^{8} 2(y_i-\mu)}{2\sigma^2} = 0$$

$$\begin{aligned}\hat{\mu} &= \sum_{i=1}^{8} y_i/8 = \bar{y}\\ &= \frac{35+57+26+45+58+40+55+48}{8}\\ &= 45.5.\end{aligned}$$

Using this result, we can obtain the MLE for σ^2 in a similar way:

$$\frac{\partial \text{ Ln}L(\mu)}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{8}(y_i-\mu)^2 = 0,$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{8}(y_i-\bar{y})^2.$$

Thus, the maximum likelihood estimators of the mean and the variance of the normal distribution turn out to be the corresponding sample characteristics. In simple models, the MLE is often the "obvious" estimator; however, the obviousness usually is lost when the model becomes slightly more complicated. The obviousness of the MLE in the simplest cases may be considered reassuring because we probably want to see the MLE in the simplest cases as simple and easily understood.

Maximum likelihood estimation is a method with an intuitive appeal. Apart from the obviousness of MLE in the simplest cases, the ML estimator is the population estimate that is most likely to produce the observed sample. More fundamentally, the popularity of MLE is due to many of the attractive ASYMPTOTIC PROPERTIES (large sample properties) that it possesses. Under broad conditions, an ML estimate is consistent, efficient, and normally distributed in its sampling distribution, with clearly defined and easily computed mean and variance.

The estimator $\hat{\theta}_n$ based on a sample of size n is said to be a consistent estimator of the population parameter θ if for any positive number ε, $\lim_{n\to\infty} P(\_\,\hat{\theta}_n - \theta\,\_ > \varepsilon) = 0$. The maximum likelihood estimator is consistent. This result says that the estimates of an ML estimator would get closer and closer to the population parameter as the sample size gets larger and larger and that the estimates would converge to the population parameter when the sample size approaches positive infinity. The consistency result is intuitive. Let's consider coin tossing. The coin is tossed n times with the unknown probability (p) of resulting in a head. If the tosses are independent, the number of heads Y in n tosses has a binomial distribution. The sample proportion Y/n is an estimator of p. It is not difficult to understand that as the number of tosses n increases, our estimates of p would get closer to the actual value of p, and that as the number of tosses approaches positive infinity, our estimates would converge to the actual p.

The efficiency of an estimator is indicated by the amount of variability associated with the estimates of the estimator. A population parameter θ may be estimated by two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$. The two estimators represent two different mathematical procedures. Given a sample, the two procedures would generate two estimates. Given a large number of samples of size n, each procedure would generate a set of estimates. Suppose both sets of estimates have a mean equal to θ, meaning that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased. Then if the variance of the estimates associated with $\hat{\theta}_1$ is smaller than the variance of the estimates associated with $\hat{\theta}_2$, we say that the estimator $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$. The well-known result by Rao-Cramér provides a lower bound for every unbiased estimator of θ. If an unbiased estimator has a variance equal to the Rao-Cramér lower bound, we know that no unbiased estimator has a smaller variance and that the estimator under consideration is the unbiased minimum variance estimator. Under fairly general conditions, it can be proved that the ML estimator $\hat{\theta}$ of θ has an approximate normal distribution with mean θ and a variance equal to the Rao-Cramér lower bound. Therefore, at least approximately $\hat{\theta}$ is an unbiased minimum variance estimator. Therefore, the ML estimator can be considered an efficient estimator in the sense that it will make good use of the sample collected.

The last property of the ML estimator we consider here is perhaps the most useful. In most regular cases, it can be shown that the ML estimator based on large samples has a normal distribution with mean θ and the variance $[I(\theta)]^{-1}$, where

$$I(\theta)^{-1} = \left(-E\left[\frac{\partial^2 \ln L(\theta)}{\partial\theta^2}\right]\right)^{-1}$$

when θ has a single parameter.

The variance can be evaluated at $\hat{\theta}$ when the expected value is available. In practice, the expected value of the second derivative of the log likelihood will seldom be available. In such cases, the empirical information matrix rather than the information is often computed:

$$\left(-\frac{\partial^2 \ln L(\hat{\theta})}{\partial\hat{\theta}^2}\right)^{-1}.$$

This result holds whether or not the data have a normal distribution. This result is the basis of the z-test in regression analysis that is so prevalent in social science research.

BAYESIAN ESTIMATION

Bayesian estimation has found much use in recent years in the mainstream statistical and biostatistical literature. A classical estimator such as the ordinary least squares estimator or the maximum likelihood estimator is determined entirely by the sample. The Bayesian estimation, in contrast, draws information from both the sample and the PRIOR DISTRIBUTION. When additional data become available, one updates one's prior distribution. The Bayesian estimation incorporates one's prior knowledge on the problem and is more subjective. For large samples and regular models, the prior distribution usually is overwhelmed by the amount of data and the Bayesian estimation yields the same estimates as does maximum likelihood estimation. Thus, Bayesian estimation is not a general substitute for maximum likelihood estimation.

There are, however, at least two situations in which Bayesian estimation can play an important role. The first is when the number of observations is small and the information about each observation is abundant. Studies in macrosociology sometimes belong to this category. When a sample consists of countries, the number of countries in the sample is necessarily limited, but the background information about each country can be huge. In such cases, taking advantage of prior knowledge about each country—that is, information outside the observed data—may contribute to the analysis. The second situation is when the models are too difficult to estimate by classical statistical approaches. Examples include nonlinear multilevel models. The log likelihood of these models frequently is involved, with multiple integrals that do not have a closed expression. MARKOV CHAIN MONTE CARLO (MCMC) METHODS developed within the Bayesian perspective bypass the problem by drawing samples from the required distribution and then describing the sample averages. There are other estimation methods not discussed here, for example, the method of moment estimation and indirect estimation techniques for demographic data.

—Guang Guo

REFERENCES

Greene, W. H. (1997). *Econometric analysis* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). New York: Macmillan.

Lee, P. M. (1997). *Bayesian statistics: An introduction* (2nd ed.). London: Arnold.

ESTIMATOR

An estimator, also known as a (sample) STATISTIC, is a rule, method, formula, or procedure that tells how one can estimate a POPULATION quantity, such as the population mean, population variance, or population CORRELATION coefficient. It generally is expressed as a function of sample values. The numerical value taken by an estimator in a given sample is known as an estimate. Because its value will differ from sample to sample, an estimator is a RANDOM VARIABLE and will have a PROBABILITY or SAMPLING DISTRIBUTION. The sampling distribution of an estimator is of great importance in assessing the reliability of the estimate in relation to its population value. All of this can be explained as follows.

Let X be a random variable with a PROBABILITY DENSITY FUNCTION (PDF) of $f(X)$. For simplicity of exposition, assume that this PDF has two PARAMETERS, μ_x (population mean) and σ_x^2 (population variance). Based on a RANDOM SAMPLING of n observations from this PDF, suppose one wants to estimate the mean value, μ_x. The most commonly used estimator of the population mean is the SAMPLE MEAN, defined as $\bar{X} = \frac{1}{n}\sum_{i=1}^{i=n} X_i$. Thus $\bar{X}$ is a rule or formula that says that to obtain the sample mean value, one simply adds all the values of X in the sample and divides the sum by n, the number of observations in the sample. Thus, $\bar{X}$ is an estimator of μ_x. (Besides the sample mean, the sample MODE and MEDIAN are the other common estimators of CENTRAL TENDENCY.)

If one is interested in estimating the population variance, the most commonly used estimator is the SAMPLE VARIANCE, denoted by S_x^2, which is defined as follows:

$$S_x^2 = \frac{1}{(n-1)}\sum_{i=1}^{i=n}(X_i - \bar{X})^2.$$

According to this formula, to estimate the sample variance, subtract the mean value of X from each individual value of X, square the difference, sum these squared differences, and divide the sum by DEGREES OF FREEDOM $(n-1)$. The sample variance thus defined is an UNBIASED estimator of the true σ_x^2.

The formulas for computing the sample mean and sample variance just given provide what are called the POINT ESTIMATES. That is, for a given sample they will provide a single (point) estimate. Because an estimator is a random variable, as its value will vary from sample to sample, how reliable is a point estimate of its true population value? In statistics, the reliability of a point estimator is measured by its STANDARD ERROR. Therefore, instead of relying on the point estimate alone, one may construct an interval around the point estimator, say within two or three standard errors on either side of the point estimator, such that this interval has, say, a 95% probability of including the true parameter value. The interval thus established provides the so-called *interval estimator.*

Knowing that the variance of X is σ_x^2, statistical theory shows that the variance of

$$\bar{X} = \frac{\sigma_x^2}{n}.$$

Therefore, the standard error of $\bar{X}$, which is the square root of the variance of $\bar{X}$, is

$$\frac{\sigma_x}{\sqrt{n}}.$$

Statistical theory also shows that if the number of observations is fairly large, the sample mean $\bar{X}$ approaches the NORMAL DISTRIBUTION with mean μ_x and variance $\frac{\sigma_x^2}{n}$ (this is as per the CENTRAL LIMIT THEOREM). If the random variable X is from a normal population to begin with, $\bar{X}$ is normally distributed with the stated parameters regardless of the sample size. In other words, for large samples,

$$\bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right);$$

that is, $\bar{X}$ is normally distributed with the stated mean and variance. Thus, the probability or sampling distribution of the sample mean is normal.

Using the properties of the normal distribution, it then follows that the interval

$$\left(\bar{X} + 1.96\frac{\sigma_x}{\sqrt{n}}\right)$$

provides a 95% confidence interval for the population mean. If 100 intervals like this are established, chances are that 95 out of 100 of them will include the true μ_x. The preceding interval is thus an interval estimator of the true mean.

There are several methods of obtaining a point estimator, the prominent ones being the method of ORDINARY LEAST SQUARES (OLS) and the method of MAXIMUM LIKELIHOOD ESTIMATION (ML). Of the two, the method of OLS is mathematically simpler to use. To illustrate this, consider the following MULTIPLE REGRESSION:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i,$$

where Y is the dependent variable, the Xs are the explanatory variables, and u_i is the stochastic error term. To estimate the model, given the data, OLS proceeds as follows: Rewriting the preceding equation, we obtain:

$$u_i = Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i} - \cdots - \beta_k X_{ki}.$$

Thus, u_i represents the difference (or error) between the actual Y values and their estimated values from the multiple regression. Instead of minimizing the sum of these errors, OLS minimizes the squares of these errors. That is, it minimizes

$$\sum u_i^2 = \sum (Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i} - \cdots - \beta_k X_{ki})^2.$$

Because the values of Y and the X variables will be provided by the sample at hand, the preceding error sum of squares will depend on the values taken by the β coefficients. OLS tries to find those values of the β coefficients that will make $\sum u_i^2$ as small as possible. The actual algebraic formulas to compute the k beta coefficients are rather tedious. Using matrix algebra, the OLS estimators can be shown as follows:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

where $\hat{\beta}$ is a $(k \times 1)$ vector of the k beta coefficients, $\mathbf{X}$ is an $(n \times k)$ data (or observation) matrix, and $\mathbf{y}$ is an $(n \times 1)$ vector of the observations on the dependent variable; the ˆ, called a hat or cap, denotes an estimator. Although β (vector) is unknown but fixed, $\hat{\beta}$ (vector) is random because its value will depend on the sample-specific values of the Y and the X variables. It can be further shown that the OLS estimator of the error variance is given by

$$\hat{\sigma}_u^2 = \frac{\sum \hat{u}_i^2}{n-k},$$

where $\hat{u}_i$ is the estimated error term and where $(n-k)$ represents the number of degrees of freedom, which is equal to the sample size n minus the number of parameters estimated in the model.

In choosing among competing estimating methods, one can consider the statistical properties of the estimators obtained from the various methods. The commonly used statistical properties are linearity, UNBIASEDNESS, minimum variance, EFFICIENCY, sufficiency, asymptotic unbiasedness (see ASYMPTOTIC PROPERTIES), asymptotic efficiency, and consistency. Not every estimation method will satisfy all these statistical properties, especially if the sample size is small.

For example, if we assume that the multiple regression considered above satisfies the assumptions of the classical LINEAR REGRESSION model, it can be proved that the OLS estimators are linear, are unbiased, and have minimum variance in the class of linear unbiased estimators; they are often described as BEST LINEAR UNBIASED ESTIMATORS (BLUE). (The classical linear regression model assumes that the values taken by the explanatory variables are fixed in repeated sampling, that the error terms u_i are not correlated, and the error variance of u_i, σ_u^2, is constant or HOMOSKEDASTIC.) It can further be shown that the OLS estimator of error variance given above is an unbiased estimator of the true σ_u^2.

If, in addition to the assumptions of the classical linear regression model being met, it is assumed that the error term u_i is normally distributed, then it can be shown that the maximum likelihood (ML) estimators of the β coefficients are identical to the OLS estimators. However, the ML estimator of the error variance of u_i is

$$\sigma_u^2 = \frac{\sum \hat{u}_i^2}{n},$$

which is different from the OLS estimator given above. As noted previously, the OLS estimator of the error variance is unbiased, which means that the ML estimator of the error variance is biased. The difference between the two estimators is due to the fact that the OLS estimator takes into account the degrees of freedom, whereas the ML estimator does not. Of course, as the sample size n increases indefinitely, the two estimators will converge. In other words, asymptotically the ML estimator of the error variance is unbiased.

—Damodar N. Gujarati

REFERENCES

Gujarati, D. N. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

Kendall, M. G., & Buckland, W. R. (1971). *A dictionary of statistical terms* (3rd ed.). Edinburgh: Oliver & Boyd.

Vogt, P. W. (1993). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences.* Newbury Park, CA: Sage.

ETA

In the social sciences, the term eta (Greek symbol η) probably is best known as it is used by Hays (1988) and Kirk (1995). The term does not have any particular meaning in the context of basic statistical training or theory. For example, eta is not found in the dictionary of statistical terms by Kendall and Buckland (1982). A search of all published journal articles in the *Current Index to Statistics*, for the years 1980 through 2002, based on the keyword eta (or eta-squared), turned up zero papers. Hays and Kirk use the term in the context of the ANALYSIS OF VARIANCE, but several recent texts devoted to this topic include no mention of eta. Some use the term eta to refer to a measure of effect size that is a function of the Student's t-statistic, and it has played a role in structural equation models. Sometimes eta-squared is used in regression and refers to the COEFFICIENT OF multiple DETERMINATION, which is more commonly labeled R-SQUARED.

Consider a standard ANALYSIS OF VARIANCE (ANOVA) model where the means of J independent groups are to be compared assuming normality and that all J groups have a common variance. The term eta-squared refers to a measure of association intended to reflect any nonlinear correlation between an independent and a dependent variable. For example, imagine that five different amounts of some medication are under investigation regarding their ability to treat depression. For illustrative purposes, assume that 10, 20, 30, 40, and 50 units of the drug are being studied. Of interest is the mean of some outcome measure that reflects depression. If the means of the five groups differ, one possibility is that with increasing amounts of medication, the means increase in a linear fashion. Another possibility is a nonlinear association in which, for example, the means increase by the same amount when increasing the amount of the drug from 10 to 20, and from 20 to 30, but then the means decrease when moving from 30 to 40 and 40 to 50. The term eta-squared refers to a type of correlation intended to reflect this, or any, nonlinear association. Generally, it is meant to reflect any departure from a linear trend.

In ANOVA, two standard quantities are the SUM OF SQUARES between groups (SSBG), which reflects the variation among the sample means, and the total sum of squares (SSTOT). The estimate of eta is the square root of SSBG divided by SSTOT. There is also a quantity that reflects the linear association that is based on a particular linear contrast of the means; it is essentially a variation of Pearson's correlation, r. If there is a linear association, then eta and r should have similar values, so a simple check for a nonlinear association is to compare these two quantities. A detailed illustration can be found in Kirk (1995, pp. 197–198).

—Rand R. Wilcox

REFERENCES

Hays, W. L. (1988). *Statistics.* Fort Worth, TX: Holt, Rinehart and Winston.

Kendall, M. G., & Buckland, W. R. (1982). *A dictionary of statistical terms.* New York: Longman.

Kirk, R. E. (1995). *Experimental design.* Pacific Grove, CA: Brooks/Cole.

ETHICAL CODES

A code of ethics specifies the proper conduct of members of a particular group. This article focuses on codes of ethics in the social sciences. Codes are controversial. They ignore the duty of a professional to simply be moral and are rarely used for guidance by practicing professionals. Codes testify to the claim that the profession recognizes special obligations to society that transcend economic self-interest or normal standards of morality. They help moral professionals resist pressures to compromise as well as legitimizing one's objections to poor work by other professionals and enhancing the profession's reputation and working environment. Professional societies may use them to censure or expel misbehaving members. Violation of ethics codes is not, per se, legally punishable, but it may be admissible as evidence in some legal proceedings. The role of codes is best understood in historical context.

HISTORY

Until about 1800, professionals were considered gentlemen who needed no written rules of behavior. Mere suggestion that a professional was dishonorable might result in law suits, pamphlet wars, or even duels. The inappropriateness of such idiosyncratic standards was recognized in 1792, when physicians at the Manchester Infirmary waged a pamphlet war and work stoppage during the outbreak of an epidemic. Outraged hospital trustees requested that Dr. Thomas Percival draft rules to prevent future breakdowns of professional morality. Percival's rules, published in 1794, listed specific duties of physicians, asserting the importance of common professional standards and establishing the moral independence of physicians over authorities who might pressure them to lower standards of medical care. In 1847, the American Medical Association was the first national professional society anywhere to adopt Percival's rules and call them a "code of ethics."

GROWTH AND DEVELOPMENT OF CODES IN THE SOCIAL SCIENCES

The American Psychological Association published the first social science code in 1953, following the now-familiar practice of convening a committee, gathering descriptions of ethical dilemmas encountered by members, and developing, through consensus, appropriate rules. Currently, there are codes of ethics for every branch of social science, as well as for subareas within them, such as public opinion research, and clinical hypnosis. Changes within society (e.g., emphasis on privacy), regulations (e.g., institutional review boards), professional roles, and research topics require frequent review and revision of codes. Typically, a professional association's ethics committee publishes articles describing emerging issues, suggests approaches to resolving them, and seeks membership input. Drafts of the emerging code are revised and published until membership consensus is reached.

USING CODES

Because codes of ethics are created in response to anticipated ethical conflicts, they are best understood and interpreted in the context of real-life ethical ambiguity. Case studies of ethical dilemmas are effective instructional tools for teaching the use of codes and are now an important part of each discipline's professional literature.

A voluminous literature on codes of ethics now exists. Collections of codes, annotated bibliographies, casebooks, research guidelines, methodology integrating tenets of valid research and ethics, as well as guidance in authoring codes and teaching their use, can be found at the URL of the Center for the Study of Ethics in the Professions, Illinois Institute of Technology (www.iit.edu/departments/csep/PublicWWW/codes/). The archives on codes of ethics, including their history, construction, use, and examples across the professions, are housed at the same Center for the Study of Ethics in the Professions.

—Joan E. Sieber

ETHICAL PRINCIPLES

Ethical discourse is conducted at various levels of abstraction. Moving from higher to lower levels of abstraction, there are ethical theories, principles, rules or norms, codes, and judgments. The nature and role of ethical principles is best understood in relation to other elements in this hierarchy.

Judgments are conclusions about a particular action. An example is the following:

> I don't think we should tell the 5-year-olds in our study of resistance to temptation that we observe whether they steal the money we leave around, because children that age readily take attractive items without realizing that they are doing something wrong. Rather, we should chat with them and make sure they feel OK about their actions and make sure their parents understand relevant aspects of child development and of our study.

This very concrete and specific judgment can be justified on the basis of more abstract and general rules, principles, and theories.

A *code of ethics* specifies the proper conduct of members of a particular group (see ETHICAL CODES). That is, a code is a set of rules or norms of a particular group. The code recognizes the group's special obligations to society—obligations that transcend normal standards of morality. For example, the code of ethics of the American Psychological Association requires appropriate debriefing and also justifies the judgment and norm discussed above concerning debriefing of young participants in deception research. It states that

> (a) Psychologists provide a prompt opportunity for participants to obtain appropriate information about the nature, results, and conclusions of the research, and psychologists attempt to correct any misconceptions participants may have.
>
> (b) If scientific or humane values justify delaying or withholding this information, psychologists take reasonable measures to reduce the risk of harm. (American Psychological Association, 1992, p. 1204)

Note that this statement is more generalized than the rule or norm stated above. It invites judgment about appropriate interpretation of the code.

Rules or norms often are informal and unwritten but generally are understood as general statements concerning what is right and wrong. For example, a researcher might justify the above judgment with the following rule or norm: "Researchers should debrief subjects, but should not do so in a way that would unnecessarily harm or worry them or their parents."

A *principle* is an even more generalized rule, such as "respect research participants," "be fair," "prevent harm," and "promote benefit." It states ideals, ignores conflicts that may arise between principles, and gives no instructions as to *how* one is to carry out these ideals. For example, the American Psychological Association sets forth Principle F as follows:

> Psychologists are aware of their professional and scientific responsibilities to the community and the society in which they work and live. They apply and make public their knowledge of psychology in order to contribute to human welfare. Psychologists are concerned about and work to mitigate the causes of human suffering. When undertaking research, they strive to advance human welfare and the science of psychology. Psychologists try to avoid misuse of their work. (American Psychological Association, 1992, p. 1600)

Ethical theories are systematically related bodies of principles and rules. The following are highly simplified descriptions of some major types of ethical theories. *Utilitarian* (consequential) theories hold that one should try to do whatever promises to maximize happiness in the world. *Deontological* ethics (Immanuel Kant's version) holds that one has specific duties that must be carried out, quite apart from their supposed consequences, because these duties are always morally right unless superseded by some higher principle such as saving a life. These might include duties such as truth-telling or treating people as ends in themselves rather than as means to ends. Similarly, some deontologists might hold that certain acts, such as killing a fetus, are always wrong. *Theist* ethics hold that one should do what will please God.

As these simple definitions and examples illustrate, different ethical theories would justify different principles, rules, and judgments; hence, ethicists may disagree about what is a morally or ethically correct course of action. Additionally, the same theories and principles may be interpreted differently by different persons or in different contexts.

Disagreements, exceptions, and conflicts. There are always exceptions to ethical theories, principles, and rules. For example, regarding exceptions to theories, deontologists who hold that killing another person is always wrong might argue that there is such a thing as a "just war" in which killing is justified, and deontologists who require truth-telling might permit lying to save an innocent person. Ethical principles are often found to be in conflict with one another or subject to exceptions. A major role of ethical theories is to help resolve conflicts between principles. For example, the hypothetical investigator who studied resistance to temptation by 5-year-olds reconciled the requirement to debrief honestly with the requirement not to cause harm, looking to the utilitarian principle of beneficence—the researcher's obligation to maximize possible benefit and minimize possible harm—for justification.

As this introduction to ethical principles and their place in the hierarchy of ethical discourse is intended to convey, there are both many ways to talk about ethics and many different positions one might take. Seemingly incomprehensible differences of opinion often can be understood by identifying the kind of theory on which opinions are based. For example, persons who are horrified at research practices that involve some degree of concealment or deception seem to be taking a deontological position. Thus, cultural anthropologists who observe the behavior of another without revealing the critical stance that will be taken in the scientific discussion of that behavior might be regarded as doing wrong because they are treating subjects as means to the end of learning about human behavior, and not as ends in themselves. A utilitarian, on the other hand, might argue that humankind is much better off understanding itself, even at the cost of minor deception of individual subjects; the end justifies the means. The ethical principles set forth under the *Belmont Report* and promulgated via the Federal Regulations for Protection of Human Subjects of Research are a mixture of utilitarian and deontological ethics.

One of the most difficult ethical problems for investigators to resolve—as they seek to conduct risky research that promises to yield highly important knowledge—arises when that research also requires some compromise of a deontological principle by resorting to a utilitarian principle. For example, research on minors who inject drugs may promise great benefit for society because it may further society's ability to curtail such drug use. However, it may be impossible for the researcher to obtain from parents their permission to conduct the research; the parents may be unaware of their child's drug use or may be dependent on their child's drug dealing for income.

Ethical principles have the advantage of being broad and general, hence of providing flexible guidelines for considering a wide range of issues. By the same token, they are vague on specifics and may be easily misinterpreted. With this proviso in mind, we turn to the "Belmont Principles" set forth by the U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (National Commission) in 1978. The National Commission had, as part of its charge, the mission to set forth basic ethical principles that should govern human research. After much deliberation, it set forth three principles that support six generally agreed upon norms of scientific research. The ethical principles are as follows:

a. Beneficence—maximizing good outcomes for science, humanity, and the individual research participants while avoiding or minimizing unnecessary risk, harm, or wrong

b. Respect—protecting the autonomy of (autonomous) persons, with courtesy and respect for individuals as persons, including those who are not autonomous (e.g., infants, mentally retarded persons, or persons afflicted with senility)

c. Justice—ensuring reasonable, non-exploitative, and fully considered procedures and their fair administration, along with fair distribution of costs and benefits among persons and groups (i.e., those who bear the risks of research should be those who benefit from it)

The six norms of scientific behavior, with letters in parentheses indicating the principles on which each norm is based, are as follow:

1. *Valid research design.* Valid research takes account of relevant prior findings, methods, and theory. Invalid research yields useless results, disrespects subjects, and wastes resources (A, B).

2. *Competence of researcher.* The researcher must be capable of carrying out research validly (A, B).

3. *Identification of consequences.* Risks and benefits should be identified. Procedures must be adjusted to respect privacy, ensure confidentiality, maximize benefit, minimize risk, and meet other goals (A, B, C).

4. *Selection of subjects.* Subjects must be appropriate to the study, representative of the population that will benefit, and appropriate in number (A, B, C).

5. *Voluntary informed consent.* Voluntary informed consent should be obtained (A, B, C).

6. *Compensation for injury.* The investigator is responsible for harm to subjects. Subjects must be informed whether harm will be compensated (A, B, C).

The Belmont Principles reflect the individualism and pragmatism of American culture. A more paternalistic or communitarian culture might place less emphasis on respect for autonomy and greater emphasis on beneficence. A more theocratic culture might discourage research on humans if it held that all behavior is according to God's will. Aside from such extreme moral viewpoints, the National Commission might have chosen a somewhat different set of fundamental principles. Many have observed that the current emphasis on respect for autonomy leaves many research participants with decisions they are ill-equipped or unwilling to make. Consequently, many subjects consent or refuse to participate without understanding (or even reading) the informed consent statement. Would a set of principles based more on caring or professional responsibility of the investigator to do no harm be more ethical? There are no easy answers.

—Joan E. Sieber

REFERENCES

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47,* 1597–1611.

National Commission for Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research* (DHEW Publication No. (OS) 78-0012). Washington, DC: Government Printing Office.

Solomon, W. D. (1978). Ethics: Rules and principles. In W. T. Reich (Ed.), *Encyclopedia of bioethics.* New York: Free Press.

ETHNOGRAPH

Ethnograph is a software program designed for computer-assisted qualitative data analysis. As such, it allows qualitative data to be coded and retrieved. It also supports GROUNDED THEORY practices, such as the generation of MEMOS.

—Alan Bryman

ETHNOGRAPHIC CONTENT ANALYSIS

Ethnographic content analysis (ECA) refers to an integrated method, procedure, and technique for locating, identifying, retrieving, and analyzing documents for their relevance, significance, and meaning (Altheide, 1987, 1996). The emphasis is on discovery and description, including search for contexts, underlying meanings, patterns, and processes, rather than mere quantity or numerical relationships between two or more variables (Altheide, 1996).

A document is defined as any symbolic representation and meaning that can be recorded and/or retrieved for analysis. Document analysis will expand as recording technologies improve and become more accessible. These technologies include those of print and electronic media, audio tapes, visuals (e.g., photos, home videos), clothing/fashion, Internet materials, information bases (e.g., Lexis/Nexis), and fieldnotes.

ECA or qualitative document analysis involves emergent and theoretical sampling (Glaser & Strauss 1967) of documents from information bases (including those developed by a researcher, such as FIELDNOTES), development of a protocol for more systematic analysis, and then comparisons to clarify themes, frames, and discourse. For example, if one is interested in studying TV violence, it is not an act of violence per se that is socially significant, but rather how that act is linked to a *course of action* or a scenario as part of an *entertainment emphasis* (e.g., "bad guys get shot by good guys in order to achieve justice"). Alternatively, the scenario might be that the use of violence is somehow linked to bravery, cunning, skill, or (of course) sex. The latter are themes or general messages that are reiterated in specific scenarios. The aim, then, is to query how behavior and events are placed in context, and what themes, frames, and discourse are being presented. The basic steps include:

- Pursuing a specific problem to be investigated
- Becoming familiar with the process and context of the information source (e.g., ethnographic studies of newspapers or television stations) and exploring possible sources (perhaps documents) of information
- Becoming familiar with several (6–10) examples of relevant documents, noting particularly the format, and selecting a unit of analysis (e.g., each article), recognizing that the unit of analysis may change
- Listing several items or categories (variables) to guide data collection and drafting a protocol (data collection sheet)
- Testing the protocol by collecting data from several documents
- Revising the protocol and selecting several additional cases to further refine the protocol

A dynamic use of ECA involves "tracking discourse," or following certain issues, words, themes, and frames over a period of time, across different issues, and across different news media. Initial manifest coding incorporates emergent coding and theoretical sampling in order to monitor changes in coverage and emphasis over time and across topics. For example, in a study of "fear" (Altheide, 2002), a protocol was constructed to obtain data about date, location, author, format, topic, sources, theme, emphasis, and grammatical use of fear (as noun, verb, or adverb). The contexts for using the word "fear" were clarified through theoretical sampling and CONSTANT COMPARISON to delineate patterns and thematic emphases. Materials were enumerated, charted, and analyzed qualitatively using a word processor and a qualitative data analysis program—NUD*IST—as well as quantitatively.

—David L. Altheide

REFERENCES

Altheide, D. L. (1987). Ethnographic content analysis. *Qualitative Sociology, 10*, 65–77.

Altheide, D. L. (1996). *Qualitative media analysis.* Newbury Park, CA: Sage.

Altheide, D. L. (2002). *Creating fear: News and the construction of crisis.* Hawthorne, NY: Aldine de Gruyter.

Glaser, B. G., & Strauss, A. L. (1967). *Discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

ETHNOGRAPHIC REALISM

Ethnographic realism holds that ETHNOGRAPHY can and should strive to document sociocultural structures, processes, and situations as existing independently of the researcher.

For much of its history, ethnography has been committed to REALISM and to methods of investigation based on that commitment. For example, anthropologists have long argued that other societies must be studied through close PARTICIPANT OBSERVATION in order for their culture and social organization to be understood properly. Within 20th-century U.S. sociology, Herbert Blumer and others recommended much the same approach, emphasizing the need to capture the interpretative and processual character of human social life, as well as criticizing methods that failed to do this. In the late 1960s, this realism—sometimes also referred to as NATURALISM—was summarized by David Matza as the obligation to remain true to the nature of the phenomenon under study.

In the 1980s and 1990s, however, ethnographic realism came under challenge. The critics raised questions about the very possibility of providing accounts of the "nature" of social phenomena, sometimes denying that these phenomena exist independently of the process of inquiry. This critique arose, to some extent, out of ethnography itself. Typically, ethnographers have portrayed people as actively making sense of the world, thereby developing distinctive cultures. If this approach is applied reflexively to ethnographic accounts themselves, the conclusion may be reached that these accounts are creative constructions that reflect the sociocultural identities of researchers, rather than being objective representations. The tendency to draw this conclusion was encouraged by the influence of some strands of continental philosophy, notably PHENOMENOLOGY, HERMENEUTICS, and POSTSTRUCTURALISM (though these positions do not all imply epistemological RELATIVISM or skepticism).

Another important component of the critique pointed out that ethnographic accounts often rely on the same literary devices that are used in realist fiction. The critics insisted that no form of writing is merely representational: All language use is constitutive and performative, and no writer should pretend otherwise.

Along with criticisms of ethnographic realism on epistemological and rhetorical grounds, there has also been an ethical and political critique. Drawing on radical Leftist challenges to sociology in the 1960s, as well as on feminism, this critique has argued that realist ethnography amounts to voyeurism and/or serves as a means of surveillance by which those in subordinate social positions are kept under control.

In line with these arguments, some critics of ethnographic realism have sought to develop new forms of writing designed to subvert any impression that ethnographic texts can provide objective accounts of the world, to display their own constructed and indeterminate character, and/or to give voice to those who are seen as being "silenced" by mainstream accounts (see Denzin, 1997).

There has been some defense of ethnographic realism against these criticisms. The self-undermining character of skeptical and relativist arguments has been reiterated. Furthermore, a distinction has been drawn between the naive realism that is sometimes ascribed to ethnographers of the past and more sophisticated forms. The latter take account of epistemological conundrums and of the socially contexted character of all research but do not deny the possibility of social scientific knowledge or insist that ethnography is inevitably implicated in political struggle supporting or resisting the status quo (see Hammersley, 1992).

—Martyn Hammersley

REFERENCES

Atkinson, P. (1990). *The ethnographic imagination: Textual constructions of reality.* London: Routledge.

Denzin, N. K. (1997). *Interpretive ethnography.* Thousand Oaks, CA: Sage.

Hammersley, M. (1992). *What's wrong with ethnography?* London: Routledge.

Matza, D. (1969). *Becoming deviant.* Englewood Cliffs, NJ: Prentice-Hall.

ETHNOGRAPHIC TALES

ETHNOGRAPHY, a long-established data collection method in anthropology, provides rich, detailed information about individuals and groups. As a qualitative research method, ethnography typically involves PARTICIPANT or NON-PARTICIPANT OBSERVATION and/or IN-DEPTH INTERVIEWS, allowing researchers to get a comprehensive look at the population of interest.

A good ethnographic account provides "thick description," or a detailed account of the social context and shared meanings of a particular group (Geertz, 2000). As such, ethnographic studies yield a vast amount of data—literally hundreds of pages of FIELDNOTES, analytic MEMOS, and/or interview transcripts. In part because of the magnitude of information gathered by ethnographic approaches, the stories social scientists construct from their data may take many different forms. Indeed, although ethnographers share the goal of providing an in-depth look at a group or social phenomenon, they may choose a number of ways to present their findings.

In his book *Tales of the field: On writing ethnography,* Van Maanen (1988) identified three main types of ethnographic tales: realist, confessional, and impressionist. The realist tale aims at presenting an objective view of the group being studied. Thus, the realist tale is akin to the more traditional ethnographic account wherein readers get a clear, concise description of the individuals or groups under study. They do not get a sense of the author's reaction to that setting, the impact the author feels he or she has had on the findings, or the reflexive thoughts the author has after returning from "the field."

The confessional tale, on the other hand, lets the reader in on what the author's fieldwork experience was like. Although some may argue that the subjective nature of this tale is less scientific and therefore flawed, confessional tales can be a valuable tool in assessing the strength of the research because they allow the reader to get a sense of the author's approach and biases, as well as the methodological lessons learned while conducting the research. REFLEXIVITY and disclosure of the *process* of research have always been important to feminist qualitative researchers who aim at demystifying the scientific research method in addition to providing meaningful findings.

The impressionist tale is an account of atypical phenomena. The impressionist tale is intended to shock and engage the reader with stark description of unique findings. In this type of tale, the researcher does not offer an *interpretation* of findings but rather brings unusual stories forward to speak for themselves.

Van Maanen's types of ethnographic tales are not mutually exclusive categories; one may present two or all three types of tales in a narrative account. As Miller (1998) explained in her research on soup kitchens for the homeless, these tales may emerge somewhat unexpectedly as authors reflect on and share their findings. In addition, a primary goal of ethnography and other qualitative methodologies is to *give a voice* to marginalized groups (Ragin, 1994). Giving a voice to individuals whose experiences are not widely recognized *and* providing sound interpretation of research findings may well necessitate the use of various ethnographic tales.

—Desirée Ciambrone

REFERENCES

Geertz, C. (2000). *The interpretation of cultures: Selected essays.* New York: Basic Books.

Miller, D. (1998). Writing and retelling multiple ethnographic tales of a soup kitchen for the homeless. *Qualitative Inquiry, 4,* 469–492.

Ragin, C. C. (1994). *Constructing social research.* Thousand Oaks, CA: Pine Forge Press.

Van Maanen, J. (1988). *Tales of the field: On writing ethnography.* Chicago: The University of Chicago Press.

ETHNOGRAPHY

Ethnography is the art and science of describing a group or culture. Its strength is description. Ethnography provides the reader with a detailed picture of what's going on, from the perspective of natives of the given culture. It also provides an insight into the processes associated with implementing a program, processes that lead to specific outcomes. One of the primary methodological tools involves fieldwork. Ethnographers immerse themselves in a culture in order to observe and record people's behavior in their natural setting. The time spent observing may range from a few months to a few years; there must be enough time to observe patterns of behavior over time. This is a form of RELIABILITY.

Ethnographers attempt to elicit the insider's or emic perspective of reality while in the field (see EMIC/ETIC DISTINCTION). The ethnographer may or may not agree with the insider's perspective or agree that it conforms to an objective perspective of reality. The aim is to be nonjudgmental, recognizing that there are real consequences for a person's perception of reality, regardless of the scientific merit of that perception. The ethnographer recognizes multiple realities or perceptions of reality based on people's roles in society. For example, the view of the same police officer may be significantly different for an upper-middle-class elderly woman and for a lower-socioeconomic African American male. The ethnographer documents multiple and often conflicting emic or insider perspectives. Most ethnographers build on that knowledge and, using an external or etic perspective, explain the relationship between these emic perspectives of reality.

Ethnography takes more time than other forms of research because it involves fieldwork—time on site with people in their natural environment. Ethnographers spend time in the field in order to place into context the data they collect. Adaptations can be made concerning the amount of time devoted to fieldwork, but there is a direct relationship between time spent in the field and the quality of the data. Ironically, it is often more efficient to be inefficient: Spending more time in the field saves time revisiting poorly understood issues after a study has been completed.

This immersion in the lives of other people allows ethnographers to more accurately interpret people's behavior. This is called cultural interpretation. Clifford Geertz (1975), borrowing Gilbert Riles' notion of THICK DESCRIPTION, provided a classic in ethnography. It is the example of the wink or blink and how it captures the significance of cultural interpretation in ethnography. The same mechanical movement of an eyelid quickly closing and then opening again might be a wink in a dark and smoky bar or a blink in a dusty furniture repair shop. It is the cultural context that defines how that same mechanical behavior is interpreted (Fetterman, 1998).

FIELDWORK

Fieldwork is one of the most characteristic features of ethnography. Ethnographers typically adopt a naturalistic approach, studying people in their own environment. The most common technique is a stratified judgmental sample, with ethnographers relying on their own judgment to select for study the most appropriate members of the culture in different categories or roles, based on the research question. Questions generally are informal, open, and nonthreatening.

Fieldwork has many phases. The first phase involves entering the field (based on the research proposal). Once introduced to the community, the ethnographer typically begins with a survey period to learn the basics, including language, history, politics, and economics. The postsurvey phase is characterized by continual data collection, theory generation, hypothesis testing, and cross-checking information. In other words, the ethnographer spends a great deal of time guessing about relationships and behavior patterns. Observations and interviews are used to verify, refute, or build on the best guesses about the situation. This approach helps to build a solid foundation of knowledge about the culture or group.

The later phase of fieldwork typically is more narrow and precise in focus. Critical concepts and themes are explored in greater depth (but often less breadth). There is often time pressure to track down missing information, gaps, and contradictions most closely associated with the critical findings or insights associated with the study.

PARTICIPANT OBSERVATION

PARTICIPANT OBSERVATION characterizes most ethnographic work and is crucial to effective fieldwork. Participant observation combines participation in the lives of the people under study with maintenance of a professional distance that allows adequate observation and recording of data. Long-term residence helps the ethnographer internalize basic beliefs, fears, hopes, and expectations of the group.

CONCEPTS

An ethnographer enters the field with an open mind but not an empty head. Ethnographers have traditional concepts guiding their observation and interpretation. A few of the most significant concepts guiding their work include culture, a holistic perspective, contextualization, emic and etic perspectives, and a nonjudgmental orientation.

Culture

Culture is the broadest ethnographic concept guiding practice. Culture comprises the ideas, beliefs, knowledge, and behavior that characterize a group of people. The concept of culture helps the ethnographer search for a logical, cohesive pattern within a group. Each culture has an identifiable value system that shapes behavior. Discovering the underlying values of a group is like unlocking the human genome. Values constitute a social DNA that helps researchers understand and potentially predict human behavior within a group.

Holistic Perspective

Ethnographers assume a holistic outlook in research, attempting to describe as much as possible about a culture or social group, including its history, religion, politics, and environment. No study can capture an entire culture; however, this orientation pushes the ethnographer to learn more about the depths and breadth of a culture and observe each event within a larger social and economic context. This conceptual lens enables the ethnographer to see interrelationships at the margins and multiple layers of meaning that might be overlooked or ignored in a more focused and narrow approach.

Contextualization

Contextualizing data involves placing observations into a larger perspective. An event or behavior is meaningfully interpreted only when placed in context. A behavior or event viewed in isolation or out of context is easily misinterpreted. Moreover, it is stripped of its multilayered meaning unless situated within its own environment or set of circumstances.

Emic and Etic Perspectives

The emic or insider's perspective is at the heart of most ethnographic research. Ethnographers build their knowledge about a group or culture based on the insiders' views, rather than working from an a priori set of assumptions about how things work. Ethnographers also typically use an etic or external social scientific posture after assembling multiple insider perspectives. The aim is to see how insiders interrelate, communicate, and function as part of a larger system or culture.

Nonjudgmental Orientation

A nonjudgmental orientation helps prevent ethnographers from making inappropriate and unnecessary value judgments about what they observe. A nonjudgmental orientation requires the ethnographer to suspend personal valuation of any given cultural practice. Maintaining a nonjudgmental orientation is similar to suspending disbelief while one watches a movie or play or reads a book—one accepts what may be an obviously illogical or unbelievable set of circumstances to allow the author to unravel a riveting story.

We all have personal beliefs, biases, and individual tastes. The ethnographer guards against the more obvious biases by making them explicit and by trying to view another culture's practices impartially. Ethnocentric behavior—the imposition of one culture's values and standards on another culture, with the assumption that one is superior to the other—is a fatal error in ethnography.

Many other concepts shape or guide an ethnographer in the field, including inter- and intracultural diversity, structure and function, symbol and ritual, micro- or macrolevel study, and operationalism. These concepts are like a theory guiding the ethnographer in practice.

METHODS

The ethnographer is a human instrument. Relying on all its senses, thoughts, and feelings, the human instrument is a most sensitive and perceptive data-gathering tool. The information this tool gathers, however, can be subjective and misleading. Ethnographic methods and techniques help to guide the ethnographer through the wilderness of personal observation and to identify and classify accurately the bewildering variety of events and actions that form a social situation.

Although fieldwork and participant observation characterize much of anthropological work, ethnographers rely on a variety of specific tools, including interviews, SURVEYS, and UNOBTRUSIVE MEASURES.

Interviews

Interviewing is the ethnographer's most important data-gathering technique. Interviews explain and put into a larger context what the ethnographer sees and experiences. General interview types include structured, semistructured, informal, and retrospective.

Structured and semistructured interviews serve comparative and representative purposes—comparing responses and putting them in the context of common group beliefs and themes. An informal interview is different from a conversation, but it typically merges with one, forming a mixture of conversation and embedded questions. The questions typically emerge from the conversation. In some cases, they are serendipitous and result from comments made by the participant. The ethnographer uses retrospective interviews to reconstruct the past, asking informants to recall personal historical information.

Types of Questions. There are many types of questions posed in an interview. A survey question or grand tour question is designed to elicit a broad picture of the participant's or native's worldview. Survey questions also help define the boundaries of a study. Once survey questions reveal a category of some significance, specific questions about that category become most useful. Specific questions probe further into an established category of meaning or activity. Whereas survey questions shape and inform a global understanding, specific questions refine and expand that understanding. Structural and attribute questions—subcategories of specific questions—often are used to further clarify the emic perspective.

In addition, ethnographers use both open-ended and closed-ended questions to pursue fieldwork. An open-ended question allows participants to interpret it. Closed-ended questions, with a limited number of answers from which to choose, are useful in trying to clarify views and quantify behavior patterns.

Key Actor and Life History

Some people are more articulate and culturally sensitive than others. These individuals make excellent "key actors," acting as important sources of information for ethnographers. They provide detailed historical data, knowledge about contemporary interpersonal relationships, and a wealth of information about the nuances of everyday life. Although the ethnographer tries to speak with as many people as possible, time is always a factor. Therefore, anthropologists traditionally have relied most heavily on one or two individuals in the group. Key actors often provide ethnographers with rich, detailed autobiographical descriptions or help identify members of the community meriting such depth. These life histories often provide an integrated picture of the culture under study.

Surveys

Surveys are an efficient means of large-scale data collection and an excellent way to tackle questions dealing with representativeness. They are the only realistic way of taking the pulses of hundreds or thousands of people. (See Fowler, 1988, and Fink and Kosecoff, 1988, for guides to conducting surveys.) In addition to surveys, ethnographers use many unobtrusive measures.

Unobtrusive Measures

Unobtrusive measures draw social and cultural inferences from physical evidence. An outcropping, or physical manifestation of an economic or psychological condition, might be graffiti or a burned-out building. Initial inferences are possible without any human interaction. Such cues by themselves, however, can be misleading. Cross-checking and additional data collection are necessary.

Other unobtrusive methods include the study of archival documentation, e-mail, school records, census records, and budgets. Proxemics help differentiate between a close familial relationship and a distant, more businesslike relationship. It also can help us interpret a social situation in which someone's personal space has been entered, with or without permission. Kinesics or body language often signals moods and dispositions (see Birdwhistell, 1970; Hall, 1974).

EQUIPMENT

Notepads, computers, tape recorders, and cameras—all tools of ethnography—are merely extensions of the human instrument, assisting and enhancing memory, capacity, and vision. These useful devices can facilitate the ethnographic mission by capturing the rich detail and flavor of the ethnographic experience and then helping to organize and analyze the data.

The most common tools ethnographers use are pen and paper. With these tools, the fieldworker records notes from interviews during or after each session, sketches the physical layout of an area, traces an organizational chart, and outlines informal social networks. Tape recorders effectively capture long verbatim quotations, essential to good fieldwork, while the ethnographer maintains a natural conversational flow. In addition, TRANSCRIPTION, although time-consuming, is a means of getting closer to the data, becoming more familiar with the patterns of speech, local references, and a person's pauses and tone.

Cameras enable the ethnographer to create a photographic record of specific behaviors. This is a form of FACE VALIDITY. The pictures also can be used as a projective technique; sharing pictures with key actors can elicit useful insights about their behaviors. Photos also can supplement and prompt the ethnographer's memory during the period of data analysis and writing. Digital cameras are now standard in the field. They enable the ethnographer to share pictures immediately—on the camera's screen, attached to e-mail, or stored on a file server for team or community members to view (with permission).

Digital video camcorders are also standard tools in the field. Ethnographers use inexpensive software to produce movies about key events. These represent visual representations or presentations of ethnographic findings. (See Heider, 1976, and Bellman and Jules-Rosette, 1977, for guides concerning ethnographic filmmaking.)

Computers are indispensable in ethnography. Laptops are used in the field to take notes during interviews. Qualitative data analysis software programs such as NUD*IST, ETHNOGRAPH, and ATLAS.ti used to code, chunk, and sort fieldnotes and identify patterns. PowerPoint presentations highlight findings in reports to sponsors and colleagues.

The Internet embodies several tools useful in ethnographic work, including e-mail, search engines, and reference pages, such as phone directories and maps. In addition, there are many free or inexpensive online survey programs on the Internet. Videoconferencing on the Internet is a way to collect data in the field or to maintain a healthy rapport with key actors while away from the field. These tools save time and money, and they can greatly facilitate communication (see Fetterman, 2002, for additional details).

WRITING

Ethnography requires good writing at every stage. Research proposals, fieldnotes, memoranda, interim reports, final reports, articles, and books, including ethnographies, are the tangible products of ethnographic work. The ethnographer can share these written works with participants to verify their accuracy and with colleagues for review and consideration.

Fieldnotes are the brick and mortar of an ethnographic edifice. These notes consist primarily of data from interviews and daily observation. Ethnographers produce summaries of the research effort during various stages of their fieldwork. The last stage in ethnographic research is writing the final report, article, or book. Ethnographies typically are characterized by thick description and verbatim quotations. Internet publishing represents an alternative to publication in print. Articles can be published and reviewed much quicker on the Internet than they can in traditional media.

ETHICS

Ethics pervade every stage of ethnographic work. Ethnographers must formally or informally seek INFORMED CONSENT to conduct their work. Taking photographs and making audio or video recordings requires the participant's permission (in writing if used for educational or commercial purposes). Ethnographers are candid about their task, explaining what they plan to study and how they plan to study it. They promise confidentiality when requested and use pseudonyms to protect the identity of the individuals with whom they work. Ethnographers need the trust of the people they work with to complete their task. Ethics guide the first and last steps of ethnography, beginning with the selection and definition of the culture and problem, and concluding with the written record of the experience. Ethnographers stand at ethical crossroads throughout the research life cycle, including inception or identification of the problem;

birth, or a funded proposal; childhood, represented by field preparation; adolescence and adulthood, best characterized as fieldwork; and retirement, when it is time to disengage from the field. Each phase poses new challenges as the ethnographer balances methodological concerns with ethical considerations. This fact of ethnographic life sharpens the senses and ultimately refines and enhances the quality of the endeavor. (See Fetterman, 1998, for additional detail; see also the American Anthropological Association's "Code of Ethics" and "Principles of Professional Responsibility" at www.aaanet.org/committees/ethics/ethics.htm.)

—David M. Fetterman

REFERENCES

Bellman, B. L., & Jules-Rosette, B. (1977). *A paradigm for looking: Cross-cultural research with visual media.* Norwood, NJ: Ablex.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays on body motion communication.* Philadelphia: University of Pennsylvania Press.

Fetterman, D. M. (1998). *Ethnography: Step by step* (2nd ed.). Thousand Oaks, CA: Sage.

Fetterman, D. M. (2002). Web surveys to digital movies: Technological tools of the trade. *Educational Researcher, 31*(6), 29–37.

Fink, A., & Kosecoff, J. (1988). *How to conduct surveys: A step-by-step guide.* Newbury Park, CA: Sage.

Fowler, F. J. (1988). *Survey research methods* (Rev. ed.). Newbury Park, CA: Sage.

Geertz, C. (1975). *The interpretation of cultures.* New York: Basic Books.

Hall, E. T. (1974). *Handbook for proxemic research.* Washington, DC: Society for the Anthropology of Visual Communication.

Heider, K. G. (1976). *Ethnographic film.* Austin: University of Texas Press.

ETHNOMETHODOLOGY

Ethnomethodology is the name of an academic field of study as well as a name for what that field studies. The word "ethnomethodology" literally means *popular* or *folk methodology.* The term was coined by Harold Garfinkel (for an account of the origin of the term, see Garfinkel, 1974). After the publication of Garfinkel's (1967) *Studies in Ethnomethodology,* it became an established, if controversial, subfield of sociology. Ethnomethodology has strong connections with sociology and was recently granted Section status by the American Sociological Association. The connection with sociology runs deep, as ethnomethodology provides a novel approach to the classic themes and problems of sociology; a new volume of Garfinkel's studies (2002) addresses itself to Emile Durkheim's fundamental conception of social facts. Ethnomethodology, however, is not limited to sociology. Together with CONVERSATION ANALYSIS—a spin-off program that investigates the constitutive organization of social interaction—ethnomethodology has made inroads into anthropology, sociolinguistics, computer-supported cooperative work, management studies, and social studies of science, among other fields.

METHODS AS SUBJECT MATTER

Ethnomethodology is not itself a method in any straightforward sense. Just as ethnomusicology is not a branch of music—a particular technique or tradition of music—ethnomethodology is not a kind of methodology. Ethnomethodological studies often make use of ethnographic procedures, and they deploy detailed records of continuous activities "in real time," but in many cases the heaviest burden on the student is to master, and give a procedural account of, specialized methods that constitute the social activities studied.

Ethnomethodology is *sociological* in its fundamental orientation, as it is the study of professional and nonprofessional methods through which social order is produced. Accordingly, ethnomethodologists write procedural descriptions of methods used in a whole range of commonplace and specialized activities. The modern prototype of *method* is the scientific account of experimental or observational procedures. As Sacks (1992, pp. 802–806) noted, scientific methods are communicated by means of vernacular descriptions and prescriptions. Scientific methods can be highly technical, but they are specialized instances of the much broader social phenomenon: instructions that enable members to reproduce a community's practices. In other words, methods and methodological accounts are reflexive features of disciplinary forms of life. Sacks imagined that ethnomethodologists would write methodology reports, but unlike the professional methodologists who write experimental and observational protocols, ethnomethodologists would delve into the rich variety of ordinary methods through which members of a society coordinate and reproduce their activities. These methods include such mundane "achievements" as crossing the street at a traffic light, coordinating the openings and closings of conversations, bringing a meeting to order, and many other routine actions in daily life.

Ethnomethodological descriptions of lay and professional methods are analogous to methodologies written by scientists and other professionals. Unlike formal methodological accounts, however, they include context-sensitive details of how plans, rules, algorithms, recipes, maps, guidelines, maxims, protocols, proverbs, and other formal instructions are used (as well as disused, misused, or abused) in actual conduct. Such descriptions are not antiformalistic, in the sense of rejecting the relevance of rules and formulas. Instead, ethnomethodological descriptions address the way participants deploy rules, plans, and so forth within singular courses of action. Ethnomethodological studies can require extensive training and PARTICIPANT OBSERVATION in specialized work settings. Often, ethnomethodologists use audio- and videotapes recorded *in situ*, in order to recover the way collective activities are assembled over time through collaborative, often improvisational, sequences of discourse and embodied practice. Typically, this orientation to the phenomenological here and now precludes the use of methods that reduce situated discourse to standard codes and aggregate patterns. Ethnomethodology is not a search for subjective meaning, however. Although informed by existential phenomenology, ethnomethodological descriptions rarely take the form of first-person experiential reflections [an exception is Sudnow's (1978, 2001) elucidation of his own developing competence with playing jazz piano]. More often, such descriptions are written in the third person, without privileging a "private" or individual vantage point. Such descriptions rely upon the intelligibility of common practices to masters of a natural language.

METHODOLOGIES AS SUBJECT MATTER

The subject matter of ethnomethodology includes *methodological* investigations as well as *methodic* practices. In other words, ethnomethodology is the study of *both* methodology (in the sense of systematic reflections about methods and efforts to design them) and methods themselves (practices performed in accordance with one or another methodological design). Such a treatment contrasts with the view that practices are performed unconsciously, requiring professional analysts to account for their systematic organization. For ethnomethodologists, accounts of practice and practices themselves are inextricable, and it is their intertwining that is of interest.

Garfinkel's program made no exception for social scientific methods, and some of his earliest studies delved into social science research practices such as coding interview responses and analyzing organizational records. These studies demonstrated that *ad hoc* judgments grounded in commonsense knowledge of the social world were crucial for constituting data points and assigning instances to nominal categories. Starting in the 1970s, Garfinkel and some of his students turned their attention to the natural sciences. Like other ethnomethodological studies, investigations of laboratory practices and high-tech workplaces focus on quotidian details, but it is necessary to develop a mastery of the specialized techniques studied in order to demonstrate their ordinariness. A puzzling feature of these studies is that the sciences already possess their own, highly developed, methodologies. In order to have something to say that is not already part of the sizable professional literatures on methodology, it is necessary for ethnomethodologists to delve into more obscure, and sometimes contentious, features of a practice that are not mentioned in its formal methodological prescriptions and reports. There is some resemblance to social constructionist approaches, but ethnomethodologists do not suggest or imply that the ad hoc production of scientific evidence detracts from the objectivity of such evidence. Instead, for ethnomethodologists, the tacit features of scientific production are unavoidable; they are a source of objectivity as much as a barrier to it.

—Michael Lynch

REFERENCES

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.

Garfinkel, H. (1974). On the origins of the term "ethnomethodology." In R. Turner (Ed.), *Ethnomethodology* (pp. 15–18). Harmondsworth, UK: Penguin.

Garfinkel, H. (2002). *Ethnomethodology's program: Working out Durkheim's aphorism.* Blue Ridge Summit, PA: Rowman and Littlefield.

Sacks, H. (1992). *Lectures on conversation* (Vol. 1). Oxford, UK: Blackwell.

Sudnow, D. (1978). *Ways of the hand.* Cambridge, MA: Harvard University Press.

Sudnow, D. (2001). *Ways of the hand* (Rev. ed.). Cambridge, MA: MIT Press.

ETHNOSTATISTICS

Ethnostatistics uses concepts from ETHNOMETHODOLOGY to study sensemaking practices that social scientists employ in the production, interpretation, and display of statistics created in social research. Ethnostatistics focuses on quantitative reasoning and language use in social contexts to understand how professionals produce numerals that create a sense of shared social reality.

Ethnostatistics contrasts with STATISTICS, which involves techniques and rule-governed procedures for counting, aggregating, and estimating data. The field of statistics does not study practical knowledge needed to produce statistics or how situations influence the behavior of the scientist. Formal statistical procedures and rules are thus incomplete guides to producing statistics. Ethnostatistics is concerned with describing and understanding the technical knowledge and procedures needed to actually do statistics.

Ethnostatistics has three levels. *First-level ethnostatistics* uses ethnography to produce THICK DESCRIPTIONS of the social production of statistics. For example, Gubrium and Buckholdt (1979) studied the professional production of quantitative measures of hospital program effectiveness and demonstrated that technical rules are insufficient to explain counting practices and statistics. Informal criteria, such as patient motivation to comply with program goals, were used in addition to technical criteria by professional staff to determine which actions to count.

Second-level ethnostatistics studies statistics at work using COMPUTER SIMULATION of the implications for the interpretation of statistics that are associated with violations of various technical assumptions. One study tested the validity of the common social science assumption that ordinal data can be transformed into interval data for statistical analysis purposes by assigning successive numerals (e.g., 1–7) to data categories. This practice assumes that data transformation does not affect the interpretation of results and findings. Ethnostatistical research shows that this transformation can produce measured values and statistics that diverge "substantially from true measures" (Gephart, 1988, p. 38). Hypothesis testing and PATH ANALYSIS interpretations of data also were affected. Second-level ethnostatistics thus shows that common measurement practices may distort results.

Third-level ethnostatistics uses RHETORIC and textual analysis methods to examine persuasive properties of statistical displays in documents and texts. One study of how quantitative methodology papers create meaning and interpret numerals found that meaning is created by the adjectives that are used to describe and interpret numerals and by contexts where statistical descriptions are embedded. The absolute value of a numeral did not govern the interpretation made of the numeral, and tests of statistical significance played a minor role. Indeed, the correlation an author termed "substantial" in a quantitative paper often was within .1 of a correlation claimed to be "relatively small" at another point in the article.

Ethnostatistics integrates the quantitative and qualitative aspects of science by examining how meaning is actually created for numbers. This re-humanizes quantitative research by relating it to qualitative foundations of social life (Gephart, 1988).

—Robert P. Gephart, Jr.

REFERENCES

Gephart, R. P. (1986). Deconstructing the defence for quantification in social science: A content analysis of journal articles on the parametric strategy. *Qualitative Sociology, 9,* 126–144.

Gephart, R. P. (1988). *Ethnostatistics: Qualitative foundations for quantitative research.* Thousand Oaks, CA.: Sage.

Gephart, R. P. (1997). Hazardous measures: An interpretive textual analysis of quantitative sensemaking during crises. *Journal of Organizational Behavior, 18,* 583–622.

Gubrium, J. F., & Buckholdt, D. R. (1979). Production of hard data in human service organizations. *Pacific Sociological Review, 22,* 115–136.

ETHOGENICS

The word "ethogenics" was coined to refer to a new paradigm for social psychological research that developed in Oxford, England, in the 1970s (Harré & Secord, 1972) in conscious reaction to the importation of methods and theories from mainstream American psychology. Ethogenics was influenced by ETHNOMETHODOLOGY (Garfinkel, 1967), by the microsociology of Goffman (1959), and by social and cultural anthropology.

The movement rejected universalistic presuppositions that had led to levels of abstraction that eliminated concrete social phenomena. The treatment of persons as passive sites for stimulus/response patterns was rejected in favor of treating persons as active agents engaged with others in carrying out projects according to local rules and conventions. Fallacious uses of statistical methods were highlighted, in particular the familiar but still common fallacy of drawing conclusions about individual propensities from statistical distributions.

Emphasis was placed on the analysis of actual episodes of social interaction as unfolding sequential structures of meanings, ordered in accordance with local rules, conventions, and customs of correct conduct. Studies of the dynamics of social action—for example, the *making* of friends—displaced studies of the conditions under which static states, such as *friendship*, were brought about. The focus on local customs and conventions led to an interest in social anthropology as an essential component of a new paradigm of social psychology.

The methodological characteristic of research within the new paradigm required the use of the act/action distinction to identify social phenomena as meanings and how those meanings were understood by the participants. Observation of real-life episodes replaced EXPERIMENTS. Analysis of participants' justificatory and interpretative accounts was used to identify the projects to the accomplishment of which a social interaction was directed, as well as the rules and conventions in accordance with which it was managed.

The upshot of research within the new paradigm was a catalog of situation-specific meanings and sets of context-sensitive rules that explained the pattern of the evolving social episode, viewing it as an actual sequence of meaningful social actions. This led to an alliance with the newly emergent field of discursive psychology, which was directed to a similar end product. The concept of "rule" stood in for a wide variety of normative constraints that could be seen to be effective in shaping the flow of action.

Ethogenics placed great emphasis on the construction of descriptive and explanatory models. The most powerful of these was the dramaturgical model. Social episodes can be looked at as if they were performances of stage plays, with scene, actor, and action mutually influencing one another. This basic model was elaborated by exploring analogies that compared certain kinds of social episodes to games and ceremonies.

In later applications of this point of view in developmental social psychology, alliances grew up with (SOCIAL) CONSTRUCTIONISM and with cultural psychology (Cole, 1996).

—Rom Harré

REFERENCES

Cole, M. (1996). *Cultural psychology: A once and future discipline.* Cambridge, MA: Belknap Press of Harvard University Press.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.

Goffman, E. (1959). *The presentation of self in everyday life.* New York: Doubleday.

Harré, R., & Secord, P. F. (1972). *The explanation of social behaviour.* Oxford, UK: Blackwell.

ETHOLOGY

Ethology is a prime example of a discipline whose models and methods were developed in nonhuman animal research and then applied to research on humans. Firmly grounded in the evolutionary thinking of Charles R. Darwin, ethology perhaps made its most valuable contribution to theoretical refinement through its commitment to understanding behavior on multiple levels: (a) proximate causation (which may be physiological or environmental in origin), (b) ultimate causation (the adaptive function of the behavior in enhancing survival and/or reproductive fitness), (c) phylogeny (the behavior's pattern in related species), and (d) ontogeny (the developmental course of the behavior) (Tinbergen, 1963). This general framework led animal researchers to leave their controlled laboratories and seek out natural habitats where species of interest could be observed in the environmental niches to which they had adapted. For example, Konrad Lorenz discovered the phenomenon of imprinting in geese by observing goslings as they hatched and began following either their mothers or Lorenz himself soon after emerging from their eggs (Lorenz, 1981). It was only after completing the descriptive phase of this observational work that Lorenz began to experiment with imprinting, for example, by introducing time since hatching as a variable, still in the natural setting of a lakeshore in Germany. Thus, ethology is often defined as the study of animal behavior in the animal's natural environment.

Describing behavior in its natural setting does not imply abandoning scientific ideals of rigorous research.

An ethologist will spend countless hours cataloging behaviors of the subject species, thus creating an "ethogram" for that species. Specific questions typically are asked: Which behaviors seem to occur together? Under what conditions in the environment? Do the behaviors seem directed toward a goal? What does the animal do if the goal is not achieved? Do the behaviors make up a fixed action pattern? How often do the behaviors appear? Are there individual differences, perhaps related to sex, age, or social status? Are there differences related to time of day, season, temperature, or other environmental signals? Throughout the process, the ethologist will ideally return to Tinbergen's four levels of analysis cited earlier, for theoretical refinement. Where appropriate, the ethologist also will utilize experimental methods to answer questions. For example, when a specific environmental stimulus is hypothesized to be a "releaser" of a behavior, the ethologist may carry out a deprivation experiment, eliminating the stimulus, to see how the behavior of interest is modified. By manipulating the timing of releasers, ethologists have identified sensitive periods (formerly called critical periods) in the development of such behaviors as imprinting in geese and socialization in dogs. Thus, ethological research may be described as moving back and forth in a systematic way between qualitative and quantitative approaches to studying behavior.

During the 1960s and 1970s, some scientists trained in classical ethology moved toward human research, as seen in the work of Irenaeus Eibl-Eibesfeldt (1989), who studied with Lorenz and documented human emotional communication around the world. It was also at this time that John Bowlby and Mary Ainsworth, influenced by the evolutionary approach of Darwin and the primate methodology of Harry Harlow, captured the attention of psychology with their work on mother-infant attachment. Many of those who admired the methodology, however, failed to appreciate Bowlby's evolutionary viewpoint, particularly in proposing infant survival as the ultimate goal of attachment.

Ethology takes a meticulous approach to OBSERVATIONAL RESEARCH. Several sources outline in great detail ethological standards for conceptualizing questions, SAMPLING subjects, recording time and space, recording events, measuring environmental stimuli, and evaluating observers' reliability (see, e.g., Blurton Jones, 1972; LaFreniere & Charlesworth, 1983; and McGrew, 1972). For example, if one is interested in how often preschoolers engage in aggressive interactions, one would first define aggression and list observable behaviors that fit the definition (push, bite, kick, etc.) Then, in preliminary stages of a study, one might note easily observed acts of aggression (*ad libitum* sampling); or one might note how many times those behaviors are observed every minute (event sampling); or one might note whether or not an aggressive behavior is observed every 15 seconds (one-zero sampling); or one might watch a certain child and note whether or not that child exhibits an aggressive behavior in a given time period (focal sampling); or one might note what comes before and after aggressive behaviors (sequence sampling). By utilizing these methods in an ethological study of nursery schoolers, McGrew (1972) demonstrated (among other findings) that aggression is related to winning/losing property fights and that teachers' ratings correlate highly with scientifically observed behavior.

At times, the need for more fine-grained analysis will necessitate the use of film or video analyzed in slow motion, or the use of sound recordings analyzed in the laboratory, or the use of chemical analysis to break down olfactory stimuli. In human research, facial muscle movements have been broken down into single units, and posture and movement quality are now being explored through computer modeling in the work of Karl Grammer, Bernhard Fink, and LeeAnn Renninger (2002).

In assessing the importance of ethology, Tinbergen (1963) commented, "It has been said that, in its haste to step into the twentieth century and to become a respectable science, Psychology skipped the preliminary descriptive stage that other natural sciences had gone through, and so was soon losing touch with the natural phenomena" (p. 411). Ethology, which has the advantage of providing both descriptive and experimental research methods, may have untapped heuristic value for the social sciences in general. In addition, ethology's underpinnings in evolutionary theory offer an inclusive approach that provides a way out of the fruitless nature/nurture dichotomy.

In recognition for their work in ethology, Konrad Lorenz, Nikolaas Tinbergen, and Karl von Frisch were awarded the Nobel Prize in Physiology or Medicine in 1973.

—Carol Cronin Weisfeld

REFERENCES

Blurton Jones, N. (Ed.). (1972). *Ethological studies of child behaviour*. London: Cambridge University Press.

Eibl-Eibesfeldt, I. (1989). *Human ethology*. Hawthorne, NY: Aldine de Gruyter.

Grammer, K., Fink, B., & Renninger, L. (2002). Dynamic systems and inferential information processing in human communication. *Neuroendocrinology Letters, 23*(Suppl. 4), 15–22.

LaFreniere, P., & Charlesworth, W. R. (1983). Dominance, attention, and affiliation in a preschool group: A nine-month longitudinal study. *Ethology and Sociobiology, 4*(2), 55–67.

Lorenz, K. (1981). *The foundations of ethology*. New York: Simon and Schuster.

McGrew, W. (1972). *An ethological study of children's behavior.* New York: Academic Press.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*, 410–433.

EVALUATION APPREHENSION

Evaluation apprehension is a form of reactive effect that has been shown to occur in EXPERIMENTS, whereby the experimental subject becomes worried that he or she is being tested by the experimenter and that he or she may perform poorly. It is sometimes recommended that the COVER STORY for an experiment should make it clear that the experiment has not been designed to evaluate the performance of participants.

—Alan Bryman

EVALUATION RESEARCH

Evaluation research is systematic, data-based inquiry to determine the merit or worth of a program, product, organization, intervention, or change effort. Evaluation research applies social science and related inquiry methods for the systematic collection of information about the activities, characteristics, and outcomes of change efforts to inform judgments about goal attainment, improve program effectiveness, identify costs and benefits, and/or inform future decisions.

Billions of dollars are spent to fight problems of poverty, disease, ignorance, joblessness, mental anguish, crime, hunger, and inequality, to name but a few of the many efforts at improving the human condition. How are programs that combat these societal ills to be judged? How does one distinguish effective from ineffective programs? How can information be gathered and reported in ways that increase program effectiveness and enhance decision making? These questions are the domain of evaluation research.

EVALUATION AND THE EXPERIMENTING SOCIETY

Edward Suchman (1967) began his seminal text on evaluation research with the observation that "one of the most appealing ideas of our century is the notion that science can be put to work to provide solutions to social problems" (p. 1). Visionaries such as Donald T. Campbell (1991) conceptualized evaluation as the centerpiece of a new kind of society: *the experimenting society.* He gave voice to this vision in his 1971 address to the American Psychological Association (1991):

> The experimenting society will be one which will vigorously try out proposed solutions to recurrent problems, which will make hard-headed and multidimensional evaluations of the outcomes, and which will move on to other alternatives when evaluation shows one reform to have been ineffective or harmful. We do not have such a society today. (p. 223)

HISTORICAL CONTEXT

Evaluation in the United States traces its formal professional roots to the federal projects spawned by the Great Society legislation of the 1960s, which were aimed at nothing less than the elimination of poverty. The creation of large-scale federal health programs, including community mental health centers, was coupled with a mandate for evaluation, often at a level of 1% to 3% of program budgets. Other major programs were created in housing, employment, services integration, community planning, urban renewal, and welfare. The rapid growth of social action programs created a demand for systematic empirical evaluation of the effectiveness of government programs. Passage of the U.S. Elementary and Secondary Education Act in 1965 contributed greatly to more comprehensive approaches to evaluation. The massive influx of federal money aimed at desegregation, innovation, compensatory education, greater equality of opportunity, teacher training, and higher student achievement was accompanied by calls for evaluation data to assess the effects on the nation's children of various programs.

Program evaluation as a distinct field of professional practice was born of two lessons from this period of large-scale social experimentation and government intervention: first, the realization that there is not enough money to do all the things that need doing; and second, even if there were enough money, it takes more than money to solve complex human and social problems. *Because not everything can be done, there must be a basis for deciding which things are worth doing.* Evaluation research was charged with informing decision making by bringing credible data to bear on judgments about effectiveness.

Early visions for evaluation, then, focused on evaluation's expected role in guiding funding decisions and differentiating the wheat from the chaff in federal programs. As evaluations were implemented, a new role emerged: *helping improve programs as they were implemented.* Evaluators were called on not only to offer final judgments about the overall effectiveness of programs but also to gather process data and provide feedback to help solve programming problems along the way.

MAJOR DEVELOPMENTS IN EVALUATION RESEARCH

1. Emergence of Evaluation as Both Discipline and Profession

Evaluation research began as the application of methods to particular problems within established disciplines such as economics, psychology, and education. As evaluation researchers began to come together to share experiences and perspectives, evaluation emerged as a distinct field of professional practice. Formal professional associations with annual meetings and professional journals were formed in the 1970s and 1980s, among them the American Evaluation Association, the Canadian Evaluation Society, and the Australasian Evaluation Society.

Evaluation also became a scholarly area of inquiry specializing in how merit, worth, value, and effectiveness are determined, thereby emerging out of the shadows of other disciplines to become a discipline in its own right, or even, as philosopher and evaluator Michael Scriven (2001) has suggested, a *transdiscipline.*

> [E]valuation is one of the elite group of transdisciplines, a term used in this context to refer to disciplines that are most notable for their service to other disciplines, although having their own autonomous status as well. These transdisciplines range from little but important ones like measurement, up through major ones like probability and statistics, key tools for most of the quantitative disciplines, to the all-encompassing ones—logic and evaluation. (p. 304)

2. Standards of Excellence for Evaluation

The professionalization of evaluation led to articulation of standards. Before the profession identified and adopted its own standards, criteria for judging evaluations could scarcely be differentiated from criteria for judging research in the traditional social sciences, namely, technical quality and methodological rigor. The most comprehensive effort at developing standards was hammered out over 5 years by a 17-member committee appointed by 12 professional organizations, with input from hundreds of practicing evaluation professionals. In 1981, the standards were published, then revised in 1994 (Joint Committee on Standards for Educational Evaluation, 1994). The standards dramatically reflected the ways in which the practice of evaluation had matured, calling for evaluations to be judged by four sets of criteria: utility, feasibility, propriety, and accuracy.

3. Increasing Attention to Evaluation Utilization

By the end of the 1960s, it was clear that evaluations of Great Society programs were largely ignored or politicized. Carol Weiss (1972) was one of the first to call attention to underutilization as one of the foremost problems in evaluation research, in a seminal review that concluded that "evaluation results have not exerted significant influence on program decisions" (pp. 10–11).

Nor was the challenge only one of increasing use. Evaluators became increasingly concerned about the misuse of findings. Marvin Alkin (1990), an early theorist of user-oriented evaluation, emphasized that evaluators must attend to *appropriate* use, not just amount of use.

Thus, evaluation faced a dual challenge: enhancing appropriate uses while also eliminating improper uses. In response to these concerns, an approach called UTILIZATION-FOCUSED EVALUATION emerged (Patton, 1997), with emphasis on conducting evaluations to achieve intended use by intended users. Utilization-focused evaluation involves designing an evaluation with careful consideration of how everything that is done, *from beginning to end*, will affect use. A psychology of use undergirds utilization-focused evaluation: Intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings; they are more likely to understand and feel ownership if they've been actively involved; and by actively involving primary intended users, the evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation at every step along the way.

Utilization-focused evaluation also brought attention to *process use* (Patton, 1997), in contrast to findings use. Process use refers to changes in thinking and behavior, as well as program or organizational changes in procedures and culture, that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process. Process use is represented by program staff saying after an evaluation, "The impact on our program came not so much from the findings but from going through the thinking process that the evaluation required."

4. Methodological Diversity Valued

Early in the development of evaluation research, an intense debate raged about the relative merits of quantitative/experimental methods versus qualitative/naturalistic methods. That debate has run out of intellectual steam (Patton, 1997). A consensus has emerged in the profession that evaluators need to know and use a variety of methods in order to be responsive to the nuances of particular evaluation questions and the idiosyncrasies of specific stakeholder needs. The focus in designing evaluation research is now the appropriateness of methods for specific purposes and intended uses, *not* adherence to some absolute orthodoxy that quantitative or qualitative methods should be inherently preferred. The Joint Committee Standards (1994) give equal weight to both kinds of methods. The field has come to recognize that, where possible, using multiple methods—both quantitative and qualitative—can be valuable because each has strengths and one approach can often overcome weaknesses of another.

5. Many Different Types of Evaluation and New Evaluator Roles

Parallel to increased methodological diversity, the practice of evaluation research has become diversified with many new approaches, models, and evaluator roles. Initially, evaluations simply assessed a program's goal attainment, making an overall judgment about merit or worth called a *summative evaluation* or *judgment-oriented evaluation*. Summative evaluations measure outcomes and impacts, and the causal linkages between program activities and outcomes that constitute a program model or theory. Gradually, more attention came to be placed on using evaluation to improve programs by getting participant feedback and identifying strengths and weaknesses, what is often called *formative evaluation* or improvement-oriented evaluation. Formative evaluations examine implementation, program processes, and program adaptations to changing clients or conditions. Sometimes formative evaluations serve to get programs ready for summative evaluation, but increasing use of improvement-oriented evaluation has become part of ongoing efforts for continuous quality improvement.

Both judgment-oriented and improvement-oriented evaluations involve the *instrumental* use of results, meaning that a decision or action follows from the evaluation. Conceptual use of findings, on the other hand, increases knowledge so that evaluation findings can influence thinking about issues, options, or policy alternatives. The knowledge generated can be as specific as clarifying a program's model, testing theory, distinguishing types of interventions, figuring out how to measure outcomes, generating lessons learned, and/or elaborating policy options. In recent years, knowledge-generating evaluations have come to be highly valued.

Another shift involved external versus internal evaluation. With the early focus on summative evaluation, most evaluations were commissioned externally to ensure independence and credibility. With greater emphasis on continuous quality improvement, internal evaluation began to flourish.

New approaches to evaluation have emerged to solve particular problems or accomplish specific objectives. *Goal-free evaluation* offers an alternative to goals-based evaluation as a way of finding out whether clients' primary needs were met and of capturing side effects and unintended impacts. *Responsive evaluation* is an approach to capturing and taking into account different and possibly conflicting stakeholder perspectives. *Empowerment evaluation* is a way to help participants tell their own stories and increase their capacity to engage in evaluation processes. *Inclusive evaluation* emphasizes the need for special efforts to include the perspectives of the disadvantaged and less privileged. *Theory-driven evaluation* formally tests social science theories in real program settings. *Realist evaluation* uses the philosophy of REALISM to build an evaluation framework. *Democratic evaluation* focuses on using evaluation to support dialogue and deliberation to enhance social justice. In *developmental evaluation*, the evaluator is part of a team whose members collaborate to conceptualize, design, and test new approaches in an ongoing process of continuous improvement.

These approaches offer but a sampling of the rich and varied menu of options that have emerged in evaluation research.

6. New Skills and Competencies Needed

New approaches to evaluation research have broadened the skills needed to be an effective evaluator. Evaluation research has developed beyond the simple application of social science research methods to assess a program's goal attainment. To ensure utility, evaluators need skills in communication, negotiation, conflict resolution, and group facilitation. To meet the standards of feasibility, evaluators need to be politically sophisticated, pragmatic, and creative, as well as able to manage evaluation processes so as to produce results on time and within budget. To meet standards of propriety, evaluators need a keen sense of ethics and sensitivity to diverse stakeholder groups. To meet standards of accuracy, evaluators need to be able to employ a variety of methods and work across disciplines. These new demands are changing the ways that evaluation researchers are trained.

7. Technology and Evaluation Research

The dominant role of technology in modern society has affected evaluation research as it has affected other endeavors. The Internet has opened up new communication channels as evaluators around the world are linked together through networks and listservs. Advances in software facilitate data analysis, including especially major developments in software for analyzing qualitative data (Patton, 2002). Advances in computer graphics and word processing software have changed how evaluation reports are written and disseminated. Placing major evaluation reports on the World Wide Web has increased access to evaluation findings, facilitated the dissemination of evidence-based practices, and enabled more work in synthesizing lessons learned across different evaluations, a form of generating knowledge. New techniques for data gathering using the Internet (e.g., Web-based surveys) are changing how data are gathered. Videotaping and photography are used in both data gathering and reporting. Management information systems are assisting programs in gathering routine monitoring data that can be aggregated for program evaluation purposes. In these and many other ways, evaluation research is influenced by technological developments.

8. Worldwide Demand for Evaluation

The first international evaluation conference in Vancouver in 1995 was planned by the three major evaluation societies active at that time: the Canadian, the Australasian, and the American. With more than 1,500 participants from 61 countries, that conference remains one of the defining moments of evaluation's history. By early in the 21st century, more than 40 national evaluation associations had been formed, in addition to the European Evaluation Society, the African Evaluation Association, and an International Evaluation Association made up of national and regional associations. International agencies have also begun using evaluation to assess the full range of development efforts under way in developing countries. Global interactions are defining the future of evaluation research, infusing the profession with new energy as evaluators around the world learn from each other and grapple with the ways in which evaluation research has to be adapted to fit different cultural, societal, and political contexts.

—Michael Quinn Patton

REFERENCES

Alkin, M. (1990). *Debates on evaluation*. Newbury Park, CA: Sage.

Campbell, D. T. (1991). Methods for the experimenting society. *Evaluation Practice, 12*, 223–260.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards.* Thousand Oaks, CA: Sage.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Rossi, P. H., Freeman, H. E., & Lipsey, M. (1998). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Scriven, M. (2001). Evaluation: Future tense. *The American Journal of Evaluation, 22,* 301–307.

Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice.* Newbury Park, CA: Sage.

Shadish, W. R., Jr., Newman, D. L., Scheirer, M. A., & Wye, C. (1995). *Guiding principles for evaluators* (New Directions for Program Evaluation, No. 66). San Francisco: Jossey-Bass.

Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs.* New York: Russell Sage.

Weiss, C. H. (1972). A treeful of owls. In C. H. Weiss (Ed.), *Evaluating action programs* (pp. 3–27). Boston: Allyn & Bacon.

Weiss, C. H., & Bucuvalas, M. (1980). Truth test and utility test: Decision makers' frame of reference for social science research. *American Sociological Review, 45,* 302–313.

EVENT COUNT MODELS

In many research contexts, the dependent variable is measured as a count. A count variable is a nonnegative integer that represents the number of observed events (or people, items, etc.). The objective is to explain the variation in observed counts by incorporating predictor variables. ORDINARY LEAST SQUARES REGRESSION analysis, which assumes a CONTINUOUS dependent VARIABLE, is not well suited for count data.

Count models are best thought of within the context of the GENERALIZED LINEAR MODEL. Any PROBABILITY distribution that allows only nonnegative integers can be used to describe the dependent variable. A careful study of count models can serve as an excellent bridge into the literature on the generalized linear model.

Suppose there is a vector of m predictor variables for each observation i: $X_i = (x_{1i}, x_{2i}, \ldots, x_{mi})$. Suppose that estimated coefficients $b = (b_1, b_2, \ldots, b_m)$, can be used to predict the expected value of the dependent variable, y_i. One's first instinct might be to use the linear model, such as

$$E(y_i|X_i) = X_i b = \sum_{j=1}^{m} b_j x_{ji}.$$

The shortcoming of this approach is that the expected value of y_i might be negative, thus violating the basic ASSUMPTION that all observed counts are 0 or greater. To circumvent that problem, a mathematical TRANSFORMATION typically is used. Most commonly, one finds the exponential transformation:

$$E(y_i|X_i) = exp(X_i b) = e^{X_i b} = e^{\sum_j b_j x_{ji}}.$$

The MAXIMUM LIKELIHOOD ESTIMATION approach is used to derive the estimates of b.

The maximum likelihood calculations require that some specific statistical distribution for y_i must be assumed. Until about 1990, by far the most commonly used distribution was the POISSON DISTRIBUTION. Given input variables X_i and parameter estimates $\hat{b}$, the probability of observing exactly y_i events is

$$\Pr(y_i|X_i) = \frac{\exp(-e^{X_i b})(e^{X_i b})^{y_i}}{y_i!}.$$

The estimates of the parameters b are chosen so as to make the sample of observations most likely.

The Poisson distribution is relatively simple and workable. If $E(y_i|X_i)$ is small, then the Poisson distribution is sharply skewed, with a very small modal value and a long tail to the right. As $E(y_i|X_i)$ increases, the shape changes and becomes rather normal in appearance. (The similarity between the Poisson distribution for large values and the normal distribution sometimes is used to justify the use of ordinary least squares regression with count data.) For a survey of the Poisson model and social science applications, consult King (1988).

The Poisson distribution has an unsuitable property: Its VARIANCE is equal to its mean. When it is fitted to observed data, one often finds the problem of overdispersion; the variance of the observed data is greater than expected according to the Poisson distribution.

Many scholars propose replacing the Poisson distribution with the NEGATIVE BINOMIAL (NB) DISTRIBUTION. The NB model has the same mean as the Poisson, but it has greater variance. It has another useful property. We can write the original model in a way that allows individual HETEROGENEITY in the form of an additive DISTURBANCE (or error) TERM. Let the (unobserved) individual variation be represented by u_i:

$$E(y_i|X_i) = \exp(X_i b + u_i) = e^{X_i b + u_i}$$
$$= e^{\sum_j b_j x_{ji} + u_i} = e^{\sum_j b_j x_{ji}} \times e^{u_i}.$$

If the variance of u_i happens to be 0, then this degenerates to a Poisson model (because $e^0 = 1$). However, if we posit that u_i has a GAMMA distribution, then the observed distribution of y_i would be the negative binomial. It is typical to assume that $E(u_i) = 0$, so that "on average" the heterogeneity observed among cases has no effect. That is to say, the expected value of a Poisson or negative binomial model is the same. However, when the variance of $u_i > 0$, the dispersion of the negative binomial distribution is greater. Long's (1997) textbook offers an excellent treatment of this issue.

Many extensions of the count model framework are available or are being pioneered in the advanced literature of statistics. The approach can be extended to situations in which one observes a greater than expected number of zeroes in the counts. An encyclopedic treatment of count models was presented by Cameron and Trivedi (1998).

—Paul E. Johnson

REFERENCES

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge, UK: Cambridge University Press.

King, G. (1988). Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science, 32,* 838–863.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

EVENT HISTORY ANALYSIS

Event history analysis is a technique that allows researchers to address not only whether an event occurs but also when it occurs. An event is a change from one state to another, and the dependent variable is the time until the event occurs. Event history analysis is ideally suited to the study of longitudinal change and can be thought of as extending LOGIT/PROBIT ANALYSIS and EVENT COUNT MODELS to take into consideration the timing of the event(s).

Two common complications arise in longitudinal data analysis that motivate the use of event history analysis. First, censored observations exist in the data when information about the duration (the amount of time an observation spends in a particular state) is incomplete. This may occur, for example, because the observation did not experience the event of interest prior to the end of the study or because the observation is lost in follow-up, perhaps because the subject moved and could not be located. Second, time-varying covariates (or independent variables) have values that change over time. For example, in a study of the timing of challenger entry in a congressional election, the amount of money raised by an incumbent legislator could be a time-varying explanatory variable across the election cycle. Event history techniques can readily incorporate censored observations and time-varying explanatory variables. The inclusion of time-varying covariates in event history analysis can lead to novel information regarding how the risk of an event occurrence changes in relation to changes in the value of that covariate.

HISTORICAL DEVELOPMENT

Event history analysis is also referred to as *duration*, *survival*, or *reliability analysis*, depending on the substantive origins of the discussion (medicine and engineering for the latter two terms, respectively). Early applications involved life table analysis by Kaplan and Meier (1958), but the historical roots can be traced back even further to the late 1600s (Hald, 1990). There was an increased use of the technique during World War II because of concerns over the expected reliability of military equipment. The path-breaking work of D. R. Cox (1972) is credited with another period of expansion in the use of event history analysis as a result of his development of semiparametric techniques. His work is likely to be heralded as one of the top statistical achievements in the 20th century. Applications in medicine and the social sciences have increased greatly as a result of the less restrictive semiparametric Cox regression model and its various extensions, which are built upon the mathematics of counting processes (Therneau & Grambsch, 2000).

PARAMETRIC AND SEMIPARAMETRIC MODELS

Both parametric and semiparametric models are available in event history analysis. Analysts studying mechanical systems typically use parametric models, which assume that the time until the event of interest follows a specific distribution, such as the exponential. Studies of human behavior and biology typically use the less restrictive semiparametric Cox model, which leaves the particular distributional form of the duration times unspecified. Blossfeld and Rohwer (2002, pp. 180, 263) argued that social science theory rarely provides the justification for a specific parametric distribution and instead advocated for use of the Cox model.

A key concept in the estimation and interpretation of an event history model is the HAZARD RATE. The hazard rate gives the rate at which observations fail (or durations end) by time t given that the observation has survived through time $t-1$. In other words, it can be interpreted as the probability that an event will occur for a particular observation at a particular time. The risk set refers to those observations that are still "at risk" of experiencing the event of interest. Once the observation experiences the event at time t, the observation exits the risk set and is no longer part of the data set being analyzed at $t+1$. The hazard rate has substantive appeal in that the event of interest is conditional on its history. For example, given that a war has lasted t periods, what is the likelihood that it will end in the subsequent period? The hazard rate for the Cox model may be written as

$$h(t|X_i) = h_0(t)e^{X_i\beta},$$

where $h_0(t)$ is an (unspecified) baseline hazard function and X_i are covariates for observation i.

MODEL ASSUMPTIONS AND MODEL FITTING

The major assumption to be checked when fitting event history models is the proportional hazards assumption. Most event history models, including the Cox model, assume that the hazard functions of any two individuals with different values on one or more covariates differ only by a factor of proportionality. Put differently, the baseline hazard rate varies with time but not across individuals, so that the ratio of the hazards for individuals i and j are independent of t and are constant for all t:

$$\frac{h_i(t)}{h_j(t)} = e^{\beta(X_j - X_i)}.$$

Estimation of Cox's model when hazards do not satisfy the proportionality assumption can result in BIASED and inefficient estimates of *all* parameters, not simply those for the covariate(s) in question. The proportional hazards assumption should be checked with Harrell's ρ for individual covariates and with Grambsch and Therneau's global test for nonproportionality (Box-Steffensmeier & Zorn, 2001; Therneau & Grambsch, 2000). If evidence of nonproportionality is found (and, in most social science research, proportionality is more the exception than the rule), then the potentially nonproportional covariates should be interacted with ln(time) or other appropriate transformations of time. Such interactions allow each interacted covariate's effect on the hazard of conflict to vary monotonically with the duration of the event being studied. Relaxing this assumption allows scholars to test whether the effects of covariates change over time and permits a more nuanced understanding of the phenomenon being studied. Moreover, nonproportionality tests, and the residuals upon which they are based, are increasingly easy to obtain in commonly used software packages for analyzing duration data.

If parametric models are estimated, the assumption of the chosen parametric distribution needs to be tested, and the proportional hazard assumption may still need to be assessed (for example, the Weibull model also assumes proportional hazards). The generalized gamma distribution is an encompassing model for several commonly used parametric distributions and thus may serve to help adjudicate among competing nested models. Because the parametric models are estimated by maximum likelihood and the properties of these estimators are well known, the standard battery of goodness-of-fit indices and statistics are directly applicable to the parametric modeling framework, for example, use of the LIKELIHOOD RATIO test or the Akaike information criterion (AIC). However, the principal advantage of the Cox model is not having to make assumptions about the nature and shape of the baseline hazard rate, and thus the Cox model should be the first choice among modeling strategies for social scientists (Box-Steffensmeier & Jones, 1997).

Another underlying assumption of almost all event history models is that all observations eventually will experience the event of interest. This assumption can be relaxed by estimating a split-population model. (These models are also known as cure models in biostatistics, a name based on the idea that part of the population is cured.) As examples, in studies of the timing of campaign contributions, split-population models do not assume that every political action committee (PAC) eventually will give to every political candidate, and in studies of criminal recidivism, the models do not assume that all former prisoners eventually will return to prison. Split-population models estimate the proportion of observations that will not experience the event, together with the parameters characterizing the hazard rate for the proportion experiencing the event. These models allow differential effects for the covariates on whether the event occurred and its timing. For example, a covariate may have a positive effect on whether a contribution is made but a negative effect on when it is made. The appeal of the log-likelihood in a split-population model is that observations that never experience the event contribute information only to part of the function. As such, the log-likelihood "splits" the two populations (Schmidt & Witte, 1988).

DIAGNOSTICS

As in the traditional regression setting, residual analysis in the event history analysis context is a method of checking specification or model adequacy. Various pseudo-residuals are defined in event history analysis for checking different aspects of a model, taking into account the complication that censoring adds to the definition of a residual. In addition to their use in testing the proportional hazards assumption, pseudo-residuals can help the researcher in assessing the model fit, functional form of the covariates, and influence of particular observations. For example, the martingale residuals can be calculated to test whether a given covariate X should be entered linearly, as a quadratic, or in one of the many other possibilities. These diagnostic methods should be used routinely in applications to ensure the integrity of the model (Therneau & Grambsch, 2000).

SINGLE AND MULTIPLE EVENTS

In addition to studying a single-event occurrence, where once the event is experienced, the observation leaves the risk set, event history analysis also can consider multiple events. Event history models for multiple events take into account the lack of independence across events, because ignoring the correlation can yield misleading variance estimates and possibly biased estimates of the coefficients.

Multiple events can be simultaneous unordered events whose risk of occurring varies. In this case, we consider one of several types of "failure," and such processes are referred to as *competing risks*. For example, a member of the U.S. House of Representatives may leave the House in a variety of substantively interesting ways that we should recognize and incorporate into our models, such as being defeated in the primary, being defeated in the general election, running for higher office, or retiring. We would expect that the hazard rate and effects of the covariates will differ across these types of departure.

One can also consider ordered multiple (or *repeated*) events. Repeated-events processes, in which subjects experience the same type of events more than once, are common in fields as diverse as public health, criminology, labor and industrial economics, demography, and political science. Failing to account for repeated events implicitly assumes that the first, second, third, and subsequent events are statistically independent of one another, a strong and usually untenable assumption. The conditional interevent (or gap) time model will be applicable for most instances of repeated events in social science. However, the nature of the means by which repeated events occur (that is, sequentially or simultaneously) and the corresponding construction of the "risk set" for each observation should provide the primary motivation for selecting one model over another (Box-Steffensmeier & Zorn, 2002).

EXAMPLE AND INTERPRETATION

Many applications of event history analysis exist in the social sciences, and the substantive realm of problems being studied is greatly expanding. Examples include studies of the duration of unemployment, peace, survey response time, criminal recidivism, marriages, public policy program implementation, and lobbying. Table 1 presents typical Cox proportional hazard estimates, using militarized conflict data. (Efron's approximation for ties is used in the estimation of the Cox model. The Breslow approximation was the first approximation developed and is not generally

Table 1 Cox Proportional Hazards Model of Conflict

	β (S.E.)	*p value*
Democracy	−0.439(0.100)	< 0.001
Growth	−3.227(1.229)	0.009
Alliance	−0.414(0.111)	< 0.001
Contiguous	1.213(0.121)	< 0.001
Capability ratio	−0.214(0.051)	< 0.001
Trade	−13.162(10.327)	0.202
Wald or LR test	272.35 ($df = 6$)	< 0.001

NOTE: $N = 20{,}448$.

recommended, whereas the exact likelihood option typically gives results extremely similar to those obtained with the Efron approximation while taking considerably longer to converge.)

Oneal and Russett's (1997) widely used data on the relationship among economic interdependence, democracy, and peace is used for the illustration. The data consist of 20,448 observations on 827 dyads (i.e., pairs of states such as the United States and Canada), between 1950 and 1985. We model the hazard of a militarized international conflict as a function of six primary covariates (some of which vary over time): a score for *democracy* (a dyadic score for the two countries which ranges from −1 to 1), the level of *economic growth* (the lesser rate of economic growth, as a percentage, of the two countries), the presence of an *alliance* in the dyad (a dummy variable indicating whether the two countries were allied), the two nations' *contiguity* (a dummy variable for geographic contiguity), their military *capability ratio* (a ratio measuring the dyadic balance of power), and the extent of bilateral *trade* in the dyad (a measure of the importance of dyadic trade to the less trade-oriented country; it is the ratio of dyadic trade to the gross domestic product of each country). (See Oneal and Russett, 1997, for details of the variables and coding.)

Liberal theory suggests that all variables except contiguity ought to decrease the hazard of a dispute, while contiguity should increase it. The likelihood ratio (LR) test at the bottom of Table 1 shows that the specified model is preferred to the null model (i.e., the null hypothesis is that there is no statistically significant difference between the specified model and the null model of no independent variables). All the coefficients are in the expected directions, and all except that for trade are statistically significant. Note that the Cox model does not have an intercept term; it is absorbed into the baseline hazard. Because the coefficients of the Cox model are parameterized in terms of the hazard rate, a positive coefficient indicates that the hazard is increasing, or "rising," with changes in the covariate (and hence survival time is decreasing), and a negative sign indicates the hazard is decreasing as a function of the covariate. For this model, the negative coefficient of −0.439 for democracy suggests that dyadic democracy reduces the likelihood of conflict; that is, dyadic democracy results in a lower hazard (and longer survival time). Box-Steffensmeier and Jones (1997) used the percentage change in the risk of experiencing the event to understand the impact of the effect (p. 1434). For a dichotomous independent variable, the percentage change in the risk of experiencing the event is

$$100[e^{(\beta_k \times 1)} - e^{(\beta_k \times 0)}]/e^{(\beta_k \times 0)}.$$

Negative coefficients produce values of $e^{(\beta_k \times 1)}$ that are less than one, and therefore produce negative percentage changes. The interpretation for a continuous independent variable is similar:

$$100[e^{[\beta_k \times (x+\delta)]} - e^{(\beta_k \times x)}]/e^{(\beta_k \times x)}.$$

This gives the percentage change in the hazard rate for a δ-unit change in the independent variable, x. So, a one-unit increase in the democracy variable corresponds to a $[(e^{(-0.439)} - 1) \times 100] = 36\%$ decrease in the hazard of conflict at any given time.

In actuality, the militarized conflict data are characterized by large numbers of repeated events; for example, Britain and Germany fought each other in both World War I and World War II. Box-Steffensmeier and Zorn (2002) used these data to illustrate repeated events duration modeling and show that important differences are uncovered by taking into account the dependence generated from repeated conflicts.

Social science theories are increasingly focused on change processes, and temporal data are becoming widely available. Event history analysis is ideally suited for leveraging these research elements. The flexibility of the techniques, recent extensions for multiple events, and the incorporation of the observation's history about the events of interest are all compelling reasons to expect the use and popularity of event history techniques to increase in the social sciences.

—Janet M. Box-Steffensmeier

REFERENCES

Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New approaches to causal analysis.* Mahwah, NJ: Lawrence Erlbaum Associates.

Box-Steffensmeier, J. M., & Jones, B. S. (1997). Time is of the essence: Event history models in political science. *American Journal of Political Science, 41,* 336–383.

Box-Steffensmeier, J. M., & Zorn, C. J. W. (2001). Duration models and proportional hazards in political science. *American Journal of Political Science, 45,* 951–967.

Box-Steffensmeier, J. M., & Zorn, C. J. W. (2002). Duration models for repeated events. *The Journal of Politics, 64*(4), 1069–1094.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B, 34,* 187–220.

Hald, A. (1990). *A history of probability and statistics and their applications before 1750.* New York: John Wiley and Sons.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53,* 457–481.

Oneal, J. R., & Russett, B. (1997). The classical liberals were right: Democracy, interdependence, and conflict, 1950–1985. *International Studies Quarterly, 41,* 267–294.

Schmidt, P., & Witte, A. D. (1988). *Predicting recidivism using survival models.* New York: Springer-Verlag.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model.* New York: Springer-Verlag.

EVENT SAMPLING

Event sampling refers to a diverse class of specific empirical methods for studying individual experiences and social processes within their natural, spontaneous context. Event sampling procedures are designed to obtain reasonably detailed accounts of thoughts, feelings, and behaviors as they occur in everyday life. Examples include the experience sampling method (ESM), in which respondents are signaled at random moments during the day and asked to describe their activity at that exact moment, and daily diaries, in which at the end of each day, for some period (typically ranging from 1 week to 1 month), respondents report their experiences on that day. In one typical ESM study, 50 college students were signaled by pagers seven times per day for 1 week. When signaled, students with higher need for intimacy (as assessed on a projective personality test) were more likely to be thinking about people and relationships, were more often engaged in conversations with others and less likely to wish to be alone, and reported more positive affect if they were socializing (McAdams & Constantian, 1983).

Event sampling methods require that participants monitor and describe their ongoing activity along dimensions and according to schedules and formats defined by the researcher. Event sampling has three fundamental rationales that differentiate it from other common research paradigms (e.g., laboratory experiments, surveys): (a) that because behavior is influenced by context, it is important to sample behavior in its natural environment; (b) that global, retrospective reports are often biased by people's limited abilities to remember and summarize numerous events over time; and (c) that accounts of seemingly ordinary, everyday experience, when properly examined, are capable of providing valuable insights about human behavior. Topics for which event sampling methods have been employed profitably include emotion, social interaction, pain, smoking, stress and coping, student motivation, exercise, eating disorders, psychopathology, self-relevant cognition, personality, intergroup relations, and evaluations of drug treatments and therapeutic interventions.

Most event sampling studies employ one of three general protocols: *time-contingent responding*, in which participants report their experiences at fixed intervals (e.g., daily, hourly); *event-contingent responding*, in which a report is solicited whenever a predefined event occurs (e.g., smoking a cigarette, conversing with a friend), and *signal-contingent responding*, in which participants describe their experiences when signaled to do so by some device (e.g., a pager or a preprogrammed portable computer). Signals may follow a fixed or random schedule. Although event sampling was originated with simple paper-and-pencil responses, recent developments in electronic recording devices [e.g., personal digital assistants (PDAs) and voice recorders], as well as in ambulatory physiological monitoring, have added considerably to the validity, flexibility, and range of these methods.

Event sampling is not limited to participant self-reports. For example, in an observational study of schoolchildren, observers might code the behavior of a target child (e.g., what the child is doing, with whom he or she is currently interacting, visible affective expressions, and so on) according to any of the above schedules (e.g., every 10 minutes, after conflict, or following a randomized schedule, respectively).

In a typical event sampling study, a researcher might obtain a series of detailed descriptions of adolescents' momentary moods and actions across a 2-week period. These records could then be used in several ways: for example, to determine how mood differs depending on what the adolescent is doing and with whom he or she is doing it, to isolate personality factors or other individual differences associated with differential patterns of mood or activity, or to predict outcomes, such as school grades, social competence, or health, from a summary of these momentary reports. Regardless of the particulars, event sampling researchers share a fascination with observing ordinary behavior in its natural context and a conviction that such observations hold important clues about key questions in the behavioral and social sciences.

—Harry T. Reis

REFERENCES

Bolger, N., Davis, A., & Rafaeli, E. (in press). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*.

Hektner, J. M., & Csikszentmihalyi, M. (2002). The experience sampling method: Measuring the context and the content of lives. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 233–243). New York: John Wiley & Sons.

McAdams, D. P., & Constantian, C. A. (1983). Intimacy and affiliation motives in daily living: An experience sampling analysis. *Journal of Personality and Social Psychology, 45*, 851–861.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). New York: Cambridge University Press.

Stone, A. A., Shiffman, S., & DeVries, M. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 27–38). New York: Russell Sage.

EXOGENOUS VARIABLE

An exogenous variable is a factor in CAUSAL MODELING or causal system whose value is independent from the states of other variables in the system; that is, it is a factor whose value is determined by factors or variables outside the causal system under study. For example, rainfall is exogenous to the causal system constituting the process of farming and crop output. There are causal factors that determine the level of rainfall—so rainfall is endogenous to a weather model—but these factors are not themselves part of the causal model one would use to explain the level of crop output.

As with endogenous variables, the status of the variable is relative to the SPECIFICATION of a particular model and causal relations among the INDEPENDENT VARIABLES. An exogenous variable is by definition one whose value is wholly causally independent from other variables in the system. The category of "exogenous" variables therefore is contrasted to those of "purely endogenous" and "partially endogenous" variables. A variable can be made endogenous by incorporating additional factors and causal relations into the model.

There are causal and statistical interpretations of exogeneity. The causal interpretation is primary and defines exogeneity in terms of the factor's causal independence from the other variables included in the model. The statistical or ECONOMETRIC concept emphasizes noncorrelation between the exogenous variable(s) and the error term(s) in the model. Typical REGRESSION models assume that all the independent variables are exogenous.

—Daniel Little

See also ENDOGENOUS VARIABLE

REFERENCES

Engle, R. F., Hendry, D. F., & Richard, J. F. (1983). Exogeneity. *Econometrica, 51*, 277–304.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Woodward, J. (1995). Causation and explanation in econometrics. In D. Little (Ed.), *On the reliability of economic models: Essays in the philosophy of economics* (pp. 9–61). Boston: Kluwer Academic.

EXPECTANCY EFFECT. *See* EXPERIMENTER EXPECTANCY EFFECT

EXPECTED FREQUENCY

An expected *value* equals the average or mean of some statistic, whereas an expected *frequency* is a mean count or mean number of times an event occurs. Events could be a category of a single discrete variable, a cell of a cross-classification of two or more discrete variables, or some other specified event (e.g., the number of goals scored by a team during a soccer game). An expected frequency is computed by multiplying the probability that an event occurs by the total number of possible times that the event could occur. For example, consider random samples of size $n = 75$ people from a population in which the probability that an individual is left-handed equals $\pi = 0.10$. The number of left-handers in a sample could be any integer between 0 and 75; however, the expected or mean number of left-handers equals $n\pi = 75(0.10) = 7.5$. Although frequencies are always integers, expected frequencies typically are not integers.

Expected frequencies are used in test statistics for assessing hypotheses and model goodness of fit. Two common test statistics are Pearson's CHI-SQUARE TEST, $\chi^2 = \sum_{i=1}^{N}(f_i - F_i)^2/F_i$, and the LIKELIHOOD RATIO STATISTIC, $L^2 = \sum_{i=1}^{N} f_i \log(f_i/F_i)$, where f_i equals the observed number of observations for category i, F_i equals the expected frequency for category i, and N equals the total number of categories. If a null hypothesis or model for data specifies a value for the probability (or probabilities), such as "the probability of being left-handed is 0.10," then an expected frequency is computed as $F_i = n\pi_i$, where n equals the total number of observations.

The probabilities of events usually are unknown and are estimated from data by computing sample proportions or by fitting a model to data. When data are used to estimate probabilities, the resulting expected frequencies are known as estimated expected frequencies. For example, consider the data in Table 1, which is reformatted from data presented by Jon Cohen (2001). The data consist of the number of smallpox cases in Liverpool, England, during 1902–1903, cross-classified according to whether a person was vaccinated in infancy and whether the person died of smallpox. Assuming that vaccinations do not affect the severity of a case of smallpox, the probability of death as a result of smallpox is estimated by computing the proportion of smallpox cases that resulted in death, that is, the total number of deaths divided by the total number of cases, which equals $\hat{\pi}_{\text{death}} = 79/1,163 = 0.0679$. The estimated expected frequency of death given vaccination equals $\hat{F}_{\text{vac,death}} = n_{\text{vac}}\hat{\pi}_{\text{death}} = 943(0.0679) = 64.0559$. The estimated expected frequencies for other cells in Table 1 could also be computed. For a hypothesis test, the estimated expected frequencies, $\hat{F}_i$, replace the expected frequencies, F_i, in Pearson's chi-square statistic or the likelihood ratio statistic.

Table 1 Smallpox Cases in Liverpool, England (1902–1903)

	Death Due to Smallpox		
	Yes	*No*	*Total*
Vaccinated in infancy	19	924	943
Not vaccinated	60	160	220
Total	79	1,084	1,163

SOURCE: Cohen (2001, p. 985).

—Carolyn J. Anderson

REFERENCES

Agresti, A. (1996). *An introduction to categorical data analysis.* New York: Wiley.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Cohen, J. (2001). Smallpox vaccinations: How much protection remains? *Science, 294,* p. 985.

EXPECTED VALUE

An expected value is the long-run average of a RANDOM VARIABLE X. More formally, the expected value $E(X)$ is a weighted average of all of X's possible values. For a DISCRETE variable, the weights are the PROBABILITIES of the X values, $p(X)$, and the average is obtained by summation:

$$E(X) = \sum_{x=-\infty}^{\infty} xp(x).$$

For a CONTINUOUS VARIABLE, the weights are the probability densities of the X values, $f(X)$, and the average is obtained by integration:

$$E(X) = \int_{x=-\infty}^{\infty} xf(x).$$

For certain distributions, the integral or sum fails to converge. In such cases, the expectation is undefined. Although undefined expectations are fairly rare, a well-known example occurs in connection with the

Cauchy distribution, which is defined as the ratio of two standard normal variables.

To illustrate the calculations, consider the flipping of a fair coin. Let X be 1 if the coin comes up heads, and 0 if it comes up tails. Each value of X has the same probability, $p(X = 1) = p(X = 0) = 0.5$, so the expected value, using the discrete variable formula, is

$$E(X) = 1(0.5) + 0(0.5) = 0.5.$$

Whenever X is a DUMMY VARIABLE, as here, the expected value is the probability that X is 1. When X is an INTERVAL variable, the expected value is the population mean.

Expectation is a linear operation: The expected value for a weighted sum of two random variables is just the weighted sum of the expectations,

$$E(aX + bY) = aE(X) + bE(Y),$$

where a and b are the weights and X and Y are the variables. For example, suppose that X and Y are dummy variables associated with two fair coins, and each variable is 1 if its coin comes up heads. If both coins are tossed, the expected number of heads is

$$E(X) + E(Y) = 0.5 + 0.5 = 1.$$

If two variables are independent, then the expectation of their product is just the product of the expectations:

$$E(XY) = E(X)E(Y),$$

but this relationship does not hold if the variables X and Y are not independent.

The expectation for a general function g of X is just a weighted average of the values of $g(X)$. For a discrete variable X, the weights are the probabilities of the X values, $p(X)$, and the average is obtained by summation:

$$E(g(X)) = \sum_{x=-\infty}^{\infty} g(x)p(x).$$

For a continuous variable X, the weights are the densities of the X values, $f(X)$, and the average is obtained by integration:

$$E(g(X)) = \int_{x=-\infty}^{\infty} g(x)f(x).$$

These formulas extend in a straightforward way to the multivariate case where X and Y are vectors of random variables, and g is a function, possibly vector-valued, of the variable X.

—Paul T. von Hippel

REFERENCE

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury Press.

EXPERIMENT

In social sciences, an experiment is a research strategy used by a social scientist to establish causal relationships between one or more independent variables and one or more dependent variables. An independent variable is a variable that is manipulated by a researcher; its causal impact on the dependent variable is investigated by the researcher. An independent variable is alternatively called the treatment variable or factor. The dependent variable is the observed phenomenon or measurement that has been affected by the manipulation of the independent variable. Dependent variables are alternatively referred to as outcome or criterion variables.

In establishing the causal relationship between independent variables and dependent variables, the researcher should design an experiment that includes the following elements:

- Manipulation of the amount (as in the case of quantitative independent variables) or the level of the independent variable (as in the case of qualitative independent variables)
- Control of nuisance (or confounding) variables using random selection and random assignment of subjects into treatment conditions
- Careful recording or observation of the change in the dependent variable

The first and the second requirements are achieved in either laboratory studies or field experiments (see FIELD EXPERIMENT); they distinguish the experimental research strategy from other research strategies such as quasi-experimental studies, surveys, or naturalistic studies. For these reasons, experiments that possess the above three characteristics are sometimes called true experiments (Campbell & Stanley, 1966).

An example of a true experiment is reported in Cordova and Lepper's (1996) study of elementary school children's learning of arithmetical order-of-operations rules. All learning activities took place in a computerized environment. First, children were randomly assigned, within gender, to either a control condition or one of four experimental conditions. In the control condition, learning materials were presented abstractly. In the four experimental conditions, identical materials were contextualized in a meaningful and visually enhanced format. For half of the students in the four experimental conditions, the learning context was individually personalized; the other half were presented with a generic format. Furthermore, in each experimental condition, half of the group was given choices regarding instructional aspects of the learning context while the other half was not. Thus, the four experimental conditions were (a) learning context personalized with choices, (b) learning context personalized without choices, (c) learning context not personalized but with choices, and (d) learning context not personalized and also without choices. In sum, three independent variables were manipulated in the study: contextualization, personalization, and choice. The dependent variables were several, including students' intrinsic motivation for learning, depth of engagement during learning, the amount of learning acquired in a fixed period of time, students' perceived self-competence, and levels of aspiration. The causal relationship between three independent variables and multiple dependent variables was established through (a) random assignment of subjects into control and experimental conditions; (b) control of miscellaneous variables such as grade level of subjects, gender, and learning materials; and (c) statistical theory of inference making.

Indeed, inferential statistical theories enable social and behavioral scientists to make sense of empirical data collected in the so-called true experiments, in which causal relationships are established not by ruling out all possible alternative explanations or controlling all possible confounding variables, but by probabilities. The logic behind inferential statistical theories is testing statistical hypotheses, in the light of data, followed by making inferences about the underlying populations in the tradition of inductive logical reasoning (Maxwell & Delaney, 1990). The principles and statistical theories that have guided experimental research studies in social and behavioral sciences were, to a great extent, formulated by Karl Pearson (1857–1936), Sir Ronald A. Fisher (1890–1962), and Jerzy Neyman (1894–1981).

Three basic research designs have been employed routinely by social scientists to carry out true experiments or, simply, experiments (Kirk, 1995). They are the completely randomized design, RANDOMIZED-BLOCKS DESIGN, and LATIN SQUARE design. Kirk refers to these as building blocks to all research designs used in experiments. Each is described below.

COMPLETELY RANDOMIZED DESIGN

This research design includes one independent variable and one dependent variable. The independent variable may have two or more levels (or treatment conditions), and the dependent variable is always continuous. If the Cordova and Lepper (1996) study cited earlier is simplified to include only one independent variable, say *contextualization*, and the dependent variable observed is *intrinsic motivation*, the simplified study design will be a bona fide completely randomized design provided that subjects are randomly assigned to either the contextualized or the abstract condition at the onset. A layout of this simplified study may look like Table 1:

Table 1

	Independent Variable	
Subject	*Contextualized Condition*	*Abstract Condition*
S_1	Y_{11}	Y_{12}
S_2	Y_{21}	Y_{22}
.	.	.
.	.	.
.	.	.
S_{15}	$Y_{1,15}$	$Y_{2,15}$

The statistical linear model for the data collected in a completely randomized design is

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \quad (i = 1, \ldots, n \text{ or 15 in this case; } j = 1, \ldots, p \text{ or 2 in this case}),$$

where

- Y_{ij} is the ith subject's score in the jth treatment condition of the independent variable
- μ is the grand mean, therefore a constant, of the population of all subjects
- α_j is the treatment effect of the jth treatment condition of the independent variable; algebraically, it equals the deviation of the population mean (μ_j) from the grand mean (μ). It is a constant for all subjects' dependent scores in the jth condition, subject to the constraint that all α_j sum to zero across all treatment conditions of the independent variable

ε_{ij} is the random error effect associated with the ith subject randomly assigned to the jth level of the independent variable; it is a random variable that is normally distributed in the underlying population and is independent of the effect α_j

The null (H_0) and alternative (H_1) hypotheses for the completely randomized design are

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0,$$
$$H_1: \text{at least one } \alpha_p \neq 0.$$

To test the null hypothesis, data are organized to form a ratio of MEAN SQUARES between (MS_b) over mean squares within (MS_w). The ratio is distributed as an F distribution provided that the normality, homogeneity of variance, and independence assumptions about the data are met. The normality assumption states that the underlying distribution of data in all treatment conditions of the independent variable is the normal distribution. The homogeneity of variance assumption is that all normal population distributions have identical variances. Finally, the independence assumption says that random error effects, ε_{ij}, are independent from each other both within and across treatment conditions. Failure to meet these assumptions results in quasi F ratios that may inflate the Type I error rate (Glass, Peckham, & Sanders, 1972). The example given above is obviously the simplest completely randomized design, with only two treatment conditions of the independent variable. In this case, the test statistic can be either an F or t, as in the comparison of two independent group means. Most applications of the completely randomized design in social sciences involve three or more treatment conditions.

If treatments of the independent variable included in a study (e.g., contextualized and abstract) are the only treatments that a researcher is interested in investigating and, consequently, making generalizations about, the corresponding independent variable is regarded as a fixed factor, and the design is a FIXED-EFFECTS DESIGN. The statistical model prescribed for a fixed-effects design is MODEL I (Kirk, 1995). If, however, treatments included in a study represent only a sample of all treatments that a researcher wishes to study, the corresponding independent variable is treated as a random factor. The design is a random-effects design, and MODEL II (ANOVA) is an appropriate statistical model for the data (Kirk, 1995).

RANDOMIZED-BLOCKS DESIGN

The use of completely randomized design requires that subjects be randomly assigned into treatments of an independent variable. Hence, subjects in a treatment condition are considered an independent sample (or subsample) from subjects in another treatment condition. This characteristic of independence is factored into the calculation of the F or t statistic when the statistical test is performed. For many studies, such as those involving repeated measurements of the same subjects, it is not possible to maintain the independence among subjects in different treatment conditions. For other research, it is desirable (as in family or marital research) to pair up subjects (such as husbands and wives or classmates) so that researchers can study the dynamics of the relationship between subjects. Researchers may purposely introduce dependence by matching subjects on certain characteristics that are related to the dependent measures to ensure that these characteristics are "controlled" in the study. For example, in the simplified Cordova and Lepper (1996) study described in the previous section, one might wish to match subjects on prior knowledge that is related to the dependent measure (i.e., *intrinsic motivation*). By so doing, the source of variance that is due to prior knowledge is planned into the study and, therefore, can be isolated from the source of variance that is due to the treatment (i.e., *contextualization*). Let's further assume that prior knowledge is differentiated at three levels: high, medium, and low. An appropriate research design to control for prior knowledge is the randomized-blocks design. Table 2 illustrates the design features.

The variable Prior Knowledge is technically called a blocking variable; its levels (high, medium, and low) are blocks. Within each block of the blocking variable, subjects (e.g., $S_1, \ldots, S_5$) are homogeneous insofar as prior knowledge is concerned. Subjects in each block are randomly assigned to treatment conditions (e.g., contextualization condition or abstract condition). The random assignment of subjects to treatments within each block and the planned differences across blocks ensure that the idiosyncratic differences of subjects are distributed randomly among treatment conditions, and differences on the blocking variable are equally

Table 2

Prior Knowledge	Subject	*Independent Variable* Contextualized Condition	Abstract Condition
High	S_1	Y_{11}	Y_{12}
	.	.	.
	.	.	.
	.	.	.
	S_5	Y_{15}	Y_{25}
Medium	S_6	Y_{16}	Y_{26}
	.	.	.
	.	.	.
	.	.	.
	S_{10}	$Y_{1,10}$	$Y_{2,10}$
Low	S_{11}	$Y_{1,11}$	$Y_{2,11}$
	.	.	.
	.	.	.
	.	.	.
	S_{15}	$Y_{1,15}$	$Y_{2,15}$

represented in each treatment condition. Thus, results are not systematically biased as a result of differences in the blocking variable. In this sense, the randomized-blocks design can isolate and attribute a portion of score variance to the blocking variable, whereas the completely randomized design cannot.

The statistical model assumed for a randomized-blocks design is

$$Y_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij} \quad (i = 1, \ldots, n \text{ or 3 in the example above; and } j = 1, \ldots, p \text{ or 2 in the example}),$$

where

- Y_{ij} is the score of the ith block and the jth treatment condition
- μ is the grand mean, therefore a constant, of the population of observations
- α_j is the treatment effect for the jth treatment condition; algebraically, it equals the deviation of the population mean ($\mu_{.j}$) from the grand mean (μ or $\mu_{..}$); it is a constant for all observations' dependent score in the jth condition, subject to the restriction that all α_j sum to zero across all treatment conditions as in a fixed-effects design, or the sum of all α_j equals zero in the populations of treatments, as in the random-effects model
- π_i is the block effect for population i and is equal to the deviation of the ith population mean ($\mu_{i.}$) from the grand mean ($\mu_{..}$); it is a constant for all observations' dependent score in the ith block, normally distributed in the underlying population
- ε_{ij} is the random error effect associated with the observation in the ith block and jth treatment condition; it is a random variable that is normally distributed in the underlying population of ith block and jth treatment condition and is independent of π_i.

The null (H_0) and alternative (H_1) hypotheses for the randomized-blocks design are identical to those for the completely randomized design. To test the null hypothesis, data are organized to form a ratio of mean squares treatment (MS_t) over mean squares residual (MS_r). Mean squares residual is a ratio of the sum of squares residual over its degrees of freedom, where the sum of squares residual is the difference of sum of squares within (SS_{within}) minus the sum of squares due to the blocking variable. The ratio is distributed as an F distribution provided that the normality, homogeneity of variance, and independence assumptions about the data are met.

The example given above is the simplest randomized block design, with only two treatment conditions of the independent variable. In this case, the test statistic can be either an F or t, as in the comparison of two dependent (or matched pair) group means. Most applications of the randomized block design in the social sciences involve three or more treatment conditions.

If treatments of the independent variable included in a study (e.g., contextualized and abstract) are regarded as fixed and the blocking variable is regarded likewise, the design is a fixed-effects design. The statistical model prescribed for a fixed-effects design is Model I (Kirk, 1995). If, however, the blocking variable is regarded as a RANDOM FACTOR, then the corresponding design is a MIXED DESIGN and the model is MODEL III (ANOVA). If treatments included in a study represent only a sample of all treatments, and hence the corresponding independent variable is a random factor, the design is a random-effects design and Model II is an appropriate statistical model for the data (Kirk, 1995).

When more than one subject is assigned to the combination of block and treatment condition, the design is referred to as the generalized randomized block design. This is the case with the example given here.

Alternatively, the same subject (or observation) can serve as a block and is subject to all treatments in a random sequence. Some authors prefer to call this specific variation of block designs a REPEATED-MEASURES DESIGN. Details on how the blocks are formed and on additional statistical assumptions uniquely associated with this design and its corresponding statistical test are found in the entries titled "Block Designs" and "Randomized-Blocks Design."

LATIN SQUARE DESIGN

The Latin square design is an experimental design or QUASI-EXPERIMENT that can isolate and attribute a portion of score variance to two blocking variables, whereas the randomized block design controls for only one blocking variable.

Much like the Rubik's cube, all Latin square designs are square (2 by 2, 3 by 3, 4 by 4, etc.); that is, the two blocking variables and the independent variable all have the same number of levels (or conditions). Let us assume that the second blocking variable we wish to control in the simplified Cordova and Lepper study described previously is gender, which has two levels. The independent variable also has two levels: contextualized and abstract. Consequently, the first blocking variable, prior knowledge, has to be redefined to comprise only two levels: high and low. Table 3 illustrates this 2 by 2 Latin square design.

Table 3

	Gender	
Prior Knowledge	*Female*	*Male*
High	Contextualized condition	Abstract condition
Low	Abstract condition	Contextualized condition

Notice that in Table 3, the treatments are shown inside the square, while the two blocking variables are presented as row and column variables. Indeed, treatment conditions occur only once in each row and once in each column.

The statistical model for the above Latin square design is

$$Y_{ijkl} = \mu + \alpha_j + \beta_k + \gamma_l + \varepsilon_{\text{pooled}},$$
$$(i = 1, \ldots, n \text{ in the example above; } j, k, \text{ and } l = 1, \ldots, p \text{ or 2 in the example}),$$

where

- Y_{ijkl} is the score of the ith subject in the jth treatment condition, kth level of the first block variable (e.g., prior knowledge) and the lth level of the second blocking variable (e.g., gender)
- μ is the grand mean, therefore a constant, of the population of observations
- α_j is the treatment effect for the jth treatment condition; algebraically, it equals the deviation of the population mean ($\mu_{j..}$) from the grand mean (μ or $\mu_{...}$); it is a constant for all subjects' dependent score in the jth condition, subject to the restriction that all α_j sum to zero across all treatment conditions as in a fixed-effects design, or the sum of all α_j equals zero in the populations of treatments, as in the random-effects model
- β_k is the block effect for the kth population and is equal to the deviation of the kth population mean ($\mu_{.k.}$) from the grand mean ($\mu_{...}$); it is a constant for all subjects' dependent score in the kth block, normally distributed in its underlying population
- γ_l is the block effect for the lth population and is equal to the deviation of the lth population mean ($\mu_{..l}$) from the grand mean ($\mu_{...}$); it is a constant for all subjects' dependent score in the lth block, normally distributed in its underlying population
- $\varepsilon_{\text{pooled}}$ is the pooled error effect associated with Y_{ijkl}; algebraically, this concept equals the deviation of the observed score Y_{ijkl} from all other parameters in the model (i.e., μ, α_j, β_k, and γ_l) and, by definition, it is statistically independent of all the parameters in the model

The null (H_0) and alternative (H_1) hypotheses for the Latin square design continue to be identical to those for the completely randomized design. To test the null hypothesis, data are organized to form a ratio of mean squares treatment (MS_t) over mean squares pooled error (MS_{pooled}). Mean squares pooled error is a ratio, consisting of the sum of squares pooled error over its degrees of freedom, where the sum of squares pooled error is the difference of the sum of squares within (SS_{within}) minus the sum of squares due to *two* blocking variables. The ratio is distributed as an F distribution provided that the normality, homogeneity of variance, and independence assumptions about the data are met. Furthermore, the F test assumes that no interaction exists among the two blocking variables and the independent variable (also the treatment). To test this interaction assumption, a Latin square design must have two or more observations per cell (i.e., $n \geq 2$). Otherwise, this assumption of no interaction is assumed for data.

If treatments of the independent variable included in a study (e.g., contextualized and abstract) are regarded as fixed and the two blocking variables are regarded likewise, the design is a fixed-effects design. The statistical model prescribed for a fixed-effects design is Model I (Kirk, 1995). If, however, one of the blocking variables is regarded as a random factor, the corresponding design is a mixed-effects design and the model is Model III. If treatments included in a study represent only a sample of all treatments, and hence the corresponding independent variable is a random factor, the design is a random-effects design and Model II is an appropriate statistical model for the data. For additional readings on Latin square designs, readers are encouraged to consult the entry titled "Latin Square Design," Kirk (1995), or Maxwell and Delaney (1990).

For the three research designs and the corresponding statistical models presented above, it is assumed that data are collected and measured without errors. If measurement errors are present in the data, RELIABILITY needs to be taken into consideration when the F statistic is computed (Cleary & Linn, 1969; Levin & Subkoviak, 1977). Furthermore, the VALIDITY of any causal relationship established between independent variables and one or more dependent variables in an experiment can be threatened by a number of factors. For details on threats to external, internal, and statistical conclusion validities, readers are encouraged to consult Campbell and Stanley (1966) and Kirk (1995).

Within the field of probability theory, an EXPERIMENT is a process or procedure that leads to a single outcome whose probability of occurrence is to be determined. An experiment can be a laboratory study (e.g., engaging students in collaborative learning to improve their social skills) or a process of observing an aspect of behavior in a sample taken from a population (e.g., noting the birth months of two randomly chosen guests at a dinner party of 50). Both the experiment and the outcome (the improvement of social skills or two guests born in the same month) should be well defined such that a researcher could formulate rules for determining the probability (likelihood) of obtaining the specific outcome, among all possible outcomes. The approach to probability determination may be based on either (a) the subjective-personalistic framework, (b) the logical or classical framework, or (c) the empirical-relative or frequency framework. In this context, an experiment may also be called a simple experiment.

—Chao-Ying Joanne Peng

See also BLOCK DESIGN, FIELD EXPERIMENTATION, LATIN SQUARE, QUASI-EXPERIMENT, RELIABILITY, VALIDITY

REFERENCES

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally College Publishing.

Cleary, T. A., & Linn, R. L. (1969). Error of measurement and the power of a statistical test. *The British Journal of Mathematical and Statistical Psychology, 22*(1), 49–55.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally College Publishing.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730.

Fisher, R. A. (1971). *Design of experiments.* New York: Hafner Press. (Original work published 1935.)

Glass, G. V, Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. *Review of Educational Research, 42,* 237–288.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Levin, J. R., & Subkoviak, M. J. (1977). Planning an experiment in the company of measurement error. *Applied Psychological Measurement, 1*(3), 331–338.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

Vogt, W. P. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (2nd ed.). Thousand Oaks, CA: Sage.

EXPERIMENTAL DESIGN

CHARACTERISTIC FEATURES OF EXPERIMENTS

Two characteristics set experimentation apart from other methods of social inquiry. First, experimentation involves a planned intervention. Unlike passive observation, in which the researcher attempts to trace the effects of naturally occurring fluctuations in putative causes, experiments create a disturbance in the environment and track its consequences. For example, an observational study of welfare reform might examine whether rates of joblessness vary with over-time or cross-jurisdiction variations in program requirements. By contrast, an experimental investigation would assign program participants to different sets of requirements in order to examine whether they affected joblessness.

The second characteristic of experiments is RANDOM ASSIGNMENT. Although the term EXPERIMENT is used loosely in common parlance to refer any type of intervention, in social science it has come to refer to studies in which units of observation are assigned at random to treatment and control conditions. Because social scientists cannot create the equivalent of a physics lab, in which all extraneous causes are eliminated, they instead rely on random assignment, which, as R. A. Fisher pointed out in his classic work *The Design of Experiments* (1935), is the only *procedure* that guarantees the comparability of treatment and control groups. So long as randomization is carried out faithfully, we can be sure that treatment and control groups differ solely as a result of chance.

Randomization therefore enables researchers to make precise statistical statements about the likelihood that any postintervention differences between treatment and control groups resulted from fortuitous differences between the groups, as opposed to the intervention in question. Precise probability statements typically are unavailable in non-experimental social science, where the independent variables are generated by an unknown process. Beyond the usual statistical uncertainty that arises from limited sample size, non-experimental research involves considerable uncertainty about whether the observed correlation between two variables, X and Y, reflects the causal influence of X on Y, of Y on X, or of some third variable Z on both X and Y.

PROBLEMS CONFRONTING EXPERIMENTAL DESIGN

Although, in principle, experiments represent the most reliable means of drawing causal inferences, in practice experiments confront several problems. Problems of INTERNAL VALIDITY arise when an experiment fails to isolate the true effects of a particular causative agent. For example, the Lanarkshire milk experiment, an early study designed to test whether distributing milk in schools improved students' growth rates, was undone when teachers reassigned underweight students in the control group to receive milk supplements. Thus, the contrast between treatment and control groups reflected both the effects of milk and the vagaries of teacher reassignments.

Concerns about EXTERNAL VALIDITY arise when the environment within which an experiment takes place or the people who participate in the experiment differ in important respects from the places and populations about whom the researcher intends to generalize. This concern arises frequently when LABORATORY EXPERIMENTS attempt to simulate economic or political environments using college students as subjects. In response to concerns about external validity, social scientists have turned increasingly to FIELD EXPERIMENTS, or experiments conducted in real-world settings. Such studies have examined the effectiveness of voter mobilization campaigns, preschool education programs, methods for encouraging compliance with tax rules, and a wide array of other interventions. Nevertheless, practical constraints limit the range of field experiments. Social scientists lack the resources and authority to manipulate large-scale causative factors, such as legislative institutions or the religiosity of the population.

Related to concerns about external validity is the issue of homogeneous treatment effects. Are all subjects equally influenced by a given intervention? If the answer is yes, the range of research opportunities expands, because one may study situations in which only some of those assigned to the treatment group actually receive treatment. For example, if a voter mobilization campaign has the same effect on everyone it contacts, but it reaches only half of the people it seeks to contact, its effect is twice as strong as a naive comparison of treatment and control groups would suggest. The assumption of homogeneous treatment effects, in other words, enables us to extrapolate easily from those who were actually exposed to a treatment to those whom researchers sought to treat.

Whether treatment effects are in fact homogeneous is an empirical question, not unlike the question of whether college students are as susceptible to social influences as those outside the university. Concerns about homogeneity and external validity underscore the importance of REPLICATING experimental findings in different settings and populations.

Ethical limitations are a further constraint. Experimental interventions may adversely affect the subjects involved, as well as the broader society. Besides the practical constraints of devising feasible interventions, researchers much follow procedural safeguards to ensure that subjects are neither harmed nor coerced. These considerations generally remove from consideration far-reaching experiments involving social or economic policy, but as Donald T. Campbell (1969) pointed out, the social costs of *not* conducting experiments must also be taken into account. Governments, firms, and organizations continually intervene in the world, and the question is whether their interventions can be structured in such a way as to generate useful knowledge.

EXPERIMENTS AND SCIENTIFIC ADVANCEMENT

Despite these limitations, experiments remain the gold standard for adjudicating causal claims. A single well-crafted experiment—conducted in a real-world setting and of sufficient size to produce statistically precise conclusions—can overshadow a large body of research based on observational data.

Part of the allure of experiments is their elegant transparency. In contrast to non-experimental data analysis, the analysis of experimental results often requires little more than elementary statistical methods, and the choice of statistical techniques seldom has a material bearing on the results. The experimental design largely dictates the manner in which the data will be analyzed statistically; committing to a plan of analysis ex ante helps guard against post hoc decisions that may bias the results in a particular direction. Experimental procedures not only lead to clearer causal inferences but also free the analyst from the moral hazards of data mining.

—Donald P. Green

See also EXPERIMENT

REFERENCES

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*, 409–429.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

Fisher, R. A. (1935). *The design of experiments.* New York: Hafner Publishing.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Wilmington, MA: Houghton Mifflin.

EXPERIMENTER EXPECTANCY EFFECT

The experimenter expectancy effect is one of the sources of artifact or error in scientific inquiry (see INVESTIGATOR EFFECT). Specifically, it refers to the unintended effect of experimenters' hypotheses or expectations on the results of their research.

Some expectation of how the research will turn out is virtually a constant in science. Social scientists, like other scientists generally, conduct research specifically to examine hypotheses or expectations about the nature of things. In the social and behavioral sciences, the HYPOTHESIS held by the investigator can lead him or her unintentionally to alter behavior toward the research participants in such a way as to increase the likelihood that participants will respond so as to confirm the investigator's hypothesis or expectations. We are speaking, then, of the investigator's hypothesis as a self-fulfilling prophecy: One prophesies an event, and the expectation of the event then changes the behavior of the prophet in such a way as to make the prophesied event more likely. The history of science documents the occurrence of this phenomenon with the case of the horse Clever Hans as a prime example (Pfungst, 1911/1965).

The first experiments designed specifically to investigate the effects of experimenters' expectations on the results of their research employed human research participants. Graduate students and advanced undergraduates in the field of psychology were employed to collect data from introductory psychology students. The experimenters showed a series of photographs of faces to research participants and asked participants to rate the degree of success or failure reflected in the photographs. Half the experimenters, chosen at random, were led to expect that their research participants would rate the photos as being of more successful people. The remaining half of the experimenters were given the opposite expectation—that their research participants would rate the photos as being of less successful people. Despite the fact that all experimenters were instructed to conduct a perfectly standard experiment, reading only the same printed instructions to all their participants, those experimenters who had been led to expect ratings of faces as being of more successful people obtained such ratings from their randomly assigned participants. Those experimenters who had been led to expect results in the opposite direction tended to obtain results in the opposite direction.

These results were replicated dozens of times employing other human research participants. They also were replicated employing animal research subjects. In the first of these experiments, experimenters were employed who were told that their laboratory was collaborating with another laboratory that had been developing genetic strains of maze-bright and maze-dull rats. The task was explained as simply observing and recording the maze-learning performance of the maze-bright and maze-dull rats. Half the experimenters were told that they had been assigned rats that were maze-bright while the remaining experimenters were told that they had been assigned rats that were maze-dull. None of the rats had really been bred for maze-brightness or maze-dullness, and experimenters were told purely at random what type of rats they had been assigned. Despite the fact that the only differences between the allegedly bright and dull rats were in the minds of the experimenters, those who believed their rats were brighter obtained brighter performance from their rats than did the experimenters who believed their rats were duller. Essentially the same results were obtained in a replication of this experiment employing Skinner boxes instead of mazes.

—Robert Rosenthal

See also PYGMALION EFFECT

REFERENCES

Pfungst, O. (1965). *Clever Hans.* Translated by C. L. Rahn. New York: Holt, Rinehart and Winston. (Original work published 1911.)

Rosenthal, R. (1976). *Experimenter effects in behavioral research: Enlarged edition.* New York: Irvington Publishers, Halsted Press Division of Wiley.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart and Winston.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research.* New York: W. H. Freeman.

EXPERT SYSTEMS

Expert systems (ES), also called *knowledge-based systems*, are ARTIFICIAL INTELLIGENCE (AI) computer programs with sophisticated decision-making and reasoning capabilities for limited domains of knowledge. Usually using ABDUCTION (reasoning toward an inference), they are more than smart databases because they mimic human thinking. For social scientists, ES offer ways to model different theoretical frameworks and assumptions for theory testing and SIMULATION. They also provide rich applications for instruction, research design, problem solving, identification, diagnosis, and control.

Typically, ES comprise four parts: (a) the knowledge base proper, which is the body of declarative knowledge (the "stuff") and procedural knowledge (how to use the "stuff") about a domain; (b) the inference engine, which provides means of reasoning; (c) the knowledge-acquisition interface, which specifies modes of entry of knowledge and the forms of rules; and (d) the user interface, which can give intuitive, human-like end-user access. Problems amenable to modeling by ES are of magnitude equivalent to those solvable by a telephone consultation with a human expert.

By way of example, the rule-based ES illustrated in Table 1 constitutes a parsimonious account to predict whether pregnant women will present early or late for first prenatal care. It was constructed from FIELDNOTES of interviews with women after their first prenatal visit. It performed comparably to a MULTIPLE REGRESSION ANALYSIS of the same data and was more easily interpretable.

Table 1

Rule 1	IF Encouraged by Influential Other = yes, THEN *present early*
Rule 2	IF Encouraged by Influential Other = no, OR Encouraged by Mother = yes, THEN *present late*
Rule 3	IF Encouraged by Influential Other = indeterminate (from interview data) AND Early Symptoms = yes, THEN *present early*, ELSE (i.e., if no) *present late*

Although frequently written in AI languages such as C++, Prolog, or Lisp, many ES are developed using programming environments called *shells*. A typical shell offers three of the four critical parts of an expert system—the inference engine, the knowledge-acquisition interface, and the user interface—lacking only the knowledge base provided by the developer.

Social scientists find creating an ES a useful, sobering exercise in formulating and checking their understandings because ES require specificity of both assumptions and heuristics (the informal rules of thumb that experts use in manipulating knowledge). Permitting abduction-based, INDUCTION-based, and PROBABILISTIC reasoning, an ES is more a MODEL than a rule-based approach.

Related methods include data mining, an inductive approach, and fuzzy systems, where class-inclusion may have fuzzy boundaries. Disadvantages are the inability to deal with unexpected relations and maintenance, although ES shells are more self-documenting than is, say, C++.

Benfer, Brent, and Furbee (1991) provided an accessible introduction for social scientists. Online resources are rich and include the sites www.aaai.org/AITopics/html/expert.html (with accessible general and how-to information) and www.pcai.com/ai_info/expert_systems.htmlr (with general information, from a business perspective, and resources including freeware and shareware).

—N. Louanna Furbee

REFERENCE

Benfer, R. A., Brent, E. E., Jr., & Furbee, L. (1991). *Expert systems*. Newbury Park, CA: Sage.

EXPLAINED VARIANCE. *See* *R*-SQUARED

EXPLANATION

A scientific explanation of an event or regularity is an argument that purports to demonstrate why the event or regularity came to pass, given other features of the world. Why was the outcome necessary or probable in the circumstances and in the presence of relevant governing laws or regularities? The most general model of a scientific explanation is that of a "covering-law explanation." A covering-law explanation is an argument consisting of one or more general laws, one or more particular statements (boundary conditions), and deductive derivation of a statement of the phenomenon to be explained (the explanandum). Such explanations are designed to show why the outcome was necessary in the circumstances, given the initial conditions and the laws of nature. A covering-law explanation subsumes the phenomenon under the general laws. A probabilistic explanation has a similar logic. Probabilistic explanations identify one or more statistical laws and subsume the phenomenon under these laws: Given probabilistic laws L_i, the probability of O is P. These approaches to the logic of scientific explanation give primacy to the role of scientific laws or laws of nature in explanation. Using this approach, we have explained an outcome when we have shown how it is necessary (or probable), given the relevant laws of nature.

A different approach to scientific explanation proceeds from the point of view that the world is a system of causal processes and mechanisms. Using this approach, we have explained an outcome when we have provided an account of the causal mechanisms and powers that led to the occurrence of the outcome. Causal explanations proceed by identifying causal mechanisms through which the initial conditions brought about the explanandum. The two approaches are related because the presence of causal mechanisms also implies the availability of lawlike generalizations that can function within covering-law explanations. The causal mechanism approach, however, comes closer to the intellectual task of scientific explanation. We want to know why the event occurred when we ask for an explanation of an outcome. Ptolemy's explanation of the observed locations of the planets in the heavens proceeded on the basis of subsumption of planetary motion under a set of lawlike generalizations (cycles and epicycles). The explanation was unsatisfactory because it rested upon a false conception of the causal processes that resulted in astronomical observations (geocentric rather than heliocentric motion of the planets).

—Daniel Little

REFERENCES

Achinstein, P. (1983). *The nature of explanation*. New York: Oxford University Press.

Hempel, C. (1965). *Aspects of scientific explanation, and other essays in the philosophy of science*. New York: Free Press.

Kitcher, P., & Salmon, W. C. (Eds.). (1989). *Scientific explanation* (Minnesota Studies in the Philosophy of Science, Vol. 13). Minneapolis: University of Minnesota Press.

Little, D. (1991). *Varieties of social explanation: An introduction to the philosophy of social science.* Boulder, CO: Westview.

Miller, R. W. (1987). *Fact and method: Explanation, confirmation and reality in the natural and the social sciences.* Princeton, NJ: Princeton University Press.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

EXPLANATORY VARIABLE

An explanatory variable is a factor that is used by the data analyst to explain a phenomenon. In a REGRESSION type of model, it is also known as an INDEPENDENT VARIABLE, a regressor, an input variable, a right-hand-side variable, or an X variable.

—Tim Futing Liao

EXPLORATORY DATA ANALYSIS

Exploratory data analysis (sometimes abbreviated as EDA) consists of an approach to data analysis that allows the data themselves to reveal their underlying structure and that gives the researcher a "feel" for the data. It relies heavily on graphs and displays to reach these goals because visual inspection offers special insight into the data, but it also uses many numerical techniques. Both the visual and numerical techniques focus on searching for knowledge that allows more effective testing of theories and hypotheses. Given that exploratory data analysis represents a general philosophy of understanding data, the techniques serve as a means to that end rather than ends in themselves.

The philosophy of exploratory data analysis can be said to value two traits: openness and skepticism (Hartwig & Dearing, 1979, pp. 9–12). Analysts should be open to unanticipated data patterns that go beyond the planned MODEL and expectations. Analysts should also be skeptical of numerical summaries of data that can conceal or misrepresent the most informative aspects of the data.

HISTORICAL DEVELOPMENT

The seminal works on exploratory data analysis come from John W. Tukey (1977) and Frederick Mosteller and John W. Tukey (1977). Tukey (1977, p. vii) noted that over the 20th century, confirmatory data analysis, which assesses how precisely SAMPLE statistics can be used to make inferences about POPULATION parameters, had come to dominate statistical research. He believed that although exploratory and confirmatory data analyses emphasize different principles, both should be used by practicing researchers. In aiming to complement the dominant methods of statistical analysis, he argued that researchers cannot get along without confirmatory data analysis but need not start with it. His central principle follows from this point (Tukey, 1977): "It is important to understand what you CAN DO before you learn to measure how WELL you seem to have DONE it" (p. v).

Consider the differences between confirmatory and exploratory data analysis (NIST/SEMATECH, 2003):

- Confirmatory approaches begin with a prespecified model and use the parameters obtained from the data analysis to evaluate the model, whereas exploratory approaches begin with analysis of the data to infer the model and specify its ASSUMPTIONS.
- Confirmatory approaches are formal and rigorous in efforts to evaluate a model, whereas exploratory approaches are more informal and active in the effort to discover meanings contained in the data.
- Confirmatory approaches impose a model on the data, whereas exploratory approaches let the data suggest appropriate models.
- Confirmatory approaches filter the data in reducing the information to a small number of parameters, whereas exploratory approaches rely on techniques that reflect all parts of the data.
- Confirmatory approaches require strong assumptions, whereas exploratory approaches make few assumptions or attempts to validate assumptions before model testing.
- Confirmatory approaches have much power to precisely test a HYPOTHESIS and estimate parameters under narrow circumstances, whereas exploratory approaches have greater generality and less sensitivity.

The two approaches represent ideal types that, in good research, both receive attention, but Tukey and others have had considerable influence in countering a belief that mere description and exploration of data were inferior to formal hypothesis testing and model confirmation. Today, most empirical social scientific research relies on confirmatory models, such as in the use of REGRESSION and analysis of variance to test theories, but exploratory techniques have become essential parts of the research process leading to the confirmatory models.

EXAMPLES

The long volumes of Tukey (1977) and Mosteller and Tukey (1977) offer dozens of techniques and statistics—as well as a new terminology—for exploratory data analysis. A few examples of the techniques and statistics, and how they contribute to the goals of exploratory data analysis, can illustrate their potential use for researchers.

To uncover hidden data structures, the use of displays and graphs in exploratory data analysis helps researchers to notice what they never expected to see. A STEM-AND-LEAF DISPLAY and a box-and-whisker plot (see BOXPLOT) each depict the full DISTRIBUTION of the values of a single VARIABLE in a visual way, and a SCATTERPLOT depicts the relationship between two variables with all values of the variables. The visual representations do much to identify any DEVIATION from normality (in a single variable) and from linearity (in the relationship between two variables), and thereby help evaluate assumptions commonly made in the use of LEAST SQUARES regression.

To counter non-normality and NONLINEARITY, exploratory data analysis searches for ways to reexpress the SCALE of measurement of variables. Reexpression alters the distance between values of a variable while maintaining the sequence or ordering. Taking powers, roots, and logs of variables normalizes distributions and straightens relationships, making the use of the transformed variables more appropriate for multivariate analyses. Techniques to smooth sequential variables similarly help to reveal underlying patterns in the data.

To best summarize distributions and relationships, exploratory data analysis relies on the computation of ROBUST and resistant measures. By reflecting the bulk of the data, resistant measures are less sensitive to DEVIANT CASE ANALYSES. Such measures are often based on a MEDIAN rather than a mean, on absolute values rather than squared values, and on least absolute deviations rather than least squared deviations. Various robust regression techniques further provide ways to make models less sensitive and discount the disproportional influence of outlying cases.

To evaluate the suitability of a model, exploratory data analysis gives much attention to OUTLIERS and RESIDUALS. Least squares analyses of relationships among variables are accompanied by visual inspections of the scatterplots and deviations of points from the regression line. Many numerical techniques also identify large outliers and residuals in a multivariate context but cannot fully replace the benefits obtained from the visual contrast of the model with the data. In turn, the attention to outliers and residuals helps determine how to reexpress variables, respecify models to better describe relationships, and gain new insights into the topics under investigation.

—Fred C. Pampel

REFERENCES

Hartwig, F., & Dearing, B. E. (1979). *Exploratory data analysis* (Sage University Papers on Quantitative Applications in the Social Sciences). Beverly Hills, CA: Sage.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics.* Reading, MA: Addison-Wesley.

NIST/SEMATECH. (2003). *E-handbook of statistical methods.* Retrieved from http://www.itl.nist.gov/div898/handbook/eda

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

EXPLORATORY FACTOR ANALYSIS.

See FACTOR ANALYSIS

EXTERNAL VALIDITY

External validity is a property that allows research findings to be generalized to a larger POPULATION. External validity is a concern when constructing experimental and non-experimental (OBSERVATIONAL) RESEARCH designs. Researchers can ensure external validity through careful construction of their research design.

There are three main threats to external validity: nonrepresentative SAMPLES, an artificial laboratory environment, and testing effects. One of the main goals of research is to generalize the findings to a larger population. Consider a study on the effect that negative campaign advertising has on voters' approval of the incumbent candidate. For the study to be feasible, it cannot be administered to all possible voters, the population of interest. Instead, a sample of the population is selected to take part in the study. Special attention needs to be paid so that the underlying causal process is the same for both the sample and the population of interest. Finding a REPRESENTATIVE SAMPLE is sometimes difficult. Many times, participants are recruited from local communities or college campuses where the study is taking place. A campaign advertising study based on a sample of college juniors is unlikely to resemble the results that would come from the population of interest. College juniors may be more easily swayed by the advertisements because they have not formed strong partisan ties, whereas older voters may be less affected by the negative advertisements.

The setting in which a study is carried out can have an impact on the findings. Experiments tend to study phenomena in a laboratory setting. This unnatural environment, where participants know they are being studied, may produce unintended results. In the negative advertising study, participants may be asked to view various advertisements, and then their opinion of the incumbent candidate would be measured. However, we know most people do not actively pay attention to commercials on television; rather, during commercial breaks, many people hold conversations, get snacks, or change stations. Thus, the artificial environment in which the advertisements are viewed could cause participants to evaluate the incumbent poorly, though the same commercials viewed at home could have no effect. Observational studies may encounter problems with artificial settings if the observation takes place in the laboratory setting versus the natural environment.

The final threat to external validity is testing. PRETESTs and the experience of being tested may change the magnitude of the treatment effect. To be able to measure the effect of negative advertising on incumbent approval, we would need some measure of approval before the participants viewed the advertisements. A pretest could make the participants become more observant or opinionated about the incumbent and produce misleading results. In observational studies, the researcher must ensure that the presence of the observer will not influence the behavior of the subjects.

Researchers increase the external validity of their studies by being mindful of the threats when creating their research design. Iyengar, Peters, and Kinder (1982) illustrated how this can be done in an exploration of the agenda-setting power of the evening news on television. They manipulated the volume of stories shown in an evening news broadcast and observed whether it had an impact on how participants viewed the issues.

Iyengar, Peters, and Kinder's sample was drawn from volunteers from New Haven, Connecticut. This could pose a problem because the population of New Haven is not representative of the population of the United States (the population of interest). To ensure that their results would be generalizable to the population of interest, they attempted to encourage diversity by soliciting participation through a classified advertisement. Increasing heterogeneity of the sample can make the research more generalizable (Frankfort-Nachmias & Nachmias, 2000). They also reported on the demographic makeup of the participants, allowing the reader to make judgments about the generalizability of the experiment. They also used RANDOM ASSIGNMENT between their control and experimental groups to ensure the absence of systematic differences between these groups.

Second, the researchers took several steps to prevent artificial findings from being generated by the process of experimentation. The participants watched the news broadcasts in small groups and were never encouraged to pay particular attention to the broadcasts. In fact, the researchers reported that many of the participants engaged in conversations during the broadcasts. This indicates that although the experiment did not take place in the participants' living rooms (the environment of interest), the researchers were successful in reproducing a similar environment, including distractions. Careful consideration should be paid to the location of the testing environment to make it resemble the natural setting as much as possible. Another option is to use a randomized FIELD EXPERIMENT (Gerber & Green, 2000).

The researchers also took several precautions to prevent spurious results from testing. Some of these included presenting a convincing COVER STORY to participants and designing the pretest in a nonleading way. A cover story may be a useful tool; however, researchers also need to be concerned with the ethical implications of using a cover story to deceive participants. It is through careful consideration of the threats to external validity that Iyengar et al. were able to construct a research design that minimized the impact of these threats.

Frankfort-Nachmias and Nachmias (2000) explained that researchers need to be conscious of the trade-off between external validity and INTERNAL VALIDITY when constructing a research design. Internal validity ensures that the INDEPENDENT VARIABLES are producing the change in the DEPENDENT VARIABLE. An experiment with strong external validity but weak internal validity contributes very little to advancing our understanding because conclusions cannot be drawn about the RELATIONSHIP being studied. To ensure internal validity, many times we need to sacrifice external validity. In Iyengar et al.'s (1982) experiment on agenda setting, having the participants watch the news in their living rooms would have provided a more natural setting. However, this would have prevented the researchers from being able to control what the participants were viewing. Thus, to ensure internal validity, they sacrificed some external validity by holding the experiment in a controlled environment. When constructing a research design, one needs to be aware of the impact that choices have on both external and internal validity.

—Heather L. Ondercin

REFERENCES

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Frankfort-Nachmias, C., & Nachmias, D. (2000). *Research methods in the social sciences* (6th ed.). New York: Worth.

Gerber, A. S., & Green, D. P. (2000). The effects of personal canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *The American Political Science Review, 94,* 653–664.

Iyengar, S., Peters, M. E., & Kinder, D. R. (1982). Experimental demonstrations of the "not-so-minimal" consequences of television news programs. *The American Journal of Political Science, 4,* 848–858.

EXTRAPOLATION

Also known as *trend extrapolation* and *curve fitting*, extrapolation is a projection technique in which the analyst plots data points from past time periods, chooses a best-fitting trend line (or curve) for this data, and then extends that trend line to project future values. At its core, extrapolation is a bivariate REGRESSION technique in which the INDEPENDENT VARIABLE is time and the variable to be projected is the DEPENDENT VARIABLE.

Smith, Tayman, and Swanson (2001) noted that the defining characteristic of the extrapolation technique is that historical values are used exclusively to project future values of the variable being studied. There are numerous variations of the extrapolation technique, ranging from simple linear regression to more complex polynomial and logistic curves, with some variations also incorporating ratios into the projection model.

Extrapolation is employed routinely by demographers and planners to project population figures, by public administrators to project budget revenues and expenditures, and by businesspersons to project sales and revenues. Extrapolation's popularity rests primarily in its simplicity and its low data requirements. Using a spreadsheet program or calculator, projections using the extrapolation technique can be generated easily. As for data requirements, as few as two data points are needed to provide a trend line to project future values. Although extrapolation is relatively simple in application, Smith et al. (2001) reported that numerous empirical studies have shown that it is as accurate as other, more complex methods, such as the cohort-component technique.

Figure 1 shows the results when extrapolation is used to project the population of a medium-sized county using three variations: a linear curve, a geometric curve, and a parabolic curve. Using data from the U.S. Census Bureau, total population figures from 1940 to 2000 were used to project the county's population for 2010, 2020, and 2030. Figure 1 shows that population projections range markedly based on the curve chosen. At face value, the parabolic method appears to be the "best fitting" curve for this situation.

The simplicity and low data requirements that make extrapolation attractive as a projection method also serve as the primary limitations to the technique. Extrapolation relies upon aggregated data inputs and yields aggregated results. Factors that may influence the variable under study remain external to the method. For example, when undertaking population projections, the extrapolation technique does not consider the effects of housing trends, economic changes, or other external pressures on population. The inputs into

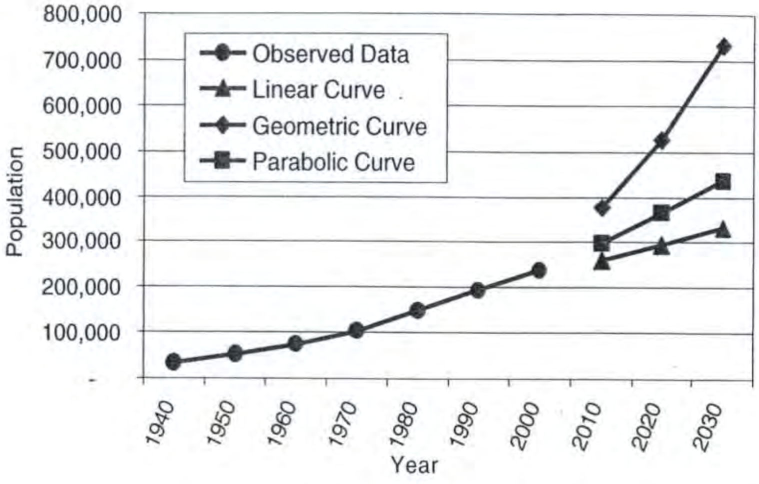

Figure 1 Example Application of the Extrapolation Technique

the model are simple, and so are the outputs. When using extrapolation to project government revenues, for example, only a total revenue figure is generated by the technique; no detail is provided on changes in revenues across different funds.

A further limitation to the model is that extrapolation uses past conditions to project future conditions. Although past conditions are a useful starting point for predicting the future, there is no assurance that past trends will continue. Last, the choice of time period and the number of data observations employed in the analysis can greatly influence the results. For these reasons, extrapolation should be employed carefully, with full understanding of its utility and limitations for projecting the future.

—Tim Chapin

REFERENCE

Smith, S. K., Tayman, J., & Swanson, D. A. (2001). *State and local population projections: Methodology and analysis.* New York: Kluwer Academic/Plenum Publishers.

EXTREME CASE

An extreme case is a CASE that has a value on a variable that is at the limit of the possible values. For example, in the study of age in a community, the most extreme high value may be 95 years, and the most extreme low value may be 18 years. These represent extreme cases, but they may or may not be OUTLIERS. An outlier necessarily falls far from the other values, but an extreme case may not do this. For example, in an age study, there may be many respondents in their 80s and 90s, which would mean that an age of 95, though extreme, does not really lie out from neighboring values. In contrast, if below 95, the next highest age was 72, then clearly 95 would be an outlier, as well as being an extreme case.

—Michael S. Lewis-Beck

F

F DISTRIBUTION

The F distribution, also called the variance-ratio distribution, is the SAMPLING DISTRIBUTION derived from the ratio of two sample variances (s^2) estimating identical population VARIANCES (σ^2). The distribution was originally proposed by Sir Ronald A. Fisher (1890–1962) and was developed by George W. Snedecor (1881–1974), who named it the F distribution in honor of Fisher.

SIGNIFICANCE TESTING WITH THE *F* DISTRIBUTION

Consistent with its development by Fisher and Snedecor, the F distribution is most often used in conjunction with the ANALYSIS OF VARIANCE. However, in its most generic SIGNIFICANCE TESTING application, the F distribution is useful for testing hypotheses about the equivalence of two population variances:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2.$$

To understand the F distribution, and its utility in testing such hypotheses, let's first imagine that H_0 is true. That is, consider two normally distributed populations with means (μ) that may differ, but with identical variances (σ^2). If we were to take a random sample of size n_1 from one population and a random sample of size n_2 from the other population, we could then estimate the population variance from each sample by constructing a sample variance:

$$\hat{\sigma}^2 = s^2 = \frac{\sum(X - \bar{X})^2}{n-1}.$$

The denominator of the sample variance ($n - 1$) is the DEGREES OF FREEDOM associated with that statistic, which we can label ν. Even though the two populations have identical variances, SAMPLING ERROR would likely lead to different sample variances. We could then construct a ratio of the two sample variances, which would be the F RATIO:

$$F = \frac{s_1^2}{s_2^2} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}.$$

With H_0 true, we could derive the sampling distribution of the F ratio by constructing the F ratio for every possible combination of samples of size n_1 and n_2, which results in the F distribution for ν_1 (i.e., $n_1 - 1$) and ν_2 (i.e., $n_2 - 1$). (With H_0 false, the distribution—called the noncentral F distribution—is not easily specified, but it would be more symmetrical than the F distribution.)

As an example of how the F distribution could be used to test a hypothesis about the equivalence of population variances, consider two samples of $n_1 = 31$ ($\nu_1 = 30$) and $n_2 = 41$ ($\nu_2 = 40$). Because we are most often interested in the tails of a distribution, wherein reside atypical values, statistics texts usually contain a table of F ratios that determine the upper percentages of the F distribution (upper 2.5%, upper 5%, etc.). If one had adopted an α level of .05, the F ratio that would cut off the upper 5% of the

F distribution with $\nu_1 = 30$ and $\nu_2 = 40$ (called F_{Critical}) would be 1.74. Thus, according to the F distribution, only 5% of the F ratios would be 1.74 or greater when H_0 is true. (Similarly, statistical programs will generate a probability value that the F ratio would have been obtained with H_0 true.) Armed with such information, we would be able to test null hypotheses that two sample variances were drawn from populations with identical variances. If the F ratio for our two samples were 1.74 or greater (or the probability less than our α level, typically .05), we would be inclined to think that the numerator sample had been drawn from a population with a greater variance than that from which the denominator sample had been drawn. If the F ratio is greater than 1.0 but less than 1.74, we would retain H_0. (If the F ratio is less than one, simply invert the ratio and compare it to F_{Critical} with 40 and 30 degrees of freedom, which would be 1.79 with $\alpha = .05$.)

CHARACTERISTICS OF THE *F* DISTRIBUTION

Given that the shape of the F distribution is driven by the degrees of freedom (ν_1 and ν_2) associated with the two sample variances, there is not a single F distribution, but a family of distributions. The F ratios in the F distribution will always fall between zero and infinity. As long as ν_2 is greater than 2, the mode of the F distribution would be

$$\left(\frac{\nu_1 - 2}{\nu_1}\right)\left(\frac{\nu_2}{\nu_2 + 2}\right).$$

As long as ν_2 is greater than 2, the mean of the F distribution would be

$$\frac{\nu_2}{\nu_2 + 2}.$$

As long as ν_2 is greater than 4, the variance of the F distribution would be

$$\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}.$$

The F distribution typically will be unimodal and positively skewed (as long as ν_1 is greater than 2). In a positively skewed distribution, the mean would be larger than the mode. For the example above, with $\nu_1 = 30$ and $\nu_2 = 40$, the modal F ratio would be .89 and the mean F ratio would be .95, consistent with the positive skew of the F distribution. From the formulas, it should be clear that the mode will always be less than the mean, although the difference will lessen as ν_1 approaches infinity.

The F distribution is related to another commonly used statistical distribution, the T-DISTRIBUTION, such that when $\nu_1 = 1$, $t^2 = F$.

—Hugh J. Foley

See also DISTRIBUTION

REFERENCES

Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.

Heyde, C. C., & Seneta, E. (Eds.). (2001). *Statisticians of the centuries*. New York: Springer.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice Hall.

Stuart, A., & Ord, J. K. (1994). *Kendall's advanced theory of statistics* (Vol. 1). London: Edward Arnold.

F RATIO

The F ratio is the ratio of variances or variance-like quantities. A simple example is the testing of whether the variance between groups is significantly larger than the variance within groups in ANALYSIS OF VARIANCE. The general notion underlying such a test is that the differences between individuals within groups are only random differences (error variance), but that the differences between group means stand out above chance (alternative hypothesis).

AN EXAMPLE

An example of calculating this can be found when the partition of variance is dealt with. A series of scores 2 2 3 4 4 5 6 7 8 9 with a mean of 5 and variation 54 is partitioned into two groups. We can think of two groups (e.g., a group of men and a group of women) for which the scores of a quantitative variable are considered—let us say, their smoking behavior measured as the number of cigarettes per day. The group of men with scores 2 2 3 4 4 has a mean of 3 and a variation of 4. The group of women with scores 5 6 7 8 9 has a mean of 7 and a variation of 10. The within-group variation is $4 + 10 = 14$. The between-group variation is defined by the group means 3 and 7. Their mean is 5 and their variation 8. The between-group variation entirely determined by the group is obtained by multiplying the variation between groups by the number of individuals per group: $8 \times 5 = 40$. It appears that $54 = 40 + 14$ or SST = SSB + SSW (sum of squares total = sum of squares between + sum of squares within). An analogous partition of the total into a between and within component can be carried out for the degrees of freedom. One degree of freedom is lost to the mean of a series. Therefore, there are 9 degrees of freedom for the total of 10 scores. For the between component, we calculate for 2 groups, so that there is 1 degree of freedom. For the within component, we calculate for 2×5 individuals, each time losing a degree of freedom to the group mean, so that $2(5 - 1) = 8$ remain. We see here, too, that total = between + within, because $9 = 1 + 8$. Should one now want to work with variances instead of variations (Mean squares, MS, instead of Sum of Squares, SS), then one must divide each of the variations concerned by the corresponding degrees of freedom: MSB = 40/1 and MSW = 14/8. The ratio of this between-variance MSB and the within-variance MSW is a statistical quantity $F = (40/1)/(14/8) = 22.86$, for which the sampling distribution for the degrees of freedom, 1 for the numerator and 8 for the denominator, is distributed as F (the symbol F was chosen to commemorate the contributions by Sir Ronald Fisher). The critical value of F for 1 degree of freedom for the numerator and 8 degrees of freedom for the denominator is found in statistical tables of the theoretical F distribution and is equal to 5.32 for a postulated α of .05. The value of the F ratio we found in our empirical example is greater, and we can therefore, even with these small numbers, reject the null hypothesis with a probability of 95%. The null hypothesis states that the ratio MSB/MSW is equal to 1; in other words, that the differences between groups (MSB) are not greater than the randomly assumed differences between individuals within groups (MSW = error variance).

MANY APPLICATIONS IN MULTIVARIATE ANALYSIS

In multivariate analysis, the F ratio is frequently used in situations more complicated than the example above. The test of difference of groups in analysis of variance, the test for the significance of the regression model in multiple REGRESSION analysis and the test for centroid differences in DISCRIMINANT ANALYSIS are just a few of the many examples of application. C. R. Rao was especially charmed by the F distribution and has developed an F variant for many applications in multivariate analysis.

RELATIONSHIP BETWEEN F AND OTHER DISTRIBUTIONS

The attentive reader will already have seen that instead of the F test for this example, we could just as easily have used the t test of difference of means. We could have situated the obtained difference (–4) between the mean of the first group (3) and the mean of the second group (7) in a sampling distribution of differences of means with the estimated standard error = $(\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2})^{1/2}$ and where s_w^2 is equal to $(4 + 10)/[(5 - 1) + (5 - 1)] = 1.75$ and $n_1 = n_2 = 5$. The t value for the difference of means found in our pair of samples would be equal to $t = (4 - 0)/[(1.75/5) + (1.75/5)]^{1/2} = 4.78$. From the t table, a significant difference would then appear between the means of the two groups. The F value of 22.86 obtained earlier is (within rounding errors) equal to the square of the t value found of 4.78, because when the degree of freedom is one, $F = t^2$.

One can also see in the F and t tables that the critical values of F for a numerator of 1 degree of freedom (only two groups!) are equal to the squares of the corresponding values of t.

There are close connections between the F DISTRIBUTION and many other statistical DISTRIBUTIONS. The F distribution belongs to a subfamily of beta distributions, sometimes called *inverted beta distributions*. There is also a connection with the t, the chi square, the NORMAL DISTRIBUTION, and others. See F DISTRIBUTION for further discussions.

—Jacques Tacq

REFERENCES

Blalock, H. M. (1972). *Social statistics.* Tokyo: McGraw-Hill Kogakusha.

Fisher, R. A. (1953). *The design of experiments.* Edinburgh, UK: Oliver and Boyd.

Hays, W. L. (1972). *Statistics for the social sciences.* New York: Holt International.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis.* London: Sage.

FACE VALIDITY

Face validity is an estimate of the degree to which a measure is clearly and unambiguously tapping the construct it purports to assess. Thus, face validity refers to the "obviousness" of a test—the degree to which the purpose of the test is apparent to those taking it. Tests wherein the purpose is clear, even to naïve respondents, are said to have *high face validity*; tests wherein the purpose is unclear have *low face validity* (Nevo, 1985). The concept of face validity is similar to *item subtlety*, but there are important differences as well. Whereas face validity describes the transparency of an entire test, item subtlety describes the transparency of individual test items (Bornstein, Rossner, Hill, & Stepanian, 1994). It is possible to construct a test wherein the purpose of individual test items is not apparent, but when these items are scrutinized as a group, the purpose of the test as a whole becomes obvious.

Face validity has contrasting effects on different types of tests. Studies indicate that high face validity can facilitate performance on intelligence, aptitude, and achievement tests: When the purpose of the test seems clear, testees are less anxious and more motivated to persevere, even when test items are highly challenging (Messick, 1995; Nevo, 1985).

High face validity can be a liability when a test is designed to assess some aspect of personality or psychopathology. In this situation, high face validity enables respondents to bias their responses to present themselves as they want to be seen by the examiner. Studies show that naïve respondents are able to "fake good" and "fake bad" more effectively on tests with high face validity than those with low face validity (Bornstein, 2002; Bornstein et al., 1994).

There are two general methods for evaluating the face validity of a test. Some psychometricians have taken a direct approach, asking participants to identify the purpose of an assessment instrument from among an array of likely choices. To the degree that participants can do this accurately, the test has high face validity. Other researchers have taken an indirect approach, asking participants to deliberately bias their answers to raise or lower their scores on the test. To the degree that participants can alter their scores in this way, the test is presumed to have high face validity and to be susceptible to self-presentation effects in vivo (Bornstein, 2002).

Given its importance for psychological tests, face validity should always be assessed and controlled during test development. Several strategies are useful. For example, test items can be worded subtlely, so the true purpose of the measure is masked. Alternatively, test items can be ordered so that those from a given subscale do not appear in proximity to each other; this has also been shown to lower the face validity of the test. Finally, distracter items unrelated to the true purpose of the test can be included so that the content of genuine test items is less obvious. This latter strategy is particularly effective, when combined with the other two approaches, for disguising the purpose of the test.

—Robert F. Bornstein

REFERENCES

Bornstein, R. F. (2002). A process dissociation approach to objective-projective test score interrelationships. *Journal of Personality Assessment, 78,* 47–68.

Bornstein, R. F., Rossner, S. C., Hill, E. L., & Stepanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment, 63,* 363–386.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749.

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement, 22,* 287–293.

FACTOR ANALYSIS

In general, the term *factor analysis* refers to any one of a number of similar but distinct multivariate statistical models that model observed variables as linear functions of a set of LATENT or hypothetical variables that are not directly observed, known as *factors*.

Factor analysis models are similar to REGRESSION models in that they possess DEPENDENT VARIABLES that are linear functions of INDEPENDENT VARIABLES. But unlike regression, the independent variables of the factor analysis models are not observed independently of the observed dependent variables.

Factor analysis models may be further distinguished according to whether the factor variables are determinate or not. Factors are determinate if they can be derived in turn as linear functions of the observed variables. Otherwise, they are indeterminate. Determinate models encompass the various component analysis models, such as PRINCIPAL COMPONENTS ANALYSIS (Hotelling, 1933; Joliffe, 1986; Pearson, 1901); weighted principal components (Mulaik, 1972); and Guttman's image analysis (Guttman, 1953). Indeterminate models are represented by the common factor model (Spearman, 1904; Thurstone, 1947), which seeks to account for the covariation between the observed variables as the result of the observed variables' sharing in varying degrees the influences of the variation of a common set of common factor variables.

Determinate factor analysis models are often useful in a data reduction role by finding a smaller number of variables that capture most of the information of variation and covariation among the observed variables. Scores on the determinate factors can be computed as linear combinations of the observed variables. These factor scores may be used as independent and dependent variables—as the case warrants—in other multivariate statistical procedures, such as multivariate regression or multivariate analysis of variance. However, for substantive theoretical work, the main drawback of these determinate component analysis models is that their factors represent statistical artifacts unique to the set of observed variables determining them (Mulaik, 1987). Change the set of observed variables, and you obtain different linear combinations. The component factors have no independent existence apart from the set of observed variables of which they are linear combinations.

In contrast, the common factors of the common factor model are indeterminate from the observed variables and are not linear combinations of them. Common factors can correspond to variables having an independent existence, and, in the theory of simple structure in common factor analysis (Thurstone, 1947), different sets of observed variables from a domain can be linear functions of the same common factors. Thus, for the purposes of discovering autonomous variables that have theoretical import as common causes of other variables, the common factor model is generally preferred to a determinate component analysis model.

However, the common factor model has limits to its application. Common factor analysis is limited to the case where there is no natural order among the observed variables. The only ordering principle permitted by the common factor model is the relation of functional dependency of observed variables on latent variables. No provision exists in the common factor model for functional dependencies between latent variables. Relations between latent variables only take the form of correlations or covariances, which are nondirectional and nonordering. But variables that are naturally ordered in time, space, or degree of some attribute may be more correlated with the variables immediately adjacent to them and have diminishing correlations with variables farther from them in the natural ordering. In these cases, in addition to functional dependencies of observed variables on latent variables, each successive latent variable corresponding to one of the observed variables in the order may be a linear function of only the immediately preceding latent variable in the order plus some new latent variable unique to it. Models with this property are known as simplex models (Jöreskog, 1979). To deal with this and other cases, STRUCTURAL EQUATION MODELING is more appropriate because it has provisions for establishing linear functional relations between latent variables. Thus, if one is to apply the common factor model properly to a set of observed variables, one must take care to determine at the outset that there is no apparent natural ordering among the variables. Although the common factor model may fit well to such data with a small number of common factors, the factors will usually defy theoretical interpretation (Jones, 1959).

Factor analysis models also can be distinguished as to whether they are exploratory or confirmatory. Exploratory common factor analysis is principally analytic, used in situations where there is minimal knowledge of the constituent factor variables of a domain of variables believed to conform to the common factor model. The aim is to analyze the observed variables to "discover" constituent variables that are more basic. Confirmatory factor analysis is principally synthetic. It is used where there is sufficient knowledge to formulate hypotheses as to how certain theoretical variables function as factors common to a number of observed variables. The theory allows one to predict within certain constraints the pattern of covariances among the observed variables and to test for this pattern (see CONFIRMATORY FACTOR ANALYSIS). The rest of this article will be about exploratory common factor analysis.

EQUATIONS OF EXPLORATORY FACTOR ANALYSIS

The equations of factor analysis are usually expressed in matrix algebra. Thus, the model equation of common factor analysis is given by the following matrix equation:

$$\mathbf{Y} = \mathbf{\Lambda X} + \mathbf{E},$$

where $\mathbf{Y}$ is a $p \times 1$ column vector of p observed random variables; Λ is a $p \times m$ matrix of factor pattern loadings λ_{ij}, which are analogous to regression coefficients; $\mathbf{X}$ is an $m \times 1$ column vector of common factor random variables; and $\mathbf{E}$ is a $p \times 1$ column vector of p unique factor random variables.

One then assumes the following: (a) The common and unique factor variables are uncorrelated, that is, $\text{cov}(\mathbf{X}, \mathbf{E}) = 0$; and (b) the unique factors are mutually uncorrelated, that is, $\text{cov}(\mathbf{E}) = \mathbf{\Theta}^2 =$ a diagonal matrix.

From the model equation and these assumptions, we are able to derive the fundamental theorem of the common factor model, expressed as

$$\text{cov}(\mathbf{Y}) = \mathbf{\Sigma} = \mathbf{\Lambda \Phi \Lambda}' + \Theta^2,$$

where $\mathbf{\Sigma}$ is the $p \times p$ variance-covariance matrix for the p observed variables, $\mathbf{\Lambda}$ is the $p \times m$ factor pattern matrix of elements λ_{ij}, $\mathbf{\Phi}$ is the $m \times m$ matrix of VARIANCES and COVARIANCES among the common factors, and $\mathbf{\Theta}^2$ is a $p \times p$ diagonal matrix of unique factor variances. Because $\mathbf{\Theta}^2$is a diagonal matrix with zero off-diagonal elements, the off-diagonal elements of the covariance matrix $\mathbf{\Sigma}$ are due only to the off-diagonal elements of $\mathbf{\Lambda \Phi \Lambda}'$, which is a function of only the common factor variables.

Table 1 Fifteen Bipolar Personality Rating Scales Used to Produce 15 Variables for the Factor Analysis Study

1.	FRIENDLY:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:UNFRIENDLY
2.	SYMPATHETIC:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:UNSYMPATHETIC
3.	KIND:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:CRUEL
4.	AFFECTIONATE:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:UNAFFECTIONATE
5.	INTELLIGENT:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:UNINTELLIGENT
6.	CAPABLE:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:INCAPABLE
7.	COMPETENT:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:INCOMPETENT
8.	SMART:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:STUPID
9.	TALKATIVE:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:UNTALKATIVE
10.	OUTGOING:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:WITHDRAWN
11.	GREGARIOUS:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:SOLITARY
12.	EXTRAVERTED:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:INTROVERTED
13.	HELPFUL:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:UNHELPFUL
14.	COOPERATIVE:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:UNCOOPERATIVE
15.	SOCIABLE:	_1_:_2_:_3_:_4_:_5_:_6_:_7_	:UNSOCIABLE

STEPS IN PERFORMING A FACTOR ANALYSIS: A WORKED EXAMPLE

Step 1: Choosing or Constructing Variables

It is best to select or construct variables representing some domain of activity in a systematic manner. L. L. Thurstone (1947), a pioneer in the field of factor analysis, believed that a domain of variables is defined by the common factors that span it. Thus, one must consider the possible common factors that might exist in the domain. Then, for each of those anticipated factors, one should select or construct at least four indicator variables that one believes are relatively pure measures of those factors. Four indicators of an anticipated factor overdetermine the factor.

As an example, Carlson and Mulaik (1993) selected 15 variables based on 15 bipolar personality rating scales shown in Table 1 that they expected would represent four factors of personality: (a) friendliness, represented by the bipolar rating scales friendly-unfriendly, sympathetic-unsympathetic, kind-unkind, and affectionate-unaffectionate; (b) ability, represented by the scales intelligent-unintelligent, capable-incapable, competent-incompetent, and smart-stupid; (c) extraversion, represented by talkative-untalkative, outgoing-withdrawn, gregarious-solitary, extraverted-introverted, and sociable-unsociable. They anticipated that variables such as helpful-cooperative and cooperative-uncooperative would be concepts representing a combination of friendliness and ability, and would have "loadings" on these factors plus loadings on (d) an additional factor unique to them alone.

Step 2: Obtaining Scores on Variables and Computing Correlation Matrix

Carlson and Mulaik (1993) then had 280 students rate randomly selected descriptions of people in a work setting on the 15 rating scales. The correlations among the 15 scales were then obtained and are given in Table 2.

Step 3: Determining Number of Factors

The extraction of factors is based on the matrix of correlations between factors. However, a decision must be made at this point as to which method of factor extraction to use to find the estimates for the factor pattern coefficients, the correlations among the common factors, and the unique factor variances, which will be the basis for the ultimate interpretation of the analysis. The program used to perform this analysis was SPSS FACTOR Version 4.0 for the Macintosh

Table 2 Correlations Among 15 Variables

SYMMETRIC MATRIX
LOWER HALF SHOWN

	1 FRIENDLY	2 SYMPATHETIC	3 KIND	4 AFFECTIONATE	5 INTELLIGENT	6 CAPABLE	7 COMPETENT	8 SMART	9 TALKATIVE	10 OUTGOING	11 GREGARIOUS	12 EXTRAVERTED	13 HELPFUL	14 COOPERATIVE	15 SOCIABLE
1 FRIENDLY	1.000														
2 SYMPATHETIC	.777	1.000													
3 KIND	.809	.869	1.000												
4 AFFECTIONATE	.745	.833	.835	1.000											
5 INTELLIGENT	.176	.123	.123	.112	1.000										
6 CAPABLE	.234	.159	.205	.183	.791	1.000									
7 COMPETENT	.243	.155	.187	.186	.815	.865	1.000								
8 SMART	.234	.190	.238	.215	.818	.841	.815	1.000							
9 TALKATIVE	.433	.319	.321	.435	.174	.209	.239	.258	1.000						
10 OUTGOING	.473	.480	.410	.527	.220	.274	.269	.261	.744	1.000					
11 GREGARIOUS	.433	.438	.406	.526	.188	.227	.242	.228	.711	.853	1.000				
12 EXTRAVERTED	.447	.396	.350	.500	.192	.221	.227	.224	.758	.846	.801	1.000			
13 HELPFUL	.649	.693	.697	.694	.283	.344	.370	.365	.443	.552	.514	.473	1.000		
14 COOPERATIVE	.662	.692	.676	.679	.311	.345	.375	.351	.431	.557	.514	.493	.740	1.000	
15 SOCIABLE	.558	.543	.510	.632	.213	.289	.287	.287	.745	.886	.820	.830	.631	.626	1.000

computer. (Later versions are essentially unchanged in these options.) The program offers several methods of factor extraction appropriate for a common factor analysis. These are principal axis factoring, which is essentially unweighted least squares; unweighted least squares (a slightly different ALGORITHM for the same result); GENERALIZED LEAST SQUARES; and MAXIMUM LIKELIHOOD. All of these methods of extraction are iterative and require that an initial decision be made as to the number of factors to extract, because this number must be fixed at this value throughout the iterations. SPSS computes a principal components analysis with every factor analysis and displays the EIGENVALUES of the unmodified CORRELATION matrix R. These can be used to estimate the number of common factors. This is done by means of a SCREE PLOT, which is a plot of the magnitude of each of the eigenvalues against its ordinal position in the descending series of eigenvalues. Connecting the dots in this plot reveals a large first eigenvalue much higher than the rest, followed by a number of lesser eigenvalues of still substantial magnitude. But there is usually a point where the rapid descent in magnitude of the eigenvalues suddenly changes to a gradual, almost linear descent for the remainder of the eigenvalues. Many believe that this is in the vicinity of the point where the variables begin to be influenced by common factors in a significant way, and so this number is used as the number of common factors to extract. Others retain only so many factors as there are eigenvalues of R greater than 1.00. But this is well known to be a weakest lowest bound for the number of common factors. There may be more of lesser influence.

This and subsequent versions of SPSS FACTOR do not report a different set of eigenvalues that would be more revealing theoretically as to the number of common factors. Given the equation of the fundamental theorem of factor analysis above, we may substitute $\mathbf{R}$ for $\mathbf{\Sigma}$ to represent a correlation matrix. Guttman (1954, 1956) showed that a strong lower bound to the number of common factors would be the number of positive eigenvalues of the matrix $\mathbf{R} - \mathbf{S}^2$ where $\mathbf{S}^2 = [\mathrm{diag}\mathbf{R}^{-1}]^{-1}$. $\mathbf{S}^2$ is a strong upper-bound approximation to $\mathbf{\Theta}^2$. This number of positive eigenvalues, however, is often on the order of $p/2$. Furthermore, many of the eigenvalues of $\mathbf{R} - \mathbf{S}^2$ are quite small, near zero, and so many researchers would like to examine a plot of these eigenvalues to determine a point at which the eigenvalues begin to assume substantial values. Developments by Harris (1962) that were adopted by Jöreskog (1967) in developing an algorithm for performing maximum likelihood factor analysis suggested that one could pre- and postmultiply $\mathbf{R} - \mathbf{S}^2$ by $\mathbf{S}^{-1}$ to obtain $\mathbf{S}^{-1}(\mathbf{R} - \mathbf{S}^2)\mathbf{S}^{-1} = \mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1} - \mathbf{I}$. This transformation does not change the number of positive eigenvalues of the resulting matrix, so the number of positive eigenvalues of this matrix would equal the number of factors to retain. But each of the eigenvalues of this matrix would correspond to an eigenvalue of the matrix $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$ minus 1. So, any eigenvalue greater than 1.0 of $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$ would correspond to a positive eigenvalue of $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1} - \mathbf{I}$. Thus, researchers could also profit from examining the eigenvalues of $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$ to determine the number of common factors to retain.

A separate calculation in this example of the eigenvalues of $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$ revealed only 7 of the

15 eigenvalues were greater than 1.00. These were 37.484, 14.467, 9.531, 1.436, 1.289, 1.116, and 1.065. Although three factors would be substantial in contribution, we will take one more because that was the theoretical expectation, and those beyond it would have a much smaller contribution (if you subtract one from each).

Step 4: Extraction of the Unrotated Pattern Matrix

The method of factor extraction used in this example was maximum likelihood. It is relatively robust, even when the variables do not have a multivariate normal distribution, and maximizes the determinant of the partial correlation matrix among the variables with the common factors partialed out, regardless of the form of the distribution. In this case, the estimate of the unrotated factor pattern matrix is given by $\hat{\mathbf{\Lambda}} = \hat{\mathbf{\Theta}}\mathbf{A}_m[\gamma_i - 1]_m^{1/2}$, where $\hat{\mathbf{\Theta}}$ is the square root of the iteratively estimated diagonal unique variance matrix, $\mathbf{A}_m$ is the $p \times m$ matrix whose columns are the first m eigenvectors of $\hat{\mathbf{\Theta}}^{-1}\mathbf{R}\hat{\mathbf{\Theta}}^{-1}$, and $[\gamma_i - 1]_m$ is a diagonal matrix of order m whose diagonal elements are formed by subtracting 1 from each of the first m largest eigenvalues γ_i of $\hat{\mathbf{\Theta}}^{-1}\mathbf{R}\hat{\mathbf{\Theta}}^{-1}$. In this solution, the common factors are mutually uncorrelated. But the unrotated factor pattern matrix is only an intermediate solution. It defines a common factor space that contains the maximum common variance for any m common factors for these variables. It is not used for factor interpretation.

Step 5: Rotation of Factors

Mathematically, the factor pattern matrix is not unique. Given the model equation $\mathbf{Y} = \mathbf{\Lambda X} + \mathbf{E}$, we can find alternative factor pattern matrices and common factors by linear transformations of them. Let $\mathbf{T}$ be an $m \times m$ transformation matrix of full rank and $\mathbf{T}^{-1}$ its matrix inverse. Then $\mathbf{Y} = \mathbf{\Lambda T T}^{-1}\mathbf{X} + \mathbf{E} = \mathbf{\Lambda}^*\mathbf{X}^* + \mathbf{E}$, with $\mathbf{\Lambda}^* = \mathbf{\Lambda T}$ and $\mathbf{X}^* = \mathbf{T}^{-1}\mathbf{X}$. This also has the form of a factor analysis equation. So, which factor pattern matrix should be our solution? L. L. Thurstone (1947) solved this problem with the idea of simple structure. He argued that if most of the observed variables were not functions of all of the common factors of their domain, then the vectors representing them in common factor space would fall in "coordinate hyperplanes" or subspaces of the full common factor space. Each hyperplane or subspace would be spanned by, at most, $m - 1$ of the common factors, and all variables within those subspaces would be functions of just the common factors spanning the hyperplane. The common factors of the full common factor space then would be found at the intersections of these "coordinate hyperplanes," because they would be among the basis vectors of each of $m - 1$ hyperplanes. If the coordinate hyperplanes could be identified by discovering subsets of the variables occupying these subspaces of lower dimension, then the common factors would be identified at their intersections. Furthermore, different selections of variables from the domain would still identify the same common factors at the intersections of the hyperplanes. Hence, the solution for the factors is not unique to the selection of variables from the domain and is an objective, invariant result over almost all selections of observed variables from the domain. To discover these hyperplanes, Thurstone proposed using hypothetical vectors called *reference axes* inserted into the common factor space. A subspace of the common factor space would be a subspace of vectors orthogonal to a reference axis and would be like a parasol whose ribs represent vectors in the subspace, with the reference axis its handle. Thus, one would move each reference axis around in the common factor space seeking different sets of variables to fall in the "parasol" orthogonal to the reference axis. These subsets of variables would define the coordinate hyperplanes. Originally, Thurstone used graphical, two-dimensional plots for each pair of factors and moved his reference axes manually in these two-dimensional spaces to discover sets of variables that would line up orthogonal to the reference axes. This was slow work. Later, with the advent of computers, analytic criteria for simple structure were formulated and the process of finding simple structure solutions automated. It is important, however, to realize that simple structure does not imply uncorrelated or orthogonal factors. The common factors may be correlated. Thus, one should avoid rotational procedures such as VARIMAX, which forces the common factors to be mutually orthogonal, if one wants a genuine simple structure solution. Good algorithms for simple structure that are generally available are direct Oblimin and Promax. Other algorithms claim a modest superiority over these, but they are not generally available in commercial factor analysis programs.

In our example, we rotated our four-factor solution to simple structure using direct Oblimin rotation. The program produced a factor pattern matrix, a factor structure matrix, and a matrix of correlations among the factors. The factor pattern matrix contains weights like regression weights for deriving the observed variables from the common factors. The matrix is the most useful in interpretation of the factors, because these weights will be invariant under restriction of range across selected subpopulations of subjects. Furthermore, the pattern weights clearly show which variables are and are not functions of which common factors. The factor structure matrix contains the correlations between the observed variables and the common factors. It is itself the matrix product $\Lambda\Phi$ of the factor pattern and factor correlation matrices. So, once these two are obtained, the factor structure matrix is redundant. Furthermore, the factor structure matrix and the matrix of correlations among factors are not invariant under restriction of range or selection of subjects.

Step 6: Factor Interpretation

The method of factor interpretation is eliminative induction. One looks down each column of the factor pattern matrix for those variables having large "loadings" on the factor. One interprets the factor as that hypothetical variable that is common to those variables with large loadings but absent in variables with near-zero loadings. In Table 3, we find something like positive orientation to others or "kindness" to be what is common to those variables having high loadings on the first factor. Ability seems to be the common element for the second factor, whereas extraversion is the common element for the third. The fourth factor indeed is something common to just helpful and cooperative, and these variables also have modest loadings as expected on kindness and ability. The correlations among factors in Table 4 suggest that kindness and extraversion share something in common and less with ability. In some cases, factor analysts will factor analyze the correlations among the factors to obtain second-order common and unique factors.

—Stanley A. Mulaik

Table 3 Factor Pattern Loadings

	1	2	3	4
FRIEND	.78317	.04561	.06936	.02296
SYMPATH	.84525	-.05340	-.01319	.12206
KIND	1.03068	.02353	-.09171	-.04496
AFFECTN	.79470	-.03680	.14617	.03676
INTELL	-.07708	.88650	-.01461	.03989
CAPABL	.02454	.92586	.01284	-.02959
CMPTNT	-.07117	.90317	.00333	.10189
SMART	.10794	.91176	.00614	-.09064
TALKTV	.01144	.03235	.84101	-.09020
OUTGO	-.04548	-.00038	.92307	.08421
GREG	.01553	-.01468	.88010	.01090
EXTRAV	-.02696	-.01428	.93815	-.03029
HELPFL	.32864	.11450	.12744	.44996
COOPER	.23477	.10052	.11966	.56880
SOCIAB	.07860	.00542	.82472	.11774

Table 4 Correlation Between Common Factors

	1	2	3	4
FACTOR 1	1.00000			
FACTOR 2	.21857	1.00000		
FACTOR 3	.52371	.28226	1.00000	
FACTOR 4	.72944	.33731	.52443	1.00000

REFERENCES

Carlson, M., & Mulaik, S. A. (1993). Trait ratings from descriptions of behavior as mediated by components of meaning. *Multivariate Behavioral Research, 28,* 111–159.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika, 18,* 277–296.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19,* 149–161.

Guttman, L. (1956). "Best possible" systematic estimates of communalities. *Psychometrika, 21,* 273–285.

Harris, C. W. (1962). Some Rao-Guttman relationships. *Psychometrika, 27,* 247–263.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24,* 417–441, 498–520.

Joliffe, I. T. (1986). *Principal component analysis.* New York: Springer-Verlag.

Jones, M. B. (1959). *Simplex theory* (U.S. Naval School of Aviation Medicine Monograph Series No. 3). Pensacola, FL: U.S. Naval School of Aviation Medicine.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32,* 443–482.

Jöreskog, K. G. (1979). Statistical models and methods for analysis of longitudinal data. In K. G. Jöreskog & D. Sörbom (Eds.), *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw-Hill.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research, 22,* 267–305.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 6*(2), 559–572.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15,* 366–374.
Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

FACTORIAL DESIGN

A factorial design involves the simultaneous study of two or more variables or factors on some other variable or variables. The factors may be manipulated or measured. Each factor consists of two or more levels or categories. The levels may differ qualitatively or quantitatively. For example, the qualitative factor of marital status may consist of the five levels of never married, married, separated, divorced, and widowed. Quantitative factors may comprise particular values (such as increasing dosages of a drug) or a range of values (such as the four age groups of 20–29, 30–39, 40–49 and 50–59).

A factorial design may be further described in terms of the number of factors, the number of factors and levels within each factor, whether cases have been randomized to factors or treatments, and whether cases are measured on more than one occasion. Factors may be referred to as "ways," so that a two-factor design may be called a two-way design, a three-factor design a three-way design, and so on. Alternatively, the design may be described in terms of the number of levels within each factor. A design that consists of three factors, two having two levels and the third having four levels, may be designated a $2 \times 2 \times 4$ ("a two by two by four") factorial design. A completely randomized factorial design is where each case has been RANDOMLY ASSIGNED to one and only one combination of factor levels or cells. A mixed factorial design comprises at least one factor where cases are measured on more than one occasion and one factor where cases are measured on only one occasion.

Ronald Fisher (1935) suggested that the factorial design had three advantages over the single-factor design. The first advantage is that it is more efficient or economical in that it requires fewer cases or observations for the same degree of precision or power. Compared with a single-factor design, a two-factor or two-way factorial design requires half as many cases, a three-way design a third as many cases, a four-way design a quarter as many cases, and so on (Snedecor, 1937). The reason for this is that the values of one factor are averaged across the values of the other factors. The second advantage is that it is more comprehensive in that it allows the interaction between two or more factors to be examined. The third advantage is that it enables greater GENERALIZABILITY of the results in that a factor has been investigated over a wider range of conditions. A fourth advantage is that it may provide a more sensitive or powerful test of a factor if that factor does not interact substantially with one or more factors in that the other factors may account for some of the unexplained or error variance (Stevens, 2001).

—Duncan Cramer

See also EXPERIMENT

REFERENCES

Fisher, R. A. (1935). *The design of experiments* (1st ed.). Edinburgh, UK: Oliver and Boyd.
Snedecor, G. W. (1937). *Statistical methods* (1st ed.). Ames: Iowa State University Press.
Stevens, J. (2001). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

FACTORIAL SURVEY METHOD (ROSSI'S METHOD)

Humans form ideas about the way the world works. They also form ideas about the way the world ought to work. These positive and normative ideas can be represented by equations, termed, respectively, *positive-belief equations* and *normative-judgment equations.* Rossi's factorial survey method makes it possible to estimate these equations-inside-the-head.

The positive-belief and normative-judgment equations are linked to two further equations—an equation describing the determinants of components of the beliefs or judgments, called a *determinants equation,* and an equation describing consequences of the belief/judgment components, called a *consequences equation.*

For example, individuals form ideas about earnings determination and, concomitantly, about the earnings they regard as just for themselves and others. Both the ideas about actual earnings determination and about just earnings determination are themselves the product of personal and social factors, such as information about the occupational structure and religion. Moreover, the ideas about actual earnings help shape many decisions and behaviors—what to study, where to work, and so on—and the ideas about just earnings help shape many other decisions and behaviors—such as voting, joining a strike, making contributions to lobbying organizations, and designing a pay schedule.

Thus, the positive-belief equation and the normative-judgment equation each may join with a determinants equation to form a multilevel system of equations. Similarly, the positive-belief equation and the normative-judgment equation each may join with a consequences equation to form another (possibly multilevel) system of equations.

The factorial survey method pioneered by Peter H. Rossi (1951, 1979) and developed with associates (Jasso & Rossi, 1977; Rossi & Anderson, 1982; Rossi & Berk, 1985; Rossi, Sampson, Bose, Jasso, & Passel, 1974) provides an integrated framework for estimating the beliefs/judgments equations-inside-the-head, testing for interrespondent homogeneity, and estimating the determinants and consequences equations. All elements of the research protocol are designed with the objective of obtaining estimates with the best possible properties of the beliefs/judgments equations.

DATA COLLECTION IN THE FACTORIAL SURVEY METHOD

Each respondent is asked to rate the level of a specified outcome variable (e.g., healthiness, wage attainment, just prison sentence) corresponding to a fictitious unit (a person, say, or a family) that is described in terms of potentially relevant characteristics such as age, gender, study or eating habits, access to medical care or housing, and the like. The descriptions are termed "vignettes" (VIGNETTE TECHNIQUE). One of Rossi's key insights was that fidelity to a very rich and complex reality can be achieved by generating the POPULATION of all logically possible combinations of all levels of potentially relevant characteristics and then drawing random samples to present to respondents. Accordingly, the vignettes are described in terms of many characteristics, each characteristic is represented by many possible realizations, and the characteristics are fully crossed. Three additional important features of the Rossi design are the following: (a) In the population of vignettes, the correlations between vignette characteristics are all zero or close to zero, thus reducing or eliminating problems associated with MULTICOLLINEARITY; (b) the vignettes presented to a respondent are under the control of the investigator (i.e., they are "fixed"), so that endogeneity problems in the estimation of positive-beliefs and normative-judgments equations arise only if respondents do not rate all of the vignettes presented to them; and (c) a large set of vignettes is presented to each respondent (typically 40 to 60), improving the precision of the obtained estimates. The rating task reflects the outcome variable, which may be a cardinal quantity, a probability, a set of unordered categories, or a set of ordered categories.

DATA ANALYSIS IN THE FACTORIAL SURVEY METHOD

The analysis protocol begins with inspection of the pattern of ratings, which in some substantive contexts may be quite informative (e.g., the proportion of workers judged underpaid and overpaid), and continues with estimation of the beliefs/judgments equation. Three main approaches are the following: (a) The classical analysis of covariance approach, in which the respondent-specific equations may have different error variances; (b) the seemingly unrelated regressions approach, in which the errors from the respondent-specific equations are correlated; and (c) the random coefficients approach, in which the respondents constitute a random sample, and the parameters of the respondent-specific equations are viewed as coming from a distribution. Under all approaches, an important step involves testing for interrespondent homogeneity.

Depending on the substantive context and on characteristics of the data, the next step is to estimate the determinants equation and the consequences equation. Again depending on the context, the determinants equation may be estimated jointly with the beliefs/judgments equation.

Rossi's method was, for many years, somewhat difficult to implement, given the computational resources required to generate the vignette population and to estimate respondent-specific equations. Recent advances in desktop computational power, however, render straightforward vignette generation, the drawing of random samples, and data analysis, thus setting the stage for a new generation of studies using Rossi's method.

FURTHER READING

Rossi's articles cited above, together with recent factorial survey studies, provide comprehensive exposition of procedures of data collection and data analysis.

—Guillermina Jasso

REFERENCES

Jasso, G., & Rossi, P. H. (1977). Distributive justice and earned income. *American Sociological Review, 42,* 639–651.

Rossi, P. H. (1951). *The application of latent structure analysis to the study of social stratification.* Unpublished doctoral dissertation, Columbia University.

Rossi, P. H. (1979). Vignette analysis: Uncovering the normative structure of complex judgments. In R. K. Merton, J. S. Coleman, & P. H. Rossi (Eds.), *Qualitative and quantitative social research: Papers in honor of Paul F. Lazarsfeld* (pp. 176–186). New York: Free Press.

Rossi, P. H., & Anderson, A. B. (1982). The factorial survey approach: An introduction. In P. H. Rossi & S. L. Nock (Eds.), *Measuring social judgments: The factorial survey approach* (pp. 15–67). Beverly Hills, CA: Sage.

Rossi, P. H., & Berk, R. A. (1985). Varieties of normative consensus. *American Sociological Review, 50,* 333–347.

Rossi, P. H., Sampson, W. A., Bose, C. E., Jasso, G., & Passel, J. (1974). Measuring household social standing. *Social Science Research, 3,* 169–190.

FALLACY OF COMPOSITION

Fallacy of composition is the failure to accept that what is true for the parts of a whole may not be true for the whole. A special case of this is referred to as the Tragedy of the Commons, where land held in common belongs to everyone collectively rather than to any one particular person. Thus, no one wants to take care of the property, and the land suffers. An everyday example is to reason that if one can see the ballet better by standing on one's seat, then it follows that if everybody stood on his or her seats, we would all have a better view.

—Tim Futing Liao

FALLACY OF OBJECTIVISM

This is the problem of falsely assuming that an entity exists independently of our consciousness. In PHENOMENOLOGY, the analyst seeks to bracket out such assumptions. In the social sciences, it may take the form of falsely assuming that our conceptual categories exist, as opposed to being ways of understanding social reality.

—Alan Bryman

FALLIBILISM

Fallibilism is the deliberate and consistent attempt by researchers to disprove the fallibilities (both methodological and substantive) inherent in the research they are conducting. After the most rigorous scrutiny possible, if some methodological approaches or conclusions remain relatively intact, then researchers feel confident in the RELIABILITY and VALIDITY of these aspects of the research. Fallibilism thus views research as an engagement in continuous critical scrutiny.

As a research approach, fallibilism draws on three intellectual traditions. From the philosophy of science, it builds on the principle of FALSIFIABILITY. As articulated by Karl Popper (1965), falsifiability requires researchers to expose forms of inquiry, and specific HYPOTHESES, to the most rigorous attempt to disprove their central contentions. From this point of view, research is a Darwin-like search for the fittest knowledge. The emphasis is on disconfirmation, on probing for error, omissions, and contradictions.

From PRAGMATISM, fallibilism borrows the idea that inquiry is not a search for final universal truth, but a contingent, future-oriented activity in which researchers accept that the best knowledge available at any one moment is always open to further reformulation. Pragmatism has an EXPERIMENTAL, antifoundational tenor that advocates using an eclectic range of methods and approaches to promote democracy. It encourages the constant critical analysis of assumptions and is skeptical of reified, standardized models of practice or inquiry. The element of fallibilistic self-criticality endemic to pragmatism means avoiding a slavish adherence to a particular methodology.

From CRITICAL THEORY, fallibilism takes that tradition's skepticism regarding dominant ideologies. Critical theory is interested in laying bare the way that ideology—the system of ideas that seems normal, natural, and obvious and that explains the workings of society to people—helps maintain the power of an unrepresentative minority. Its main intellectual project—ideology critique—seeks to expose the way that people learn to embrace ideas and practices that end up harming them, without their realizing that this is happening. Like fallibilism, critical theory assumes that what people view as universal norms, uncontested beliefs, and taken-for-granted explanations are, in fact, humanly constructed and infinitely malleable. Although fallibilism does not share critical theory's concern to expose the way in which dominant ideology legitimizes an unjust social order, it does share its skepticism regarding conventional truth.

A fallibilistic orientation always seeks to expose what is unwarranted, unexamined, and contradictory regarding a research method or set of findings. A good example of such a research effort is Gary Cale's (2001) analysis of how the attempt to encourage critical, oppositional thinking among community college students actually served to increase their resistance to this process.

—Stephen D. Brookfield

REFERENCES

Bronner, S. E., & Kellner, D. M. (1989). *Critical theory and society: A reader.* New York: Routledge.

Cale, G. (2001). *When resistance becomes reproduction: A critical action research study.* Proceedings of the 42nd Adult Education Research Conference. East Lansing: Michigan State University.

Cherryholmes, C. H. (1999). *Reading pragmatism.* New York: Teachers College Press.

Popper, K. E. (1965). *The logic of scientific discovery.* New York: Harper & Row.

FALSIFICATIONISM

Falsificationism is a philosophy of science, also known as critical rationalism, that uses the logic of deduction to provide the foundation for the HYPOTHETICO-DEDUCTIVE METHOD. It was developed in the 1930s by Karl Popper (1959) to deal with the deficiencies of POSITIVISM. The positivist position is rejected in favor of a different logic of explanation based on a critical method of trial and error in which theories are tested against "reality." Falsificationism shares some aspects of positivism's ONTOLOGY but rejects its EPISTEMOLOGY. It adopts the position that the natural and social sciences differ in their content but not in the logical form of their methods.

Although Popper was not a member of the Vienna circle, the group of scientists and philosophers that was responsible for developing logical positivism, he had a close intellectual contact with it. He shared with this tradition the view that scientific knowledge, imperfect though it may be, is the most certain and reliable knowledge available to human beings. However, he was critical of positivism—particularly logical positivism—and was at pains to distance himself from the circle. He rejected the idea that observations provide the foundation for scientific theories, and he recognized the important historical role played by metaphysical ideas in the formation of scientific theories.

Popper argued that observation is used in the service of deductive reasoning; theories are invented to account for observations, not derived from them, as in INDUCTION. Rather than scientists waiting for nature to reveal its regularities, they must impose regularities on the world and, by a process of trial and error, use observation to try to reject false theories (deduction). Theories that survive this critical process are provisionally accepted but never proven to be true. All knowledge is tentative and subject to ongoing critical evaluation. This critical attitude makes use of both verbal argument and observation; observation is used in the interest of argument.

The question of whether theories or observations come first was not a problem for Popper. He accepted that a hypothesis is preceded by observations, particularly those that it seeks to explain. However, these observations presuppose a frame of reference, that is, one or more theories.

In addressing the appropriate logic for the social sciences in his later work, Popper (1961) summarized what he called his main thesis. The logic of the social sciences, like that of the natural sciences, consists in trying out tentative solutions to certain problems; solutions are proposed and criticized. A solution that is not open to criticism must be excluded as unscientific. The ultimate criticism is to attempt to refute the theory. If the theory is refuted, another must be tried. However, if the theory withstands the testing, we can accept it temporarily, although it should be subjected to further discussion and criticism. Hence, the method of science is one of tentative attempts to solve problems by making conjectures that are controlled by severe criticism. The so-called objectivity of science lies in the objectivity of the critical method.

Popper argued that although science is a search for truths about the world or universe, we can never establish whether these theories are true. All that can be done is to eliminate false theories by a process of conjecture and refutation. Some theories will be rejected and some tentatively accepted (corroborated). This process allows us to get as near the truth as possible, but we never know when we have produced a true theory. Some scientist in the future may test the theory in some other circumstances and find it to be false. Therefore, theories must always be regarded as tentative—they may be refined or refuted in the future. All that we can do is get rid of theories that do not match reality. Popper believed that there is no more rational procedure than the method of conjecture and refutation. It is a process of boldly proposing theories, of trying our best to show that these are wrong, and of accepting them tentatively if our critical efforts are unsuccessful.

Criticisms made of the logic of deduction also apply to falsificationism.

—Norman Blaikie

REFERENCES

Blaikie, N. (1993). *Approaches to social enquiry.* Cambridge, UK: Polity.

Chalmers, A. F. (1982). *What is this thing called science?* St. Lucia: University of Queensland Press.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Popper, K. R. (1961). *The poverty of historicism.* London: Routledge & Kegan Paul.

FEMINIST EPISTEMOLOGY

Feminist epistemology consists of principles of EPISTEMOLOGY that have been informed by feminism.

—Alan Bryman

See also FEMINIST RESEARCH, STANDPOINT EPISTEMOLOGY

FEMINIST ETHNOGRAPHY

ETHNOGRAPHY that has been informed by the principles of FEMINIST RESEARCH. As such, feminist ethnography documents women's lives and activities, conducts research and analysis from women's perspectives, and appreciates women's positions and perspectives contextually. There has been some debate about the prospects for feminist ethnography on the grounds that it is difficult to conduct truly according to feminist principles, because it is difficult to conduct in a totally nonexploitative way. Against such a viewpoint, feminist researchers argue that injecting reciprocity into the interactions and transactions between researchers and participants can help to offset an exploitative relationship between researchers and the women they study.

—Alan Bryman

FEMINIST RESEARCH

Feminist research developed as a response to two perceived related failings in western social sciences. The first was the relative invisibility of women and a lack of concern with the gender-specific issues that influenced their lives. The second concerned the practices of social research and the processes through which knowledge was constructed. It was argued that the social world had been studied from the perspective of male interests and concerns, and in ignorance of the different picture that emerged when focusing on women's lives and ways of seeing. Knowledge, which was presented as neutral, objective, and value-free, was, instead, partial and gendered. The purpose of feminist research was to bring women's experiences more fully into view. This necessitated challenging conventional research practices, as well as radically reviewing many of the taken-for-granted assumptions about the nature of social science. The overall aim was to understand better the nature of gender inequalities.

It is not the focus on women, gender, or gendered lives per se, however, that makes a research project feminist. At first, feminist research was defined as such if it was seen to be about, by, and for women. More recently, however, it has been suggested that feminist approaches to research can be identified through their framing by theories of gender and power, their normative frameworks and notions of justice (however conceived), their focus on transformation and social change, and ideas about ethics and accountability (Ramazanoglu, 2002). Views about these issues are not held uniformly, and identification of when research might be feminist is never an open-and-shut case. There is also overlap between the defining features of feminist work and other approaches to social investigation.

Early on in their discussions about doing social research, feminists distinguished between method, methodology, and EPISTEMOLOGY (Harding, 1987). Method was used to refer to the techniques and tools used in conducting research. Methodology was concerned with how those methods were actually operationalized and put into practice. Epistemology focused on the philosophical arguments for deciding what kinds of knowledge are possible and how to ensure that they are adequate and legitimate. Feminists have continually refined their ideas about these issues, and there has been extensive debate in relation to them.

METHODS OF RESEARCH

During the early 1980s, much of the debate about feminist research was concerned with what constituted the appropriate methods to be used. Concerns were expressed about the greater legitimacy that seemed to be afforded to SURVEY research and to approaches that emphasized the importance of scientific MEASUREMENT and OBJECTIVITY at that time. It was argued that such research fractured people's lives, resulting in the production and measurement of atomistic facts, the significance of which had been decided in advance of the research itself. This was seen as problematic when focusing on women and, particularly, for research into those areas of their lives that had remained relatively hidden. With issues such as violence, sexuality, and motherhood, for instance, there was little existing knowledge from which survey questions might be written.

As a result, many feminists argued that qualitative methods, founded on an INTERPRETIVIST philosophical position and using sensitive, flexible, and open-ended approaches to data generation, were more appropriate than quantitative ones. Methods such as IN-DEPTH INTERVIEWS, LIFE HISTORY INTERVIEW, and ETHNOGRAPHY focused on the meanings and interpretations of those being researched, thereby enabling the researcher to see the social world through participants' eyes. It was also felt that qualitative approaches were more in keeping with feminist principles, which promote equality and are against objectification and subordination. These arguments became something of an orthodoxy among feminist researchers, although even then, not everyone agreed with this position. For instance, awareness of the prevalence of sexual violence or the poverty of lone mothers is dependent, in part, on the availability of data that can indicate the extent of the problem.

In recent years, the old orthodoxy about QUALITATIVE RESEARCH has been breaking down. Some commentators have claimed that such methods may be more exploitative than quantitative ones because they rely on assumptions about trust and RAPPORT, which may mask relations of power and inequality. Feminists have been adopting a more pragmatic approach toward mixing methods and even adopting quantitative ones in a bid to develop a more emancipatory social science that can inform policy development. Indeed, some have suggested that there may be circumstances (e.g., investigating sexual abuse) where respondents may find completing a questionnaire to be less stressful than being interviewed. However, limitations of funding and resourcing often mean that relatively small-scale qualitative projects remain the main way in which many feminists are able to contribute to the research agenda.

METHODOLOGY

The debate about methods of research has been related to a second issue concerning feminist methodology or research practice. It has been suggested that what distinguishes feminist research from other forms has less to do with methods per se and more to do with the kinds of questions that are asked, the relationship of the researcher to the research and its conduct, and the intentions and purpose behind the work. There has been particular concern about how to ensure that research takes place in nonexploitative and ethically responsible ways. A central issue here has been the possible existence of structural inequalities between researchers and their participants and the extent to which these might be minimized. It has been argued that feminists who replicate, during their research, the kind of power relations of which they are critical elsewhere are hypocritical, and such action undermines the knowledge produced.

A range of strategies for dealing with such issues has been debated, with some filtering into the more mainstream practices of nonfeminist research. These have involved, for instance, developing good rapport and communication with participants in research by telling them as much as possible about the content and aims of the research; responding to participants' questions, problems, and concerns as honestly and sensitively as possible; and making available the names of organizations or publications that can provide any necessary information or social support. Some feminists advocate obtaining feedback from the researched on the interpretation and analysis of research data in order to ensure that women's voices are heard without distorting or exploiting them. Emphasis has been placed on including participants in the design and implementation of a research project, as well as ensuring that findings are disseminated to those who have been involved.

However, although some feminist researchers seem to assume that it is possible to remove most aspects of power relations from the research process, many are less sanguine. It may be all too easy for an educated researcher to persuade participants to talk conversationally, in relaxed circumstances, about their lives. But with the benefit of hindsight, they may feel it would have been better to remain silent. It is also easy for researchers to believe that they have established relations of harmony and equality, when the perceptions of those whom they are studying is otherwise. Some feminists have also suggested that the power of researchers can be overstated, with the researched being ascribed a passive, victimized status, where they have no apparent space for resistance.

One way in which the issue of power in research has been addressed is through the concept of REFLEXIVITY, which has become something of a central principle in feminist methodological concerns. Reflexivity has two distinct aspects. The first involves making explicit how power might be exercised in and affects the research process despite strategic efforts to minimize its impact. It focuses on the assumptions and ethical judgments that frame the research and how the research agenda and process have been constructed. It thereby demonstrates how researchers can account for the knowledge they produce (Ramazanoglu, 2002). The idea is to turn the critical methods of social researching upon the practice of social researching itself.

The second form of reflexivity explores the social situatedness of the researcher. This refers to how the personal biography and position of the researcher relate to the project in which she is engaged. The argument is that researchers cannot be separate from their work, which has no existence apart from their involvement in it. Therefore, it is important for a critical reflection on this relationship to be included as part of any discussion, analysis, or interpretation of research findings.

In fact, feminists have claimed that this two-fold reflexive commentary on research is essential to making knowledge accountable. They are critical of what is referred to as "weak" objectivity, where attempts are made to purge knowledge, along with the research on which it is based, of all values (Harding, 1991). This form of objectivity is weak because, by pretending that such values do not exist in research, they are included in hidden and unexplicated ways. In contrast, what is known as "strong" objectivity necessarily involves critical scrutiny of all evidence marshalled as part of the research process (Harding, 1991). Strong objectivity, with its emphasis on reflexivity, includes the systematic examination of background beliefs and cultural agendas. By making this part of the research process itself, it is argued that the knowledge produced becomes more rigorous, robust, and defensible.

There is debate, however, as to how much reflexivity and strong objectivity might actually be achieved. Being reflexive in the ways suggested can be a long and drawn-out business, with difficulty in establishing guidelines as to how far or for how long the process should continue. There are dangers that accounts of the research procedure and of researchers' involvement come to dominate the analysis and findings of the research. This could result in the complexities of participants' lives being reduced to the researcher's own autobiographical history.

One issue that has significantly affected feminists' views on methodology relates to the concept of difference. Early assumptions that there is an essential womanhood and that women's oppression is universal, homogeneous, and shared have been challenged. An emphasis on difference draws attention to diversities among women, particularly those deriving from ethnicity, sexuality, disability, class, and age. This means that just being a woman researching women cannot be sufficient in the effort to establish rapport and minimize hierarchy. These other axes of experience will also impinge on the research process, and this is why it has been suggested that researcher and researched might be matched in terms of shared characteristics. However, it is not always easy to establish which differences are likely to have an influence in a particular study and what the precise nature of this impact might be. Overall, then, shared gender characteristics do not themselves promote shared understandings, and it cannot be assumed that trust can be established easily on the basis of gender alone.

EPISTEMOLOGY

The feminist concern with epistemology has centered on who knows what, about whom, and how this knowledge is legitimated. Early writers were critical of those who simply conducted their studies of women within already existing frameworks. In order to make the meaning of women's lives more visible, it was necessary to analyze it from their point of view. This has come to be known as STANDPOINT EPISTEMOLOGY (Hartsock, 1998). There are different ways of thinking about taking a feminist standpoint, but the key is that women's particular gendered experiences produce distinctive and privileged understandings. This access to different knowledge makes it possible to reveal the existence of forms of human relationships that may not be visible from a man's position. Therefore, standpoint epistemology offers the possibility of new and more reliable insights into women's lives because it is grounded in women's experiences, including emotions and embodiment.

Standpoint epistemology is also based upon a REALIST approach to knowledge; the idea that systematic or structural factors influence LIVED EXPERIENCE, and that these structures are knowable. The whole point of feminist research has been to produce continually better understandings of women's lives. This view, however, has been challenged by developments in postmodern thinking, and the standpoint perspective is often contrasted with postmodern accounts.

As with all forms of POSTMODERNISM, feminist thinking has different forms and emphases. However, it shares most of the characteristics of other (non-feminist) variants. There is the criticism of generalizations, universal categories, the existence of a stable and coherent self, the transparency of language, and the ability of scientific rationality to produce "truth" (Nicholson, 1990). The focus, instead, is on discourse, textuality, fragmentation, multiple subjectivities, and flux. Whereas the standpoint position acknowledges that accounts are selective constructions, while also agreeing that they are able to represent independent phenomena with some degree of rigorousness and reliability, feminist postmodernism rejects this. Instead, some have claimed that feminist postmodernism sweeps away the foundations of feminist methodology (Ramazanoglu, 2002). It does this by arguing that because social phenomena are only to be apprehended through the use of discourse, and because the very practice of discourse serves to invoke forms of sociality, it is therefore impossible to understand the social world without also simultaneously constructing it. The latter then becomes the task of what it means to "know." The emphasis switches from trying to investigate the nature of social relations from accounts of women's experiences to the discourse and texts through which things come to be known. Indeed, the very idea of experience itself, as something that people have, and that can be taken at face value unquestioningly, becomes problematic. All of this causes major difficulties for social researchers who wish to explore critically the nature of the social world; from postmodern perspectives, this is impossible. While not denying their own material embodiment or that they live in a material world, feminist postmodernists claim that particular links between experience and material realities cannot be specified (Ramazanoglu, 2002). Rather, feminist research is transformed into the continuing deconstruction of identities, subjectivities, and selves, which is then subjected to further deconstruction.

Postmodern thinking has had a considerable impact on feminist research, with some regarding any attempts at any empirical work as misguided and defective. Others argue that it is possible to use some of the insights to be gained from a postmodern approach, particularly when researching culture, signification, and modes of representation. The problem for the feminist social researcher is how to strike a balance between an empirical focus on embodied and material differences, power, and inequality, and critical reflections on how knowledge is produced. The question as to whether it is possible to combine concerns about the relationship between subjectivity and the production of texts/interpretations with empirical methods of social investigation is still being debated. It produces differences among feminists, who respond with various and divergent answers.

—Mary Maynard

REFERENCES

Harding, S. (Ed.). (1987). *Feminism and methodology.* Milton Keynes, UK: Open University Press.

Harding, S. (1991). *Whose science? Whose knowledge?* Milton Keynes, UK: Open University Press.

Hartsock, N. C. M. (1998). *The feminist standpoint revisited and other essays.* Boulder, CO: Westview.

Nicholson, L. (Ed.). (1990). *Feminism/postmodernism.* London: Routledge.

Ramazanoglu, C., with Holland, J. (2002). *Feminist methodology.* London: Sage.

FICTION AND RESEARCH

Until the mid-20th century, social research tended to reduce fiction to determining the social contexts of its literary production—a POSITIVIST legacy in evolutionism, structural-functionalism, and Marxism. An alternative, *Literaturwissenschaft* approach, analyzes fiction for knowledge about the society of its production and to which it refers, and informs analyses of fiction by cultural studies (Milner, 1996). It combines two distinct approaches: Extrinsic approaches analyze fiction as a critical reflection on the sociocultural conditions of its production, distribution, and reception; intrinsic approaches analyze it as a reflexively constitutive feature of its society, in which linguistic structures of TEXTS are analyzed as structural homologies of social relations. Intrinsic approaches seek to avoid reductions of fiction to illustrations of social realities by emphasizing its textual linguistic subtleties and complexities, but they raise research questions of representativeness of majority experience in complex modern societies and are applied more effectively to the fiction of both minority, selective cultural traditions and avant garde modernism. Extrinsic approaches provide more effective analyses of popular fiction because of their lack of sustained comparative concern with social experience in other cultures and periods (Loewenthal, 1989).

More recent sociolinguistic and neo-Marxian STRUCTURALIST research on relations between language and social structure as constitutive of texts (Filmer, 1998) has focused on the HERMENEUTIC issue of how the language of fiction indicates possible readings. This requires openness to a text's capacity to invent both itself and the reality to which it refers. The following approaches address this.

Semiotics and Structural Linguistics

SEMIOTICS and STRUCTURAL LINGUISTICS provide theories of signs required to make sense of the language of the text and its relations to languages of everyday life and other critical discourses (Eco, 1981).

Phenomenological Sociology

Phenomenological sociology analyzes relations between meanings of idiomatic expressions and connotations contained within the text's version of reality and those of the external realities of readers' worlds of experience and discourse (Schutz, 1964).

Structures of Feeling

Structures of feeling analyzes the underlying symbolic and semantic organization of texts as expressions of emergent experiential senses of changing contexts, prefiguring possible new social structures (Prendergast, 1995).

Genetic Structuralism

Genetic structuralism applies structuralist Marxist analysis of fiction to articulate world visions of emergent social groups. This shared consciousness is analyzed for its potential in enabling the groups developing it to change the existing institutional social and political order (Goldmann, 1981).

All of these approaches engage with fiction as a reflexive relation between language, subjective imagination, and social structures to establish the adequacy of fiction as a resource on fundamental topics of social research: structures of relations between individuals and society, processes of social change, and meanings as well as causes of social action. In doing so, they seek to transcend the specificities of personal biographies, particular cultures, and historical periods.

—Paul Filmer

REFERENCES

Eco, U. (1981). *The role of the reader: Explorations in the semiotics of texts.* London: Hutchinson.

Filmer, P. (1998). Analysing literary texts. In C. Seale (Ed.), *Researching society and culture.* London: Sage.

Goldmann, L. (1981). *Method in the sociology of literature* (W. Boelhower, Trans. & Ed.). Oxford, UK: Basil Blackwell.

Loewenthal, L. (1989). Sociology of literature in retrospect. In P. Desan et al. (Eds.), *Literature and social practice.* Chicago and London: University of Chicago Press.

Milner, A. (1996). *Literature, culture and society.* London: UCL Press.

Prendergast, C. (1995). *Cultural materialism: On Raymond Williams.* Minneapolis: University of Minnesota Press.

Schutz, A. (1964). Don Quixote and the problem of reality. In *Collected papers,* Vol. 2. The Hague: Martinus Nijhoff.

Williams, R. (1977). *Marxism and literature.* Oxford, UK: Oxford University Press.

FIELD EXPERIMENTATION

Field experiments, as distinct from LABORATORY EXPERIMENTS, are randomized interventions that take place in naturalistic settings. The nature of these interventions varies widely. Examples include randomized school voucher programs, voter mobilization campaigns, police raids on crack houses, job training programs, income tax schedules, and health insurance plans. The unit of analysis in field experimentation also varies. Most field interventions target individuals, but many studies randomly assign treatments to groups or institutions.

In contrast to ethnographic or descriptive research, field experiments are principally designed to establish CAUSAL relationships. Well-executed field experiments combine the strengths of randomized designs with the EXTERNAL VALIDITY of field studies. Whereas a laboratory study might examine the effects of commercial advertising by exposing TREATMENT and CONTROL GROUPS to different advertising stimuli within an artificial setting and gauging their purchasing preferences within the context of a simulated marketplace, field experiments randomly manipulate the content and timing of actual advertisement campaigns and attempt to link these varied interventions to observed patterns of consumer demand. Both types of studies use randomization, but the latter has the advantage of linking cause and effect in terms that have direct, real-world applicability.

Field experimentation is also a valuable tool in program evaluation, particularly when used to gauge the effects of a new program or alternative versions of existing programs. Suppose, for example, that one sought to evaluate the effectiveness of school-based drug prevention programs. Some schools might receive drug awareness programs and others might not, the sequence in which schools receive drug awareness programs might be varied randomly, or the entire population of schools might receive alternative drug awareness programs. By comparing outcomes in treatment and control groups that were initially formed by random chance, the evaluator may make an unbiased assessment of the intervention's effects.

One attractive feature of experiments in natural settings is the transparency of the data analysis. In contrast to nonexperimental data analysis, where results often vary markedly depending on the model the researcher imposes on the data and where researchers often fit a great many models in an effort to find the right one, experimental data analysis tends to be quite robust. Simple comparisons between control and treatment groups often suffice to give an unbiased account of the treatment effect, and additional analysis with COVARIATES merely estimates the causal parameters with greater precision. This is not to say that experimental research is free from data mining, but the transformation of raw data into statistical results involves less discretion and therefore fewer moral hazards.

In principle, field experimentation represents the strongest basis for sound causal inference available to social scientists, but in practice, this methodology faces important ethical and pragmatic constraints. Rarely do social scientists have the opportunity to manipulate the variables of most interest to them, such as culture, political systems, economic prosperity, and so on. Even in situations where random interventions are attempted, they are sometimes undone by the people charged with implementing them. The well-intentioned physician who smuggles the sickest patients into the treatment group may cause the experiment to produce misleading results.

The challenge of orchestrating field experiments and maintaining the integrity of the randomization means that the experiments tend to occur in a small number of sites that are chosen for reasons of logistics rather than representativeness. This constraint raises the issue of whether the study's conclusions apply only to the types of people who actually participate in an experiment. In general, REPLICATION is the appropriate response to concerns about drawing conclusions based on studies of particular times, places, and people.

A further complication arises when experimental subjects refuse to participate in the study or cannot be reached for treatment. Although noncompliance and attrition diminish the power of an EXPERIMENTAL DESIGN, they are remediable problems as long as the decision to participate is unrelated to the strength of the treatment effect. The statistical correction is to perform an instrumental variables regression in which the independent variable is whether a subject was actually treated and the INSTRUMENTAL VARIABLE is whether a subject was originally assigned to the TREATMENT group. The resulting effect estimate is termed the "effect of treatment on the treated." Whether the effects of the treatment do, in fact, vary may be studied empirically by experimentally manipulating participation rates (e.g., by varying inducements offered to subjects).

In light of these practical and ethical concerns, social scientists often turn to NATURAL EXPERIMENTS, where treatment and control groups are formed in ways that resemble random assignment. For example, if municipal leaders increase the size of the police force during election years in an effort to placate a crime-conscious citizenry, the electoral cycle provides a means for studying the effects of policing on crime rates. Similarly, if public schools admit students to kindergarten on the basis of their birthdates, one may study the effect of kindergarten enrollment on intellectual development by comparing children whose birthdays fall immediately before and after the cutoff date. Like field experimentation, these approaches seize upon opportunities to learn about the consequences of near-exogenous variation in naturalistic settings. Unlike experimentation, the exogeneity of the independent variables is not ensured by means of randomization procedures.

—Donald P. Green

REFERENCES

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91,* 444–455.

Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research.* Washington, DC: Brookings Institution.

Green, D. P., & Gerber, A. S. (2002). Reclaiming the experimental tradition in political science. In I. Katznelson & H. V. Milner (Eds.), *Political science: State of the discipline* (3rd ed., pp. 805–832). New York: W. W. Norton.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives, 9,* 85–110.

Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural "natural experiments" in economics. *Journal of Economic Literature, 38,* 827–874.

FIELD RELATIONS

The definitive characteristic of social science field research is the encounter between researcher(s) and researched. Encounters in the field may vary from those that are short-lived to those that are sufficiently long-lived to form the basis of ongoing relationships. In the latter case, the nature of relations between the researcher(s) and the researched is central to the progress and success—or otherwise—of the research.

Perhaps the most obvious point is that all intensive field research requires the careful and reflexive cultivation of ongoing good relationships with a wide range of people. Two of the most important field relations are with *gatekeepers* and *key informants.* These are vital—indeed, probably unavoidable—if long-term field research is to be successful. Gatekeepers facilitate initial ACCESS to the field and to information, situations, and individuals within it, but their continued sponsorship remains important throughout the research period. Typical gatekeepers include politicians, archivists, senior managers within organizations, community leaders, and core members of informal networks. Key informants, of whom the archetypal example is William Foote Whyte's "Doc" in *Street Corner Society* (Whyte, 1955), may also be gatekeepers, and often are, but the reverse is not necessarily true. Key informants and gatekeepers become disproportionately significant if the fieldworker lacks working fluency in the local language.

Among the problems raised by the importance of gatekeepers and key informants are how to identify the most useful gatekeepers; how to prevent oneself from becoming overidentified with gatekeepers and key informants; how to prevent gatekeepers from exercising proprietary control over the research; and how to avoid privileging the accounts and knowledge of a few, typically not disinterested, key informants over the points of view of other research subjects. None of these has a straightforward, textbook solution.

Relationships in the field are particularly important when the researcher has an already existing *membership role* within the social setting under investigation (Adler & Adler, 1987). There is much to gain in terms of privileged access and understanding from insider status, whether that be total or partial, and the realities of much social research today, particularly evaluative work, are that this is increasingly common. However, in such situations, it is possible to identify a number of issues that demand attention, including maintaining critical distance and objectivity, dealing with the demands made by other members in terms of their personal or corporate interests, balancing ethical responsibilities, and managing subsequent reentry into the situation as a nonresearcher.

Finally, perhaps the most pervasive and routine sense in which relations between fieldworker(s) and research subjects are consequential for research outcomes reflects the crucial importance of *gender* in the everyday human world (Warren & Hackney, 2000). This is not surprising, given that everyday life, no matter what the setting or culture, is organized around gender. Who can go where, with whom, and who can speak to whom are gender issues that are critically important for the conduct and outcome of all field research, whether the topic of the research is gender-oriented or not.

—Richard Jenkins

REFERENCES

Adler, P., & Adler, P. A. (1987). *Membership roles in field research* (Qualitative Research Methods Series vol. 6). Newbury Park, CA: Sage.

Warren, C. A. B., & Hackney, J. K. (2000). *Gender issues in ethnography* (Qualitative Research Methods Series vol. 9, 2nd ed.). Thousand Oaks, CA: Sage.

Whyte, W. F. (1955). *Street corner society* (2nd ed.). Chicago: University of Chicago Press.

FIELD RESEARCH

The expression "field research," or "fieldwork," derives in the first instance from natural science, where it denotes the collection of data in a naturalistic setting that is outside the controlled environment of the laboratory. In our context, despite its recent overidentification with ethnography and other qualitative approaches (e.g., Burgess, 1982; Whyte, 1984), most social research can probably lay claim to the name. Whether it be the fleeting doorstep exchanges of the large-scale social SURVEY or the long-term intimacies made possible by PARTICIPANT OBSERVATION, social research involving primary data typically involves at least some fieldwork—defined simply as encounters between researchers and subjects in uncontrolled environments—of one sort or another (with the most significant exceptions being ARCHIVAL RESEARCH, media studies, and FOCUS GROUP studies).

Why should social researchers depend so heavily on fieldwork? Or, to ask the question more directly, why do social researchers rarely, if ever, engage in experiments? There is no single answer. The ethics of the broad sociological tradition may provide one key constraint to experimentation on humans (although the fact that ethically sensitive disciplines such as medicine and psychology continue to rely on experimental evidence suggests that ethical issues are unlikely to be determinate). Three other factors are, perhaps, more to the point.

The first is the difficulty in creating sufficiently authentic social conditions out of their actual context (with the exception of certain small-scale and definitely bounded situations). Simulation is, after all, simulation, and it is probably of little relevance for our understanding of the everyday human world. Second—and it is a related point—it is difficult to maintain properly controlled experimental conditions in social research contexts. Each of these problems means that for most purposes, there is no social research substitute for NATURALISM.

Third, and perhaps most important, there is the matter of predictability. Humans are reflexive beings. Those aspects of human behavior with which social research is concerned derive from a mixture of emotion, rational decision making, habit, and compulsion. For this reason, humans cannot be relied upon to do in the future what they have done consistently in the past. Because the goal of the experimental method is to accumulate sufficient accurate observations of an event and the controlled conditions under which it occurs to permit reliable *prediction* (i.e., the formulation of probabilistic generalizations), it clearly has limited application in social research.

This is not to say, however, that there are no approximate analogues of experimentation in field research. In particular, STRUCTURED INTERVIEWS have something in common with the experimental method, in that the object of the exercise is to discover what a sample of subjects will say in answer to the same questions asked in the same controlled way, or as close as is possible. Rigorous sampling, too, may be seen as another proxy for experimental control. Furthermore, projective questions of the "What if?" kind have an essential experimental quality to them (they certainly have something in common with what philosophers call "thought experiments").

Nevertheless, with its necessarily uncontrolled environment, and in the intractability of the human behavior with which social research is concerned, field research is very different from the laboratory experiment. Epistemologically, field research also has some distinctive characteristics. The most obvious is its utter reliance on researchers, rather than instruments, for the perception and recording of data. The researcher *is* the research instrument, even when using the most structured schedules or protocols. As a consequence, it can be difficult to say where data collection finishes and its interpretation begins.

This also means that in field research, the line between method and data can sometimes blur. For example, when Howard Newby studied English farm workers, the fact that the only way he could find and contact a sample of laborers was via their employers was not merely an interesting detail of method, it was a telling indicator of their social isolation (Newby, 1977). Similarly, that female researchers typically have better access to the lives of young men than their male counterparts have to young women says a great deal about the asymmetrical character of public and private gendered, everyday worlds.

These examples suggest a further point: that the nature, extent, and boundaries of the field are not necessarily self-evident. Unless one is talking about research settings that are clearly delimited territorially and/or institutionally—as in the small settlements and workplaces that have hosted some of the classic ethnographies—then defining the field of study is one of the most important, and often least straightforward, preliminary tasks of fieldwork. In addition to the issues of sampling and gender already mentioned, size and complexity are necessary considerations. Such matters affect which methods are appropriate and what the costs are of adopting one approach rather than another. For example, most modern settings are too extensive, too heterogeneous, and too multifaceted to be easily accessible "in the round" via participant observation (Jenkins, 1984). Thorough fieldwork commonly demands a degree of flexibility and openness with respect to method.

What one can find out in the field also depends on the politics of the research process. With respect to politics, the key issues involve questions of researcher identity, the role of gatekeepers, and (over)reliance on particular sources. Who the researcher is, via whom access to the field is negotiated, and with whom researchers develop key informant relationships are all factors that will influence what can be discovered, to whom one can talk, who will respond to requests for cooperation, and which local points of view are represented (see FIELD RELATIONS).

Given the above issues, why do social researchers continue to undertake field research? To return to the point of departure, with a few notable exceptions, the collection of primary social research data allows for no alternatives. Field research is a continued necessity if social research is to maintain a critical engagement with the human world and its changes. Archives, the media, and focus groups can contribute only so much to that task, after which there is no substitute, whatever methods we adopt, for encountering humans in their everyday settings.

—Richard Jenkins

REFERENCES

Burgess, R. G. (Ed.). (1982). *Field research: A sourcebook and field manual*. London: Allen & Unwin.

Jenkins, R. (1984). Bringing it all back home: An anthropologist in Belfast. In C. Bell & H. Roberts (Eds.), *Social researching: Politics, problems, practice* (pp. 147–164). London: Routledge and Kegan Paul.

Newby, H. (1977). In the field: Reflections on a study of Suffolk farm workers. In C. Bell & H. Newby (Eds.), *Doing sociological research* (pp. 108–129). London: Allen & Unwin.

Whyte, W. F. (1984). *Learning from the field: A guide from experience*. Beverly Hills, CA: Sage.

FIELDNOTES

Fieldnotes constitute the ETHNOGRAPHIC record of ethnographic or participant observation research and are arguably the most essential part of the field research enterprise. Comprising the chronological log of experiences in the field, they include descriptions of people, events, the fabric of the setting, conversations with people, observed interactions, and sequences and duration of events, as well as the researcher's experiences connected to the investigation. It is important, too, that the observer note anything about the general state of him- or herself that might have bearing on the form or quality of the data. These notes may also contain interpretations of observed events and interactions and insights into the culture studied. In short, fieldnotes comprise the contents from which the analysis derives.

Fieldnotes should be written up as completely and comprehensively as soon as possible. Researchers generally agree that fieldnotes are more vivid and detailed the sooner they are recorded. However, in many situations, it is impossible, inconvenient, or inappropriate to record observations immediately. In such cases, it is usual to make temporary notes first, which essentially provide a summary that could be used to write up more detailed notes later. Such condensed notes, which typically include phrases, single words, and unconnected sentences, can be recorded quickly and even inconspicuously, and serve as memory triggers to help reconstruct an event. Then, the researcher can expand on this condensed version by filling in the details and recalling things that were not recorded on the spot. To be sure, field notes are comprehensive only in the relative sense. Even permanent notes are altered, added to, corrected, and updated as the researcher becomes more familiar with the range of data at his or her disposal and with emerging hypotheses and interpretations.

Labor-intensive and requiring considerable self-discipline, some fieldnotes may, at first, seem obvious and trivial. However, researchers should not assume in advance what kinds of information fall within this domain. At least initially, the observer's mind-set should be that everything occurring in the field is potentially important and, therefore, worth recording. As the research becomes more focused, the researcher is more selective in what is written down.

Because researchers are selective in what is observed and recorded, and the specific words and terminology used, the language employed necessarily, even if inadvertently, imposes a structure on the social world under study. To mitigate against the obtrusiveness of preconceptions, cultural and otherwise, researchers should record verbatim what people say and avoid writing summaries of people's remarks or restating them in a more familiar language.

Fieldnotes should be reviewed frequently because this process allows them to be corrected and completed. Such a review also assists in the tasks of data categorization and analysis, which are best pursued simultaneously with data collection. Thus, for example, fieldnotes may contain analytic MEMOS that complement those of a more substantive and methodological nature and that form the core of the preliminary analysis. Such memos may include summaries that are written at the end of a day in the field in which the researcher indicates themes that have emerged and concepts that might be developed, together with preliminary thoughts about the analytic framework. Moreover, the review also provides a guide to areas where additional data are required. If their contribution to the emerging analysis is not considered systematically, and if accumulated for their own sake, fieldnotes become progressively less useful and, perhaps, even useless.

Generating and organizing fieldnotes is a personal activity, so researchers rely upon a format that suits their taste and requirements. Generally, it includes some variation of the following: Each set of notes begins with the date, time, and place of observation, and when the notes were recorded; where appropriate, pseudonyms are used for names of people and places; and verbatim descriptions of people, events, and situations should be distinguished from those that are paraphrased. Because the objective is to record information as correctly and quickly as possible, there is little need to be overly concerned with grammar and spelling. The literature includes stylistic suggestions distinguishing between verbatim accounts and those based on reasonable and approximate recall.

Proficiency at writing up fieldnotes is gained over time and with practice. Experience, however, does not diminish the reality that writing fieldnotes is tedious, requiring patience and perseverance. However, the work is not characterized solely by drudgery: Writing fieldnotes is frequently accompanied by flashes of insight, excitement, and understandings about the social world under investigation.

Although most observers traditionally rely on their memories for recording data, it is not unusual for researchers to use recording devices. Despite the absence of absolute rules regarding such practices, a general consensus is that the researcher's obtrusiveness may detract from the quality of the data that are collected, recording devices should be used advisedly, and then only after some measure of familiarity with the setting and its key participants has been achieved.

Prior to the advent of computer programs, the task of categorizing fieldnotes was arduous. Computer programs for qualitative data can now perform many of the tedious, time-consuming, and mechanical operations associated with traditional analytic procedures as well as allow for the more rapid retrieval of field data. These operations can be performed with greater efficiency and accuracy, thus allowing the researcher to devote more time to the interpretive phase of data analysis. However, in and of itself, the technology adds little to nothing to the organization and analysis of fieldnotes without the researcher's conceptual input.

Because a good ethnography is only as good as the field notes upon which it is based, it is surprising that published accounts of field research methods pay relatively little attention to how the notes are written and organized. Their importance is underscored by the generally accepted view that field researchers may as well not spend time in the research setting if they fail to record their fieldnotes in a timely manner.

—William Shaffir

REFERENCES

Berg, B. L. (2001). *Qualitative research methods for the social sciences.* Boston: Allyn and Bacon.

Lofland, J., & Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis.* Belmont, CA: Wadsworth.

Taylor, S. J., & Bogdan, R. (1998). *Introduction to qualitative research methods: A guide & resource.* New York: Wiley.

FILE DRAWER PROBLEM

The file drawer problem represents a threat to the VALIDITY of the results and conclusions of a META-ANALYSIS or research synthesis. This phenomenon, also known as *publication bias*, or occasionally as *retrieval bias*, reflects a concern resulting from the fact that meta-analysts generally are not able to retrieve and include the entire POPULATION of studies that were conducted on the topic of interest. The term itself was coined by Robert Rosenthal (1979) to suggest the possibility that the studies that remained tucked away in researchers' file drawers contained results that were different from those included in the meta-analysis, and could therefore potentially invalidate the results of the meta-analysis.

The file drawer problem is based on the belief that there is a SELECTION BIAS operating such that studies overestimating the magnitude of an effect are more likely to find their way into a meta-analysis than are studies underestimating the magnitude of the effect. This is because, for any given sample size, the study that overestimates the TREATMENT effect is more likely to have a significant *P* VALUE and is thus more likely to be published. Because published studies are easier to locate and retrieve than unpublished ones, the possibility arises that a meta-analysis based primarily on published studies may be unrepresentative of the body of research as a whole.

Three kinds of techniques have been developed to help meta-analysts deal with publication bias. One set of techniques is designed to detect publication bias and is best exemplified by a graphical diagnostic called the funnel plot (Light & Pillemer, 1984), although it also includes explicit statistical tests such as those proposed by Mattias Egger and his colleagues (Egger, Smith, Schneider, & Minder, 1997). The second set of techniques is usually referred to as file drawer analyses because it typically involves imputing the number of EFFECT SIZE estimates with zero effects (corresponding to the unpublished studies left in researchers' file drawers) that would be necessary to reduce the observed meta-analytic result to zero, or, alternatively, to a theoretically or clinically meaningless level (Orwin, 1983). These are based on Rosenthal's (1979) original suggestion of computing the number of "missing" studies with zero effect that would be necessary to reduce the meta-analytic result to statistical insignificance. This number has come to be known as the Fail-safe *N*. When the required number of "missing" studies is implausibly large, the meta-analytic findings may be considered robust to the potential effects of publication bias. Meta-analyses based on small numbers of studies are particularly vulnerable to the file drawer problem. In this case, a small number of unretrieved studies with results that differ from those that were retrieved and included could weaken or overturn the meta-analytic findings.

The third set of techniques is designed to adjust effect size estimates for the possible effects of publication bias under some explicit model of publication selection. The best known of these techniques is the trim-and-fill method developed by Sue Duvall and Richard Tweedie (2000).

Prospectively, the file drawer problem may be ameliorated by a comprehensive search strategy that includes unpublished studies, by research registries, and by changes in editorial policies.

—Hannah R. Rothstein

REFERENCES

Duvall, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 276-284.

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634.

Light, R., & Pillemer, D. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157-159.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638-641.

FILM AND VIDEO IN RESEARCH

Since the early 20th century, film and, more recently, video have been used across the social sciences to visually record and represent human behavior and culture. Film was first used in QUALITATIVE RESEARCH at the end of the 19th century, when, for example, the anthropologist A. C. Haddon used film in an expedition to the Torres Straits. Anthropologists produced film as archive reference footage or as filmic representations of culture, developing the documentary genre of *ethnographic film.* Robert Flaherty's film *Nanook of the North* (1921), a dramatic depiction of the lives of North American Eskimos, is known as the first ethnographic film. Today, ETHNOGRAPHIC film usually follows an observational film style. Examples include the late-20th-century Granada Television *Disappearing World* series. More recently, technological developments in digital video that allow higher image quality and less costly editing processes have attracted ethnographic filmmakers to video rather than film. For example, David MacDougall (2001) has argued that digital video allows ethnographers to produce films that correspond better to academic principles, in contrast to the high budget films that conformed to the demands of TV broadcast standards. MacDougall discusses how his use of digital video in his research in India has allowed him to explore new issues and questions in greater depth and to collaborate more closely with the subjects of his films.

The use of film and video as a research method also continued throughout the 20th century. Film and video continued to be used to visually record research activities such as FOCUS GROUPS, and to study body movement and gestures. This has included extending the use of video in PARTICIPANT OBSERVATION and IN-DEPTH INTERVIEWING (both already aspects of ethnographic filmmaking). For example, in research with a dance group, Thomas (1997) participated by videotaping rehearsals, which the group later viewed. Barnes, Taylor-Brown, and Weiner (1997) videotaped spoken statements made by HIV-positive mothers for their children. Like Thomas, Barnes et al. also produced visual materials that were useful to their sociological analysis but also had a purpose for the subjects of their research. In this case, the videos were to be viewed by the women's children in the future. Similarly, researchers might ask informants to produce their own video diaries of aspects of their everyday lives, experiences, and identities. Such methods allow researchers to produce both a visual and an oral record of an interview or research situation, as well as a visual text that represents the experience of research and that can therefore be interrogated to understand the processes through which knowledge was produced in research.

Developments in digital video technology have made video research methods increasingly popular in ethnographic research since the late 1990s. Indeed, the possibilities offered by digital video suggest that it will become a popular visual research tool. New digital technologies also allow video to be transferred directly from a camera to a personal computer. This facilitates either digital video editing of ethnographic films or the digital analysis and storage of video and other research materials. Thus, with limited technical skills, contemporary visual researchers can work relatively independently.

—Sarah Pink

REFERENCES

Barnes, D. B., Taylor-Brown, S., & Weiner, L. (1997). "I didn't leave y'all on purpose": HIV-infected mothers' videotaped legacies for their children. *Qualitative Sociology, 20*(1), 7–32.

MacDougall, D. (2001). Renewing ethnographic film: Is digital video changing the genre? *Anthropology Today, 17*(3), 15–21.

Pink, S. (2001). *Doing visual ethnography: Images, media and representation in research.* London: Sage.

Thomas, H. (1997). Dancing: Representation and difference. In J. McGuigan (Ed.), *Cultural methodologies.* London: Sage.

FIRST-ORDER. *See* ORDER

FISHING EXPEDITION

Fishing expedition describes a situation in which the researcher has no particular aims for the research at hand. Generally, the term may refer (often in a negative way) to situations where researchers do not form hypotheses based on prior research but simply search for results to discuss as if on a "fishing expedition." More specifically, it may also describes situations in which statistical data are crunched without any particular guiding hypotheses in the hope of finding statistically significant relationships because, if one uses the .05 level of significance, there is a 1 in 20 chance that a significant result will be found.

—Tim Futing Liao

FIXED EFFECTS MODEL

ANALYSIS OF VARIANCE (ANOVA) models can be used to study how a dependent variable Y depends on one or several factors. A factor here is defined as a categorical independent variable; in other words, an explanatory variable with a nominal LEVEL OF MEASUREMENT. A fixed effects ANOVA model (also called a Type I ANOVA model) has fixed (i.e., nonrandom) parameters expressing the effect of the categories of each factor. Such a model is appropriate if the statistical inference aims at finding conclusions that hold for precisely the categories of the factors that are present in the data set at hand. This contrasts with RANDOM-EFFECTS MODELS, which are described in another article and can be appropriate if the observations of the categories of the factor are a sample from a population. An example of a fixed effects ANOVA model could be a study about work satisfaction of employees depending on gender and a classification in a small number of job categories.

Fixed effects ANOVA models are special cases of the GENERAL LINEAR MODEL, which is basic to REGRESSION analysis. The effects of the categories of a factor in an ANOVA can be expressed as regression coefficients of dummy variables in a regression analysis. The general linear model can be formulated as

$$Y_i = b_0 + b_1 X_{1i} + \cdots + b_p X_{pi} + E_i,$$

where i indicates the case, Y is the dependent variable, X_1 to X_p are the independent (or explanatory) variables, and E is the unexplained part, usually called the residual or error term. The quantity b_h is the regression coefficient of variable X_h. The regression coefficients are fixed (i.e., nonrandom) quantities and also are called fixed effects. They are characteristics of the population for which the regression model is defined. Fixed effects occur similarly in other, more complicated models, such as GENERALIZED LINEAR MODELS or nonlinear regression models.

There is another reason for modeling effects of a given factor as fixed rather than random, irrespective of whether the categories of the factor can be regarded as a sample from a population. This is for factors or other variables that are not of primary interest for the researcher but that have to be included in the model to achieve a good model specification; such variables are called *nuisance variables* or *control variables*. In many cases (depending on the correct model specification), such nuisance variables are better controlled for by including them with fixed effects than with random effects. "Better control" means that the parameters of primary interest can be estimated with less bias, or without bias altogether. One class of examples is panel data, used in econometrics, where, in some studies, the time variable (often the year date) has the role of a nuisance variable, which is controlled for by including it as a fixed effect (see Greene, 2003). Another example is the study about work satisfaction of employees depending on gender and job categories mentioned earlier, now conducted in a (small or large) number of different companies. The companies, if regarded as a nuisance variable, could be treated as a fixed effect. On the other hand, if there is interest in the amount and kind of variability between the companies, they could be included as a random effect, which would lead to a MIXED-EFFECTS MODEL.

—Tom A. B. Snijders

REFERENCES

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Greene, W. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–001, 2nd ed.). Newbury Park, CA: Sage.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. W. (1996). *Applied linear regression models*. New York: Irwin.

FLOOR EFFECT

A floor effect occurs when a measure possesses a distinct lower limit for potential responses and a large concentration of participants score at or near this limit (the opposite of a CEILING EFFECT). Scale attenuation is a methodological problem that occurs whenever variance is restricted in this manner. The problem is commonly discussed in the context of experimental research, but it could threaten the validity of any type of research in which an outcome measure demonstrates almost no variation at the lower end of its potential range.

Floor effects can be caused by a variety of factors. If the outcome measure involves a performance task with a lower limit (such as number of correct responses), participants may find the task too difficult and make no correct responses. Rating scales also possess lower limits, and participants' responses may fall outside the lower range of the response scale.

Due to the lack of variance, floor effects pose a serious threat to the validity of both experimental and nonexperimental studies, and any study containing a floor effect should be interpreted with caution. If a dependent variable suffers from a floor effect, the experimental or quasi-experimental manipulation will appear to have no effect. For example, a common assessment of memory is percent of items recalled. If the memory task is too difficult, all participants may demonstrate no recall of items. A study involving the effect of sensory modality on recall may produce null findings, when, in fact, the floor effect will not allow differences to be found.

Floor effects are especially troublesome when interpreting INTERACTION EFFECTS. Many interaction effects involve failure to find a significant difference at one level of an independent variable and a significant difference at another. Failure to find a significant difference may be due to a floor effect, and such interactions should be interpreted with caution.

Floor effects are also a threat for nonexperimental or correlational research. Most inferential statistics are based on the assumption of a normal distribution of the variables involved. It may not be possible to compute a significance test, or the restricted range of variance may increase the likelihood of rejecting the null hypothesis when it is actually true (TYPE II ERROR).

The best solution to the problem of floor effects is pilot testing, which allows the problem to be identified early. If a floor effect is found, the problem may be addressed several ways. If the outcome measure is task performance, the task can be made easier in order to increase the range of potential responses. For rating scales, the number of potential responses can be expanded. An additional solution involves the use of extreme anchor stimuli in rating scales. Prior to assessing the variable of interest, participants can be asked to rate a stimulus that will require them to use the extreme ends of the distribution. This will encourage participants to use the middle range of the rating scale, decreasing the likelihood of a floor effect.

—Robert M. Hessling, Tara J. Schmidt, and Nicole M. Traxel

FOCUS GROUP

A focus group is a research interviewing process specifically designed to uncover insights from a small group of subjects. The group interview is distinctive in that it uses a set of questions deliberately sequenced or focused to move the discussion toward concepts of interest to the researcher. The focus group consists of a limited number of homogeneous participants, discussing a predetermined topic, within a permissive and nonthreatening environment. A skilled moderator or interviewer guides the discussion, exercising limited control over the discussion and moving it from one question to another.

Although it is called a focus group interview, it is really more like a focused discussion. The moderator engages the participants in a conversation, and participants are encouraged to direct their comments to others in the group. The primary responsibilities of the moderator are to keep the discussion on topic and on time, and to engage the participants in the conversation.

Focus group interviews are used in a variety of ways and have gained popularity because the procedure has helped researchers understand behaviors, customs, and insights of the target audience. This information has been used to understand how customers respond to new programs or products as well as factors influencing consumer behavior. More recently, focus groups have moved beyond consumer research to include academic research, program evaluation, social marketing, needs assessments, and related studies. Focus groups are increasingly used in research studies to complement and augment other methods, such as the development of a SURVEY instrument or the discovery of alternative interpretations.

DEVELOPMENT OF FOCUS GROUP INTERVIEWING

The term *focus group interview* is derived from Robert Merton's research just before World War II. Merton and Paul Lazarfeld, sociologists at Columbia University, introduced group interviews that had a focused questioning sequence that elicited comments about behavior, attitudes, and opinions. Merton, along with Patricia Kendall and Marjorie Fisk, wrote about their studies on propaganda materials and troop morale in *The Focused Interview* (1956/1990).

Focus groups gained popularity during the late 1950s as market researchers sought to understand consumer purchasing behavior. From the 1950s through the 1980s, focus groups were used mostly to uncover consumer psychological motivations. Many prominent focus group practitioners were trained as psychologists and sought to uncover the unconscious connections to consumer behavior. At this time, focus groups were composed of 10 to 12 participants, often consisting of strangers, and conducted behind one-way mirrors to allow the sponsor opportunity to observe.

Since the 1980s, academic researchers have used focus groups as a type of qualitative research. Academics used the foundation work of Merton and his colleagues and adapted procedures used by market researchers. These more recent groups place greater emphasis on conducting the group in a natural environment (homes, public meeting spaces); with fewer participants (often 6 to 8 people); and with additional rigor in the analysis (systematic and verifiable procedures).

Focus group research is an evolving field of study. Over the years, researchers have adapted and changed the techniques to fit particular situations. Telephone or Internet focus groups are gaining in popularity as ways to involve people who are separated by distance. Focus groups have shown considerable versatility in international environments where literacy can impede other forms of inquiry.

QUALITY OF FOCUS GROUP STUDIES

As focus groups grew in popularity, the quality began to suffer. To a casual observer, focus groups can look deceptively easy, but in fact, they are complicated to plan, conduct, and analyze. When researchers conducted focus groups without adequate preparation, planning, or skills, the results were less dependable. Much of the quality of the focus group depends on four key factors:

- Planning, which includes identifying the right audience and getting those individuals to gather together
- Asking the right questions
- Moderating the discussion without leading the conversation into predetermined answers
- Using systematic and verifiable analysis strategies in preparing the report

Planning a focus group study consists of determining the number of groups, developing a strategy for recruitment, creating incentives, and anticipating other logistic concerns. The basic design strategy is to conduct 3 or 4 focus groups for each audience category that is of interest to the researcher. A statistical formula is not appropriate for determining sample size. Instead, researchers use the concepts called *redundancy* or THEORETICAL SATURATION. With redundancy or theoretical saturation, the researcher continues interviewing until no new insights are presented. In effect, the researcher has heard the range of views on the topic. Continuing the interviews would only yield more of what the researcher already knows. In focus group research, this tends to occur regularly after 3 or 4 groups with one audience.

Recruiting begins by identifying as precisely as possible the characteristics of the target audience. Participants are selected and invited because they have certain experiences or qualities in common. People are invited who meet these "screens" or qualifications of the study.

There are three distinct qualities of successful recruiting. First, the process should be personalized. This means that each invited person feels that he or she has been personally asked to attend and share his or her opinions. Second, the invitation process is repetitive. Invitations are given not just once, but two or three times. Third, an incentive is used to encourage participation. Incentives could include money, food, gifts, or intangibles, such as the opportunity to help a worthy research effort, assist a neighborhood organization, or further a cause to which the individual is committed. The incentive may vary from one study to another because of the varying interests of the participants.

Questions must be developed for the focus group. These questions are designed to be conversational and easy for the participants to understand. In a 2-hour focus group, there would be approximately a dozen questions. Care is needed in the phrasing and sequencing of the questions. Focus group questions are distinctive in their focus, their emphasis on concrete experiences, and their tendency to elicit conversation.

Here are factors often considered when developing questions:

- Use OPEN-ENDED QUESTIONS to allow the participant as much latitude as possible in responding to the question.
- Avoid questions that can be answered with a "yes" or "no" because these answers do not provide the detailed explanation often needed for the study.
- "Why?" is rarely asked because "why" questions seem demanding and make people defensive. Instead, ask about attributes and/or influences.
- Use "think back" questions to take people back to a specific time to get information based on actual experience.
- Use questions that get participants involved, such as comparing, rating, sorting, drawing, and so on.
- Focus the questions, sequencing them from general to specific so that the questions move seamlessly toward the most important topics.
- Use ending questions to bring closure to the group and get summary comments from participants.

Moderating is the process of guiding the discussion. It begins with helping people feel comfortable enough to share what they think and how they feel in the group. The moderator establishes a trusting, permissive, and comfortable environment that removes barriers to communication. The moderator introduces the study and helps the participants get acquainted with each other. At the beginning of the focus group, the moderator describes the PROTOCOL of the discussion and indicates that the conversation will be recorded. The participants are assured of confidentiality and told that although quotes might be used, no names will be attached in later reports. After setting up a few ground rules to guide the discussion, the moderator begins to ask the questions. The moderator faces several concerns as the conversation continues. The moderator must know when to wait for more information and when to move on to the next question. The moderator must be able to control dominant speakers and encourage hesitant participants. The moderator must also respect the participants, listen to what they have to say, and thank them for their views, even when he or she may personally disagree. The moderator guides the participants through the questions, carefully monitoring the time and moving from question to question when sufficient responses have been provided.

Analysis is the process of making sense out of the series of conversations. Not all studies require the same level of analysis. Occasionally, analysis can be done quickly because the data are clear and patterns are evident. But other times, the analytic process can be extremely time-consuming. The analysis process consists of reviewing the data as captured by multiple means (memory, field notes, tape recording, and/or video recording) and looking for the major themes that cut across groups. Often, the focus group conversation is transcribed, and this transcript is used in later analysis.

Focus group analysis should be systematic and verifiable. Being systematic means that the analyst has a protocol that follows a predetermined sequence for capturing and handling the data. The process is verifiable in that there is a trail of evidence that could be replicated.

WHEN TO USE FOCUS GROUPS

As with all research methods, there are situations where focus groups can be particularly effective and situations where other methods should be used.

Focus groups work particularly well for the following:

- Discovering what prompts certain behavior, such as purchasing, using, or taking action
- Discovering the barriers that impede certain behaviors
- Pilot testing ideas, campaigns, surveys, products, and so on. Focus groups can be used to get reactions to plans before big amounts of money are spent in implementation.
- Understanding how people think or feel about something, like an idea, a behavior, a product, or a service
- Developing other research instruments, such as surveys or CASE STUDIES
- Evaluating how programs or products are working and how they might be improved
- Understanding how people see needs and assets in their lives and communities

In contrast, focus group interviews are not meant to be used as

- A process for getting people to come to consensus.
- A way to teach knowledge or skills.
- A test of knowledge or skills.

Focus groups do not work well when

- The purpose is to make statistical inferences to a population. The numbers involved in focus group research are too small for statistical projections.
- The topic is so highly controversial that participants want to argue instead of discuss.
- The environment does not foster trust and respect. If the sponsoring agency or organization is not considered trustworthy, or if events in the community make people hesitant or cynical, then the study results will be jeopardized.
- The topic is not appropriate for group discussion and should be discussed individually. Note: Although the researcher has promised confidentiality, in reality, he or she cannot guarantee that the participants will maintain confidentiality.

APPROACHES TO FOCUS GROUP INTERVIEWING

The focus group interview is an evolving methodology that has been adapted to different situations and environments. What follows are some of the most common forms.

Market Research Focus Group Approach

In these groups, professional moderators conduct groups with 10 to 12 homogeneous participants who are paid to participate on a topic relating to a consumer product, service delivery, or organizational identity. These groups are often conducted in special focus group rooms with one-way mirrors, around an oval table for a 2-hour period. Results and reports are rarely shared with participants. The participants typically do not know the identity of the sponsoring organization. Often, the results are needed within weeks.

Academic Focus Group Approach

In these groups, the moderator is an academic researcher who conducts groups of 6 to 8 homogeneous participants on a topic of scholarly interest that might provide insight into a theoretical concept or an understanding of the reality of the participants. These groups are usually held in convenient and comfortable community settings instead of special rooms with one-way mirrors. Participants might receive a financial incentive for participating. The results are later published in scholarly journals or in dissertations. Considerable emphasis is placed on careful data collection and systematic analytical processes. These studies are typically conducted over a period of 6 to 12 months.

Public/Nonprofit Focus Group Approach

In these groups, the moderator could be an internal staff person, a volunteer, or an outside consultant. The groups are smaller (often 6 to 8 participants) than the market research approach. The groups are conducted in community settings, and one-way mirrors usually are avoided because of the rental cost and also the sensitivity to being observed. The topics are designed to improve services or products, evaluate programs and policies, understand customer satisfaction, or evaluate other needs of public, religious, educational, or service organizations. Almost always, the results are shared back with community members. These studies are often conducted over a period of several months.

Participatory Focus Group Approach

These focus groups are led by volunteers, community members, or possibly staff who are willing to work together as a research team. Often, a professional researcher serves as a mentor and coach, helping to prepare the team, develop the questions, and take on the complicated task of analysis. The volunteer researchers go into their communities, recruit participants, and conduct focus groups using predetermined questions. These studies have been effective in gaining insight into prevention efforts, health and safety concerns, and community needs assessments. The primary advantage of these groups is that the volunteers have a rapport within the community and can quickly establish trust and candor in the conversations. These results are shared back with community members and are often gathered over a period of several months.

Telephone Focus Group Approach

The telephone focus group is similar to a conference phone call. It is led by a moderator and includes 4 to 6 participants who are linked together by telephone. The primary advantage is the convenience. A researcher can conduct a group with people scattered across the nation at a fraction of the cost of having them together in person. Usually, fewer questions (5–8) are sent out in advance of the group. These groups often last about 1 hour.

Children Focus Groups

These groups are modified to fit the interests and abilities of the children. The participants are close in age (usually less than 2 years apart) because of developmental differences. Participants could be as young as 10 years of age, but they must be able to listen to others in the group and carry on a conversation. The group size is small (5–7), and the groups may be a bit shorter in length. The questions usually involve more activities and exercises than what is present in an adult group.

Internet Focus Groups and Chat Rooms

These groups are evolving and have considerable variation. Some groups are similar to chat lines with a small number of participants who are logged on to a central chat room. The moderator poses questions online, and the participants offer their answers in writing. Another variation occurs when the participants log on to a special Web site and receive a different question each day for a series of days. The questions often build on each other, and participants can see the comments of other participants. As technology advances, there will undoubtedly be additional options for these electronic focus groups.

—Richard A. Krueger

REFERENCES

Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). Thousand Oaks, CA: Sage. [This one-volume (215 page) overview of focus group interviewing provides a description of what focus groups are, how they are used, and strategies that make them successful.]

Merton, R. K., Fiske, M., & Kendall, P. L. (1990). *The focused interview* (2nd ed.). New York: Free Press. (Originally published in 1956, Merton and his colleagues describe the development of the focused interview.)

Morgan, D. L., & Krueger, R. A. (Eds.). (1998). *Focus group kit.* Thousand Oaks, CA: Sage. (A collection of six books, each approximately 100–140 pages, on focus group interviewing. Specific volumes include: *The focus group guidebook, Planning focus groups, Developing questions for focus groups, Moderating focus groups, Involving community members in focus groups,* and *Analyzing and reporting focus group results.*)

FOCUSED COMPARISONS. *See*

STRUCTURED, FOCUSED COMPARISON

FOCUSED INTERVIEWS

Most commonly known as the FOCUS GROUP method, this form of interview was originally formulated for use with both individuals and groups (Merton, 1987; Merton & Kendall, 1946). The explicit objective of the focused interview is to test, appraise, or produce hypotheses about a particular concrete situation in which the respondent(s) have been involved (e.g., a shared event or salient experience). The *focus* of the interview is circumscribed by relevant theory and evidence and involves skilled facilitation of the process (in a one-on-one or group forum) using an INTERVIEW GUIDE, allowing for unanticipated views to also be uncovered and explored. Optimal use of the method involves an appreciation of the paradox involved in balancing the quest for authentic subjective information through free-flowing discussion with the need for methodological rigor.

The purpose of the focused interview is to go beyond the summary judgments made by respondents about their experiences (e.g., unpleasant, stimulating) to "discover precisely what further feelings were called into play" (Merton & Kendall, 1946, p. 541). Hence, a key to the technique is the skilled use of verbal prompts to activate a concrete response to questions rather than vague generalizations, coupled with an ability to continually monitor and adjust the interview process to maximize the likelihood of self-disclosure. Applied to groups, the process facilitation challenge includes harnessing the synergy potential of multiple respondents to ensure that both a breadth and depth of evidence is derived (Millward, 2000). Important design considerations in this instance include group size (e.g., large enough to elicit a broad range of views while not producing inhibitions), group composition (e.g., educational homogeneity), and spatial arrangements (e.g., circular seating) (Merton, 1987).

The focused one-on-one interview is now an important part of the methodological repertoire, especially in clinical and health research contexts. It was not until the late 1980s, however, that the research potential of the focused group interview was formally acknowledged by social scientists, beyond its use in marketing quarters (Morgan, 1988). Both types of focused interview can be used either as a primary means of qualitative data collection (e.g., for hypothesis formulation) or as an adjunct to other methods (e.g., to aid interpretation of findings). Theory can also be used as a focusing vehicle affording constructs around which to anchor either one-on-one or group dialogue (e.g., protection motivation theory), particularly in association with complex or sensitive issues (e.g., examination of risky adolescent sexual behavior).

Depending on respondent consent, interview responses are audiotaped or videotaped and then transcribed in preparation for some form of qualitative data analysis. At its most basic level, CONTENT ANALYSIS involves inducing categories of content. Other, more specialized forms of analysis applicable to focused interview data include, for example, interpretative phenomenological analysis and DISCOURSE ANALYSIS. The latter examines both the process and the content of discussion. Ultimately, choice of analytic strategy will depend on a combination of both epistemological and theoretical factors, including precisely what question is being asked (Millward, 2000).

—Lynne Millward

REFERENCES

Greenbaum, T. L. (1998). *The handbook for focus group research.* London: Sage.

Merton, R. K. (1987). The focused interview and focus groups: Continuities and discontinuities. *Public Opinion Quarterly, 51*, 550–566.

Merton, R. K., Fiske, M., & Kendall, P. L. (1990). *The focused interview: A manual of problems and procedures* (2nd ed.). New York: Free Press.

Merton, R. K., & Kendall, P. L. (1946). The focused interview. *American Journal of Sociology, 51*(6), 514–557.

Millward, L. J. (2000). Focus groups. In G. M. Breakwell, C. Fife-Schaw, & S. Hammond (Eds.), *Research methods in psychology* (2nd ed., pp. 303–324). London: Sage.

Morgan, D. L. (1988). *Focus groups as qualitative research.* Newbury Park, CA: Sage.

Morgan, D. L. (1993). *Successful focus groups: Advancing the state of the art.* London: Sage.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice* (Applied Social Research Methods Series, Vol. 20). Newbury Park, CA: Sage.

FORCED-CHOICE TESTING

Forced-choice testing, in which test takers must choose among alternatives, is distinguished from free-response testing, in which test takers provide a response of their own construction. In the social sciences, choosing between these formats is generally dictated by the goals of testing; forced-choice testing lends itself to rapid assessment over a broad content area, resulting in data readily quantifiable for interpretation. Free-response testing is directed toward qualitative analysis of response and generally requires much more time to interpret responses.

Forced-choice testing comprises a variety of formats. Multiple-choice tests present the test taker with two or more choices for response. Two-alternative, forced-choice tests include "true-false," "yes-no," or "correct-incorrect" answer pairs. Multiple-choice ability tests include one correct response; the remainder are incorrect alternatives. Attitudinal measures include LIKERT-type formats in which the responder chooses a measure of agreement with a statement, typically from a numerically weighted five-point spread ranging from "strongly agree" to "strongly disagree"; the midpoint choice is typically "neutral" or "undecided." A Q-sort technique is useful for attitudinal and self-concept measurement. Individuals sort attitudinal statements or personality traits into groups ranging from least characteristic to most characteristic of themselves. Typically, the distribution of groups is forced, preventing a skewing of responses.

Personality tests have adopted strategies of assessing the consistency of responding (whether similarly worded items are answered in similar fashion), the presence of random responding ("content nonresponsive"), and under- or overreporting of psychopathology (Greene, 2000). In certain examination settings (e.g., personnel selection and child custody evaluations), test takers may be motivated to underreport psychopathology. In pretrial criminal examinations, personal injury litigation, and disability determinations, test takers may be motivated to overreport psychopathology. Simple counts compared to normative data allow quick determination of response sets.

In personality testing, random responding can be assessed by consistency-of-response scales. In ability testing, completely random responding usually is not at issue except when the test taker perceives an advantage in being seen as impaired. Forced-choice testing allows an analysis of the contribution of guessing toward the total number of correct responses. A good estimate of the true number of items known can be computed by [(number correct) − (number incorrect)]/$(n-1)$, where n is the number of alternatives to choose between or among (Cronbach, 1990).

Forced-choice testing enables systematic analysis of feigned impairment on cognitive and sensory tests. Pankratz (1979) popularized the technique of comparing the number of correct responses to that expected by random responding; scoring "below chance" indicates malingering. Frederick and Crosby (2000) showed that Cronbach's technique could result in an estimate of true cognitive capacity for overt malingering. Furthermore, Frederick and Crosby demonstrated that plotting average response accuracy against average response difficulty for two-alternative, forced-choice tests with a hierarchy of difficulty allowed classification of responding as compliant, inconsistent, random, or "suppressed" (malingered).

—Richard I. Frederick

REFERENCES

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: HarperCollins.

Frederick, R. I., & Crosby, R. D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior, 24,* 59–82.

Greene, R. L. (2000). *The MMPI-2: An interpretive manual.* Boston: Allyn & Bacon.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology, 47,* 409–410.

FORECASTING

It is far better to foresee even without certainty than not to foresee at all.

Henri Poincaré, *The Foundations of Science*, 1913, p. 129

A forecast is a statement about an unknown and uncertain event—most often, but not always, a future event. If time is involved, a forecast (or prediction) is an assertion about a future outcome that is based on observed regularities among events in the past. We assume that observations on a single time series become available at consecutive time periods. The observations may be monthly sales data, quarterly GNP figures, yearly homicides, or daily stock price closings. Past observations $y_1, y_2, \ldots, y_n$ are used to calculate a forecast $\hat{y}_n(r)$ of the future observation y_{n+r}. Here, n is the forecast origin, r is the forecast lead time, and $\hat{y}_n(r)$ is the r-step-ahead forecast from time origin n. The question is, how should the available observations be used when calculating the forecast? Here, we assume that there are no other series that can help with the forecast of future ys.

Exponential SMOOTHING and forecasting with ARIMA (autoregressive integrated moving average) time series models represent two successful approaches to univariate time series forecasting. We describe these methods in detail and point to useful extensions.

EXPONENTIAL SMOOTHING PROCEDURES

One can distinguish three basic procedures: simple exponential smoothing for series without trend, Holt's exponential smoothing for data exhibiting trend, and Winters exponential smoothing for data that include trend and seasonal components.

Simple Exponential Smoothing

An intuitive approach to the prediction of a series without trend is to use the historic average as prediction for all future periods. That is,

$$\hat{y}_n(r) = \bar{y} = \frac{1}{n}[y_n + y_{n-1} + \cdots + y_1].$$

Equal weighting of the observations, irrespective of their age, requires stability of the series over time.

Because most data series lack this stability, it is better to either average observations over a short, recent time window, or consider a weighted average that discounts observations according to their age. Simple exponential smoothing uses an exponential (geometric) weighted average:

$$\hat{y}_n(r) = L_n = \alpha[y_n + (1-\alpha)y_{n-1} + (1-\alpha)^2 y_{n-2} + (1-\alpha)^3 y_{n-3} + \ldots].$$

L_n is the estimate of the level at time n. It becomes the forecast of all future observations. The smoothing constant $\alpha(0 \le \alpha \le 1)$ gives more weight to recent observations and less weight to observations in the past. When $\alpha = 1$, the forecast $\hat{y}_n(r) = L_n = y_n$ makes use of the last observation only. It is called the "naïve" or "RANDOM WALK" forecast. When α approaches 0, the resulting forecast $\hat{y}_n(r) = L_n = (y_n + y_{n-1} + \cdots + y_1)/n$ puts equal weight on all observations.

The level is updated according to $L_n = \alpha y_n + (1-\alpha)L_{n-1}$; only the last observation and the previous level estimate need to be stored. This is important if many series must be predicted. It is common to start the recursion at the beginning of the series, with a certain starting value for the level at Time 1. Typically, one uses the first observation ($L_1 = y_1$), or an average of the first few observations. From this starting value, one calculates $L_2 = \alpha y_2 + (1-\alpha)L_1$, $L_3 = \alpha y_3 + (1-\alpha)L_2, \ldots$, until one reaches $L_n = \alpha y_n + (1-\alpha)L_{n-1}$. The last level estimate becomes the forecast of all future observations: $\hat{y}_n(r) = L_n$.

The calculations require a value for the smoothing constant. The smoothing constant is estimated by minimizing the sum of the squared one-step-ahead forecast errors,

$$\mathrm{SSE} = \sum_{t=2}^{n} [y_t - \hat{y}_{t-1}(1)]^2 = \sum_{t=2}^{n} (y_t - L_{t-1})^2.$$

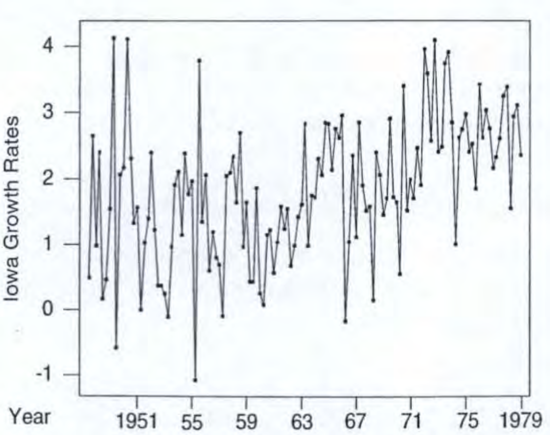

Figure 1 Quarterly Iowa Growth Rates (1948 through 1979)

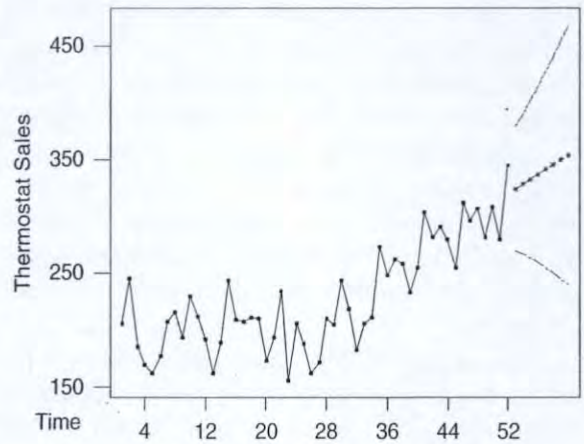

Figure 2 Observations, Forecasts, and 95% Forecast Intervals

We illustrate simple exponential smoothing with the quarterly Iowa nonfarm growth rates in Figure 1. The data ($n = 127$; from 1948/2 through 1979/4) were first analyzed in Abraham and Ledolter (1983). The time series graph shows no obvious trends, but indicates that the level of the series shifts with time. Table 1 illustrates the calculations with smoothing constant $\alpha = 0.1$. The iterations start with $L_1 = y_1 = 0.50$. Repeated use of the updating equation gives $L_2 = (0.1)(2.65) + (0.9)(0.5) = 0.715$, $L_3 = (0.1)(0.97) + (0.9)(0.715) = 0.741$, until $L_{127} = (0.1)(2.35) + (0.9)(2.681) = 2.648$. Forecasts of all future observations are given by $\hat{y}_{127}(r) = 2.648$. The one-step-ahead forecast errors are given in the fifth column: $y_2 - \hat{y}_1(1) = 2.65 - 0.50 = 2.15$, $y_3 - \hat{y}_2(1) = 0.97 - 0.715 = 0.255, \ldots, y_{127} - \hat{y}_{126}(1) = 2.35 - 2.681 = -0.331$; the sum of their squares is SSE $= 122.61$. The sum of squares depends on the smoothing constant, because a different α implies different forecasts. In this example, any smoothing constant different from 0.1 increases the sum of the squared one-step-ahead forecast errors.

Holt's Exponential Smoothing

Holt extends the exponential smoothing approach to the situation where the series follows a linear

Table 1 Simple Exponential Smoothing for Iowa Growth Rates

Time	*Growth Rate*	*Smoothed Value*	*One-Step-Ahead Forecast*	*One-Step-Ahead Forecast Error*
1	$y_1 = 0.50$	$L_1 = 0.50$		
2	$y_2 = 2.65$	$L_2 = 0.715$	$\hat{y}_1(1) = 0.50$	$y_2 - \hat{y}_1(1) = 2.15$
3	$y_3 = 0.97$	$L_3 = 0.741$	$\hat{y}_2(1) = 0.715$	$y_3 - \hat{y}_2(1) = 0.255$
4	$y_4 = 2.40$	$L_4 = 0.906$	$\hat{y}_3(1) = 0.741$	$y_4 - \hat{y}_3(1) = 1.659$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
124	$y_{124} = 1.55$	$L_{124} = 2.602$	$\hat{y}_{123}(1) = 2.719$	$y_{124} - \hat{y}_{123}(1) = -1.169$
125	$y_{125} = 2.93$	$L_{125} = 2.635$	$\hat{y}_{124}(1) = 2.602$	$y_{125} - \hat{y}_{124}(1) = 0.328$
126	$y_{126} = 3.10$	$L_{126} = 2.681$	$\hat{y}_{125}(1) = 2.635$	$y_{126} - \hat{y}_{125}(1) = 0.465$
127	$y_{127} = 2.35$	$L_{127} = 2.648$	$\hat{y}_{126}(1) = 2.681$	$y_{127} - \hat{y}_{126}(1) = -0.331$

trend. The r-step-ahead forecast of y_{n+r} is calculated from

$$\hat{y}_n(r) = L_n + rT_n,$$

where L_n and T_n reflect the level and the trend of the series at time n. Estimates of these quantities are needed, and update recursions need to be specified. Holt considers the recursions

$$L_n = \alpha_1 y_n + (1 - \alpha_1)(L_{n-1} + T_{n-1}),$$
$$T_n = \alpha_2(L_n - L_{n-1}) + (1 - \alpha_2)T_{n-1}.$$

The two smoothing constants, α_1 and α_2, reflect the weights that are given to current ("new") and previously available ("old") information. The most recent information on the level at time n is given by y_n. Prior to this observation, we have available L_{n-1} and T_{n-1} and our best guess of the next level is $L_{n-1} + T_{n-1}$. The updating equation for the level specifies a weighted average of these two components. The second equation for the trend specifies a weighted average of the most recent trend estimate, $L_n - L_{n-1}$, and the previously available estimate T_{n-1}.

Large smoothing constants put more weight on recent observations. Smoothing constants $\alpha_1 = \alpha_2 = 1$ imply $L_n = y_n$, $T_n = y_n - y_{n-1}$, and $\hat{y}_n(r) = y_n + r(y_n - y_{n-1})$. All forecasts lie on a straight line through the last two observations. For small smoothing constants, the observations are weighted equally, similar to a regression of the observations on time.

Updating calculations starts with reasonable initial values such as $L_2 = y_2$ and $T_2 = y_2 - y_1$. The recursions are used to compute L_3 and T_3, L_4 and $T_4, \ldots,$ until L_n and T_n are reached. Forecasts follow from $\hat{y}_n(r) = L_n + rT_n$. Smoothing constants are estimated by minimizing the sum of the squared one-step-ahead forecast errors.

Holt's method is applied to the thermostat sales data analyzed in Abraham and Ledolter (1983). Figure 2 shows the $n = 52$ observations, the next eight forecasts, and approximate 95% forecast intervals. Minitab is used for the estimation of the smoothing constants and the calculation of the forecasts.

Winters Seasonal Exponential Smoothing

Many series include seasonal components. Sales are typically seasonal, with large sales around Christmas. For seasonal data, one extends Holt's approach with an additional equation and a third smoothing constant. Assuming monthly data with a yearly seasonal component, the forecasts are given by

$$\begin{aligned}\hat{y}_n(r) &= (L_n + rT_n)S_{n+r-12} \quad \text{for } r = 1, 2, \ldots, 12 \\ &= (L_n + rT_n)S_{n+r-24} \quad \text{for } r = 13, \ldots, 24 \\ &= \cdots.\end{aligned}$$

The difference to the forecasts in Holt's approach is the adjustment of trend forecasts by appropriate seasonal factors. Here, the adjustment is multiplicative, but additive adjustments are also possible.

Update equations for the estimates of level (L_n), trend (T_n), and seasonal components (S_n) are needed.

Again, these equations specify weighted averages of "new" and "old":

$$L_n = \alpha_1(y_n/S_{n-12}) + (1-\alpha_1)(L_{n-1}+T_{n-1})$$
$$T_n = \alpha_2(L_n - L_{n-1}) + (1-\alpha_2)T_{n-1}$$
$$S_n = \alpha_3(y_n/L_n) + (1-\alpha_3)S_{n-12}.$$

The three smoothing constants, α_1, α_2, and α_3, determine how observations are weighted. The first two equations are similar to Holt's equations, with the only difference being that the most recent observation is adjusted for seasonality (y_n/S_{n-12}). The update of the seasonal factor in the last equation is a weighted average of the last available "old" seasonal factor (S_{n-12}) and the "new" information on the seasonal factor, y_n/L_n.

Smoothing constants are obtained by minimizing the sum of the squared one-step-ahead forecast errors. The calculations are started with initial values for the level, trend, and seasonal components; estimates are updated recursively until the last available time period is reached; and forecasts are calculated. The computations are more complicated, but conceptually not different from the simple versions.

Forecast Intervals

Point forecasts are important, but they mean little if one cannot attach a level of CONFIDENCE. An approximate 95% r-step-ahead forecast interval for the future observation y_{n+r} is given by

$$\hat{y}_n(r) \pm 2s_r$$

where the ROOT MEAN SQUARE error

$$s_r = \sqrt{\sum_{t=r+1}^{n} [y_t - \hat{y}_{t-r}(r)]^2/(n-r)}$$

is calculated from the "observed" r-step-ahead forecast errors, $y_{r+1} - \hat{y}_1(r), y_{r+2} - \hat{y}_2(r), \ldots, y_n - \hat{y}_{n-r}(r)$.

BOX-JENKINS ARIMA MODELS

Exponential smoothing is a procedure that, when applied to a time series, results in forecasts. The BOX-JENKINS ARIMA approach is *model-based.* First, one determines the model for the underlying stochastic process that has generated the series, and then one uses the model to derive the optimal forecasts.

Autoregressive integrated moving average (ARIMA) models view the autocorrelated observations y_t as the result of a linear filter on uncorrelated random variables a_t. The filter is written as a ratio of polynomials, leading to the difference equation

$$\varphi(B)y_t = \phi(B)(1-B)^d y_t = \theta_0 + \theta(B)a_t,$$

where

- B is the backshift operator, $B^m y_t = y_{t-m}$
- a_t are uncorrelated, unobserved errors with constant variance
- $\varphi(B) = \phi(B)(1-B)^d = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p+d}B^{p+d}$ is the generalized autoregressive operator of order $p+d$
- d is the degree of differencing needed to transform a nonstationary series to stationarity
- $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is a stationary autoregressive operator of order p
- $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ is an invertible moving average operator of order q
- θ_0 is a constant that introduces deterministic trends when $d \geq 1$.

Special Cases

The first-order ARIMA(1, 0, 0) model, $(1-\phi B)y_t = \theta_0 + a_t$ or $y_t = \theta_0 + \phi y_{t-1} + a_t$, is a regression of the current observation on its lag. In the ARIMA$(p, 0, 0)$ model, the observation at time t is regressed on the p previous lags. The ARIMA(0, 1, 0) random walk model, $(1-B)y_t = \theta_0 + a_t$ or $y_t = \theta_0 + y_{t-1} + a_t$, is useful for modeling stochastic nonstationarity; the constant θ_0 introduces additional deterministic drift. The ARIMA (0, 1, 1) model, $(1-B)y_t = (1-\theta B)a_t$ or $y_t = y_{t-1} + a_t - \theta a_{t-1}$, is another commonly used model. Model extensions to cover seasonal situations are available.

Box and Jenkins (1976) study the properties of these models and outline a useful model-building strategy. The model order (p, d, q) is obtained from the patterns in the AUTOCORRELATIONS and partial autocorrelations of the series, the parameters in the models are estimated by MAXIMUM LIKELIHOOD, and the adequacy of the models is assessed through residual diagnostics.

Forecasts Implied by ARIMA Models

Optimal (minimum mean square error) forecasts of future observations are obtained by taking *conditional expectations* of future observations. These conditional expectations can be derived from the difference equation of the ARIMA model. We illustrate this on the second-order autoregressive model,

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t.$$

The conditional expectations of future observations $(y_{n+1}, y_{n+2}, \ldots)$ are

$$\hat{y}_n(1) = \theta_0 + \phi_1 y_n + \phi_2 y_{n-1}$$

$$\hat{y}_n(2) = \theta_0 + \phi_1 \hat{y}_n(1) + \phi_2 y_n$$

$$\hat{y}_n(r) = \theta_0 + \phi_1 \hat{y}_n(r-1) + \phi_2 \hat{y}_n(r-2) \quad \text{for } r \geq 3.$$

Better understanding of the forecasts can be obtained through the *eventual forecast function*. The eventual forecast function of the ARIMA(p, d, q) model is given by

$$\begin{aligned}\hat{y}_n(r) &= f_1(r)\beta_1^{(n)} + f_2(r)\beta_2^{(n)} + \cdots + f_{p+d}(r)\beta_{p+d}^{(n)} \\ &= f(r)^{\text{tr}}\beta^{(n)} \quad \text{for } r > [q-(p+d)],\end{aligned}$$

where $f(r) = [f_1(r), f_2(r), \ldots, f_{p+d}(r)]^{\text{tr}}$ is a column vector of forecast functions that depend only on the generalized autoregressive operator (tr denotes the vector transpose). For example, in the ARIMA$(0, 2, q)$ model, the forecast functions are given by $f_1(r) = 1$ and $f_2(r) = r$, and the eventual forecast function is linear.

The vector of coefficients $\beta^{(n)} = [\beta_1^{(n)}, \beta_2^{(n)}, \ldots, \beta_{p+d}^{(n)}]^{\text{tr}}$ is updated with each new observation according to

$$\beta^{(n)} = L\beta^{(n-1)} + h[y_n - \hat{y}_{n-1}(1)].$$

The transition matrix L is a consequence of the forecast functions $f(r)$, and the vector h depends on the parameters in the ARIMA model.

Relationships Among ARIMA Forecasts and Exponential Smoothing

There are similarities of the eventual forecast functions implied by ARIMA models and the forecast functions in exponential smoothing. Autoregressive components and the degree of differencing imply the forecast functions of the ARIMA forecasts. Updates of the coefficients in the eventual forecast function and in exponential smoothing can be made the same if certain special ARIMA models are selected.

ARIMA models are more general. Forecasts from constrained ARIMA models coincide with exponential smoothing forecasts. Forecasts from the ARIMA $(0, 1, 1)$ model, $(1-B)y_t = (1-\theta B)a_t$, coincide with those from simple exponential smoothing provided that $\theta = 1 - \alpha$. Forecasts from the ARIMA$(0, 2, 2)$ model, $(1-B)^2 y_t = (1 - \theta_1 B - \theta_2 B^2)a_t$, are the same as the forecasts in Holt's exponential smoothing provided that $\theta_1 = 2 - \alpha_1(1+\alpha_2)$ and $\theta_2 = -(1-\alpha_1)$. Relationships among the seasonal forecast procedures are more complicated (see Abraham & Ledolter, 1986).

Despite these theoretical results, forecast competitions show that in many practical applications, simple exponential smoothing procedures outperform the more general ARIMA models (see Makridakis & Hibbon, 2000). A lack of robustness of more complicated and in all likelihood incorrect models, and problems with fitting nonparsimonious models with unnecessary parameters can explain such findings.

IMPORTANT ISSUES NOT COVERED IN THIS REVIEW

Many other univariate forecast procedures have been developed but are not covered in this review. The structural time series models described in Harvey (1990) and the stochastic models developed for finance applications (see Tsay, 2002) should be mentioned. Our review deals with quantitative methods but ignores judgmental forecast approaches that are important if little or no data are available (see Armstrong, 2001).

Univariate time series forecasts use only the series's own past history. One would expect that forecasts of Y are improved when incorporating the information from other related series X. Transfer function models can be used if y_t is affected by current or lagged xs, and forecast improvements will result if there is a strong lead-lag relationship. A contemporaneous association of y_t with x_t alone is usually of little benefit, because one needs a forecast for X in order to predict Y. Multivariate time series models, particularly multiple autoregressive models, are appropriate if feedback among the series is present. ECONOMETRIC models use systems of many equations to represent the relationships among economic variables. The purpose of econometric models is to provide conditional, or "what-if," forecasts, as compared to the unconditional forecasts provided by univariate forecast procedures.

Many different forecast methods are available. Each method has its strengths and weaknesses, and relies on ASSUMPTIONS that are often violated. Theoretical and empirical evidence suggests that a combination of forecasts is advantageous (*International Journal of Forecasting*, No. 4, 1989).

COMPUTER SOFTWARE

Excellent software is available for time series forecasting. All major statistical software packages (such as SAS, SPSS, and Minitab) include useful time series forecasting modules. In addition, several special-purpose programs (such as ForecastPro and SCA) include "automatic" procedures that can identify the most appropriate techniques without much user intervention. This is important in large-scale forecast applications where one must predict thousands of series.

SUGGESTED READINGS

The books by Abraham and Ledolter (1983), Newbold and Bos (1994), and Diebold (2001) give a good introduction to this area. *Principles of Forecasting*, edited by Armstrong (2001), includes summaries and references to commonly used forecast methods and covers both quantitative and qualitative (judgmental) forecast procedures. The *Journal of Forecasting* and the *International Journal of Forecasting* are two journals specializing in forecasting issues.

—Johannes Ledolter

REFERENCES

Abraham, B., & Ledolter, J. (1983). *Statistical methods for forecasting*. New York: Wiley.

Abraham, B., & Ledolter, J. (1986). Forecast functions implied by ARIMA models and other related forecast procedures. *International Statistical Review, 54*, 51–66.

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting*. Boston: Kluwer.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (rev. ed.). San Francisco: Holden-Day.

Diebold, F. X. (2001). *Elements of forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Harvey, A. C. (1990). *Forecasting, structural time series models, and the Kalman filter.* New York: Cambridge University Press.

Makridakis, S., & Hibbon, M. (2000). The M3-competition: Results, conclusions, and implications. *International Journal of Forecasting, 16*, 451–476.

Newbold, P., & Bos, T. (1994). *Introductory business forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Tsay, R. S. (2002). *Analysis of financial time series*. New York: Wiley.

FORESHADOWING

The term is used by some qualitative researchers to refer to the process of specifying issues in which one is interested in advance of entering the field. As the British social anthropologist Bronislaw Malinowski wrote: "Preconceived ideas are pernicious in any scientific work, but foreshadowed problems are the main endowment of the scientific thinker, and these problems are first revealed to the observer by his theoretical studies" (Malinowski, 1922, pp. 8–9, cited in Hammersley & Atkinson, 1995, pp. 24–25). Thus, unlike the open-ended approach that is sometimes advocated by qualitative researchers, Malinowski and others who have followed him recommended coming to the field with an awareness of certain issues that require investigation.

—Alan Bryman

REFERENCES

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in practice*. London: Routledge.

Malinowski, B. (1922). *Argonauts of the Western Pacific*. London: Routledge & Kegan Paul.

FOUCALDIAN DISCOURSE ANALYSIS

Michel Foucault developed a complex and nuanced theory of discourse in a series of books that includes *The History of Sexuality* (1998) and *The Archeology of Knowledge* (2002). Although his work covered a broad array of empirical topics and substantial shifts in method, three concepts recur throughout his work and provide some coherence to his varied interests: power, knowledge, and subjectivity. Understanding Foucauldian discourse analysis begins with an understanding of Foucault's CONCEPTUALIZATION of these three concepts and their relationship.

Foucault defines discourses, or discursive formations, as bodies of knowledge that form the objects of which they speak. In other words, discourses do not simply describe the social world; they constitute it by bringing certain phenomena into being through the way in which they categorize and make sense of an otherwise meaningless reality. The nature of the discursive formation in place at any point in time is the source of the relations of power (and resistance), the social objects and identities, and the possibilities for speaking and acting that exist at any point in time.

Each discourse is defined by a set of rules or principles—the "rules of formation"—that lead to the appearance of particular objects that make up recognizable social worlds. Discourse lays down the "conditions of possibility" that determine what can be said, by whom, and when. No statement occurs accidentally, and the task for the discourse analyst is to analyze how and why one particular statement appeared rather than another.

For Foucault, discourse—or at least the knowledge that it instantiates—is inseparable from power. Power is embedded in knowledge, and any knowledge system constitutes a system of power, as succinctly summarized in his "power/knowledge" couplet. In constructing the available identities, ideas, and social objects, the context of power is formed. Power, in turn, brings into being new forms of knowledge and produces new social objects.

Power is not something connected to agents but represents a complex web of relations determined by systems of knowledge constituted in discourse. It is thus the discursive context, rather than the subjectivity of any individual actor, that influences the nature of political strategy. In fact, for Foucault, the notion of agents acting purposefully in some way not determined by the discourse is antithetical.

In determining the conditions of possibility in this way, discourse "disciplines" subjects in that actors are known—and know themselves—only within the confines of a particular discursive context. Thus, discourse shapes individuals' subjectivity. Foucault decenters the subject: There is no essence or true subject but only a power/knowledge system that constituted the subjectivities of the individuals embedded within it.

Foucault's work has had a tremendous influence on social studies and, in particular, on understandings of the relation between discourse and power. At the same time, his empirical method is highly idiosyncratic and has found only limited footing as yet beyond his own work at an empirical level. However, his work is highly influential and stands to make an even more significant contribution to social science in the future.

—Nelson Phillips

REFERENCES

Foucault, M. (1998). *The history of sexuality: The will to knowledge* (Vol. 1). London: Penguin.

Foucault, M. (2002). *Archeology of knowledge.* London: Routledge.

FRAUD IN RESEARCH

Fraud in the natural sciences (especially the biological and medical) is not uncommon (Broad & Wade, 1982). Fraud in social scientific research, however, is said to be comparatively rare. But of course, it all depends on what is meant by "fraud." This is a contested concept, both in general and in particular cases. This inherent uncertainty has two sources: the shifting and fuzzy character of activities that may come under this description, and the "paradox of success."

Typologies of research fraud abound. Charles Babbage's 1830 description of the illegitimacies of "trimming," "cooking," and "forging" (Broad & Wade, 1982, pp. 29–30) is an early and often-cited example of a typology with an "action-type" criterion. The alternative kind uses an "intention-type" criterion to distinguish, for example, the fraud from the hoax, where the former is perpetrated with serious and permanent intent, and the latter is done as a joke and is intended to be temporary. Thus, the perpetrator of fraud is treated with greater opprobrium than the hoaxer.

The intentional fraud is a paradoxical object, however. Because any instance can be known only through its discovery, and because its discovery achieves its failure, *successful* frauds cannot be known to exist—or not to exist. For some commentators, like Robert Joynson, one of Cyril Burt's later defenders (see below), this can be problematic: "'The hallmark of a *successful* confidence-trickster is precisely that people are persuaded of his honesty.' If this strange argument is accepted, everyone who is trusted must be regarded as a successful confidence-trickster" (Joynson, 1989, p. 302). Strictly, what this "paradox of success" specifies is that honest work and action cannot be distinguished from their fraudulent, and undiscovered, counterparts. The implication of this is that it is vital to understand the work of its discovery (debunking) if we wish to understand the phenomenon of fraud (Ashmore, 1995).

What accounts for the relative rarity of full-blooded fraud in social research? One reason is that opportunities for its *discovery* are relatively scarce, and the resources required for such work are consequently large. Although the biologist John Darsee can be, apparently, observed "one evening in May 1981 flagrantly [forging] raw data" (Broad & Wade, 1982, p. 14), no such simple debunking techniques are available to convict a cultural anthropologist like Margaret Mead (see "Special Section," 1983) or even a quantitative psychologist like Cyril Burt. Although Burt was indeed accused, by Leon Kamin and others, of flagrantly forging raw data, this debunking effort was unstable—vulnerable to later "meta-debunking" work that accused the accusers of equivalent crimes (Joynson, 1989; see Mackintosh, 1995, for the current "not proven" consensus). The more qualitative and discursive the style of social research, the less likely it is that fraud, with the exception of plagiarism, will be discoverable.

—Malcolm Ashmore

REFERENCES

Ashmore, M. (1995). Fraud by numbers: Quantitative rhetoric in the Piltdown forgery discovery. *South Atlantic Quarterly, 94*(2), 591–618.

Broad, W., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science.* New York: Simon & Schuster.

Joynson, R. B. (1989). *The Burt affair.* London: Routledge.

Mackintosh, N. J. (1995). *Cyril Burt: Fraud or framed?* Oxford, UK: Oxford University Press.

Special section: Speaking in the name of the real: Freeman and Mead on Samoa. (1983). *American Anthropologist, 85,* 908–947.

FREE ASSOCIATION INTERVIEWING

The Free Association Narrative Interview (FANI) is a qualitative method for the production and analysis of interview data that combines a narrative emphasis with the psychoanalytic principle of free association. Questions that elicit narratives encourage interviewees to remember specific events in their experiences. These, unlike generalized accounts, possess emotional resonance. The principle of free association is based on the idea that it is the unconsciously motivated links between ideas, rather than just their contents, that provide insight into the emotional meanings of interviewees' accounts (see PSYCHOANALYTIC METHODS). Therefore, it is particularly appropriate for exploring questions about interviewees' identities or that touch emotions and sensibilities (in contrast to methods designed to elicit opinions or facts).

In most empirical social science methods, research subjects are assumed to know themselves well enough to give reliable accounts. The FANI method instead posits defenses against anxiety as a core feature of research subjects, as of all subjects (including researchers). Because anxiety characterizes the research relationship and will be more or less prominent depending on the topic, the setting, and the degree of rapport, defenses against anxiety will potentially compromise interviewees' ability to know the meaning of their actions, purposes, and relations. Such unconscious defenses are one feature of a psychosocial subject, part of the psychic dimension produced by a unique biography structured by defenses, desires, and mental conflict. These influence and are influenced by the social dimension, which consists of a shared social world that is understood as a combination of intersubjective processes, real events, and the discourses through which events are rendered meaningful.

The method was developed as an adaptation of the biographical interpretative method (see BIOGRAPHIC NARRATIVE INTERPRETIVE METHOD) in the context of a project investigating the relationship between anxiety and fear of crime (Hollway & Jefferson, 2000). A narrative question format about specific events ("Can you tell me about a time when . . .?") was adopted in order to discourage responses of a generalized and rationalized kind, as frequently elicited by a SEMISTRUCTURED INTERVIEW. OPEN-ENDED questions were asked, and "why" questions were avoided. Interviewees' answers were followed up in their order of telling, using their phrasing. These protocols are designed to keep within participants' own meaning frames rather than the interviewer's.

The FANI method, using gestalt principles, entails holding the whole data set in mind when interpreting the meaning of an extract. Consistent with psychoanalytic principles, rather than expect coherence, the researcher is alert to signs of incoherence and conflict in the text, such as changes in emotional tone, contradictions, and avoidances.

During interviews and data analysis, interviewers use their own feelings as data, following psychoanalytic principles of transference and countertransference, in order to identify their own emotional investments. Each participant is interviewed twice.

Two researchers listen to the first audiotape before the second interview, and the process of noticing the significance of links, associations, and contradictions contributes to the formulation of narrative questions for the second interview.

—Wendy Hollway and Tony Jefferson

REFERENCE

Hollway, W., & Jefferson, T. (2000). *Doing qualitative research differently: Free association, narrative and the interview method*. London: Sage.

FREQUENCY DISTRIBUTION

Frequency distribution shows the number of observations falling into each of the categories or ranges of values of a VARIABLE. Frequency distributions are often portrayed by way of a HISTOGRAM, a (FREQUENCY) POLYGON, or a frequency table. When such distribution is expressed in terms of proportions or percentages, it is known as RELATIVE FREQUENCY DISTRIBUTION. When the distribution represents incremental frequencies, it is known as cumulative frequency distribution.

—Tim Futing Liao

FREQUENCY POLYGON

A frequency polygon is one of several types of graphs used to describe a relatively large set of quantitative data. An example using a set of adult male heights is given in Figure 1.

To create a frequency polygon, data are first sorted from high to low and then grouped into contiguous intervals. The intervals are represented on the x-axis of a graph by their midpoints. The y-axis provides a measure of frequency. It shows the number of data points falling in each interval. It is scaled from a minimum of 0 to a maximum that exceeds the frequency of the modal interval. Within the field of the figure, straight lines connect points that mark the frequencies for each interval. These lines should extend to the x-axis, which

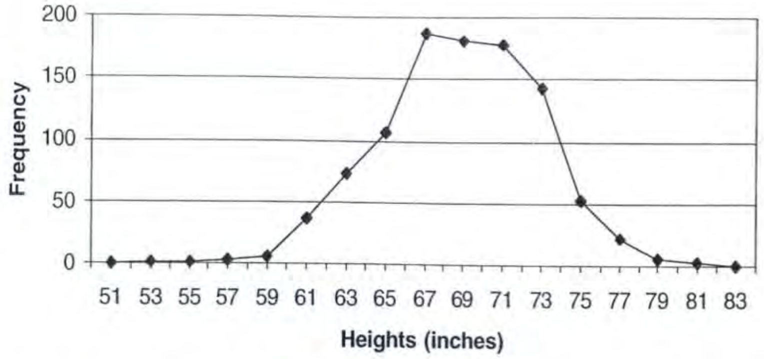

Figure 1 Adult Male Heights Displayed According to the Frequency for Each 2-Inch Interval Between 52 and 82 Inches

means that empty intervals need to be included at the low and high end of the data set.

Frequency polygons are used to represent quantitative data when the data intervals do not equal the number of distinct values in the data set. This is always the case for continuous data. When the data are discrete, it is possible to represent every value in the dataset on the x-axis. In this case, a HISTOGRAM would be used instead of a frequency polygon. The importance of the distinction is that points are connected with lines in the polygon to allow interpolation of frequencies for x-axis values not at the midpoints. When data have no values other than those represented on the x-axis, frequencies are simply marked by bars (or columns).

If the y-axis in a frequency polygon is used to denote proportions rather than frequencies, the graph is known as a relative frequency polygon. For some data displays, it is more useful to show cumulative frequencies for data intervals rather than absolute frequencies. In this case, a cumulative frequency polygon (or ogive) would be used to describe the data set.

—Kenneth O. McGraw

FRIEDMAN TEST

The Friedman (1937) test provides a test for the NULL HYPOTHESIS of equal TREATMENT effects in a two-way layout. This EXPERIMENTAL DESIGN is also known as the balanced complete BLOCK DESIGN, randomized complete block design, and REPEATED-MEASURES DESIGN.

Suppose we have t treatments and b blocks and a total of N experimental units, where N is a multiple of the product bt. For convenience, we assume that $N = bt$ so that there are exactly t experimental units within each block. The t treatments are RANDOMLY ASSIGNED to the t units within the ith block to obtain observations $X_{i1}, X_{i2}, \ldots, X_{it}$. To carry out the test, these observations in the ith block are ranked from smallest to largest using the integer ranks $1, 2, \ldots, t$. This is repeated for each block. Let R_{ij} denote the rank assigned to the experimental unit in block i that received treatment j. The complete set of ranked data can then be presented in the following two-way table:

	Treatment				
Block	*1*	*2*	$\cdots$	t	*Sum*
1	R_{11}	R_{12}	$\cdots$	R_{1t}	$t(t+1)/2$
2	R_{21}	R_{22}	$\cdots$	R_{2t}	$t(t+1)/2$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
b	R_{b1}	R_{b2}	$\cdots$	R_{bt}	$t(t+1)/2$
Sum	R_1	R_2	$\cdots$	R_t	N

We note that the row sums are each equal to $1 + 2 + \cdots + t = t(t+1)/2$. If the treatment effects are, indeed, all the same, as under the null hypothesis, the ranks in each row are simply a random permutation of $1, 2, \ldots, t$, and the column sums will tend to be equal. But the total of the column sums must equal $N = bt(t+1)/2$, and therefore, each column sum must be equal to $bt(t+1)/2$ divided by t, or $b(t+1)/2$. The test statistic S is the sum of squares of the deviations between the actual column sums and the sum expected under the null hypothesis, or

$$S = \sum_{i=1}^{t} [R_i - b(t+1)/2]^2$$
$$= \sum_{i=1}^{t} R_i^2 - b^2 t(t+1)^2/4.$$

If there is a definite ordering among the treatment effects, all of the ranks in some column will be equal to 1, all of the ranks in some other column will be equal to 2, and so on, and the column sums will be some permutation of the numbers $1b, 2b, \ldots, tb$. It can be shown that these column sums produce the largest possible value of S. Therefore, the null hypothesis of equal treatment effects should be rejected in favor of the alternative of a definite ordering among treatments for large values of S.

Table 1 Data on Currency Fluctuations

Currency	*Monday*	*Tuesday*	*Wednesday*	*Thursday*	*Friday*
AD	2	3	5	4	1
BP	2	3	4	5	1
JY	1	4	3	5	2
EE	1	3	4	5	2
Total	6	13	16	19	6

The exact null distribution of S has been tabled for small values of b and t and is reproduced in Gibbons (1993, 1997). For larger sample sizes, the test statistic is $Q = 12S/bt(t+1)$, which is approximately chi-square distributed with $t - 1$ DEGREES OF FREEDOM.

As a numerical example, suppose we want to compare median daily fluctuations in foreign currency exchange relative to the U.S. dollar for the Australian dollar (AD), British pound (BP), Japanese yen (JY) and European euro (EE). The fluctuations are measured by a skewness coefficient calculated for each day of the week. The relative values are shown in Table 1, where rank 1 = smallest value. We have $t = 5$ treatments (the days of the week) and $b = 4$ blocks (the currencies). We calculate $S = 138$ and $Q = 13.8$ with 4 degrees of freedom. The approximate p value is $p < .01$, and we reject the null hypothesis of equal median fluctuations. It appears that there is a significant day effect, with the largest fluctuations occurring in the middle of the week and the smallest fluctuations at the beginning and end of the week.

—Jean D. Gibbons

REFERENCES

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association, 32*, 675-701.

Gibbons, J. D. (1993). *Nonparametric statistics: An introduction.* Newbury Park, CA: Sage.

Gibbons, J. D. (1997). *Nonparametric methods for quantitative analysis* (3rd ed.). Columbus, OH: American Sciences Press.

FUNCTION

Scientific accounts of the way a particular phenomenon or process works typically begin by identifying the fundamental actors and the fundamental quantities involved—that is, the variables and the entities to which those variables are attached. Immediately, the challenge is to specify the relations among the fundamental quantities. Many of these relations are summarized and embodied in mathematical functions of the form

$$y = f(x_1, \ldots, x_n),$$

where y and the xs are the fundamental quantities, now termed the DEPENDENT and INDEPENDENT VARIABLES, respectively.

As one begins to think about the relations among the quantities, the steps include thinking about which variables operate as inputs for which outcomes; and restricting attention to one function, whether the outcome increases or decreases as each input increases; and whether the rate of increase or decrease is constant, increasing, or decreasing. Of course, one also thinks about the domain and the range of the function, elements that are closely linked to the mathematical representation of the fundamental quantities. Moreover, sometimes the relation between an input and an outcome is nonmonotonic, with the effect changing direction at some point.

Increased understanding about the way a particular process works is manifested in progress from no function to a general function to a specific function, as it becomes clear that some variables belong in the equation (as xs) and others do not; that y increases with some xs, decreases with others, and is nonmonotonically related to still others; and that the rates of change are different with respect to different xs. Typically, the theoretical and empirical literature, even at the prefunction stage, is rich with insights about these features. One approach to specifying the function is to identify in the literature conditions that a function representing the particular relation should satisfy. Indeed, even when a specific function appears on the scene via intuition or empirical work, it is important to strengthen its foundation by building arguments that link its features to the desiderata in the literature.

EXAMPLE

The study of distributive-retributive justice identifies two actors—an *observer* and a *rewardee* (who may, but need not, be the observer)—and three quantities—the rewardee's *actual reward*; the observer's idea of the *just reward* for the rewardee; and the observer's judgment that the rewardee is justly or unjustly rewarded and, if unjustly rewarded, whether overrewarded or underrewarded and to what degree, called the *justice evaluation.* The literature provided many insights for specifying the justice evaluation function. First, the literature made it clear that (for goods) the justice evaluation increases with the actual reward and decreases with the just reward. Second, the literature included a representation of the justice evaluation by real numbers, with zero representing the point of perfect justice and the negative and positive numbers representing, respectively, underreward and overreward. Third, the literature also suggested two desirable properties for a justice evaluation function: (a) scale invariance, so that the justice evaluation is independent of the reward's units of measurement; and (b) additivity, in the sense that the effect of the actual reward on the justice evaluation is independent of the level of the just reward, and conversely. The first two insights led to a general function, and the third to a proof that, in the case of cardinal rewards, one functional form uniquely satisfies both the scale invariance and additivity conditions (Jasso, 1990).

The discipline of reasoning about a relation in order to specify the function has a high payoff. Even when it does not lead to a unique functional form, it serves to clarify scientific understanding of the way the process works.

The functions widely used in social science include psychophysical functions in psychology; utility, production, and consumption functions in economics; earnings functions in economics and sociology; perception and status functions in sociology and social psychology. As knowledge grows, more and more relations become specified as functions. Homans's (1967, p. 18) vivid challenge to specify the function is always timely.

Functions can serve as the starting ASSUMPTIONS for theories and theoretical models; they can also be used directly in empirical work (Jasso, in press). Thus, functions lend coherence to topical domains.

FURTHER READING

Regular perusal of functions in mathematical handbooks (such as Bronshtein and Semendyaev, 1979/1985) and calculus texts (such as Courant, 1934/1988) girds the mind for confrontation with vague relations (and provides immediate pleasure). Reading about the famous functions in social science and their development, such as Fechner's (1860, 1877) work on the sensation function and Mincer's (1958, 1974) on the earnings function, does the same.

—Guillermina Jasso

REFERENCES

Bronshtein, I. N., & Semendyaev, K. A. (1985). *Handbook of mathematics* (K. A. Hirsch, Trans. & Ed.). New York: Van Nostrand Reinhold. (Originally published in 1979.)

Courant, R. (1988). *Differential and integral calculus* (2 vols.) (E. J. McShane, Trans.). New York: Wiley. (Originally published in 1934.)

Fechner, G. T. (1860). *Elemente der psychopysik.* Leipzig: Breitkopf & Härtel.

Fechner, G. T. (1877). *In sachen der psychophysik.* Leipzig: Breitkopf & Härtel.

Homans, G. C. (1967). *The nature of social science.* New York: Harcourt, Brace, & World.

Jasso, G. (1990). Methods for the theoretical and empirical analysis of comparison processes. *Sociological Methodology, 20,* 369–419.

Jasso, G. (in press). The tripartite structure of social science analysis. *Sociological Theory.*

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy, 66,* 281–302.

Mincer, J. (1974). *Schooling, experience and earnings.* New York: Columbia University Press.

FUZZY SET THEORY

Fuzzy set theory deals with sets or categories whose boundaries are blurry or "fuzzy." For instance, an object is not necessarily just red or not red, it can be "reddish" or even a "warm green" (green with a tinge of red in it). Likewise, a political scientist may rate one government as more or less democratic than another. Fuzzy set theory permits membership in the set of red objects or democratic governments to be a matter of degree. It offers an analytic framework for handling concepts that are simultaneously categorical and dimensional. Below, the basic ideas behind fuzzy set theory and fuzzy logic are introduced and, where possible, illustrated with examples from the social sciences.

CONCEPT OF A FUZZY SET

Membership in a set may be represented by values in the [0,1] interval. A classical *crisp* set permits only the values 0 (nonmembership) and 1 (full membership). Treating the set of males as a crisp set would entail assigning each person a 1 if the criteria for being a male were satisfied and a 0 otherwise. A *fuzzy* set, on the other hand, permits values in between 0 and 1. Thus, a 21-year-old might be accorded full membership in the set of "young adults," but a 28-year-old might be given a degree of membership somewhere between 0 and 1. Conventionally, 1/2 denotes an element that is neither completely in nor out of the set. A *membership function* maps values or states from one or more *support* variables onto the [0,1] interval. Figure 1 shows two hypothetical membership functions using age as the support: "young adult" and "adolescent." Although adolescents are generally younger than young adults, these two sets overlap and permit nonzero degrees of membership in each.

Although there is no universally accepted interpretation of what degree of membership in a fuzzy set means, consensus exists about several points. First, a degree of membership is not a probability and, indeed, may be assigned with complete certainty. Moreover, unlike probabilities, degrees of membership need not sum to 1 across an exhaustive set of alternatives. Membership in some contexts is defined in terms of similarity to a prototype, but in others, it may connote a degree of compatibility or possibility in relation to a concept. An example of the latter is "several," considered to be a fuzzy verbal number. The integer 6 might be considered completely compatible with "several," whereas 3 would be less so and therefore assigned a lower membership value.

The concept of possibility also has been used to characterize degree of membership. *Possibility* provides an upper envelope on probability. For example, if 45% of the people in a community own a bicycle, then in a random survey of that community, the maximum probability of finding a person who has used his or her bicycle on a given day will be 0.45, but

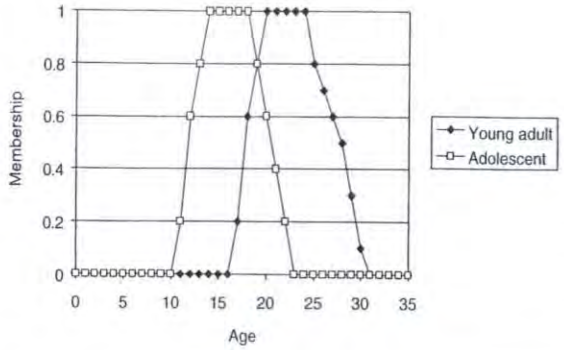

Figure 1 Membership Functions for "Young Adult" and "Adolescent"

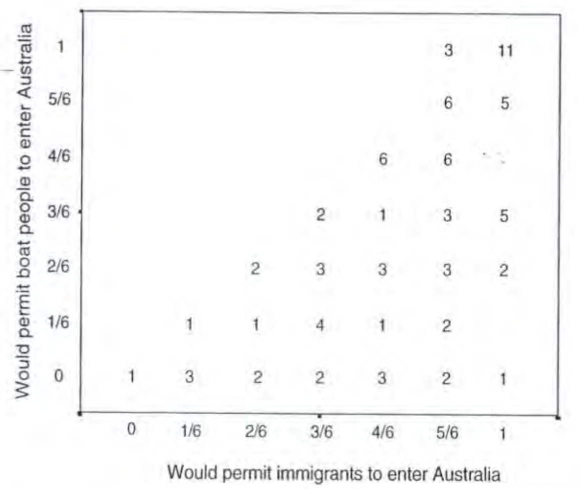

Figure 2 Example of Fuzzy Set Inclusion

of course, the actual probability may well be lower. *Possibility theory* is a framework based on fuzzy set theory and closely related to theories of imprecise probabilities (e.g., lower and upper probabilities and belief functions).

There has been some debate over the meaning of *fuzziness.* It is widely agreed that fuzziness is not ambiguity, vagueness, nonspecificity, unreliability, or disagreement. Instead, fuzziness refers to a precise kind of gradedness. The color judgments of an artist, for instance, are not ambiguous, nonspecific, or vague in any pejorative sense. Various measures of fuzziness have been proposed, ranging from generalizations of information-theoretic entropy to relative variation. Common to all of them is the notion that a set is less fuzzy as more of its elements have membership values near 0 or 1 and fuzzier as more membership values are near 1/2.

The measurement level of a membership function also has been debated, although important aspects of fuzzy set theory may be used even for ordinal-level membership functions. In many applications, a reasonable case can be made for the endpoints (0 and 1), but it is more difficult to argue for equal-interval intermediate membership values. A useful way of representing membership at the ordinal level is via *level sets,* which are crisp sets nested by ordered cutting points on the membership scale. Six of these are represented in Figure 1 by the horizontal gridlines. At a membership level of 0.6, for example, the set of young adults includes ages 18 to 27, whereas at membership level 1, the set of young adults shrinks to ages 20 to 24. Level sets are closely related to (de)cumulative distribution functions.

Finally, membership functions may be conditioned by states or membership in other sets. An example of the first is the fuzzy distance "close by" when we are on foot versus traveling by automobile. An example of the second is "tall" when applied to men versus women.

BASIC OPERATIONS ON FUZZY SETS

Let $m_A(x)$ denote the degree of membership element x has in set A. The subscript or (x) may be dropped when there is no ambiguity. Let $\sim A$ denote the complement of A. Fuzzy *negation* is the same as it is in PROBABILITY:

$$m_{\sim A} = 1 - m_A.$$

According to the membership functions in Figure 1, the degree of membership for a 20-year-old in the set of adolescents is 0.6, so the degree of membership in the complementary set is $1 - 0.6 = 0.4$.

Membership in the *intersection* of A and B is defined by

$$m_{A \cap B} = \min(m_A, m_B)$$

and membership in the *union* of A and B is defined by

$$m_{A \cup B} = \max(m_A, m_B).$$

In Figure 1, a 20-year-old is both an adolescent and young adult in degree $\min(0.6, 1) = 0.6$. Fuzzy set unions and intersections are noncompensatory. Unlike a weighted linear combination, for instance, if $m_A > m_B$, then an increase in m_A will not compensate for the impact of a decrease in m_B on membership in $m_{A \cup B}$. A crucial assumption implicit in these definitions is that the membership functions of sets A and B have a complete joint ordering, so that the comparative operations min and max are permissible.

Set B is said to be a subset of (included by) set A if, for all x, $m_A(x) > m_B(x)$. Note that this definition also requires the joint ordering assumption. Fuzzy set *inclusion* is conceptually a generalization of crisp set inclusion and thereby related directly to GUTTMAN and RASCH scaling. Although it is seldom the case that this inequality is perfectly satisfied, real examples may be readily found where it holds to quite a high degree.

Figure 2 shows one such instance, in which 83 survey respondents rated their degree of agreement with the propositions that "Australia should permit immigrants to enter the country" and "Boat people should be allowed to enter Australia and have their claims processed." The scatterplot shows only three exceptions to the inclusion relationship, in the upper right-hand corner. The set of people agreeing with the second statement is almost completely a subset of those agreeing with the first statement.

The concepts of inclusion and fuzziness have been used to interpret various transformations of membership functions in terms of approximate natural-language hedges such as "very." A transformation *concentrates* a fuzzy set if it lowers membership values and *dilates* the set if it increases them. The set of "very tall" men would be said to be a concentration of the set of "tall" men, and the original proposal for a concentration operator was $[m_A(x)]^2$. Although such operators have been strongly criticized as inadequate models of verbal hedges, they provide interpretations of certain transformations.

FUZZY LOGIC

Classical Boolean logic is two-valued, with every proposition being either true or false. Multiple-valued logics permit at least a third state: "indeterminate" or "possible." In Boolean logic, truth values are limited to 0 (false) and 1 (true), and three-valued logics assign a truth value of 1/2 to the third state. Fuzzy logic generalizes this by allowing any truth value in the [0, 1] interval.

The original versions of fuzzy logic use the min and max operators to compute logical conjunctions and disjunctions. Denoting the truth value of propositions P and Q by t_P and t_Q, the usual logical primitives are defined by

$$\sim \mathrm{P} = 1 - t_P,$$
$$\mathrm{P} \wedge \mathrm{Q} = \min(t_P, t_Q),$$
$$\mathrm{P} \vee \mathrm{Q} = \max(t_P, t_Q),$$

$$\mathrm{P} \Rightarrow \mathrm{Q} = \sim \mathrm{P} \vee (\mathrm{P} \wedge \mathrm{Q}) \quad \text{(Maxmin Rule)} \quad \text{or}$$
$$= \sim \mathrm{P} \vee Q \quad \text{(Arithmetic Rule), and}$$

$$\mathrm{P} \Leftrightarrow \mathrm{Q} = (\mathrm{P} \Rightarrow \mathrm{Q}) \wedge (\mathrm{Q} \Rightarrow \mathrm{P}).$$

The Maxmin and Arithmetic Rules yield different truth values for P $\Rightarrow$ Q only when $t_P < t_Q$, as Table 1 shows.

The result is that Arithmetic Rule truth values for implication are greater than or equal to Maxmin Rule truth values. Among the more popular alternative fuzzy logics, the Lukasiewicz (L_1) version yields the most generous truth values for implication. In L_1, the min and max operators are used for P $\wedge$ Q and P $\vee$ Q, but P $\Rightarrow$ Q $= \min(1, 1 - t_P + t_Q)$. Thus, in L_1, the truth value of P $\Rightarrow$ Q $= 1$ whenever $t_P < t_Q$, which agrees with the definition of subsethood in the previous section.

As an example, let us return to the example illustrated in Figure 2 and consider P $\Rightarrow$ Q where P = "Would permit boat people to enter Australia" and Q = "Would permit immigrants to enter Australia." Figure 3 displays the truth values produced by the Maxmin and Arithmetic Rules for the appropriate positions in the scatterplot from Figure 2. It turns out that if we sum the truth values for the 83 cases in Figure 2, the Maxmin Rule gives a total of 68.33, whereas the Arithmetic Rule gives 75.00. L_1 would assign a truth value of 1 to all but the three cases that violate $t_P < t_Q$, assigning them 5/6 instead, for a sum of 82.50.

Fuzzy logical operations may be combined to produce compound implication statements, which, in turn,

Table 1 Truth Tables for Maxmin and Arithmetic Rules

		Maxmin	*Arithmetic*
t_P	t_Q	P $\Rightarrow$ Q	P $\Rightarrow$ Q
0	0	1	1
0	1	1	1
1	0	0	0
1	1	1	1
$p >$	q	$\max(1-p, q)$	$\max(1-p, q)$
$p <$	q	$\max(1-p, p)$	$\max(1-p, q)$

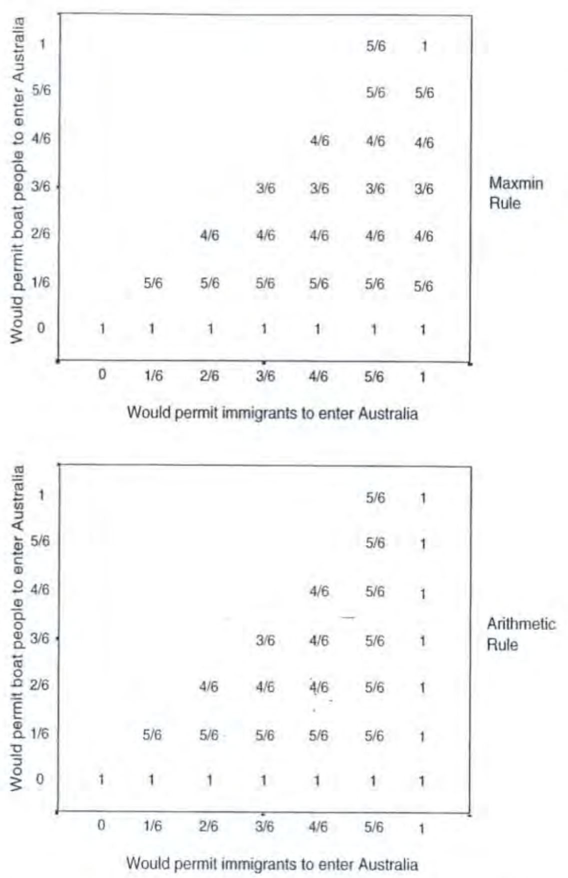

Figure 3 Maxmin and Arithmetic Rule Truth Values

may form the basis for NONLINEAR MODELS. Consider three propositions:

I = It is important to provide information to others.
T = It is important to maintain the trust of a confidant.
C = I am confident that I made the correct decision (to disclose or withhold information).

In an occupational survey, 229 respondents, presented with a dilemma over whether to disclose information confided to them by a colleague, rated each of these statements on 7-point scales that will be treated here as having truth values ranging from 0 to 1. We will examine the fuzzy logical model

$$C \Rightarrow I \vee T.$$

This model proposes that if the decision maker is confident, then she or he rates I or T highly (or both). This is a nonlinear model because the max operator is not linearizable, and, in fact, is not even smooth or differentiable.

The first two subtables in Table 2 show the bivariate distributions of I with C and T with C. The third subtable shows I $\vee$ T versus C. Visual inspection suggests that in the third subtable, the high-confidence cases have been pulled into the high I $\vee$ T region, whereas the low-confidence cases are fairly evenly scattered along the range of I $\vee$ T.

Using the Arithmetic Rule for fuzzy logical implication, the model C $\Rightarrow$ I achieves a total truth value of 159.7 out of 229, and the model C $\Rightarrow$ T achieves 162.3. An averaging model (i.e., C $\Rightarrow$ [I + T]/2) does not improve much, obtaining a truth value total of 166.7. However, the truth value of C $\Rightarrow$ I $\vee$ T is 198.00, which is more than a 22% improvement. A somewhat more conventional approach (although not as appropriate) would be to use CORRELATION and REGRESSION to evaluate I $\vee$ T versus a LINEAR model incorporating I and T for predicting C. The correlation between I and C

Table 2 Sufficiency and Joint Necessity

	I = Importance of providing information to others							
	0	*1/6*	*2/6*	*3/6*	*4/6*	*5/6*	*1*	*Total*
1	8	6	4	4	3	7	36	68
5/6		4	12	7	8	28	9	68
4/6		4	4		11	18	5	42
3/6	2	1	9	7	5	8	2	34
2/6				3	2	5	1	11
1/6		1			1			2
0			1	1	1		1	4
Total	10	16	30	22	31	66	54	229
	T = Importance of maintaining trust of colleague							
	0	*1/6*	*2/6*	*3/6*	*4/6*	*5/6*	*1*	*Total*
1	8	4	3	5	4	11	33	68
5/6	1	5	8	9	14	20	11	68
4/6		2	5	9	5	16	5	42
3/6	1			7	8	11	7	34
2/6				2	4	3	2	11
1/6			1			1		2
0	1			1	1		1	4
Total	11	11	17	33	36	62	59	229
	$I \vee T$							
	0	*1/6*	*2/6*	*3/6*	*4/6*	*5/6*	*1*	*Total*
1				1	1	13	53	68
5/6			1		11	37	19	68
4/6		1			10	24	7	42
3/6	1			4	7	14	8	34
2/6				2	1	6	2	11
1/6			1			1		2
0				1	2		1	4
Total	1	1	2	8	32	95	90	229

is .114, and between T and C, it is .025. A linear regression model combining both predictors yields a multiple correlation of .126, again not much of an improvement on the largest of the zero-order correlations. However, the correlation between $I \vee T$ and C is .430.

ORIGINS AND DEVELOPMENTS

Fuzzy set theory was invented by Lotfi Zadeh, a professor of engineering at the University of California at Berkeley, and first publicized in his classic 1965 paper. Forerunners include work on multiple-valued logics in the 1920s and 1930s, and Max Black's pioneering 1937 paper on reasoning with vague concepts. The fundamentals of fuzzy logic and possibility theory were worked out in the 1970s. By the late 1980s, axiomatizations of fuzzy set theory and fuzzy logic had been constructed, and debates about their relationship with probability frameworks were well under way, with an emerging consensus becoming apparent toward the end of the 20th century.

Partly because of its origins, the earliest and most extensive applications were made in control engineering. During the 1980s and 1990s, these applications burgeoned into a lucrative industry. In the past 15 years, fuzzy set theory and fuzzy logic have been incorporated along with BAYESIAN probability and techniques such as NEURAL NETS and genetic algorithms in the "soft computing" approach to constructing intelligent systems.

Applications in the human sciences have been slower to develop. Empirical work in cognitive psychology first appeared the 1970s, along with a few applications in the social sciences. By the end of the 1980s, a number of applications had emerged in subfields ranging from perception and memory to human geography, including at least two books in addition to a sizeable number of journal articles. Although opinion remains divided on the utility of fuzzy set theory for the human sciences, several commentators in recent years have suggested that its potential has not yet been fully realized.

—Michael Smithson

REFERENCES

Klir, G. J., & Folger, T. A. (1988). *Fuzzy sets, uncertainty, and information.* Englewood Cliffs, NJ: Prentice Hall.

Smithson, M. (1987). *Fuzzy set analysis for behavioral and social sciences.* New York: Springer-Verlag.

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8,* 338–353.

Zetenyi, T. (Ed.). (1988). *Fuzzy sets in psychology.* Amsterdam: North-Holland.

Appendix
Bibliography

Abbott, A., & Forrest, J. (1986). Optimal matching methods for historical sequences. *Journal of Interdisciplinary History, 16,* 471–494.

Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data. *American Journal of Sociology, 96,* 144–185.

Abdi, H., Dowling, W. J., Valentin, D., Edelman, B., & Posamentier, M. (2002). *Experimental design and research methods.* Unpublished manuscript, University of Texas at Dallas, Program in Cognition and Neuroscience.

Abdi, H., Valentin, D., & Edelman, B. (1999). *Neural networks.* Thousand Oaks, CA: Sage.

Abdi, H., Valentin, D., Edelman, B., & O'Toole, A. J. (1996). A Widrow-Hoff learning rule for a generalization of the linear auto-associator. *Journal of Mathematical Psychology, 40,* 175–182.

Abell, P. (1987). *The syntax of social life: The theory and method of comparative narratives.* Oxford, UK: Clarendon.

Abraham, B., & Ledolter, J. (1983). *Statistical methods for forecasting.* New York: Wiley.

Abraham, B., & Ledolter, J. (1986). Forecast functions implied by ARIMA models and other related forecast procedures. *International Statistical Review, 54,* 51–66.

Achen, C. H. (1975). Mass political attitudes and the survey response. *American Political Science Review, 69,* 1218–1231.

Achen, C. H. (1977). Measuring representation: Perils of the correlation coefficient. *American Journal of Political Science, 21,* 805–815.

Achen, C. H. (1982). *Interpreting and using regression.* Beverly Hills, CA: Sage.

Achen, C., & Shively, W. P. (1995). *Cross-level inference.* Chicago: University of Chicago Press.

Achinstein, P. (1983). *The nature of explanation.* New York: Oxford University Press.

Achinstein, P., & Barker, S. F. (1969). *The legacy of logical positivism.* Baltimore: Johns Hopkins University Press.

Adair, J. G., Sharpe, D., & Huynh, C. L. (1989). Hawthorne control procedures in educational experiments: A reconsideration of their use and effectiveness. *Review of Educational Research, 59,* 215–228.

Aday, L. A. (1996). *Designing and conducting health surveys: A comprehensive guide* (2nd ed.). San Francisco: Jossey-Bass.

Adler, P. A. (1985). *Wheeling and dealing.* New York: Columbia University Press.

Adler, P. A., & Adler, P. (1991). *Backboards and blackboards.* New York: Columbia University Press.

Adler, P. A., & Adler, P. (1994). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 377–392). Thousand Oaks, CA: Sage.

Adler, P. A., & Adler, P. (1998). *Peer power.* New Brunswick, NJ: Rutgers University Press.

Adler, P., & Adler, P. A. (1987). *Membership roles in field research* (Qualitative Research Methods Series vol. 6). Newbury Park, CA: Sage.

Adorno, T., & Horkheimer, M. (1979). *Dialectic of enlightenment* (J. Cumming, Trans.). London: Verso. (Original work published 1944).

Afifi, A. A., & Clark, V. A. (1996). *Computer-aided multivariate analysis* (3rd ed.). Boca Raton, FL: Chapman & Hall.

Agar, M. (1973). *Ripping and running.* New York: Academic Press.

Aggresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Agresti, A. (1984). *An introduction to categorical data analysis.* New York: John Wiley.

Agresti, A. (1984). *Analysis of ordinal categorical data.* New York: John Wiley.

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley.

Agresti, A. (1996). *An introduction to categorical data analysis* (2nd ed.). New York: John Wiley.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley Interscience.

Agresti, A., & Cato, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician, 54*(4), 280–288.

Agresti, A., & Finlay, B. (1997). *Statistical methods for social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Albrow, M. (1990). *Max Weber's construction of social theory.* London: Macmillan.

Aldrich, J. H., & Cnudde, C. F. (1975). Probing the bounds of conventional wisdom: A comparison of regression, probit, and discriminant analysis. *American Journal of Political Science, 19,* 571–608.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit and probit models.* Beverly Hills, CA: Sage.

Algina, J., Oshina, T. C., & Lin, W.-Y. (1994). Type I error rates for Welch's test and James's second order test under non-normality and inequality of variance when there are two groups. *Journal of Educational and Behavorial Statistics, 19,* 275–291.

Alkin, M. (1990). *Debates on evaluation.* Newbury Park, CA: Sage.

Allison, P. D. (2003). Event history analysis. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis.* London: Sage.

Allison, P. D. (1978). Measures of inequality. *American Sociological Review, 43,* 865–880.

Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. C. Clogg (Ed.), *Sociological methodology 1987.* San Francisco: Jossey-Bass.

Allison, P. D. (1999). *Multiple regression: A primer.* Thousand Oaks, CA: Pine Forge.

Allison, P. D. (2001). *Missing data.* Thousand Oaks, CA: Sage.

Allport, G. (1942). *The use of personal documents in psychological science.* New York: Social Science Research Council.

Alonso, E. (2002). AI and agents: State of the art. *AI Magazine, 23*(3), 25–29.

Altheide, D. L. (1987). Ethnographic content analysis. *Qualitative Sociology, 10,* 65–77.

Altheide, D. L. (1996). *Qualitative media analysis.* Newbury Park, CA: Sage.

Altheide, D. L. (2002). *Creating fear: News and the construction of crisis.* Hawthorne, NY: Aldine de Gruyter.

Altheide, D., & Johnson, J. (1994). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 485–499). Thousand Oaks, CA: Sage.

Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (2000). *Statistics with confidence: Confidence intervals and statistical guidelines* (2nd ed.). London: British Medical Journal Books.

Altman, M., Gill, J., & McDonald, M. P. (2003). *Numerical methods in statistical computing for the social sciences.* New York: Wiley.

Alvarez, R. M., & Glasgow, G. (2000). Two-stage estimation of nonrecursive choice models. *Political Analysis, 8,* 147–165.

Alveson, M., & Sköldberg, K. (2000). *Reflexive methodology.* London: Sage.

Alvesson, M. (2002). *Postmodernism and social research.* Buckingham, UK: Open University Press.

Alvesson, M., & Deetz, S. (2000). *Doing critical management research.* London: Sage.

Amemiya, T. (1985). *Advanced econometrics.* Cambridge, MA: Harvard University Press.

Amemiya, T. (1994). *Introduction to statistics and econometrics.* Cambridge, MA: Harvard University Press.

American Association for Public Opinion Research (AAPOR). (2000). *Standard definitions: Final dispositions for case codes and outcome rates for surveys.* Ann Arbor, MI: Author.

American Educational Research Association, American Psychological Association, & National Council on Educational Measurement. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47,* 1597–1611.

American Psychological Association. (2002, October 28). *PsycINFO database records* [electronic database]. Washington, DC: Author.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Anastasi, A., & Urbina, S. (1996). *Psychological testing* (7th ed.). New York: Prentice Hall.

Anderberg, M. R. (1973). *Cluster analysis for applications.* New York: Academic Press.

Anderson, N. H. (1981). *Methods of information integration theory.* New York: Academic Press.

Anderson, S., Auquier, A., Hauck, W. W., Oakes, D., Vandaele, W., & Weisberg, H. I. (1980). On the use of matrices in certain population mathematics. In *Statistical methods for comparative studies: Techniques for bias reduction.* New York: Wiley.

Andersson, B. E., & Nilsson, S. G. (1964). Studies in the reliability and validity of the critical incident technique. *Journal of Applied Psychology, 48*(1), 398–403.

Andrews, F. M., Morgan, J. N., & Sonquist, J. A. (1967). *Multiple classification analysis: A report on a computer program for multiple regression using categorical predictors.* Ann Arbor, MI: Survey Research Center, Institute for Social Research.

Andrews, L., & Nelkin, D. (1998). Do the dead have interests? Policy issues of research after life. *American Journal of Law & Medicine, 24,* 261.

Angle, J. (1986). The Surplus Theory of Social Stratification and the size distribution of personal wealth. *Social Forces, 65,* 293–326.

Angle, J. (2002). The statistical signature of pervasive competition on wage and salary incomes. *Journal of Mathematical Sociology, 26,* 217–270.

Angrasino, M. V., & Mays de Perez, K. (2000). Rethinking observation: From method to context. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 673–702). Thousand Oaks, CA: Sage.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91,* 444–455.

Angst, D. B., & Deatrick, J. A. (1996). Involvement in health care decisions: Parents and children with chronic illness. *Journal of Family Nursing, 2*(2), 174–194.

Anselin, L. (1988). *Spatial econometrics.* Boston: Kluwer Academic.

Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models, with an introduction to spatial econometrics. In A. Ullah & D. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237–289). New York: Marcel Dekker.

Antoni, M. H., Baggett, L., Ironson, G., LaPerriere, A., August, S., Kilmas, N., et al. (1991). Cognitive-behavioral stress management intervention buffers distress responses and immunologic changes following notification of HIV-1 seropositivity. *Journal of Consulting and Clinical Psychology, 59*(6), 906–915.

Arabie, P., & Boorman, S. A. (1973). Multidimensional scaling of measures of distance between partitions. *Journal of Mathematical Psychology, 10,* 148–203.

Arbuckle, J. (1997). *Amos users' guide version 3.6.* Chicago: Smallwaters Corporation.

Arch, D. C., Bettman, J. R., & Kakkar, P. (1978). Subjects' information processing in information display board studies. In K. Hunt (Ed.), *Advances in consumer research, Vol. 5* (pp. 555–559). Ann Arbor, MI: Association for Consumer Research.

Archer, M. (1995). *Realist social theory: The morphogenetic approach.* Cambridge, UK: Cambridge University Press.

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds.). (1998). *Critical realism: Essential readings.* London: Routledge Kegan Paul.

Aristotle. (1977). *Metaphysica.* Baarn: Het Wereldvenster.

Armacost, R. L., Hosseini, J. C., Morris, S. A., & Rehbein, K. A. (1991). An empirical comparison of direct questioning, scenario, and randomized response methods for obtaining sensitive business information. *Decision Sciences, 22,* 1073–1090.

Armitage, P., & Colton, T. (Eds.). (1998). *Encyclopedia of biostatistics.* New York: John Wiley.

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting.* Boston: Kluwer.

Aronow, E., Reznikoff, M., & Moreland, K. (1994). *The Rorschach technique.* Needham Heights, MA: Allyn & Bacon.

Aronson, E., Carlsmith, M., & Ellsworth, P. C. (1990). *Methods of research in social psychology.* New York: McGraw-Hill.

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New Haven, CT: Yale University Press.

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy.* Ann Arbor: University of Michigan Press.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258–290.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177–190). Pittsburgh, PA: Carnegie Press.

Asher, H. B. (1983). *Causal modeling* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–003). Beverly Hills, CA: Sage.

Ashmore, M. (1989). *The reflexive thesis: Wrighting sociology of scientific knowledge.* Chicago: University of Chicago Press.

Ashmore, M. (1995). Fraud by numbers: Quantitative rhetoric in the Piltdown forgery discovery. *South Atlantic Quarterly, 94*(2), 591–618.

Atkinson, J. M., & Heritage, J. (Eds.). (1984). *Structures of social action: Studies in conversation analysis.* Cambridge, UK: Cambridge University Press.

Atkinson, P. (1990). *The ethnographic imagination: Textual constructions of reality.* London: Routledge.

Atkinson, P., & Hammersley, M. (1994). Ethnography and participant observation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 248–261). Thousand Oaks, CA: Sage.

Atkinson, P., & Silverman, D. (1997). Kundera's "immortality": The interview society and the invention of self. *Qualitative Inquiry, 3*(3), 304–325.

Atkinson, P., Coffey, A., & Delamont, S. (2001). Editorial: A debate about our canon. *Qualitative Research, 1,* 5–21.

Atkinson, R. (1998). *The life story interview* (Qualitative Research Methods Series, Vol. 44). Thousand Oaks, CA: Sage.

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Axelrod, R. (1995). A model of the emergence of new political actors. In N. Gilbert & R. Conte (Eds.), *Artificial societies.* London: UCL.

Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration.* Princeton, NJ: Princeton University Press.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211,* 1390–1396.

Ayer, A. J. (1936). *Language, truth and logic.* London: Victor Gollancz.

Babbie, E. (1979). *The practice of social research.* Belmont, CA: Wadsworth.

Babbie, E. R. (1995). *The practice of social research* (7th ed.). Belmont, CA: Wadsworth.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics, 73,* 5–59.

Bainbridge, E. E., Carley, K. M., Heise, D. R., Macy, M. W., Markovsky, B., & Skvoretz, J. (1994). Artificial social intelligence. *Annual Review of Sociology, 20,* 407–436.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ.* New York: Cambridge University Press.

Baker, B. O., Hardyck, C. D., & Petrinovich, L. F. (1966). Weak measurement vs. strong statistics: An empirical critique of

S.S. Stevens' proscriptions on statistics. *Educational and Psychological Measurement, 26,* 291–309.

Baltagi B. H. (2001). *Econometric analysis of panel data.* Chichester, UK: Wiley.

Banerjee, A., Dolado, J. J., Galbraith, J., & Hendry, D. F. (1993). *Cointegration, error correction and the econometric analysis of nonstationary series.* New York: Oxford University Press.

Banks, M. (2001). *Visual methods in social research.* London: Sage.

Bargh, J. A., & Chartrand, T. L. (2000). Mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253–285). New York: Cambridge University Press.

Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika, 86,* 948–955.

Barndt, D. (1997). Zooming out/zooming in: Visualizing globalization. *Visual Sociology, 12*(2), 5–32.

Barnes, D. B., Taylor-Brown, S., & Weiner, L. (1997). "I didn't leave y'all on purpose": HIV-infected mothers' videotaped legacies for their children. *Qualitative Sociology, 20*(1), 7–32.

Barnett, V. (1974). *Elements of sampling theory.* Sevenoaks, UK: Hodder & Stoughton.

Barr, R. (1996). A comparison of aspects of the US and UK censuses of population. *Transactions in GIS, 1,* 49–60.

Barthes, R. (1967). *Elements of semiology.* London: Jonathan Cape.

Barthes, R. (1977). *Image music text.* London: Fontana.

Barthes, R. (1977). Rhetoric of the image. In *Image, music, text* (S. Heath, Trans.). London: Fontana.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis.* London: Arnold.

Bartlett, M. S. (1947). Multivariate analysis. *Journal of the Royal Statistical Society, Series B, 9,* 176–197.

Bartunek, J. M., & Louis, M. R. (1996). *Insider-outsider team research.* Thousand Oaks, CA: Sage.

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications.* New York: John Wiley.

Bateson, G., & Mead, M. (1942). *Balinese character: A photographic analysis.* New York: New York Graphic Society.

Baudrillard, J. (1998). *The consumer society.* London: Sage. (Originally published in 1970)

Bauman, Z. (1978). *Hermeneutics and social science.* London: Hutchinson.

Baumrind, D. (1964). Some thoughts on ethics after reading Milgram's "Behavioral study of obedience." *American Psychologist, 19,* 421–423.

Bayer, A. E., & Smart, J. C. (1991). Career publication patterns and collaborative "styles" in American academic science. *Journal of Higher Education, 62,* 613–636.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society, 53,* 370–418.

Bayes, T. (1958). An essay towards solving a problem in the doctrine of chances. *Biometrika, 45,* 293–315.

Baym, N. (2000). *Tune in, log on: Soaps, fandom and online community.* Thousand Oaks, CA: Sage.

Beck, N. (1991). Comparing dynamic specifications: The case of presidential approval. *Political Analysis, 3,* 51–87.

Beck, N., & Jackman, S. (1998). Beyond linearity by default: Generalized additive models. *American Journal of Political Science, 42,* 596–627.

Beck, N., & Katz, J. N. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review, 89,* 634–647.

Becker, H. S. (1967). Whose side are we on? *Social Problems, 14,* 239–247.

Becker, H. S. (1974). Photography and sociology. *Studies in the Anthropology of Visual Communication, 1*(1), 3–26.

Becker, H. S. (1986). *Doing things together.* Evanston, IL: Northwestern University Press.

Becker, H. S. (1998). *Tricks of the trade.* Chicago: University of Chicago Press.

Becker, H. S., Hughes, E. C., & Strauss, A. L. (1961). *Boys in white.* Chicago: University of Chicago Press.

Becker, S., & Bryman, A. (Eds.). (2004). *Understanding research for social policy and practice: Themes, methods and approaches.* Bristol, UK: Policy Press.

Bell, C. (1969). A note on participant observation. *Sociology, 3,* 417–418.

Bell, S. E. (1999). Narratives and lives: Women's health politics and the diagnosis of cancer for DES daughters. *Narrative Inquiry, 9*(2), 1–43.

Bellman, B. L., & Jules-Rosette, B. (1977). *A paradigm for looking: Cross-cultural research with visual media.* Norwood, NJ: Ablex.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity.* New York: John Wiley.

Beltrami, E. J. (1999). *What is random? Chance and order in mathematics and life.* New York: Copernicus.

Benfer, R. A., Brent, E. E., Jr., & Furbee, L. (1991). *Expert systems.* Newbury Park, CA: Sage.

Bennet, D. J. (1998). *Randomness.* Cambridge, MA: Harvard University Press.

Bennett, A. (1999). *Condemned to repetition? The rise, fall, and reprise of Soviet-Russian military interventionism, 1973–1996.* Cambridge: MIT Press.

Benton, T. (1977). *Philosophical foundations of the three sociologies.* London: Routledge Kegan Paul.

Benton, T. (1998). Realism and social science. In M. Archer, R. Bhaskar, A. Collier, T. Lawson, & A. Norrie (Eds.), *Critical realism: Essential readings.* London: Routledge Kegan Paul.

Benzécri, J.-P. (1973). *Analyse des Données. Tome 2: Analyse des Correspondances* [Data analysis: Vol. 2. Correspondence analysis]. Paris: Dunod.

Berelson, B. (1954). Content analysis. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 1, pp. 488–522). Reading, MA: Addison-Wesley.

Berg, B. L. (2001). *Qualitative research methods for the social sciences.* Boston: Allyn and Bacon.

Berger, J. O., & Wolpert, R. L. (1984). *The likelihood principle: A review, generalizations, and statistical implications.* Hayward, CA: Institute of Mathematical Statistics.

Berger, J., Fisek, H., Norman, R., & Zelditch, M. (1977). *Status characteristics and social interaction: An expectation states approach.* New York: Elsevier.

Berger, P., & Luckmann, T. (1967). *The social construction of reality.* Harmondsworth, UK: Penguin.

Bergner, R. M. (1974). The development and evaluation of a training videotape for the resolution of marital conflict. *Dissertation Abstracts International, 34,* 3485B. (UMI No. 73–32510)

Bergsma, W. (1997). *Marginal models for categorical data.* Tilburg, The Netherlands: Tilburg University Press.

Berk, R. A., & Freedman, D. A. (1995). Statistical assumptions as empirical commitments. In T. G. Blomberg & S. Cohen (Eds.), *Law, punishment, and social control: Essays in honor of Sheldon Messinger* (pp. 245–258). New York: Aldine de Gruyter.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* New York: Wiley.

Bernstein, R. (1983). *Beyond objectivism and relativism: Science, hermeneutics and praxis.* Oxford, UK: Basil Blackwell.

Berry, J. (2002). Validity and reliability issues in elite interviewing. *PS—Political Science and Politics, 35,* 679–682.

Berry, W. D. (1984). *Nonrecursive causal models.* Beverly Hills, CA: Sage.

Berry, W. D. (1993). *Understanding regression assumptions.* Newbury Park, CA: Sage.

Beverley, J. (1989). The margin at the center: On testimonio. *Modern Fiction Studies, 35*(1), 11–28.

Beverley, J., & Zimmerman, M. (1990). *Literature and politics in the Central American revolutions.* Austin: University of Texas Press.

Bhaskar, R. (1978). *A realist theory of science.* Hassocks, UK: Harvester Press.

Bhaskar, R. (1979). *The possibilities of naturalism: A philosophical critique of the contemporary human sciences.* Brighton, UK: Harvester.

Bhaskar, R. (1986). *Scientific realism and human emancipation.* London: Verso.

Bhaskar, R. (1997). *A realist theory of science.* London: Verso. (Original work published 1978)

Bhaskar, R. (1998). *The possibility of naturalism* (2nd ed.). Hemel Hempstead, UK: Harvester Wheatsheaf. (Original work published 1979)

Bickman, L., & Rog, D. J. (Eds.). (1998). *Handbook of applied social research methods.* Thousand Oaks, CA: Sage.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds.). (1991). *Measurement errors in surveys.* New York: Wiley.

Biggs, D., De Ville, B., & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics, 18,* 49–62.

Bijleveld, C. C. J. H., & van der Kamp, L. J. Th. (with Mooijaart, A., van der Kloot, W. A., van der Leeden, R., & van der Burg, E.). (1998). *Longitudinal data analysis: Designs, models, and methods.* London: Sage.

Billig, M. (1996). *Arguing and thinking.* Cambridge, UK: Cambridge University Press.

Bimler, D., & Kirkland, J. (1999). Capturing images in a net: Perceptual modeling of product descriptors using sorting data. *Marketing Bulletin, 10,* 11–23.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays on body motion communication.* Philadelphia: University of Pennsylvania Press.

Birnbaum, A. L. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Birnbaum, M. H. (2001). *Introduction to behavioral research on the Internet.* Upper Saddle River, NJ: Prentice Hall.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford, UK: Oxford University Press.

Bishop, G., & Smith, A. (2001). Response-order effects and the early Gallup split-ballots. *Public Opinion Quarterly, 65,* 479–505.

Black, D. (1958). *The theory of committees and elections.* Cambridge, UK: Cambridge University Press.

Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement, and statistics.* London: Sage.

Blaikie, N. (1993). *Approaches to social enquiry.* Cambridge, UK: Polity.

Blaikie, N. (2000). *Designing social research: The logic of anticipation.* Cambridge, UK: Polity.

Blalock, H. (1961). *Causal inference in nonexperimental research.* Chapel Hill: University of North Carolina Press.

Blalock, H. M. (1960). *Social statistics.* New York: McGraw-Hill.

Blalock, H. M. (1961). Theory, measurement and replication in the social sciences. *American Journal of Sociology, 66*(1), 342–347.

Blalock, H. M. (1972). *Social statistics* (2nd ed.). Tokyo: McGraw-Hill Kogakusha.

Blalock, H. M. (1979). *Social statistics* (rev. 2nd ed.). New York: McGraw-Hill.

Blalock, H. M., Jr. (1969). *Theory construction: From verbal to mathematical formulations.* Englewood Cliffs, NJ: Prentice Hall.

Blanck, P. D. (Ed.). (1993). *Interpersonal expectations: Theory, research, and applications.* New York: Cambridge University Press.

Blascovich, J. (2000). Psychophysiological indexes of psychological processes. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 117–137). Cambridge, UK: Cambridge University Press.

Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 1–51). New York: Academic Press.

Blau, P. M. (1974). Parameters of social structure. *American Sociological Review. 39,* 615–635.

Bloor, M. (1997). Addressing social problems through qualitative research. In D. Silverman (Ed.), *Qualitative research: Theory, method and practice* (pp. 221–238). London: Sage.

Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New approaches to causal analysis.* Mahwah, NJ: Lawrence Erlbaum Associates.

Blumen, I. M., Kogan, M., & McCarthy, P. J. (1955). *The industrial mobility of labor as a probability process.* Ithaca, NY: Cornell University Press.

Blumen, I., Kogan, M., & McCarthy, P. J. (1968). Probability models for mobility. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science* (pp. 318–334). Cambridge: MIT Press.

Blumer, H. (1956). Sociological analysis and the "variable." *American Sociological Review, 21,* 683–690.

Blumer, H. (1969). *An empirical appraisal of Thomas and Znaniecki (1918–20) The Polish Peasant in Europe and America.* New Brunswick, NJ: Transaction Books. (Original work published 1939)

Blumer, H. (1969). *Symbolic interactionism: Perspective and method.* Englewood Cliffs, NJ: Prentice Hall.

Blundell, R. W., & Smith, R. J. (1989). Estimation in a class of simultaneous equation limited dependent variable models. *Review of Economic Studies, 56,* 37–58.

Blurton Jones, N. (Ed.). (1972). *Ethological studies of child behaviour.* London: Cambridge University Press.

Bobko, P. (1995). *Correlation and regression : Principles and applications for industrial organizational psychology and management.* New York: McGraw-Hill.

Bogartz, R. S. (1994). *An introduction to the analysis of variance.* Westport, CT: Praeger.

Bogdewic, S. P. (1999). Participant observation. In B. J. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (pp. 47–69). Thousand Oaks, CA: Sage.

Bohman, J. (1991). *New philosophy of social science: Problems of indeterminacy.* Cambridge, UK: Polity.

Boje, D. M. (2001). *Narrative methods for organizational and communication research.* Thousand Oaks, CA: Sage.

Bolger, N., Davis, A., & Rafaeli, E. (in press). Diary methods: Capturing life as it is lived. *Annual Review of Psychology.*

Bollen, K. (1989). *Structural equations with latent variables.* New York: John Wiley.

Bollen, K. A., & Curran, P. J. (in press). *Latent curve models: A structural equation approach.* New York: John Wiley.

Bollen, K. A., & Jackman, R. W. (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In J. Fox & J. S. Long (Ed.), *Modern methods of data analysis* (pp. 257–291). Newbury Park, CA: Sage.

Bonnie, R. J., & Wallace, R. B. (2002). *Elder mistreatment: Abuse, neglect, and exploitation in an aging America.* Washington, DC: Joseph Henry Press.

Boomsma, A., Van Duijn, M. A. J., & Snijders, T. A. B. (Eds.). (2001). *Essays on item response theory.* New York: Springer.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling.* New York: Springer-Verlag.

Borgatta, E. F., & Bohrnstedt, G. W. (1980). Level of measurement: Once over again. *Sociological Methods and Research, 9,* 147–160.

Borgatta, E. F., & Jackson, D. J. (Eds.). (1980). *Aggregate data: Analysis and interpretation.* Beverly Hills, CA: Sage.

Bornstein, R. F. (2002). A process dissociation approach to objective-projective test score interrelationships. *Journal of Personality Assessment, 78,* 47–68.

Bornstein, R. F., Rossner, S. C., Hill, E. L., & Stepanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment, 63,* 363–386.

Boruch, R. F. (1997). *Randomized experiments for planning and evaluation: A practical guide.* Thousand Oaks, CA: Sage.

Boruch, R. F., & Cecil, J. S. (1979). *Assuring the confidentiality of social research data.* Philadelphia: University of Pennsylvania Press.

Boster, J. S. (1994). The successive pile sort. *Cultural Anthropology Methods, 6*(2), 7–8.

Bosworth, M. (1999). *Engendering resistance: Agency and power in women's prisons.* Aldershot, UK: Ashgate/Dartmouth.

Bottorff, J. L. (1994). Using videotaped recordings in qualitative research. In J. M. Morse (Ed.), *Critical issues in qualitative research* (pp. 224–261). Newbury Park, CA: Sage.

Boucke, O. F. (1923). The limits of social science: II. *American Journal of Sociology, 28*(4), 443–460.

Bourdieu, P. (1977). *Outline of a theory of practice.* Cambridge, UK: Cambridge University Press.

Bourdieu, P. (2000). *Pascalian meditations* (R. Nice, Trans.). Cambridge, UK: Polity.

Bourdieu, P., & Waquant, L. J. D. (1992). *An invitation to reflexive sociology.* Chicago: University of Chicago Press.

Bourque, L. B., & Clark, V. A. (1992). *Processing data: The survey example* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–085). Newbury Park, CA: Sage.

Bourque, L. B., & Fielder, E. P. (2003). *How to conduct self-administered and mail surveys* (2nd ed.). Thousand Oaks, CA: Sage.

Bourque, L. B., & Russell, L. A. (with Krauss, G. L., Riopelle, D., Goltz, J. D., Greene, M., McAffee, S., & Nathe, S.) (1994, July). *Experiences during and responses* *to the Loma Prieta earthquake*. Oakland: Governor's Office of Emergency Services, State of California.

Bourque, L. B., Shoaf, K. I., & Nguyen, L. H. (1997). Survey research. *International Journal of Mass Emergencies and Disasters, 15,* 71–101.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society (B), 26*(2), 211–252.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (rev. ed.). San Francisco: Holden-Day.

Box, G. E. P., & Tiao, G. C. (1965). A change in level of nonstationary time series. *Biometrika, 52,* 181–192.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. New York: Wiley.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association, 70,* 70–92.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters*. New York: Wiley.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd ed.). New York: Prentice Hall.

Box-Steffensmeier, J. M., & Jones, B. S. (1997). Time is of the essence: Event history models in political science. *American Journal of Political Science, 41,* 336–383.

Box-Steffensmeier, J. M., & Zorn, C. J. W. (2001). Duration models and proportional hazards in political science. *American Journal of Political Science, 45,* 951–967.

Box-Steffensmeier, J. M., & Zorn, C. J. W. (2002). Duration models for repeated events. *The Journal of Politics, 64*(4), 1069–1094.

Boyd, J. P. (1990). *Social semigroups: A unified theory of scaling and bockmodeling as applied to social networks*. Fairfax, VA: George Mason University Press.

Boyd, J. P., & Jonas, K. J. (2001). Are social equivalences ever regular? Permutation and exact tests. *Social Networks, 23,* 87–123.

Bracken, B. (Ed.). (1995). *Handbook of self concept: Developmental, social and clinical consequences*. New York: Wiley.

Bradley, J. V. (1985). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice Hall.

Brady, I. (2000). Anthropological poetics. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 949–979). Thousand Oaks, CA: Sage.

Brady, I. (2003). *The time at Darwin's Reef: Poetic explorations in anthropology and history*. Walnut Creek, CA: AltaMira.

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–054) Beverly Hills, CA: Sage.

Breen, R. (1996). *Regression models: Censored, sample selected, or truncated data* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–111). Thousand Oaks, CA: Sage.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth.

Brennan, R. L. (1983). *Elements of generalizability*. Iowa City, IA: American College Testing Program.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Brewer, K. (2002). *Combined survey sampling inference: Weighing Basu's elephants*. London: Arnold.

Briggs, C. (1986). *Learning how to ask*. Cambridge, UK: Cambridge University Press.

Broad, W., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science*. New York: Simon & Schuster.

Brockmeyer, E., Halstrom, H. L., & Jensen, A. (1960). *The life and works of A. K. Erlang*. Kobenhavn: Akademiet for de Tekniske Videnskaber.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Bronner, S. E., & Kellner, D. M. (1989). *Critical theory and society: A reader*. New York: Routledge.

Bronshtein, I. N., & Semendyaev, K. A. (1985). *Handbook of mathematics* (K. A. Hirsch, Trans. & Ed.). New York: Van Nostrand Reinhold. (Originally published in 1979)

Brown, C. (1991). *Ballots of tumult: A portrait of volatility in American voting*. Ann Arbor: University of Michigan Press.

Brown, C. (1995). *Chaos and catastrophe theories*. Thousand Oaks, CA: Sage.

Brown, C. (1995). *Serpents in the sand: Essays on the nonlinear nature of politics and human destiny*. Ann Arbor: University of Michigan Press.

Brown, H. I. (1987). *Observation and objectivity*. New York: Oxford University Press.

Brown, S. R., & Melamed, L. E. (1998). *Experimental design and analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–074). Thousand Oaks, CA: Sage.

Brown, S. R., Durning, D. W., & Selden, S. C. (1999). Q methodology. In G. J. Miller & M. L. Whicker (Eds.), *Handbook of research methods in public administration* (pp. 599–637). New York: Dekker.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 12,* 296–322.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, UK: Cambridge University Press.

Browne, M. W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

Brownlee, K. (1965). *Statistical theory and methodology in science and engineering*. New York: John Wiley.

Bruner, J. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Bruner, J. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Brunk, G. G., Caldeira, G. A., & Lewis-Beck, M. S. (1987). Capitalism, socialism, and democracy: An empirical inquiry. *European Journal of Political Research, 15,* 459–470.

Brunswik, E. (1943). Organismic achievement and environmental probability. *The Psychological Review, 50,* 255–272.

Bryant, C. G. A. (1985). *Positivism in social theory and research.* London: Macmillan.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

Bryman, A. (2001). *Social research methods.* Oxford, UK: Oxford University Press.

Bryman, A., & Cramer, D. (1997). *Quantitative data analysis with SPSS for Windows: A guide for social scientists.* New York: Routledge Kegan Paul.

Bryun, T. S. (1966) *The human perspective in sociology: The methodology of participant observation.* Englewoood Cliffs, NJ: Prentice Hall.

Bulmer, M. (Ed.). (1982). *Social research ethics.* London: Macmillan.

Bulmer, M., & Solomos, J. (Eds.). (2003). *Researching race and racism.* London: Routledge Kegan Paul.

Bunge, M. (1959). *Causality: The place of the causal principle in modern science.* Cambridge, MA: Harvard University Press.

Burawoy, M. (1998). The extended case method. *Sociological Theory, 16*(1), 4–33.

Burgess, R. (1984). *In the field.* London: Allen & Unwin.

Burgess, R. G. (Ed.). (1982). *Field research: A sourcebook and field manual.* London: Allen & Unwin.

Burkhart, R. E. (2000). Economic freedom and democracy: Post-cold war tests. *European Journal of Political Research, 37,* 237–253.

Burkhart, R. E., and Lewis-Beck, M. S. (1994). Comparative democracy: The economic development thesis. *American Political Science Review, 88,* 903–910.

Burr, V. (1995). *An introduction to social constructionism.* London: Routledge.

Burton, M. L. (1975). Dissimilarity measures for unconstrained sorting data. *Multivariate Behavioral Research, 10,* 409–424.

Bury, M. (2001). Illness narratives: Fact or fiction? *Sociology of Health and Illness, 23*(3), 263–285.

Buttny, R., & Morris, G. H. (2001). Accounting. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 285–301). Chichester, England: Wiley.

Byrne, D. (1998). *Complexity theory and the social sciences: An introduction.* London: Routledge.

Byrne, D. (2002). *Interpreting quantitative data.* London: Sage.

Byrne, D. S., McCarthy, P., Harrison, S., & Keithley, J. (1986). *Housing and health.* Aldershot, UK: Gower.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic specificity during simple physical and attitudinal tasks: Location and topographical features of integrated EMG responses. *Biological Psychology, 18,* 85–121.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). *Handbook of psychophysiology* (2nd ed.). Cambridge, UK: Cambridge University Press.

Cahill, S. (1987). Children and civility: Ceremonial deviance and the acquisition of ritual competence. *Social Psychology Quarterly, 50,* 312–321.

Cain, C. (1991). Personal stories: Identity acquisition and self-understanding in Alcoholics Anonymous. *Ethos, 19,* 210–253.

Cale, G. (2001). *When resistance becomes reproduction: A critical action research study.* Proceedings of the 42nd Adult Education Research Conference. East Lansing: Michigan State University.

Callendar, J. C., & Osburn, H. G. (1977). A method for maximizing split-half reliability coefficients. *Educational and Psychological Measurement, 37,* 819–825.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data.* New York: Cambridge University Press.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24,* 409–429.

Campbell, D. T. (1988). *Methodology and epistemology for social science: Selected papers* (E. S. Overman, Ed.). Chicago: University of Chicago Press.

Campbell, D. T. (1991). Methods for the experimenting society. *Evaluation Practice, 12,* 223–260.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts.* New York: Guilford.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago: Rand McNally.

Campbell, D. T., Boruch, R. F., Schwartz, R. D., & Steinberg, J. (1977). Confidentiality-preserving modes of access to files and to interfile exchange for useful statistical analysis. *Evaluation Quarterly, 1,* 269–300.

Cannell, C., Fowler, J., & Marquis, K. (1968). *The influence of interviewer and respondent psychological and behavioral variables on the reporting of household interviews* (Vital and Health Statistics Series 2, No. 6). Washington, DC: National Center for Health Statistics.

Capdevila, R., & Stainton Rogers, R. (2000). If you go down to the woods today . . . : Narratives of Newbury. In H. Addams & J. Proops (Eds.), *Social discourse and environmental policy: An application of Q methodology* (pp. 152–173). Cheltenham, UK: Elgar.

Carless, S. A. (1998). Assessing the discriminant validity of the transformational leadership behaviour as measured by the MLQ. *Journal of Occupational and Organizational Psychology, 71,* 353–358.

Carley, K., & Prietula, M. (Eds.). (1994). *Computational organization theory.* Hillsdale, NJ: Lawrence Erlbaum.

Carlin, B. P., & Louis, T. A. (1998). *Bayes and empirical Bayes methods for data analysis.* London: Chapman & Hall/CRC.

Carlson, M., & Mulaik, S. A. (1993). Trait ratings from descriptions of behavior as mediated by components of meaning. *Multivariate Behavioral Research, 28,* 111–159.

Carmines, E. G., & McIver, J. P. (1981). *Unidimensional scaling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–024). Beverly Hills, CA: Sage.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Beverly Hills, CA: Sage.

Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2003). *Models and methods in social network analysis.* New York: Cambridge University Press.

Carroll, J. B. (2002). The five-factor personality model: How complete and satisfactory is it? In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 97–126). Mahwah, NJ: Lawrence Erlbaum.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology, 31,* 607–649.

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via a N-way generalisation of Eckart-Young decomposition. *Psychometrika, 35,* 283–299, 310–319.

Carroll, J. M. (1997). Human-computer interaction: Psychology as a science of design. *Annual Review of Psychology, 48,* 61–83.

Cartwright, N. (1989). *Nature's capacities and their measurement.* Oxford, UK: Oxford University Press.

Casella, G., & Berger, R. L. (1990). *Statistical inference.* Belmont, CT: Duxbury.

Casella, G., & Berger, R. L. (2001). *Statistical inference* (2nd ed.). New York: Wadsworth.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276.

Chadhuri, A., & Mukerjee, R. (1987). *Randomized response: Theory and techniques.* New York: Marcel Dekker.

Chalmers, A. (1982). *What is this thing called science?* St. Lucia, Australia: University of Queensland Press.

Chalmers, A. (1999). *What is this thing called science?* (3rd ed.). Buckingham, UK: Open University Press.

Chamberlayne, P., Bornat, J., & Wengraf, T. (2000). *The turn to biographical methods in social science.* London: Routledge Kegan Paul.

Chamberlayne, P., Rustin, M., & Wengraf, T. (Eds.). (2002). *Biography and social exclusion in Europe: Experiences and life journeys.* Bristol, UK: Policy Press.

Charemza, W. W., & Deadman, D. F. (1997). *New directions in econometric practice* (2nd ed.). Cheltenham, UK: Edward Elgar.

Charlton, J., Patrick, D. L., Matthews, G., & West, P. A. (1981). Spending priorities in Kent: A Delphi study. *Journal of Epidemiology and Community Health, 35,* 288–292.

Charlton, T., Gunter, B., & Hannan, A. (2002). *Broadcast television effects in a remote community.* Hillsdale, NJ: Lawrence Erlbaum.

Charmaz, K. (2000). Grounded theory: Constructivist and objectivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509–535). Thousand Oaks, CA: Sage.

Chatfield, C. (1983). *Statistics for technology* (3rd ed.). London: Chapman & Hall.

Chatfield, C. (1996). *The analysis of time series: An introduction.* New York: Chapman & Hall.

Chavez, L., Hubbell, F. A., McMullin, J. M., Marinez, R. G., & Mishra, S. I. (1995). Structure and meaning in models of breast and cervical cancer risk factors: A comparison of perceptions among Latinas, Anglo women and physicians. *Medical Anthropology Quarterly, 9,* 40–74.

Checkland, P. B. (1981). *Systems thinking, systems practice.* Chichester, UK: Wiley.

Chell, E. (1998). Critical incident technique. In G. Symon & C. Cassell (Eds.), *Qualitative methods and analysis in organizational research: A practical guide* (pp. 51–72). London: Sage.

Chell, E., & Baines, S. (1998). Does gender affect business performance? A study of micro-businesses in business services in the U.K. *International Journal of Entrepreneurship and Regional Development, 10*(4), 117–135.

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures.* Thousand Oaks, CA: Sage.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*(2), 367–405.

Cherkassky, V., & Mulier, F. (1998). *Learning from data.* New York: Wiley.

Chernick, M. R. (1999). *Bootstrap methods: A practioner's guide.* New York: Wiley-Interscience.

Cherryholmes, C. H. (1999). *Reading pragmatism.* New York: Teachers College Press.

Chiang, A. C. (1984). *Fundamental methods of mathematical economics* (3rd ed.). New York: McGraw-Hill.

Chiari, G., & Nuzzo, M. L. (1996). Psychological constructivisms: A metatheoretical differentiation. *Journal of Constructivist Psychology, 9,* 163–184.

Chipman, J. S. (1979). Efficiency of least squares estimation of linear trend when residuals are autocorrelated. *Econometrica, 47,* 115–128.

Chouliaraki, L., & Fairclough, N. (1999). *Discourse in late modernity.* Edinburgh, UK: Edinburgh University Press.

Chow, G. (1961). Tests of equality between sets of regression coefficients in linear regression models. *Econometrica, 28*(3), 591–605.

Chrisman, N. (1997). *Exploring geographic information systems.* New York: Wiley.

Christians, C. (2000). Ethics and politics in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (2nd ed., pp. 133–155). Thousand Oaks, CA: Sage.

Cicerelli, V. G., & Associates. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development* (2 vols.). Athens: Ohio University and Westinghouse Learning Corp.

Cicourel, A. V. (1974). *Theory and method in a critique of Argentine fertility.* New York: John Wiley.

Clark, J. A., & Mishler, E. G. (1992). Attending to patients' stories: Reframing the clinical task. *Sociology of Health and Illness, 14,* 344–370.

Clarke, H. D., Norpoth, H., & Whiteley, P. F. (1998). It's about time: Modeling political and social dynamics. In E. Scarborough & E. Tanenbaum (Ed.), *Research strategies in the social sciences* (pp. 127–155). Oxford, UK: Oxford University Press.

Clarke, K. A. (2001). Testing nonnested models of international relations: Reevaluating realism. *American Journal of Political Science, 45,* 724–744.

Clayman, S. E., & Maynard, D. (1995). Ethnomethodology and conversation analysis. In P. ten Have & G. Psathas (Eds.), *Situated order: Studies in the social organization of talk and embodied activities* (pp. 1–30). Washington, DC: University Press of America.

Clayman, S., & Heritage, J. (2002). *The news interview: Journalists and public figures on the air.* Cambridge, UK: Cambridge University Press.

Cleary, T. A., & Linn, R. L. (1969). Error of measurement and the power of a statistical test. *The British Journal of Mathematical and Statistical Psychology, 22*(1), 49–55.

Clegg, C. W., & Walsh, S. (1998). Soft systems analysis. In G. Symon & C. M. Cassell (Eds.), *Qualitative methods and analysis in organisational research: A practical guide.* London: Sage.

Cleveland, W. S. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74,* 829–836.

Cleveland, W. S. (1993). *Visualizing data.* Summit, NJ: Hobart.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83,* 596–610.

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture: The poetics and politics of ethnography.* Berkeley: University of California Press.

Clogg, C. C. (1978). Adjustment of rates using multiplicative models. *Demography, 15,* 523–539.

Clogg, C. C. (1982). Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association, 77,* 803–815.

Clogg, C. C. (1982). Using association models in sociological research: Some examples. *American Journal of Sociology, 88,* 114–134.

Clogg, C. C., & Eliason, S. R. (1988). A flexible procedure for adjusting rates and proportions, including statistical methods for group comparisons. *American Sociological Review, 53,* 267–283.

Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical models for ordinal variables.* Thousand Oaks, CA: Sage.

Clogg, C. C., Shockey, J. W., & Eliason, S. R. (1990). A general statistical framework for adjustment of rates. *Sociological Methods & Research, 19,* 156–195.

Clough, T. P. (1998). *The end(s) of ethnography: From realism to social criticism* (2nd ed.). New York: Peter Lang.

Cobb, G. W. (1998). *Introduction to design and analysis of experiments.* New York: Springer-Verlag.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika, 37,* 256–266.

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics, 13*(3), 261–281.

Cochran, W. G. (1965). The planning of observational studies in human populations. *Journal of the Royal Statistical Society, Series A, 128,* 134–155.

Cochran, W. G. (1953). *Sampling techniques.* New York: Wiley.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley.

Cochrane, D., & Orcutt, G. H. (1949). Application of least squares relationships containing autocorrelated error terms. *Journal of the American Statistical Association, 44,* 32–61.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies.* Thousand Oaks, CA: Sage.

Cohen, A., Doveh, E., & Eick, U. (2001). Statistical properties of the $r_{WG(J)}$ index of agreement. *Psychological Methods, 6,* 297–310.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Cohen, J. (1978). Partialed products are interactions; partialed powers are curve components. *Psychological Bulletin, 85,* 858–866.

Cohen, J. (1988). *Statistical power analysis in the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (2001). Smallpox vaccinations: How much protection remains? *Science, 294,* p. 985.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cole, M. (1996). *Cultural psychology: A once and future discipline.* Cambridge, MA: Belknap Press of Harvard University Press.

Coleman, C., & Moynihan, J. (1996). *Understanding crime data: Haunted by the dark figure.* Buckingham, UK: Open University Press.

Collett, D. (1991). *Modelling binary data.* New York: Chapman and Hall.

Collican, H. (1999). *Research methods and statistics in psychology.* London: Hodder & Staughton.

Collier, A. (1994). *Critical realism.* London: Verso.

Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice.* London: Sage.

Collins, P. H. (1997). Comment on Heckman's "Truth and method: Feminist standpoint theory revisited": Where's the power? *Signs, 22*(21), 375–381.

Collins, R. (1995). Prediction in macrosociology: The case of the Soviet collapse. *American Journal of Sociology, 100*(6), 1552–1593.

Colson, F. (1977). *Sociale indicatoren van enkele aspecten van bevolkingsgroei.* Doctoral dissertation, Katholieke Universiteit Leuven, Department of Sociology, Leuven.

Comstock, D. E. (1994). A method for critical research. In M. Martin & L. C. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 625–639). Cambridge: MIT Press.

Confidentiality and Data Access Committee. (1999). *Checklist on disclosure potential of proposed data releases.* Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office. Retrieved from www.fcsm.gov/committees/cdac/checklist_799.doc

Conger, A. J. (1974). Revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement, 34,* 35–46.

Connell, R. W. (2002). *Gender.* Cambridge, UK: Polity.

Conover, W. J. (1980). *Practical nonparametric statistics.* New York: Wiley.

Conover, W. J. (1998). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

Conte, R., & Dellarocas, C. (2001). Social order in info societies: An old challenge for innovation. In R. Conte & C. Dellarocas (Eds.), *Social order in multi-agent systems* (pp. 1–16). Boston: Kluwer Academic.

Converse, J. (1987). *Survey research in the United States.* Berkeley: University of California Press.

Converse, J., & Presser, S. (1986). *Survey questions.* Beverly Hills, CA: Sage.

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. Apter (Ed.), *Ideology and discontent* (pp. 206–261). New York: Free Press.

Converse, P. E., & Markus, G. B. (1979). Plus ca change . . . : The new CPS Election Study Panel. *American Political Science Review, 73,* 32–49.

Cook, D. (1996). On the interpretation of regression plots. *Journal of the American Statistical Association, 91,* 983–992.

Cook, D. R., & Weisberg, S. (1999). *Applied regression including computing and graphics.* New York: Wiley.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression.* New York: Chapman and Hall.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally College Publishing.

Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research.* Washington, DC: Brookings Institution.

Cook, T. D., Campbell, D. T., & Peracchio, L. (1990). Quasi-experimentation. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organisational psychology* (2nd ed., Vol. 1, pp. 491–576). Chicago: Rand McNally.

Cooke, B., & Kothari, U. (Eds.). (2001). *Participation: The new tyranny?* London: Zed.

Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning, 24,* 119–128.

Cooley, W., & Lohnes, P. (1971). *Multivariate data analysis.* New York: Wiley.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Copi, I. M., & Cohen, C. (1990). *Introduction to logic* (8th ed.). New York: Macmillan.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730.

Cormack, R. (2001). Population size estimation and capture-recapture methods. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 17). Amsterdam: Elsevier.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to algorithms* (2nd ed.). Cambridge: MIT Press.

Correll, S. (1995). The ethnography of an electronic bar: The lesbian cafe. *Journal of Contemporary Ethnography, 24*(3), 270–298.

Cortazzi, M. (2001). Narrative analysis in ethnography. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography.* London: Sage.

Cortazzi, M., Jin, L., Wall, D., & Cavendish, S. (2001). Sharing learning through narrative communication. *International Journal of Language and Communication Disorders, 36,* 252–257.

Corter, J. E. (1996). *Tree models of similarity and association* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–112). Thousand Oaks, CA: Sage.

Corti, L. (1993). Using diaries in social research (*Social Research Update,* Iss. 2). Guildford, UK: University of Surrey, Department of Sociology.

Corti, L., Foster, J., & Thompson, P. (1995). *Archiving qualitative research data* (Social Research Update No. 10). Surrey, UK: Department of Sociology, University of Surrey.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98–104.

Costantino, G., Malgady, R. G., & Vazquez, C. (1981). A comparison of the Murray–TAT and a new Thematic

Apperception Test for urban Hispanic children. *Hispanic Journal of Behavior Sciences, 3,* 291–300.

Costigan, P., & Thomson, K. (1992). Issues in the design of CAPI questionnaires for complex surveys. In A. Westlake, R. Banks, C. Payne, & T. Orchard (Eds.), *Survey and statistical computing* (pp. 47–156). London: North Holland.

Costner, H. L. (1965). Criteria for measures of association. *American Sociological Review, 30,* 341–353.

Couper, M. P. (2000). WEB surveys: A review of issues and approaches. *Public Opinion Quarterly, 64,* 464–494.

Couper, M. P., Baker, R. P., Bethlehem, J., Clark, C. Z. F., Martin, J., Nichols, W. L., et al. (Eds.). (1998). *Computer assisted survey information collection.* New York: John Wiley.

Courant, Richard. (1988). *Differential and integral calculus* (2 vols.) (E. J. McShane, Trans.). New York: Wiley. (Originally published in 1934)

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement, 61,* 229–248.

Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (pp. 245–266). Dordrecht, The Netherlands: Kluwer Academic.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions in Evaluation, 80,* 5–23.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, B, 34,* 187–220.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data.* New York: Chapman and Hall.

Cox, R. T. (1990). Probability, frequency, and reasonable expectation. In G. Shafer & J. Pearl (Eds.), *Readings in uncertain reasoning* (pp. 353–365). New York: Morgan Kaufmann. (Original work published 1946)

Coxon, A. P. M. (1982). *The user's guide to multidimensional scaling.* London: Heinemann Educational.

Coxon, A. P. M. (1999). *Between the sheets: Sexual diaries and gay men's sex in the era of AIDS.* London: Cassell.

Coxon, A. P. M. (1999). *Sorting data: Collection and analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–127). Thousand Oaks, CA: Sage.

Cramer, J. S. (1986). *Econometric applications of maximum likelihood methods.* Cambridge, UK: Cambridge University Press.

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Crapanzano, V. (1980). *Tuhami: Portrait of a Moroccan.* Chicago: University of Chicago Press.

Crenshaw, K., Gotanda, N., Peller, G., & Thomas, K. (Eds.). (1995). *Critical race theory.* New York: The New York Press.

Cressey, D. (1950). The criminal violation of financial trust. *American Sociological Review, 15,* 738–743.

Cressey, D. (1953). *Other people's money.* Glencoe, IL: Free Press.

Cressie, N. (1993). *Statistics for spatial data.* New York: Wiley.

Cressie, N., & Read, T. (1984). Multinomial goodness of tests. *Journal of Royal Statistical Society Series B, 46,* 440–464.

Creswell, J. W. (1995). *Research design: Quantitative and qualitative approaches.* Thousand Oaks, CA: Sage.

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory.* New York: Harcourt Brace Jovanovich.

Croll, P. (1986). *Systematic classroom observation.* Lewes, UK: Falmer.

Cromwell, J. B., Labys, W., & Terraza, M. (1994). *Univariate tests for time series models.* Thousand Oaks, CA: Sage.

Cronbach, L. J. (1943). On estimates of test reliability. *Journal of Educational Psychology, 34,* 485–494.

Cronbach, L. J. (1946). A case study of the split-half reliability coefficient. *Journal of Educational Psychology, 37,* 473–480.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52,* 177–193.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: HarperCollins.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles.* New York: Wiley.

Croon, M. A., Bergsma, W., & Hagenaars, J. A. (2000). Analyzing change in categorical variables by generalized log-linear models. *Sociological Methods & Research, 29,* 195–229.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process.* London: Sage.

Crowne, D., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence.* New York: Wiley.

Cruse, D. A. (1986). *Lexical semantics.* Cambridge, UK: Cambridge University Press.

Cullen, C. G. (1990). *Matrices and linear transformations* (2nd ed.). New York: Dover.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly, 64,* 413–428.

Czaja, R., Blair, J., & Sebestile, J. P. (1982). Respondent selection in a telephone survey: Comparison of three techniques. *Journal of Marketing Research, 21,* 381–385.

Czyzewski, M. (1994). Reflexivity of actors versus reflexivity of accounts. *Theory, Culture and Society, 11,* 161–168.

D'Andrade, R. (1995). *The development of cognitive anthropology.* Cambridge, UK: Cambridge University Press.

Dale, A., & Marsh, C. (Eds.). (1993). *The 1991 census user's guide.* London: HMSO.

Dale, A., Arber, S., & Procter, P. (1988). *Doing secondary analysis*. London: Unwin Hyman.

Dale, A., Fieldhouse, E., & Holdsworth, C. (2000). *Analyzing census microdata*. London: Arnold.

Dandridge, T. C., Mitroff, I., & Joyce, W. F. (1980). Organizational symbolism: A topic to expand organizational analysis. *Academy of Management Review, 5,* 77–82.

Daniel, W. W. (1993). *Collecting sensitive data by randomized response: An annotated bibliography* (2nd ed.). Atlanta: Georgia State University Business Press.

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika, 35,* 129–135.

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.

Darnell, A. C. (1995). *A dictionary of econometrics*. Cheltenham, UK: Edward Elgar.

Darroch, J. N., Lauritzen, S. L., & Speed, T. P. (1980). Markov fields and log-linear models. *Annals of Mathematical Statistics, 43,* 1470–1480.

David, H. A., & Moeschberg, M. L. (1978). *The theory of competing risks* (Griffin's Statistical Monograph #39). New York: Macmillan.

David, P. (1985). Clio and the economics of QWERTY. *American Economic Review, 75,* 332–337.

Davidson, J., Hendry, D. F., Srba, F., & Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal, 88,* 661–692.

Davidson, R., & MacKinnon J. G. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.

Davis, J. A. (1985). *The logic of causal order* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–055). Beverly Hills, CA: Sage.

Davis, J. A., & Smith, T. W. (1992). *The NORC General Social Survey: A user's guide*. Newbury Park, CA: Sage.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.

Day, N. E. (1969). Estimating the components of a mixture of two normal distributions. *Biometrika, 56,* 463–474.

De Beauvoir, S. (1970). *The second sex*. New York: Alfred A. Knopf. (Original work published 1949)

de Leeuw, J., & Kreft, I. G. G. (Eds.). (in press). *Handbook of quantitative multilevel analysis*. Dordrecht, The Netherlands: Kluwer Academic.

de Leeuw, J., & van der Heijden, P. G. M. (1988). The analysis of time-budgets with a latent time-budget model. In E. Diday (Ed.), *Data analysis and informatics* (Vol. 5, pp. 159–166). Amsterdam: North-Holland.

de Saussure, F. (1979). *Cours de linguistique générale* (T. de Mauro, Ed.). Paris: Payot.

de Vaus, D. (Ed.). (2002). *Social surveys* (4 vols.). London: Sage.

de Vaus, D. A. (1995). *Surveys in social research* (4th ed.). Sydney, Australia: Allen & Unwin.

de Vaus, D. A. (2001). *Research design in social research*. London: Sage.

Deacon, D., Bryman, A., & Fenton, N. (1998). Collision or collusion? A discussion of the unplanned triangulation of quantitative and qualitative research methods. *International Journal of Social Research Methodology, 1,* 47–63.

Deacon, D., Pickering, M., Golding, P., & Murdock, G. (1999). *Researching communications*. London: Arnold.

DeBoef, S., & Granato, J. (2000). Testing for cointegrating relationships with near-integrated data. *Political Analysis, 8,* 99–117.

Deckner, D. F., Adamson, L. B., & Bakeman, R. (2003). Rhythm in mother-infant interactions. *Infancy, 4,* 201–217.

DeGroot M. H., & Schervish, M. J. (2002). *Probability and statistics* (3rd ed.). Reading, MA: Addison-Wesley.

Dehejia, R. H., & Wahba, S. (1999). Causal effects in non-experimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association, 94,* 1053–1062.

Del Monte, K. (2000). Partners in inquiry: Ethical challenges in team research. *International Social Science Review, 75,* 3–14.

Delanty, G. (1997). *Social science: Beyond constructivism and realism*. Buckingham, UK: Open University Press.

DeMaris, A. (1992). *Logit modeling*. Newbury Park, CA: Sage.

Deming, W. E. (1997). *Statistical adjustment of data*. New York. Dover.

DeNavas-Walt, C., & Cleveland, R. W. (2002). *Money income in the United States* (United States Census Bureau, Current Population Reports, P60–218). Washington, DC: Government Printing Office. Available: www.census.gov/prod/2002pubs/p60–218.pdf

Denzin, N. (1983). Interpretive interactionism. In G. Morgan (Ed.), *Beyond method: Strategies for social research* (pp. 128–142). Beverly Hills, CA: Sage.

Denzin, N. (1992). *Symbolic interactionism: The politics of interpretation*. Oxford, UK: Blackwell.

Denzin, N. K. (1970). *The research act in sociology*. Chicago: Aldine.

Denzin, N. K. (1989). *Interpretive biography*. Newbury Park, CA: Sage.

Dezin, N. K., (1989). *Interpretive interactionism*. Newbury Park, CA: Sage.

Denzin, N. K., (1997). *Interpretive enthnography*. Thousand Oaks, CA: Sage.

Denzin, N. K., & Lincoln, Y. S. (2000). The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 1–28). Thousand Oaks, CA: Sage.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Denzin, N. K. & Lincoln, Y. S. (2000). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 1–28). Thousand Oaks, CA: Sage.

Derrida, J. (1976). *Of grammatology* (G. Spivak, Trans.). Baltimore, MD: Johns Hopkins University Press.

Derrida, J. (1995). *The gift of death.* Chicago: The University of Chicago Press.

Derrida, J., & Ferraris, F. (2001). *A taste for the secret.* Cambridge, UK: Polity.

DeVellis, R. F. (1991). *Scale development: Theory and applications.* Newbury Park, CA: Sage.

Devore, J. L. (1991). *Probability and statistics for engineering and the sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Dexter, L. A. (1970). *Elite and specialized interviewing.* Evanston, IL: Northwest University Press.

Dey, I. (1993). *Qualitative data analysis.* London: Routledge Kegan Paul.

Diamond, S. S. (2000). Reference guide on survey research, in *Reference manual on scientific evidence.* (2nd ed., pp. 229–276). Washington, D. C.: Federal Judicial Center.

Dickens, P. (2000). *Social Darwinism.* Buckingham, UK: Open University Press.

Diebold, F. X. (2001). *Elements of forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review, 96,* 84–100.

Diesing, P. (1991). *How does social science work?* Pittsburgh, PA: University of Pittsburgh Press.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41,* 417–440.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method.* New York: Wiley.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method.* New York: Wiley.

Dilthey, W. (1985). *Poetry and experience: Selected works* (Vol. 5). Princeton, NJ: Princeton University Press.

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24,* 285–290.

DiPrete, T. A., & Forristal, J. D. (1994). Multilevel models: Methods and substance. *Annual Review of Sociology, 20,* 331–357.

Dixon-Woods, M., Fitzpatrick, R., & Roberts, K. (2001). Including qualitative research in systematic reviews: Problems and opportunities. *Journal of Evaluation in Clinical Practice, 7,* 125–133.

Doksum, K. A., & Sievers, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika, 63,* 421–434.

Doreian, P., Batagelj, V., & Ferligoj, A. (1994). Partitioning networks based on generalized concepts of equivalence. *Journal of Mathematical Sociology, 19,* 1–27.

Douglas, J. D. (1985). *Creative interviewing.* Beverly Hills, CA: Sage.

Douglass, B., & Moustakas, C. (1985). Heuristic inquiry: The internal search to know. *Journal of Humanistic Psychology, 25,* 39–55.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 175–197). Malden, MA: Blackwell.

Downs, A. (1957). *An economic theory of democracy.* New York: HarperCollins.

Doyle, P., Martin, B., & Moore, J. (2000). *Improving income measurement.* The Survey of Income Program Participation (SIPP) Methods Panel. Washington, DC: U.S. Bureau of the Census.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: Wiley.

Dreher, A. U. (2000). *Foundations for conceptual research in psychoanalysis* (Psychoanalytic Monograph 4). London: Karnac.

Drew, P., & Heritage, J. (Eds.). (1992). *Talk at work: Interaction in institutional settings.* Cambridge, UK: Cambridge University Press.

Dryzek, J. S., & Holmes, L. T. (2002). *Post-communist democratization.* Cambridge, UK: Cambridge University Press.

Dubin, J. A., & Rivers, D. (1989). Selection bias in linear regression, logit and probit models. *Sociological Methods & Research, 18,* 360–390.

Duda, R., Hart, P. E., & Stork, D. G. (2001). *Pattern classification.* New York: Wiley.

Dukes, R. L., Ullman, J. B., & Stein, J. A. (1995). An evaluation of D.A.R.E. (Drug Abuse Resistance Education), using a Solomon four-group design with latent variables. *Evaluation Review, 19*(4), 409–435.

Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology, 72,* 1–16.

Duncan, O. D. (1975). *Introduction to structural equation models.* New York: Academic Press.

Duncan, O. D., Haller, A., & Portes, A. (1968). Peer influence on aspiration: A reinterpretation. *American Journal of Sociology, 75,* 119–137.

Dupré, J. (2001). *Human nature and the limits of science.* Oxford, UK: Oxford University Press.

Dupré, J., & Cartwright, N. (1988). Probability and causality: Why Hume and indeterminism don't mix. *Nous, 22,* 521–536.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regressions I. *Biometrika, 37,* 409–428.

Durkheim, E. (1951). *Suicide.* Glencoe, IL: Free Press.

Durkheim, E. (1952). *Suicide.* London: Routledge Kegan Paul. (Original work published 1896)

Durkheim, E. (1964). *The rules of scientific method.* Glencoe, IL: Free Press.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 276–284.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good but . . . A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110,* 109–128.

Easterby-Smith, M., Thorpe, R., & Holman, D. (1996). Using repertory grids in management. *Journal of European Industrial Training, 20,* 1–30.

Easterby-Smith, M., Thorpe, R., & Lowe, A. (2002). *Management research: An introduction* (2nd ed.). London: Sage.

Eatwell, J., Milgate, M., & Newman, P. (1987). *The new Palgrave: A dictionary of economics.* London: Macmillan.

Eckstein, H. (1975). Case study and theory in political science. In F. I. Greenstein & N. W. Polsby (Eds.), *Handbook of political science: Vol. 7, Strategies of inquiry* (pp. 79–137). Reading, MA: Addison-Wesley.

Eco, U. (1972). Towards a semiotic inquiry into the television message. In *Working papers in cultural studies* (Vol. 3, pp. 103–121). Birmingham: Centre for Contemporary Cultural Studies. (Original work published 1965)

Eco, U. (1976). *A theory of semiotics.* Bloomington: Indiana University Press.

Eco, U. (1981). *The role of the reader: Explorations in the semiotics of texts.* London: Hutchinson.

Economic and Social Research Council Research Centre on Micro-Social Change. (2001, February 28). *British Household Panel Survey* [computer file] (Study Number 4340). Colchester, UK: The Data Archive [distributor].

Eden, D. (1990). *Pygmalion in management: Productivity as a self-fulfilling prophecy.* Lexington, MA: D. C. Heath.

Edwards, A. (1948). On Guttman's scale analysis. *Educational and Psychological Measurement, 8,* 313–318.

Edwards, A. L. (1957). *Techniques of attitude construction.* New York: Appleton- Century-Crofts.

Edwards, A. W. F. (1972). *Likelihood: An account of the statistical concept of likelihood and its application to scientific inference.* Cambridge, UK: Cambridge University Press.

Edwards, D. (1997). *Discourse and cognition.* London: Sage.

Edwards, D. (2000). *Introduction to graphical modelling* (2nd ed.). New York: Springer-Verlag.

Edwards, D., & Potter, J. (1992). *Discursive psychology.* London: Sage.

Edwards, D., Ashmore, M., & Potter, J. (1995). Death and furniture: The rhetoric, politics, and theology of bottom line arguments against relativism. *History of the Human Sciences, 8,* 25–49.

Edwards, W. S., Winn, D. M., Kurlantzick, V. et al. (1994). *Evaluation of National Health Interview Survey diagnostic reporting.* National Center for Health Statistics. Vital Health Stat 2(120).

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7,* 1–26.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans.* Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician, 40,* 1–5.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Egger, M., Davey Smith, G., & Altman, D. G. (2001). *Systematic reviews in health care: Meta-analysis in context* (2nd ed.). London: BMJ Books.

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634.

Eibl-Eibesfeldt, I. (1989). *Human ethology.* Hawthorne, NY: Aldine de Gruyter.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14,* 532–550.

Eisenhart, M. (1998). On the subject of interpretive reviews. *Review of Educational Research, 68*(4), 391–399.

Eisner, E. W. (1997). The new frontier in qualitative research methodology. *Qualitative Inquiry, 3,* 259–273.

Elder, G., Pavalko, E. K., & Clipp, E. C. (1993). *Working with archival lives: Studying lives* (Sage University Papers on Quantitative Applications on the Social Sciences, 07–088). Newbury Park, CA: Sage.

Eliason, S. R. (1993). *Maximum likelihood estimation: Logic and practice.* Newbury Park, CA: Sage.

Ellis, B. D., & Lierse, C. (1994). Dispositional essentialism. *Australasian Journal of Philosophy, 72*(1), 27–45.

Ellis, C. (1995). *Final negotiations: A story of love, loss, and chronic illness.* Philadelphia: Temple University Press.

Ellis, C. (2003). *The ethnographic "I": A methodological novel on doing autoethnography.* Walnut Creek, CA: AltaMira.

Ellis, C. S. (1991). Sociological introspection and emotional experience. *Symbolic Interaction, 14,* 23–50.

Ellis, C., & Bochner, A. P. (2000). Autoethnography, personal narrative, reflexivity: Researcher as subject. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 733–768). Thousand Oaks, CA: Sage.

Ellis, C., & Bochner, A. P. (Eds.). (1996). *Composing ethnography: Alternative forms of qualitative writing.* Walnut Creek, CA: AltaMira.

Elmasri, R. A., & Navathe, S. B. (2001). *Fundamentals of database systems.* New York: Addison-Wesley.

Elster, J. (1989). *Nuts and bolts for the social sciences.* Cambridge, UK: Cambridge University Press.

Elster, J. (1989). *The cement of society: A study of social order.* Cambridge, UK: Cambridge University Press.

Elster, J. (Ed.). (1986). *Rational choice.* Oxford, UK: Basil Blackwell.

Ember, C. R., & Ember, M. (2001). *Cross-cultural research methods.* Walnut Creek, CA: AltaMira.

Embree, L., Behnke, E. A., Carr, D., Evans, J. C., Huertas-Jourda, J., Kockelmans, J. J., et al. (Eds.). (1996). *The encyclopedia of phenomenology.* Dordrecht, The Netherlands: Kluwer.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum.

Emerson, R. M., & Pollner, M. (1988). On the use of members' responses to researchers' accounts. *Human Organization, 47,* 189–198.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (1995). *Writing ethnographic fieldnotes.* Chicago: University of Chicago Press.

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology, 103*(4), 962–1023.

Emmison, M., & Smith, P. (2000). *Researching the visual.* London: Sage.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica, 55,* 251–276.

Engle, R. F., Granger, C. W. J. (1991). *Long run relationships: Readings in cointegration.* New York: Oxford University Press.

Engle, R. F., Hendry, D. F., & Richard, J. F. (1983). Exogeneity. *Econometrica, 51,* 277–304.

Epley, N., & Huff, C. (1998). Suspicion, affective response, and educational benefit of deception in psychology research. *Personality and Social Psychology Bulletin, 24,* 759–768.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up.* Cambridge: MIT Press.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge: MIT Press.

Erikson, E. H. (1975). *Life history and the historical moment.* New York: Norton.

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational mobility in three Western European societies. *British Journal of Sociology, 30.*

Erlang, A. K. (1935). *Fircifrede logaritmetavler og andre regnetavler til brug ved undervisning og i praksis.* København: G.E.C. Gads.

Ermarth, M. (1978). *Wilhelm Dilthey: The critique of historical reason.* Chicago: University of Chicago Press.

Escofier, B., & Pagès, J. (1988). *Analyses factorielles multiples* [Multiple factor analyses]. Paris: Dunod.

Essed, Ph., & Goldberg, T. D. (Eds.). (2002). *Race critical theories: Text and context.* Malden, MA: Blackwell.

Eubank, R. L. (1988). Quantiles. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 7, pp. 424–432). New York: Wiley.

European Social Survey Central Co-ordinating Team. (2001). *European Social Survey (ESS): Specification for participating countries.* London: Author.

Eurostat. (1998). *Labour force survey: Methods and definitions* (1998 ed.). Luxembourg: Author.

Evans, J. L. (2000). *Early childhood counts.* Washington, DC: World Bank.

Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions* (3rd ed.). New York: Wiley.

Everitt, B. S. (1992). *The analysis of contingency tables.* London: Chapman & Hall.

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). Oxford, UK: Oxford University Press.

Exner, J. E. (2002). *The Rorschach: A comprehensive system* (Vol. 1). New York: John Wiley.

Fairclough, N. (2000). *New labour, new language?* London: Routledge Kegan Paul.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. In T. van Dijk (Ed.), *Discourse as social interaction.* Thousand Oaks, CA: Sage.

Fals-Borda, O., & Anisur Rahman, M. (Eds.). (1991). *Action and knowledge: Breaking the monopoly with participatory action-research.* New York: Apex.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications.* London: Chapman & Hall.

Faugier, J., & Sargeant, M. (1997). Sampling hard to reach populations. *Journal of Advanced Nursing, 26,* 790–797.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54,* 297–327.

Featherman, D. L., & Hauser, R. M. (1978). *Opportunity and change.* New York: Academic Press.

Featherstone, M. (1988). In pursuit of the postmodernism: An introduction. *Theory, Culture & Society, 5,* 195–215.

Fechner, G. T. (1860). *Elemente der Psychophysik.* Leipzig: Breitkopf & Härtel.

Fechner, G. T. (1877). *In sachen der Psychophysik.* Leipzig: Breitkopf & Härtel.

Federal Committee on Statistical Methodology. (1994). *Report on statistical disclosure limitation methodology* (Statistical Policy Working Paper #22). Prepared by the Subcommittee on Disclosure Limitation Methodology. Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office. Retrieved from www.fcsm.gov/working-papers/wp22.html

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). New York: Macmillan.

Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed.). New York: John Wiley.

Ferrell, J., & Hamm, M. S. (1998). True confessions: Crime, deviance, and field research. In J. Ferrell & M. Hamm (Eds.), *Ethnography at the edge: Crime, deviance and field research* (pp. 2–19). Boston: Northeastern University Press.

Fetterman, D. M. (1998). *Ethnography: Step by step* (2nd ed.). Thousand Oaks, CA: Sage.

Fetterman, D. M. (2002). Web surveys to digital movies: Technological tools of the trade. *Educational Researcher, 31*(6), 29–37.

Feyerabend, P. (1978). *Against method.* London: Verso.

Feynman, R. P., & Weinberg, S. (1987). *Elementary particles and the laws of physics: The 1986 Dirac Memorial Lectures.* Cambridge, UK: Cambridge University Press.

Field, A. (1998). A bluffer's guide to . . . sphericity. *British Psychological Society: Mathermatical, Statistical & Computing Newsletter, 6,* 13–22.

Fielding, J. (1993). Coding and managing data. In N. Gilbert (Ed.), *Researching social life* (pp. 218–238). Thousand Oaks, CA: Sage.

Fielding, N. G., & Lee, R. M. (1998). *Computer analysis and qualitative research.* London: Sage.

Fienberg, S. E. (1977). *The analysis of cross-classified categorical data.* Cambridge: MIT Press.

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: MIT Press.

Fillmore, C. (1975). An alternative to checklist theories of meaning. In C. Cogen, H. Thompson, G. Thurgood, K. Whistler, & J. Wright (Eds.), *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society* (pp. 123–131). Berkeley, CA: Berkeley Linguistics Society.

Filmer, P. (1998). Analysing literary texts. In C. Seale (Ed.), *Researching society and culture.* London: Sage.

Fink, A., & Kosecoff, J. (1988). *How to conduct surveys: A step-by-step guide.* Newbury Park, CA: Sage.

Firebaugh, G. (2001). Ecological fallacy. *International encyclopedia for the social and behavioral sciences* (Vol. 6, pp. 4023–4026). Oxford, UK: Pergamon.

Fisher, C. B., & Wallace, S. A. (2000). Through the community looking glass: Reevaluating the ethical and policy implications of research on adolescent risk and psychopathology. *Ethics & Behavior, 10*(2), 99–118.

Fisher, R. A. (1925). *Statistical methods for research workers* (1st ed.). Edinburgh, UK: Oliver & Boyd.

Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society, 22,* 700–725.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London, Series A, 144,* 285–307.

Fisher, R. A. (1935). *The design of experiments* (1st ed.). Edinburgh, UK: Oliver and Boyd.

Fisher, R. A. (1971). *Design of experiments.* New York: Hafner Press. (Original work published 1935).

Fisher, R. A., & Yates, F. (1934). The six by six Latin squares. *Proceedings of the Cambridge Philosophical Society, 30,* 492–507.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327–358.

Flick, U. (2002). *An introduction to qualitative research* (2nd ed.). London: Sage.

Foley, D. E. (1990). *Learning capitalist culture: Deep in the heart of Texas.* Philadelphia: University of Pennsylvania Press.

Fontana, A., & Frey, J. H. (1994). Interviewing: The art of science. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 361–376). Thousand Oaks, CA: Sage.

Fontana, A., & Frey, J. H. (2000). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 645–672). Thousand Oaks, CA: Sage.

Forrester, J. W. (1961). *Industrial dynamics.* Cambridge: MIT Press.

Forster, E., & McCleery, A. (1999). Computer assisted personal interviewing: A method of capturing sensitive information. *IASSIST Quarterly, 23*(2), 26–38.

Foster, J., & Sheppard, J. (1995). *British archives: A guide to archival resources in the United Kingdom* (3rd ed.). London: Macmillan.

Foucault, M. (1985). *Power/knowledge: Selected interviews and other writings: 1972–1977* (C. Gordon, Ed.). New York: Pantheon.

Foucault, M. (1991). Questions of method. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality* (pp. 73–86). London: Harvester Wheatsheaf.

Foucault, M. (1991). *Remarks on Marx: Conversations with Duccio Trombadori* (R. J. Goldstein & J. Cascaito, Trans.). New York: Semiotext(e).

Foucault, M. (1995). *Discipline and punish: The birth of the prison.* New York: Vintage Books.

Foucault, M. (1998). *The will to knowledge: The history of sexuality* (Vol. 1). London: Penguin.

Foucault, M. (2002). *Archeology of knowledge.* London: Routledge.

Fowler, F. J., Jr. (1988). *Survey research methods* (Rev. ed.). Newbury Park, CA: Sage.

Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation.* Thousand Oaks, CA: Sage.

Fowler, F. J., Jr. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Fowler, F. J., Jr., & Mangione, T. W. (1990). *Standardized survey interviewing.* Newbury Park, CA: Sage.

Fowler, R. L. (1985). Point estimates and confidence intervals in measures of association. *Psychological Bulletin, 98,* 160–165.

Fowler, R. L. (1985). Testing for substantive significance in applied research by specifying non-zero effect null-hypotheses. *Journal of Applied Psychology, 70,* 215–218.

Fox, J. (1991). *Regression diagnostics* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–079). Newbury Park, CA: Sage.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Thousand Oaks, CA: Sage.

Fox, J. (2000). *Nonparametric simple regression: Smoothing scatterplots.* Thousand Oaks, CA: Sage.

Fox, J. (2000). *Multiple and generalized nonparametric regression.* Thousand Oaks, CA: Sage.

Fox, J. (2002). *An R and S-PLUS companion to applied regression.* Thousand Oaks, CA: Sage.

Fox, J. A., & Tracy, P. E. (1986). *Randomized response: A method for sensitive surveys.* Beverly Hills, CA: Sage.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association, 87,* 178–183.

Fox, W. (1998). *Social statistics* (3rd ed.). Bellevue, WA: MicroCase.

Frank, I. E., & Friedman, J. H. (1993). A statistical view of chemometrics regression tools. *Technometrics, 35,* 109–148.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association, 81,* 832–842.

Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research methods in the social sciences* (5th ed.). New York: St. Martin's.

Frankfort-Nachmias, C., & Nachmias, D. (2000). *Research methods in the social sciences* (6th ed.). New York: Worth.

Fransella, F., & Bannister, D. (1977). *A manual for repertory grid technique.* London: Academic Press.

Franses, P. H. (1998). *Time series models for business and economic forecasting.* Cambridge, UK: Cambridge University Press.

Franzese, R. (2002). *Macroeconomic policy of developed democracies.* Cambridge, UK: Cambridge University Press.

Franzosi, R. (2003). *From words to numbers.* Cambridge, UK: Cambridge University Press.

Fraser, N. (1997). *Justice interruptus: Critical reflections on the 'postsocialist' condition.* London: Routledge Kegan Paul.

Frederick, B. N. (1999). Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 305–318). Stamford, CT: JAI.

Frederick, R. I., & Crosby, R. D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior, 24,* 59–82.

Freedman, D. A. (2001). Ecological inference and the ecological fallacy. *International encyclopedia for the social and behavioral sciences* (Vol. 6, pp. 4027–4030). Oxford, UK: Pergamon.

Freedman, D. A., & Wachter, K. W. (2001). *On the likelihood of improving the accuracy of the census through statistical adjustment* (Tech. Rep. 612). Berkeley: University of California, Department of Statistics.

Freedman, D. A., Klein, S. P., Ostland, M., & Roberts, M. R. (1998). Review of "A solution to the ecological inference problem." *Journal of the American Statistical Association, 93,* 1518–1522. (Discussion appears in Vol. 94, pp. 352–357)

Freedman, D. A., Pisani, R., & Purves, R. A. (1998). *Statistics.* 3rd ed. New York: W. W. Norton, Inc.

Freeman, D. (1983). *Margaret Mead and Samoa.* Cambridge, MA: Harvard University Press.

Freeman, J. F. (1983). Granger causality and the time series analysis of political relationships. *American Journal of Political Science, 27,* 325–355.

Frey, J. H. (1989). *Survey research by telephone* (2nd ed.). Newbury Park, CA: Sage.

Frey, J. H., & Fontana, A. (1991). The group interview in social research. *The Social Science Journal, 28,* 175–187.

Frey, J. H., & Oishi, S. M. (1995). *How to conduct interviews by telephone and in person.* Thousand Oaks, CA: Sage.

Fricker, R. D., & Schonlau, M. (2002). Advantages and disadvantages of Internet research surveys: Evidence from the literature. *Field Methods, 14*(4), 347–365.

Friedkin, N. (1998). *A structural theory of social influence.* New York: Cambridge University Press.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association, 32,* 675–701.

Friedman, M. (1962). The interpolation of time series by related series. *Journal of the American Statistical Association, 57,* 729–757.

Friedman, M. (1991). The re-evaluation of logical positivism. *Journal of Philosophy, 88*(10), 505–519.

Friedricks, R. (1970). *A sociology of sociology.* New York: Free Press.

Friendly, M. (1992). Mosaic displays for loglinear models. In American Statistical Association (Ed.), *Proceedings of the Statistical Graphics Section* (pp. 61–68). Alexandria, VA: American Statistical Association.

Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association, 89,* 190–200.

Friendly, M. (1998). Mosaic displays. In S. Kotz, C. Reed, & D. L. Banks (Eds.), *Encyclopedia of statistical sciences* (Vol. 2, pp. 411–416). New York: John Wiley.

Friendly, M. (1999). Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *Journal of Computational and Graphical Statistics, 8*(3), 373–395.

Friendly, M. (2000). *Visualizing categorical data.* Cary, NC: SAS Institute.

Friendly, M. (2002). A brief history of the mosaic display. *Journal of Computational and Graphical Statistics, 11*(1), 89–107.

Fu, V., Mare, R., & Winship, C. (2003). Sample selection bias models. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis.* London: Sage.

Fuller, D. (1999). Part of the action, or "going native"? Learning to cope with the politics of integration. *Area, 31*(3), 221–227.

Fuller, W. A. (1987). *Measurement error models.* New York: John Wiley.

Fuss, D. (1989). *Essentially speaking.* New York: Routledge.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika, 58,* 453–467.

Gadamer, H.-G. (1975). *Truth and method.* New York: Seabury.

Gadamer, H-G. (1989). *Truth and method* (rev. 2nd ed.). New York: Crossroad.

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin, 87,* 564–567.

Gallmeier, C. P. (1991). Leaving, revisiting, and staying in touch: Neglected issues in field research. In W. B. Shaffir & R. A. Stebbins (Eds.), *Fieldwork experience: Qualitative approaches to social research* (pp. 224–231). London: Sage.

Galton, F. (1886). Regression toward mediocrity in hereditary stature. *Journal of the Anthropological Institute, 15,* 246–263.

Galton, F. (1889) *Natural inheritance.* London: Macmillan.

Galton, M., Simon, B., & Croll, P. (1980). *Inside the primary classroom.* London: Routledge and Kegan Paul.

Galunic, D. C., & Eisenhardt, K. M. (2001). Architectural innovation and modular corporate forms. *Academy of Management Journal, 44*(6), 1229–1249.

Game, A. (1991). *Undoing the social: Towards a deconstructive sociology.* Buckingham, UK: Open University Press.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.

Garfinkel, H. (1974). On the origins of the term "ethnomethodology." In R. Turner (Ed.), *Ethnomethodology* (pp. 15–18). Harmondsworth, UK: Penguin.

Garfinkel, H. (2002). *Ethnomethodology's program: Working out Durkheim's aphorism.* Blue Ridge Summit, PA: Rowman and Littlefield.

Garman, M. (1990). *Psycholinguistics.* Cambridge, UK: Cambridge University Press.

Garson, G. D. (2002, June). *Statnotes: An online textbook* [Online]. Retrieved from http://www2.chass.ncsu.edu/garson/pa765/statnote.htm.

Gaventa, J. (1980). *Power and powerlessness: Quiescence and rebellion in an Appalachian valley.* Chicago: University of Chicago Press.

Gee, J. P. (1991). A linguistic approach to narrative. *Journal of Narrative and Life History, 1,* 15–39.

Geertz, C. (1973). *The interpretation of cultures.* New York: Basic Books.

Geertz, C. (1983). *Local knowledge.* New York: Basic Books.

Geertz, C. (1988). *Works and lives: The anthropologist as author.* Stanford, CA: Stanford University Press.

Geertz, C. (2000). *The interpretation of cultures: Selected essays.* New York: Basic Books.

Geladi, P., & Kowlaski, B. (1986). Partial least square regression: A tutorial. *Analytica Chemica Acta, 35,* 1–17.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85,* 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* London: Chapman & Hall.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* 721–741.

Gensler, H. J. (2002). *Introduction to logic.* London: Routledge Kegan Paul.

Gentle, J. (2003). *Computational statistics.* New York: Springer-Verlag.

George, A. L. (1979). Case studies and theory development: The method of structured, focused comparison. In P. G. Lauren (Ed.), *Diplomacy: New approaches in history, theory, and policy* (pp. 43–68). New York: Free Press.

George, A. L., & McKeown, T. J. (1985). Case studies and theories of organizational decision making. *Advances in Information Processing in Organizations, 2,* 21–58.

George, A. L., & Smoke, R. (1974). *Deterrence in American foreign policy: Theory and practice.* New York: Columbia University Press.

Georges, R. A., & Jones, M. O. (1995). *Folkloristics: An introduction.* Bloomington: Indiana University Press.

Gephart, R. P. (1986). Deconstructing the defence for quantification in social science: A content analysis of journal articles on the parametric strategy. *Qualitative Sociology, 9,* 126–144.

Gephart, R. P. (1988). *Ethnostatistics: Qualitative foundations for quantitative research.* Newbury Park, CA: Sage.

Gephart, R. P. (1997). Hazardous measures: An interpretive textual analysis of quantitative sensemaking during crises. *Journal of Organizational Behavior, 18,* 583–622.

Gerber, A. S., & Green, D. P. (2000). The effects of personal canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *The American Political Science Review, 94,* 653–664.

Gergen, K. J. (1994) *Realities and relationships.* Cambridge, MA: Harvard University Press.

Gergen, K. J. (1999). *An invitation to social construction.* London: Sage.

Gerhardt, U. (1985). Erzähldaten und Hypothesenkonstruktion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 37,* 230–256.

Gerhardt, U. (1986). *Patientenkarrieren: Eine idealtypen-analytische Studie.* Frankfurt: Suhrkamp.

Gerhardt, U. (1994). The use of Weberian ideal-type methodology in qualitative data interpretation: An outline for ideal-type analysis. *BMS Bulletin de Methodologie Sociologique* (International Sociological Association, Research Committee 33), *45,* 76–126.

Gerhardt, U. (1999). *Herz und Handlungsrationalität: Eine idealtypen-analytische Studie.* Frankfurt: Suhrkamp.

Gershuny, J. (2000). *Changing times: Work and leisure in post-industrial societies.* Oxford, UK: Oxford University Press.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences.* San Francisco: W. H. Freeman.

Gibbons, J. D. (1988). Truncated data. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 9, p. 355). New York: Wiley.

Gibbons, J. D. (1993). *Nonparametric statistics: An introduction.* Newbury Park, CA: Sage.

Gibbons, J. D. (1997). *Nonparametric methods for quantitative analysis* (3rd ed.). Columbus, OH: American Sciences Press.

Gibbons, J. D., & Chakraborti, S. (1992). *Nonparametric statistical inference* (3rd ed.). New York: Marcel Dekker.

Gibson, N., Gibson, G., & Macaulay, A. C. (2001). Community-based research: Negotiating agendas and evaluating outcomes. In J. Morse, J. Swanson, & A. J. Kuzel (Eds.), *The nature of qualitative evidence* (pp. 160–182). Thousand Oaks, CA: Sage.

Giddens, A. (1976). *New rules of sociological method.* London: Hutchinson.

Giddens, A. (1979). *Central problems in social theory.* Berkeley: University of California Press.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Cambridge, UK: Polity.

Giele, J. Z., & Elder, G. H. (1998). *Methods of life course research: Qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Gifi, A. (1990). *Nonlinear multivariate analysis.* New York: Wiley.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669.

Gilbert, G. N., & Mulkay, M. (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse.* Cambridge, UK: Cambridge University Press.

Gilbert, N., & Troitzsch, K. G. (1999). *Simulation for the social scientist.* Milton Keynes, UK: Open University Press.

Gill, J. (2000). *Generalized linear models: A unified approach.* Thousand Oaks, CA: Sage.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach.* London: Chapman & Hall.

Gintis, H. (2000). *Game theory evolving.* Princeton, NJ: Princeton University Press.

Girden, E. R. (1992). *ANOVA: Repeated measures* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–084). Newbury Park, CA: Sage.

Glaser, B. (1992). *Emergence vs. forcing: Basics of grounded theory analysis.* Mill Valley, CA: Sociology Press.

Glaser, B. G. (1978). *Theoretical sensitivity.* Mill Valley, CA: Sociology Press.

Glaser, B. G. (2002). Grounded theory and gender relevance. *Health Care for Women International, 23,* 786–793.

Glaser, B. G., & Strauss, A. L. (1967). *Discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston: Allyn & Bacon.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and convariance. *Review of Educational Research, 42,* 237–288.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Research, 5,* 3–8.

Glass, G. V. (1977). Integrating findings. *Review of Research in Education, 5,* 351–379.

Glass, G. V., & Smith, M. L. (1978). Meta-analysis of research on the relationship of class size and achievement. *Educational Evaluation and Policy Analysis, 1,* 2–16.

Glenn, N. D. (2003). Distinguishing age, period, and cohort effects. In J. Mortimer & M. Shanahan (Eds.), *Handbook of the life course* (pp. 465–476). New York: Kluwer Academic/Plenum.

Gluck, S. B., & Patai, D. (Eds.). (1991). *Women's words: The feminist practice of oral history.* London: Routledge.

Glymour, C. N. (1980). *Theory and evidence.* Princeton, NJ: Princeton University Press.

Goffman, E. (1959). *The presentation of self in everyday life.* Garden City, NY: Doubleday.

Goffman, E. (1983). The interaction order. *American Sociological Review, 48,* 1–17.

Gold, R. L. (1958). Roles in sociological field observation. *Social Forces, 36,* 217–223.

Gold, R. L. (1969). Roles in sociological field observations. In G. J. McCall & J. L. Simmons (Eds.), *Issues in participant observation* (pp. 30–38). Reading, MA: Addison-Wesley.

Goldberger, A. S. (1964). *Econometric theory.* New York: John Wiley & Sons.

Goldmann, L. (1981). *Method in the sociology of literature* (W. Boelhower, Trans. & Ed.). Oxford, UK: Basil Blackwell.

Goldstein, H. (1987). *Multilevel models in educational and social research.* London: Griffin.

Goldstein, H. (1995). *Multilevel statistical models.* London: Edward Arnold.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Hodder Arnold.

Golub, G. H., & Van Loan, G. F. (1996). *Matrix computations* (3rd ed.). Baltimore: Johns Hopkins University Press.

Gomm, R., Hammersley, M., & Foster, P. (2000). Case study and generalisation. In R. Gomm, M. Hammersley, & P. Foster (Eds.), *Case study method: Key issues, key texts.* London: Sage.

Gomm, R., Hammersley, M., & Foster, P. (Eds.). (2000). *Case study method.* London: Sage.

Good, I. J. (1988). The interface between statistics and philosophy of science. *Statistical Science, 3,* 386–397.

Good, P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses.* New York: Springer.

Goode, W. J. (1978). *The celebration of heroes: Prestige as a control system.* Berkeley: University of California Press.

Goodenough, W. (1957). Cultural anthropology and linguistics. In P. L. Garvin (Ed.), *Report of the 7th Annual Roundtable on Linguistics and Language Study* (pp. 167–173). Washington, DC: Georgetown University Press.

Goodman, A., Johnson, P., & Webb, S. (1997). *Inequality in the UK.* Oxford, UK: Oxford University Press.

Goodman, L. (1953). Ecological regression and the behavior of individuals. *American Sociological Review, 18,* 663–664.

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association, 56,* 841–868.

Goodman, L. A. (1973). Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology, 78,* 1135–1191.

Goodman, L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika, 60,* 179–192.

Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable: I. A modified latent structure approach. *American Journal of Sociology, 79,* 1179–1259.

Goodman, L. A. (1978). *Analyzing qualitative/categorical data.* Cambridge, MA: Abt.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association, 74,* 537–552.

Goodman, L. A. (1984). *The analysis of cross-classified data having ordered categories.* Cambridge, MA: Harvard University Press.

Goodman, L. A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review, 54,* 243–309.

Goodman, L. A., & Hout, M. (1998). Understanding the Goodman-Hout approach to the analysis of differences in association and some related comments. In A. E. Raftery (Ed.), *Sociological methodology* (pp. 249–261). Washington, DC: American Sociological Association.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association, 49,* 732–764.

Goodman, S. N. (1999). Toward evidence-based medical statistics: 1. The *p* value fallacy. *Annals of Internal Medicine, 130,* 995–1004.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York: Academic Press.

Gopaul-McNicol, S.-A., & Armour-Thomas, E. (2002). *Assessment and culture: Psychological tests with minority populations.* San Diego: Academic Press.

Gottman, J. M. (1981). *Time-series analysis: A comprehensive introduction for social scientists.* Cambridge, UK: Cambridge University Press.

Gottschalk, L., Kluckhohn, C., & Angell, R. (1945). *The use of personal documents in history, anthropology, and sociology.* New York: Social Science Research Council.

Gouldner, A. V. (1973). *For sociology.* Harmondsworth, UK: Penguin.

Gouldner, A. W. (1962). Anti-minotaur: The myth of a value-free sociology. *Social Problems, 9,* 199–213.

Gower, J. C., & Hand, D. J. (1996). *Biplots.* London: Chapman & Hall.

Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification, 5,* 5–48.

Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling, 10,* 142–153.

Grammer, K., Fink, B., & Renninger, L. (2002). Dynamic systems and inferential information processing in human communication. *Neuroendocrinology Letters, 23*(Suppl. 4), 15–22.

Gramsci, A. (1971). *Selections from the prison notebooks.* London: Lawrence and Wishart.

Granger, C. W. J. (1964). *Spectral analysis of economic time series.* Princeton, NJ: Princeton University Press.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37,* 24–36.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric models specification. *Journal of Econometrics, 16,* 121–130.

Granger, C. W. J. (1983). *Co-integrated variables and error-correcting models.* Discussion Paper No. 1983–13, University of California, San Diego.

Granger, C. W. J. (1989). *Forecasting in business and economics.* Boston: Academic Press.

Granger, C. W. J., & Newbold, P. (1986). *Forecasting economic time series* (2nd ed.). New York: Academic Press.

Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences* (5th ed.). Belmont, CA: Wadsworth.

Gray, J., & Reuter, A. (1992). *Transaction processing: Concepts and techniques.* San Francisco: Morgan Kaufmann.

Green, D. P., & Gerber, A. S. (2002). Reclaiming the experimental tradition in political science. In I. Katznelson & H. V. Milner (Eds.), *Political science: State of the discipline* (3rd ed., pp. 805–832). New York: W.W. Norton.

Green, D. P., Gerber, A. S., & De Boef, S. L. (1999). Tracking opinion over time: A method for reducing sampling error. *Public Opinion Quarterly, 63,* 178–192.

Green, P. (1978). *Analyzing multivariate data.* Hinsdale, IL: Dryden.

Green, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River: Prentice Hall.

Greenacre, M. J. (1984). *Theory and method of correspondence analysis.* London: Academic Press.

Greenacre, M. J. (1993). *Correspondence analysis in practice.* London: Academic Press.

Greenacre, M. J., & Blasius, J. (1994). *Correspondence analysis in the social sciences.* London: Academic Press.

Greenbaum, T. L. (1998). *The handbook for focus group research.* London: Sage.

Greene, R. L. (2000). *The MMPI-2: An interpretive manual.* Boston: Allyn & Bacon.

Greene, W. H. (1993). *Econometric analysis* (2nd ed.). New York: Macmillan.

Greene, W. H. (1997). *Econometric analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Greenfield, P. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52*(10), 1115–1124.

Greenhouse, G. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 55,* 431–433.

Greenland, S. (2003). Quantifying biases in causal models. *Epidemiology, 14,* 300–306.

Greenland, S., & Brumback, B. A. (2002). An overview of relations among causal modelling methods. *International Journal of Epidemiology, 31,* 1030–1037.

Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology, 15,* 413–419.

Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science, 14,* 29–46.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3–25.

Greenwood, D., & Levin, M. (1998). *Introduction to action research: Social research for social change.* Thousand Oaks, CA: Sage.

Greimas, A. J., & Courtés, J. (1982). *Semiotics and language: An analytical dictionary.* Bloomington: Indiana University Press.

Griffin, L., Caplinger, C., Lively, K., Malcom, N. L., McDaniel, D., & Nelsen, C. (1997). Comparative-historical analysis and scientific inference: Disfranchisement in the U.S. South as a test case. *Historical Methods,* 30, 13–27.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica, 25,* 501–522.

Grinyer, A. (2002). The anonymity of research participants: Assumptions, ethics, and practicalities. *Social Research Update, 36,* 1–4.

Grofman, B., & Davidson, C. (1992). *Controversies in minority voting: The voting rights act in perspective.* Washington, DC: Brookings Institution.

Gross, D., & Harris, C. M. (1998). *Fundamentals of queueing theory* (3rd ed.). New York: Wiley.

Groves, E. R., & Ogburn, W. F. (1928). *American marriage and family relationships.* New York: Holt.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: Wiley.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys.* New York: Wiley.

Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (Eds.). (2001). *Survey nonresponse.* New York: Wiley.

Guba, E. G. (1978). *Toward a methodology of naturalistic inquiry in educational evaluation.* CSE Monograph Series in Evaluation, No. 8. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communications and Technology Journal, 29,* 75–92.

Guba, E. G., & Lincoln, Y. S. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Guba, E., & Lincoln, Y. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage.

Gubrium, J. F., & Buckholdt, D. R. (1979). Production of hard data in human service organizations. *Pacific Sociological Review, 22,* 115–136.

Gubrium, J. F., & Holstein, J. A. (1997). *The new language of qualitative method.* New York: Oxford University Press.

Gubrium, J. F., & Holstein, J. A. (2002). From the individual interview to the interview society. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 3–32). Thousand Oaks, CA: Sage.

Gujarati, D. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Gujarati, D. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

Gulliksen, H. (1950). *The theory of mental tests.* New York: Wiley.

Guttman, L. (1941). An outline for the statistical theory of prediction. In P. Horst (with collaboration of P. Wallin & L. Guttman) (Ed.), *The prediction of personal adjustment* (Bulletin 48, Supplementary Study B-1, pp. 253–318). New York: Social Science Research Council.

Guttman, L. (1944). A basis for scaling quantitative data. *American Sociological Review, 9,* 139–150.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika, 18,* 277–296.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19,*149–161.

Guttman, L. (1956). "Best possible" systematic estimates of communalities. *Psychometrika, 21,* 273–285.

Haack, S. (1993). *Evidence and inquiry: Toward a reconstruction of epistemology.* Oxford, UK: Blackwell.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica, 12*(Suppl.), preface, p. iii.

Haberman, S. J. (1979). *Analysis of qualitative data: Vol. 2. New developments.* New York: Academic Press.

Hacking, I. (1965). *The logic of statistical inference.* Cambridge, UK: Cambridge University Press.

Hacking, I. (2001). *An introduction to probability and inductive logic.* Cambridge, UK: Cambridge University Press.

Hagenaars, J. A. (1990). *Categorical longitudinal data: Log-linear analysis of panel, trend and cohort data.* Newbury Park, CA: Sage.

Hagenaars, J. A. (1993). *Loglinear models with latent variables.* Newbury Park, CA: Sage.

Hagenaars, J. A. (1998). Categorical causal modeling: Latent class analysis and directed log-linear models with latent variables. *Sociological Methods and Research, 26,* 436–486.

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis.* Cambridge, UK: Cambridge University Press.

Hagle, T. M., & Mitchell, G. E., II. (1992). Goodness-of-fit measures for probit and logit. *American Journal of Political Science, 36*, 762–784.

Haight, B. K., Michel, Y., & Hendrix, S. (1998). Life review: Preventing despair in newly relocated nursing home residents: Short- and long-term effects. *International Journal of Aging and Human Development, 47*(2), 119–142.

Hajek, J., & Sidák, Z. (1967). *Theory of rank tests*. New York: Academic Press.

Hakim, C. (1982). *Secondary analysis of social research*. London: George Allen and Unwin.

Haladyna, T. M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Lawrence Erlbaum.

Hald, A. (1952). *Statistical theory with engineering applications*. New York: John Wiley.

Hald, A. (1990). *A history of probability and statistics and their applications before 1750*. New York: John Wiley and Sons.

Halfpenny, P. (1982). *Positivism and sociology: Explaining social life*. London: Allen & Unwin.

Hall, E. T. (1974). *Handbook for proxemic research*. Washington, DC: Society for the Anthropology of Visual Communication.

Hall, S. (1974). The television discourse—Encoding and decoding. In *Education and culture* (Vol. 35, pp. 8–14). Paris: United Nations Educational, Social, and Cultural Organization.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hamilton, D. (1980). Some contrasting assumptions about case study research and survey analysis. In H. Simons (Ed.), *Towards a science of the singular: Essays about case study in educational research and evaluation* (pp. 78–92). Norwich, UK: Centre for Applied Research in Education University of East Anglia.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Hamilton, L. C. (1990). *Modern data analysis: A first course in applied statistics*. Pacific Grove, CA: Brooks/Cole.

Hamilton, L. C. (1992). *Regression with graphics: A second course in applied statistics*. Pacific Grove, CA: Brooks/Cole.

Hamlyn, D. W. (1967). Empiricism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (pp. 499–505). New York: Macmillan.

Hammersley, M. (1989). *The dilemma of qualitative method: Herbert Blumer and the Chicago school of sociology*. London: Routledge and Kegan Paul.

Hammersley, M. (1989). The problem of the concept: Herbert Blumer on the relationship between concepts and data. *Journal of Contemporary Ethnography, 18*, 133–160.

Hammersley, M. (1990). *Reading ethnographic research*. London: Longman.

Hammersley, M. (1992). *What's wrong with ethnography*. London: Routledge Kegan Paul.

Hammersley, M. (1995). *The politics of social research*. London: Sage.

Hammersley, M. (1995). Theory and evidence in qualitative research. *Quality and Quantity, 29*, 55–66.

Hammersley, M. (1997). Qualitative data archiving: Some reflections on its prospects and problems. *Sociology, 31*(1), 131–142.

Hammersley, M. (1999). Sociology, what's it for? A critique of the grand conception. *Sociological Research Online, 4*(3). Retrieved from http://www.socresonline.org.uk/socresonline/4/3/hammersley.html

Hammersley, M. (2000). *Taking sides in social research*. London: Routledge.

Hammersley, M. (2001). On "systematic" reviews of research literatures: A "narrative" reply to Evans and Benefield. *British Educational Research Journal, 27*(5), 543–554.

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in practice*. London: Routledge.

Hammond, M., Howarth, J., & Keat, R. (1991). *Understanding phenomenology*. Oxford, UK: Blackwell.

Handcock, M. S., & Morris, M. (1999). *Relative distribution methods in the social sciences*. New York: Springer-Verlag.

Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology, 1*, 69–97.

Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample survey methods and theory*. New York: John Wiley & Sons.

Hansen, P. R., & Johansen, S. (1998). *Workbook on cointegration*. Oxford, UK: Oxford University Press.

Hantrais, L., & Mangen, S. (Eds.). (1996). *Cross-national research methods in the social sciences*. London: Pinter.

Harary, F., Norman, D., & Cartwright, D. (1965). *Structural models for directed graphs*. New York: Free Press.

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimating equations*. London: Chapman & Hall.

Harding, S. (1991). *Whose science? Whose knowledge?* Milton Keynes, UK: Open University Press.

Harding, S. (1992). After the neutrality ideal: Science, politics, and "strong objectivity." *Social Research, 59*, 568–587.

Harding, S. (Ed.). (1987). *Feminism and methodology*. Milton Keynes, UK: Open University Press.

Hardy, M. A. (1993). *Regression with dummy variables* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–093). Newbury Park, CA: Sage.

Hardy, M. A., & Reynolds, J. (2003). Incorporating categorical information into regression models: the utility of dummy variables. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis*. London: Sage.

Hare, R. D. (1966). Temporal gradient of fear arousal in psychopaths. *Journal of Abnormal and Social Psychology, 70*, 442–445.

Harper, D. (2001). *Changing works: Visions of a lost agriculture*. Chicago: University of Chicago Press.

Harper, D. (2002). Talking about pictures: A case for photo elicitation. *Visual Studies, 17*(1), 13–26.

Harré, R. (1961). *Theories and things*. London: Sheed & Ward.

Harré, R. (1986). *Varieties of realism*. Oxford, UK: Blackwell.

Harré, R., & Secord, P. F. (1972). *The explanation of social behaviour.* Oxford, UK: Basil Blackwell.

Harris, C. W. (1962). Some Rao-Guttman relationships. *Psychometrika, 27,* 247–263.

Harris, J. A. (1913). On the calculation of intraclass and interclass coefficients of correlation from class moments when the number of possible combinations is large. *Biometrika, 9,* 446–472.

Harris, M. (1976). History and significance of the emic/etic distinction. *Annual Review of Anthropology, 5,* 329–350.

Hart, C. (1998). *Doing a literature review: Releasing the social science research imagination.* London: Sage.

Hartigan, J. A., & Kleiner, B. (1981). Mosaics for contingency tables. In W. F. Eddy (Ed.), *Computer science and statistics: Proceedings of the 13th Symposium on the Interface* (pp. 268–273). New York: Springer-Verlag.

Hartsock, N. C. M. (1998). *The feminist standpoint revisited and other essays.* Boulder, CO: Westview.

Hartwig, F., & Dearing, B. E. (1979). *Exploratory data analysis* (Sage University Papers on Quantitative Applications in the Social Sciences). Beverly Hills, CA: Sage.

Harvey, A. (1990). *The econometric analysis of time series* (2nd ed.). Cambridge: MIT Press.

Harvey, A. C. (1990). *Forecasting, structural time series models, and the Kalman filter.* New York: Cambridge University Press.

Harvey, A. S. (1999). Guidelines. In W. E. Pentland, A. S. Harvey, M. P. Lawton, & M. A. McColl (Eds.), *Time use research in the social sciences* (pp. 19–45). New York: Kluwer.

Harvey, D. (1973). *Social justice and the city.* Baltimore: Johns Hopkins University Press.

Harvey, D. (1989). *The condition of postmodernity.* Oxford, UK: Basil Blackwell.

Harvey, L. (1987). *Myths of the Chicago school of sociology.* Aldershot, England: Avebury.

Hastie, T. J. (1992). Generalized additive models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 249–307). Pacific Grove, CA: Wadsworth and Brooks/Cole.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models.* New York: Chapman & Hall.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning.* New York: Springer-Verlag

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains. *Biometrika, 57,* 97–109.

Hauser, R. M. (1978). A structural model of the mobility table. *Social Forces, 56,* 919–953.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica, 46,* 1251–1271.

Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica, 52,* 1219–1240.

Hayano, D. (1979). Auto-ethnography: Paradigms, problems and prospects. *Human Organization, 38*(1), 99–104.

Hayes, S. C., White, D., & Bissett, R. T. (1998). Protocol analysis and the "silent dog" method of analyzing the impact of self-generated rules. *Analysis of Verbal Behavior, 15,* 57–63.

Hays, W. L. (1972). *Statistics for the social sciences.* New York: Holt.

Hays, W. L. (1988). *Statistics.* New York: Holt, Rinehart and Winston.

Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.

Headland, T. N., Pike, K. L., & Harris, M. (1990). *Emics and etics: The insider/outsider debate.* Newbury Park, CA: Sage.

Healey, J. F. (1995). *Statistics: A tool for social research* (3rd ed.). Belmont, CA.: Wadsworth.

Heaton, J. (1998). Secondary analysis of qualitative data [Online]. *Social Research Update, 22.* Retrieved April 28, 2002, from http://www.soc.surrey.ac.uk/sru/SRU22.html.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47,* 153–161.

Heckman, J. J., & Singer, B. (1982). Population heterogeneity in demographic models. In K. Land & A. Rogers (Eds.), *Multidimensional mathematical demography* (pp. 567–599). New York: Academic Press.

Heckman, J. J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica, 52*(2), 271–320.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives, 9,* 85–110.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedstrom, P., & Swedberg, R. (1998). *Social mechanisms: An analytical approach to social theory.* Cambridge, UK: Cambridge University Press.

Heider, K. G. (1976). *Ethnographic film.* Austin: University of Texas Press.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences.* Thousand Oaks, CA: Sage.

Heise, D. (1971). The semantic differential and attitude research. In G. F. Summers (Ed.), *Attitude measurement* (pp. 235–253). Chicago: Rand McNally.

Heise, D. R. (1975). *Causal analysis.* New York: Wiley.

Heise, D. R. (1987). Affect Control Theory: Concepts and model. *Journal of Mathematical Sociology, 13,* 1–33.

Hektner, J. M., & Csikszentmihalyi, M. (2002). The experience sampling method: Measuring the context and the content of lives. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 233–243). New York: John Wiley & Sons.

Helland, I. S. (1990). PLS regression and statistical models. *Scandinavian Journal of Statistics, 17,* 97–114.

Hempel, C. (1965). *Aspects of scientific explanation, and other essays in the philosophy of science.* New York: Free Press.

Hempel, C. (1965). Confirmation, induction, and rational belief. In C. Hempel (Ed.), *Aspects of scientific explanation.* New York: Free Press.

Hempel, C. G. (1966). *Philosophy of natural science.* Englewood Cliffs, NJ: Prentice Hall.

Hendrick, T., Bickman, L., & Rog, D. J. (1993). *Applied research design: A practical guide.* Newbury Park, CA: Sage.

Hendry, D. F. (1995). *Dynamic econometrics.* New York: Oxford University Press.

Henkel, R. E. (1976). *Tests of significance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–004). Beverly Hills, CA: Sage.

Henle, M., & Hubble, M. B. (1938). Egocentricity in adult conversation. *Journal of Social Psychology, 9,* 227–234.

Henry, G. T. (1990). *Practical sampling.* Newbury Park, CA: Sage.

Henry, G. T., & Gordon, C. S. (2001). Tracking issue attention: Specifying the dynamics of the public agenda. *Public Opinion Quarterly, 65*(2), 157–177.

Hepburn, A. (2000). On the alleged incompatibility between feminism and relativism. *Feminism and Psychology, 10,* 91–106.

Heritage, J. (1984). *Garfinkel and ethnomethodology.* Cambridge, UK: Polity.

Herrnstein, R., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Hess, I. (1985). *Sampling for social research surveys 1947–1980.* Ann Arbor, MI: Institute for Social Research.

Hewitt, J. (1989). *Dilemmas of the American self.* Philadelphia: Temple University Press.

Heyde, C. C., & Seneta, E. (Eds.). (2001). *Statisticians of the centuries.* New York: Springer.

Hibbs, D. A. (1982). The dynamics of political support for American presidents among occupational and partisan groups. *American Journal of Political Science, 26,* 312–332.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13,* 141–154.

Hill, A. B. (1953). Observation and experiment. *New England Journal of Medicine, 248,* 995–1001.

Hill, M. M. (1993). *Archival strategies and techniques* (Qualitative Research Methods Series No. 31). Newbury Park, CA: Sage.

Hill, M. S. (1992). *The panel study of income dynamics: A user's guide.* Newbury Park, CA: Sage.

Hilsum, S., & Cane, B. (1971). *The teacher's day.* Windsor, UK: National Foundation for Educational Research.

Himmelfarb, G. (1987). *The new history and the old.* Cambridge, MA: Harvard University Press.

Hinchman, L. P., & Hinchman, S. K. (Eds.). (1997). *Memory, identity, community: The idea of narrative in the human sciences.* Albany: State University of New York Press.

Hinds, P. S., Vogel, R. J. . & Clarke-Steffen, L. (1997). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research, 7,* 408–424.

Hine, C. (2000). *Virtual ethnography.* London: Sage.

Hinton, P. R. (1996). *Statistics explained: A guide for social science students.* London: Routledge.

Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra.* New York: Academic Press.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (Eds.). (1983). *Understanding robust and exploratory data analysis.* New York: Wiley.

Hobcraft, J., Menken, J., & Preston, S. (1982). Age, period, and cohort effects in demography: A review. *Population Index, 42,* 4–43.

Hobert, J. P., & Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association, 91,* 1461–1473.

Hobson-West, P., & Sapsford, R. (2001). *The Middlesbrough Town Centre Study: Final report.* Middlesbrough, UK: University of Teesside, School of Social Sciences.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures.* New York: John Wiley.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression. *Technometrics, 12*(1), 55–67, 69–82.

Hoey, M. (2001). *Textual interaction.* London: Routledge.

Hofmann, H. (2000). Exploring categorical data: Interactive mosaic plots. *Metrika, 51*(1), 11–26.

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). New York: Macmillan.

Hogg, R. V., & Tanis, E. A. (1988). *Probability and statistical inference.* New York: Macmillan.

Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical inference.* New York: Wiley.

Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods* (2nd ed.). New York: Wiley.

Hollis, M. (1977). *Models of man.* Cambridge, UK: Cambridge University Press.

Hollis, M., & Smith, S. (1990). *Explaining and understanding international relations.* New York: Oxford University Press.

Hollway, W., & Jefferson, T. (2000). *Doing qualitative research differently: Free association, narrative and the interview method.* London: Sage.

Holmes, D. (1976). Debriefing after psychological experiments: Effectiveness of post-experimental desensitizing. *American Psychologist, 32,* 868–875.

Holmwood, J. (2001). Gender and critical realism: A critique of Sayer. *Sociology, 35,* 947–965.

Holstein, J. A., & Gubrium, J. F. (1995). *The active interview.* Thousand Oaks, CA: Sage.

Holstein, J. A., & Gubrium, J. F. (2000). *The self we live by: Narrative identity in a postmodern world.* New York: Oxford University Press.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities.* Reading, MA: Addison-Wesley.

Homan, R. (1992). *The ethics of social research.* London: Longman.

Homans, G. C. (1964). Contemporary theory in sociology. In R. E. L. Faris (Ed.), *Handbook of modern sociology* (pp. 951–977). Chicago: Rand McNally.

Homans, G. C. (1967). *The nature of social science.* New York: Harcourt, Brace, & World.

Honneth, A. (1996). *The struggle for recognition: The moral grammar of social conflicts* (J. Anderson, Trans.). Cambridge: MIT Press.

Hopkins, K. D. (1982). The unit of analysis: Group means versus individual observations. *American Educational Research Journal, 19*(1), 5–18.

Horn, R. (1996). Negotiating research access to organizations. *Psychologist, 9*(12), 551–554.

Horst, P. (1941). The role of the predictor variables which are independent of the criterion. *Social Science Research Council, 48,* 431–436.

Horton, M., & Freire, P. (2000). *We make the road by walking* (Bell, B., Gaventa, J., & Peters, J., Eds.). Philadelphia: Temple University Press.

Horton, N. J., & Lipsitz, S. R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *American Statistician, 55,* 244–254.

Horvitz, D. G., Shah, B. U., & Simmons, W. R. (1967). The unrelated question randomized response model. In E. D. Goldfield (Ed.), *Proceedings of the Social Statistics Section* (pp. 65–72).Washington, DC: American Statistical Association.

Hoskins, J. (1998). *Biographical objects: How things tell the stories of people's lives.* London: Routledge.

Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics, 2,* 211–228.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24,* 417–441, 498–520.

Hougaard, P. (2000). *Analysis of multivariate survival data.* New York: Springer-Verlag.

House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research.* Thousand Oaks, CA: Sage.

Hout, M. (1983). *Mobility tables.* Beverly Hills, CA: Sage.

Howell, D. C. (1999). *Fundamental statistics for the behavioral sciences* (4th ed.). Belmont, CA: Duxbury.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Duxbury, UK: Thomson Learning.

Hox, J. J. (2002). *Multilevel analysis, techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsely & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 465–497). New York: Academic Press.

Hsiao, C. (1986). *Analysis of panel data.* Cambridge, UK: Cambridge University Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1–55.

Huber, P. (1981). *Robust statistics.* New York: Wiley.

Huberman, A. M., & Miles, M. B. (1994). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428–444). Thousand Oaks, CA: Sage.

Huck, S. (2000). *Reading statistics and research* (3rd ed.). New York: Addison Wesley Longman.

Huff, A. S. (2000). Changes in organizational knowledge production. *Academy of Management Review, 25*(2), 288–293.

Hughes, J. A. (1990). *The philosophy of social research* (2nd ed.). Harlow, UK: Longman.

Hughes, K., MacKintosh, A. M., Hastings, G., Wheeler, C., Watson, J., & Inglis, J. (1997). Young people, alcohol, and designer drinks: A quantitative and qualitative study. *British Medical Journal, 314,* 414–418.

Hughes, R. (1998). Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of Health and Illness, 20,* 381–400.

Hujala, E. (1998). Problems and challenges in cross-cultural research. *Acta Universitatis Ouluensis, 35,* 19–31.

Hume, D. (1786). *Treatise on human nature.* Oxford, UK: Claridon.

Humphreys, L. (1975). *Tearoom trade: Impersonal sex in public places.* Chicago: Aldine.

Hunt, J. C. (1989). *Psychoanalytic aspects of fieldwork* (University Paper Series on Qualitative Research Methods 18). Newbury Park, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86,* 721–735.

Husserl, E. (1970). *Logical investigations* (Vols. 1–2). London: Routledge Kegan Paul.

Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models.* London: Sage.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association, 65,* 1582–1589.

Hyde, R. (2003). *The art of assembly language.* San Francisco: No Starch Press.

Hyman, H. (1955). *Survey design and analysis.* New York: Free Press.

Hyman, H. H. (1972). *Secondary analysis of sample surveys.* Glencoe, IL: Free Press.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis.* New York: Wiley.

ICPSR. (2002). *Guide to social science data preparation and archiving* [Online]. Available: http://www.ifdo.org/archiving_distribution/datprep_archiving_bfr.htm.

IFDO. (n.d.). [Online]. Available: http://www.ifdo.org/.

Imber, J. B. (Ed.). (2001). Symposium: Population politics. *Society, 39,* 3–53.

Inkeles, A., & Sasaki, M. (Eds.). (1996). *Comparing nations and cultures.* Englewood Cliffs, NJ: Prentice Hall.

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDRB). (1992). *Demographic surveillance system—Matlab: Registration of demographic events—1985* (Scientific Report 68). Dhaka, Bangladesh: Author.

Inter-University Consortium for Political and Social Research. (2002). *Guide to social science data preparation and archiving* [Online]. Supported by the Robert Wood Johnson Foundation. Available: www.ICPSR.umich.edu

Iversen, G. R. (1973). Recovering individual data in the presence of group and individual effects. *American Journal of Sociology, 79,* 420–434.

Iversen, G. R. (1984). *Bayesian statistical inference* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–043). Beverly Hills, CA: Sage.

Iversen, G. R. (1991). *Contextual analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–081). Newbury Park, CA: Sage.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–001). Newbury Park, CA: Sage.

Iversen, G., & Gergen, M. (1997). *Statistics: The conceptual approach.* New York: Springer-Verlag.

Iyengar, S., Peters, M. E., & Kinder, D. R. (1982). Experimental demonstrations of the "not-so-minimal" consequences of television news programs. *The American Journal of Political Science, 4,* 848–858.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–118). Thousand Oaks, CA: Sage.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–72). Thousand Oaks, CA: Sage.

Jaccard, J., Turrisi, R, & Wan, C. K. . (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Jackman, R. W. (1973). On the relationship of economic development to political performance. *American Journal of Political Science, 17,* 611–621.

Jackson, D. N. (2002). The constructs in people's heads. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 3–18). Mahwah, NJ: Lawrence Erlbaum.

Jackson, J. E. (1991). *A user's guide to principal component analysis.* New York: John Wiley.

Jackson, S., & Brashers, D. E. (1994). *Random Factors in ANOVA* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–098). Thousand Oaks, CA: Sage.

Jacobs, H. A. (1861). *Incidents in the life of a slave girl: Written by herself.* Boston.

Jacoby, L. L., Debner, J. A., & Hay, J. F. (2001). Proactive interference, accessibility bias, and process dissociations: Valid subject reports of memory. *Journal of Experimental Psychology: Memory, Learning, and Cognition, 27,* 686–700.

Jacoby, W. G. (1991). *Data theory and dimensional analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–078). Newbury Park, CA: Sage.

Jacoby, W. G. (1997). *Statistical graphics for univariate and bivariate data.* Thousand Oaks, CA: Sage.

Jaeger, R. M. (1993). *Statistics: A spectator sport* (2nd ed.). Newbury Park, CA: Sage.

Jahoda, G. (1977). In pursuit of the emic-etic distinction: Can we ever capture it? In Y. H. Poortinga (Ed.), *Basic problems in cross-cultural research* (pp. 55–63). Amsterdam: Swets & Zeitlinger.

James, J. B., & Sørensen, A. (2000, December). Archiving longitudinal data for future research: Why qualitative data add to a study's usefulness. *Forum Qualitative Sozialforschung* [Forum: Qualitative Social Research] [Online journal], *1*(3). Available: http://qualitative-research.net/fqs/fqs-eng.htm

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69,* 85–98.

James, W. (1952). *The varieties of religious experience: A study in human nature.* London: Longmans, Green & Co. (Original work published 1902)

Jameson, F. (1983). Postmodernism and consumer society. In H. Foster (Ed.), *Postmodern culture.* London: Pluto.

Jasso, G. (1988). Principles of theoretical analysis. *Sociological Theory, 6,* 1–20.

Jasso, G. (1990). Methods for the theoretical and empirical analysis of comparison processes. *Sociological Methodology, 20,* 369–419.

Jasso, G. (2001). Formal theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 37–68). New York: Kluwer Academic/Plenum.

Jasso, G. (2001). Rule-finding about rule-making: Comparison processes and the making of norms. In M. Hechter & K.-D. Opp (Eds.), *Social norms* (pp. 348–393). New York: Russell Sage.

Jasso, G. (2001). Comparison theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 669–698). New York: Kluwer Academic/Plenum.

Jasso, G. (2001). Studying status: An integrated framework. *American Sociological Review, 66,* 96–124.

Jasso, G. (2002). Seven secrets for doing theory. In J. Berger & M. Zelditch (Eds.), *New directions in contemporary sociological theory* (pp. 317–342). Boulder, CO: Rowan & Littlefield.

Jasso, G. (in press). The tripartite structure of social science analysis. *Sociological Theory.*

Jasso, G., & Rossi, P. H. (1977). Distributive justice and earned income. *American Sociological Review, 42,* 639–651.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Clarendon.

Jenkins, G. M., & Watts, D. G. (1968). *Spectral analysis and its applications.* San Francisco: Holden-Day.

Jenkins, R. (1984). Bringing it all back home: An anthropologist in Belfast. In C. Bell & H. Roberts (Eds.), *Social researching: Politics, problems, practice* (pp. 147–164). London: Routledge and Kegan Paul.

JMP Introductory Guide. (2000). Cary, NC: SAS Institute.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control, 12,* 231–254.

Johansen, S. (1991). Estimation and hypotheses testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica, 59,* 1551–1580.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford, UK: Oxford University Press.

John, K. E. (1989). The polls—A report. *Public Opinion Quarterly, 53,* 590–605.

John, N. R., & Draper, J.A. (1980). An alternative family of transformations. *Applied Statistics, 29*(2), 190–197.

Johnson, C. (1982). Risks in the publication of fieldwork. In J. E. Sieber (Ed.), *The ethics of social research: Fieldwork, regulation and publication* (pp. 71–92). New York: Springer-Verlag.

Johnson, C. (1994). Gender, legitimate authority, and leader-subordinate conversations. *American Sociological Review, 59,* 122–135.

Johnson, J. C. (1990). *Selecting ethnographic informants.* Newbury Park, CA: Sage.

Johnson, J. C., & Weller, S. C. (2002). Elicitation techniques for interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *The handbook of interview research.* Thousand Oaks, CA: Sage.

Johnson, N. L., & Kotz, S. (1969). *Discrete distributions.* Boston: Houghton-Mifflin.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions, Vol. 1.* New York: Wiley.

Johnson, R. A., & Bhattacharyya, G. K. (2001). *Statistics: Principles and methods* (4th ed.). New York: Wiley.

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis.* Upper Saddle River, NJ: Prentice Hall.

Johnson, T., Dandeker, C., & Ashworth, C. (1984). *The structure of social theory.* London: Macmillan.

Johnston, J., & Dinardo, J. (1997). *Econometric methods* (4th ed.). New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards.* Thousand Oaks, CA: Sage.

Joliffe, I. T. (1986). *Principal component analysis.* New York: Springer-Verlag.

Jones, J. M. G., & Hunter, D. (1995). Consensus methods for medical and health services research. *British Medical Journal, 311,* 376–380.

Jones, K. (1997). Multilevel approaches to modelling contextuality: From nuisance to substance in the analysis of voting behaviour. In G. P. Westert & R. N. Verhoeff (Eds.), *Places and people: Multilevel modelling in geographical research* (Nederlandse Geografische Studies 227). Utrecht, The Netherlands: The Royal Dutch Geographical Society and Faculty of Geographical Sciences, Utrecht University.

Jones, M. B. (1959). *Simplex theory* (U.S. Naval School of Aviation Medicine Monograph Series No. 3). Pensacola, FL: U.S. Naval School of Aviation Medicine.

Jones, M. O. (1996). *Studying organizational symbolism: What, how, why?* Thousand Oaks, CA: Sage.

Jones, S. R. (1992). Was there a Hawthorne effect? *American Journal of Sociology, 98,* 451–468.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32,* 443–482.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Academic Press.

Jöreskog, K. G. (1979). Statistical models and methods for analysis of longitudinal data. In K. G. Jöreskog & D. Sörbom (Eds.), *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt.

Josselson, R., & Lieblich, A. (Eds.). (1993–1999). *The narrative study of lives* (Vols. 1–6). Newbury Park, CA: Sage.

Josselsyn, R. (Ed.). (1996). *Ethics and process in the narrative study of lives.* Thousand Oaks, CA: Sage.

Joynson, R. B. (1989). *The Burt affair.* London: Routledge.

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1). Boston: McGraw-Hill.

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations* (6th ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

Judge, G. G., Griffiths, W. E., Hill, R. C., & Lee, T.-C. (1980). *The theory and practice of econometrics* (Wiley Series in Probability and Mathematical Statistics). New York: John Wiley.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1985). *The theory and practice of econometrics* (2nd ed.). New York: John Wiley.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H., & Lee, T. (1988). *Introduction to the theory and practice of econometrics* (2nd ed.). New York: John Wiley.

Julian, D. A. (1997). The utilization of the logic model as a system level planning and evaluation device. *Evaluation and Program Planning, 20*(3), 251–257.

Juster, F. T., & Stafford, F. P. (1985). *Time,goods and well-being.* Ann Arbor, MI: Institute for Social Research

Kalichman, S. C. (1999). *Mandated reporting of suspected child abuse: Ethics, law and policy.* Washington, DC: American Psychological Association.

Kalton, G. (1983). *Introduction to survey sampling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–035). Beverly Hills, CA: Sage.

Kan, M. (2002). Reinterpreting the Multifactor Leadership Questionnaire. In K. W. Parry & J. R. Meindl (Eds.), *Grounding leadership theory and research: Issues, perspectives and methods* (pp. 159–173). Greenwich, CT: Information Age Publishing.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice, 21*(1), 31–41.

Kant, I. (1781). *Kritik der reinen Vernunft* [Critique of pure reason]. Hamburg: Felix Meiner.

Kanuha, V. K. (2000). "Being" native versus "going native": Conducting social work research as an insider. *Social Work, 45*(5), 439–447.

Kaplan, A. (1943). Content analysis and the theory of signs. *Philosophy of Science, 10,* 230–247.

Kaplan, A. (1964). *The conduct of inquiry.* New York: Chandler.

Kaplan, D. (1990). Evaluation and modification of covariance structure models: A review and recommendation. *Multivariate Behavioral Research, 25,* 137–155.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions.* Thousand Oaks, CA: Sage.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53,* 457–481.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 451–477). New York: Cambridge University Press.

Kasprzyk, D., Duncan, G. J., Kalton, G., & Singh, M. P. (1989). *Panel surveys.* New York: John Wiley.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119–127.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91,* 1343–1370.

Kazdin, A. E. (2003). *Research design in clinical psychology, 4th Edition.* Boston: Allyn & Bacon.

Keat, R., & Urry, J. (1975). *Social theory as science.* London: Routledge Kegan Paul.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly, 64,* 125–148.

Keleman, M., & Bansal, P. (2002). The conventions of management research and their relevance to management practice. *British Journal of Management, 13,* 97–108.

Kelle, U. (Ed.). (1995). *Computer-aided qualitative data analysis: Theory, methods and practice.* London: Sage.

Kelly, G. (1955). *The psychology of personal constructs.* New York: Norton.

Kendall, G., & Wickham, G. (1999). *Using Foucault's methods.* London: Sage.

Kendall, M. G. (1962). *Rank correlation methods* (3rd ed.). London: Griffin.

Kendall, M. G., & Buckland, W. R. (1971). *A dictionary of statistical terms* (3rd ed.). Edinburgh: Oliver & Boyd.

Kendall, M., & Gibbons, J. D. (1990). *Rank correlation methods* (5th ed.). New York: Oxford University Press.

Kendall, M., & Stuart, A. (1969). *The advanced theory of statistics: Vol. 2. Inference and relationship.* London: Griffin.

Kendall, P. L., & Lazarsfeld, P. F. (1950). Problems of survey analysis. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research: Studies in the scope and method of "the American soldier"* (pp. 160–176). Glencoe, IL: Free Press.

Kennedy, P. (1992). *A guide to econometrics* (3rd ed.). Cambridge: MIT Press.

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

Kennedy, P. E. (2002). More on Venn diagrams for regression. *Journal of Statistics Education, 10*(1) [Online]. Retrieved from http://www.amstat.org/publications/jse/v10n1/kennedy.html

Kenny, D. A. (1979). *Correlation and causality.* New York: John Wiley.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 142–182). Orlando, FL: Academic Press.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Keren, G., & Lewis, C. (Eds.). (1993). *A handbook for data analysis in the behavioral sciences: Methodological issues.* Hillsdale, NJ: Lawrence Erlbaum.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research.* Ft. Worth, TX: Harcourt.

Kerr, A. W., Hall, H. K., & Kozub, A. (2002). Descriptive statistics. Chapter 2 in *Doing statistics with SPSS.* London: Sage.

Keselman, H. J., Lix, L. M., & Kowalchuk, R. K. (1998). Multiple comparison procedures for trimmed means. *Psychological Methods, 3,* 123–141.

Kessler, R., & Greenberg, D. (1981). *Linear panel analysis.* New York: Academic Press.

Keyfitz, N. (1985). *Applied mathematical demography* (2nd ed.). New York: Wiley.

Keynes, J. M. (1936). *The general theory of employment, interest, and money.* New York: Harcourt Brace Jovanovich.

Khurana, B. (1995). *The older spouse caregiver: Paradox and pain of Alzheimer's disease.* Unpublished doctoral dissertation, Center for Psychological Studies, Albany, CA.

Kiel, L. D., & Elliott, E. (1996). *Chaos theory in the social sciences.* Ann Arbor: University of Michigan Press.

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues.* Beverly Hills, CA: Sage.

Kinal, T., & Lahiri, K. (1983). Specification error analysis with stochastic regressors. *Econometrica, 54,* 1209–1220.

Kincaid, H. (1994). Defending laws in the social sciences. In M. Martin & L. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 111–130). Cambridge: MIT Press.

Kincaid, H. (1996). *Philosophical foundations of the social sciences: Analyzing controversies in social research.* Cambridge, UK: Cambridge University Press.

King, G. (1988). Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science, 32,* 838–863.

King, G. (1989). *Unifying political methodology: The likelihood theory of statistical inference.* New York: Cambridge University Press.

King, G. (1990). Stochastic variation: A comment on Lewis-Beck and Skalaban's "The R-Squared." *Political Analysis, 2,* 185–200.

King, G. (1997). *A solution to the ecological inference problem.* Princeton, NJ: Princeton University Press

King, G. (1998). *Unifying political methodology: The likelihood theory of statistical inference.* Ann Arbor: University of Michigan Press.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis, 9,* 137–163.

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research.* Princeton, NJ: Princeton University Press.

King, J. A. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation, 80,* 57–67.

Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research.* Beverly Hills, CA: Sage.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Kirk, R. E. (1999). *Statistics: An introduction* (4th ed.). Orlando, FL: Harcourt Brace. .

Kirkman, B. L., Rosen, B., Gibson, C. B., Tesluk, P. E., & McPherson, S. O. (2003). Five challenges to virtual team success: Lessons from Sabre, Inc. *Academy of Management Executive, 16*(3), 67–79.

Kish, L. (1949). A procedure for objective respondent selection within a household. *Journal of the American Sociological Association, 44,* 380–387.

Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association, 57,* 92–115.

Kish, L. (1965). *Survey sampling.* New York: Wiley.

Kish, L. (1987). *Statistical design for research.* New York: Wiley.

Kish, L. (1995). Methods for design effects. *Journal of Official Statistics, 11*(1), 55–77.

Kitagawa, E. (1955). Components of a difference between two rates. *Journal of the American Statistical Association, 50,* 1168–1194.

Kitcher, P., & Salmon, W. C. (Eds.). (1989). *Scientific explanation* (Minnesota Studies in the Philosophy of Science, Vol. 13). Minneapolis: University of Minnesota Press.

Kivlahan, D. R., Marlatt, G. A., Fromme, K., Coppel, D. B., & Williams, E. (1990). Secondary prevention with college drinkers: Evaluation of an alcohol skills training program. *Journal of Consulting and Clinical Psychology, 58*(6), 805–810.

Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences.* Hoboken, NJ: Wiley.

Kleinbaum, D. G., Kupper, L. L., & Morgensten, H. (1982). *Epidemiologic research: Principles and quantitative methods.* New York: Van Nostrand Reinhold.

Kleinman, A. L., Eisenberg, L., & Good, B. J. (1978). Culture, illness and care: Clinical lessons from anthropologic and cross-cultural research. *Annals of Internal Medicine, 88,* 251–258.

Kleinman, S., & Copp, M. A. (1993). *Emotions and fieldwork.* Thousand Oaks, CA: Sage.

Kline, P. (1994). *An easy guide to factor analysis.* Thousand Oaks, CA: Sage.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Klir, G. J., & Folger, T. A. (1988). *Fuzzy sets, uncertainty, and information.* Englewood Cliffs, NJ: Prentice Hall.

Kmenta, J. (1997). *Elements of econometrics* (2nd ed.). Ann Arbor: University of Michigan Press.

Knoke, D., & Burke, P. J. (1980). *Log-linear models.* Beverly Hills, CA: Sage.

Knoke, D., Bohrnstedt, G. W., & Mee, A. P. . (2002). *Statistics for social data analysis* (4th ed.). Itasca, IL: Peacock.

Knuth, D. E. (1997). *Fundamental algorithms: The art of computer programming* (Vol. 1, 3rd ed.). Reading, MA: Addison-Wesley.

Knuth, D. E. (1997). *The art of computer programming, volume 2: Seminumerical algorithms.* Reading, MA: Addison-Wesley.

Koch, W., Schulz, E. M., Wright, R., Smith, R. M., Lang, S. et al. (1996). What is a ratio scale? *Rasch Measurement Transactions, 9,* 457.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To err is human: Building a safer health system.* Washington, DC: National Academy Press.

Kondo, D. K. (1990). *Crafting selves: Power, gender and discourses of identity in a Japanese workplace.* Chicago: University of Chicago Press.

Korczynski, M. (2000). The political economy of trust. *Journal of Management Studies, 37*(1), 1–22.

Kornberg, A. (1997). *Basic research, the lifeline of medicine.* Retrieved from http://www.nobel.se/medicine/articles/research/

Körner, S. (1955). *Kant.* Harmondsworth, UK: Penguin.

Kotz, S., & Johnson, N. (Eds.). (1983). *Encyclopedia of statistical sciences.* New York: Wiley.

Kotz, S., Johnson, N. L., & Read, C. B. (1982). Censoring. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 1, p. 396). New York: Wiley.

Koutsoyiannis, A. (1978). *Theory of econometrics: An introductory exposition of econometric methods* (2nd ed.). London: Macmillan.

Kozlowski, S. W., & Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. *Journal of Applied Psychology, 77,* 161–167.

Kracauer, S. (1952–1953). The challenge of qualitative content analysis. *Public Opinion Quarterly, 16,* 631–642.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35–60). New York: Academic Press.

Krieger, N. (1994). Epidemiology and the web of causation: Has anyone seen the spider? *Social Science & Medicine, 39*(7), 887–903.

Krieger, S. (1983). *The mirror's dance: Identity in a women's community.* Philadelphia: Temple University Press.

Krippendorf, K. (1980). *Content analysis: An introduction to its methodology.* Beverly Hills, CA: Sage.

Kristeva, J. (1991). *Strangers to ourselves.* New York: Columbia University Press.

Kroenke, D. M. (2001). *Database processing: Fundamentals, design and implementation.* Upper Saddle River, NJ: Prentice Hall.

Krosnick, J. A. (1999). Survey methodology. *Annual Review of Psychology, 50,* 537–567.

Krueger, R. A. (1994). *Focus groups: A practical guide for applied research* (2nd ed.). Thousand Oaks, CA: Sage.

Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). Thousand Oaks, CA: Sage.

Kruskal, J. B. (1964). Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika, 29,* 1–27, 115–129.

Kruskal, J. B. (1968). Transformation of data. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (Vol. 15, pp. 182–192). New York: Macmillan.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association, 53,* 814–861.

Kuhn, M., & McPartland, T. S. (1954). An empirical investigation of self attitudes. *American Sociological Review, 19,* 68–76.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press. (Original work published 1962)

Kuran, T. (1995). The inevitability of future revolutionary surprises. *American Journal of Sociology, 100*(6), 1528–1551.

Kuusela, H., Spence, M. T., & Kanto, A. J. (1998). Expertise effects on prechoice decision processes and final outcomes: A protocol analysis. *European Journal of Marketing, 32,* 5–6, 559.

Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing.* Thousand Oaks, CA: Sage.

Kvale, S. (1999). The psychoanalytic interview as qualitative research. *Qualitative Inquiry, 5*(1), 87–113.

Labov, W. (1982). Speech actions and reactions in personal narrative. In D. Tannen (Ed.), *Analyzing discourse: Text and talk.* Washington, DC: Georgetown University Press.

Lachenbruch, P.A. (1975). *Discriminant analysis.* New York: Hafner.

LaFreniere, P., & Charlesworth, W. R. (1983). Dominance, attention, and affiliation in a preschool group: A nine-month longitudinal study. *Ethology and Sociobiology, 4*(2), 55–67.

Lagopoulos, A. Ph., & Boklund-Lagopoulou, K. (1992). *Meaning and geography: The social conception of the region in northern Greece.* Berlin: Mouton de Gruyter.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association, 73,* 805–811.

Lakatos, I. (1978). *The methodology of scientific research programmes: Philosophical papers. Vol. 1.* Cambridge, UK: Cambridge University Press.

Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of knowledge.* Cambridge, UK: Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics, 34,* 1–14.

Langeheine, R., & van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods & Research, 18,* 416–441.

Langeheine, R., & Van de Pol, F. (1994). Discrete-time mixed Markov latent class models. In A. Dale & R. B. Davies (Eds.), *Analyzing social and political change: A casebook of methods* (pp. 171–197). London: Sage.

Langellier, K. M. (2001). Personal narrative. In M. Jolly (Ed.), *Encyclopedia of life writing: Autobiographical and biographical forms* (Vol. 2). London: Fitzroy Dearborn.

Langellier, K. M., & Peterson, E. E. (2003). *Performing narrative: The communicative practice of storytelling.* Philadelphia: Temple University Press.

Lapadat, J. C., & Lindsay, A. C. (1999). Transcription in research and practice: From standardization of technique to interpretive positionings. *Qualitative Inquiry, 5*(1), 64–86.

Larsen, R. J., & Marx, M. L. (1990). *Statistics.* Englewood Cliffs, NJ: Prentice Hall.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium* (pp. 112–136). New York: John Wiley.

Latham, G. P., Skarlicki, D., Irvine, D., & Siegel, J. P. (1993). The increasing importance of performance appraisals to employee effectiveness in organizational settings in North America. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational* *psychology 1993* (Vol. 8, pp. 87–132). Chichester, UK: Wiley.

Lather, P. (1993). Fertile obsession: Validity after post-structuralism. *Sociological Quarterly, 35,* 673–694.

Lather, P. (1999). To be of use: The work of reviewing. *Review of Educational Research, 69*(1), 2–7.

Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth.* Berkeley: University of California Press.

Lauder, M. (2003). Covert participant observation of a deviant community: Justifying the use of deception. *Journal of Contemporary Religion,* 18(2), 185–196.

Laufer, R. S., & Wolfe, M. (1977). Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues, 33,* 44–87.

Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics, 15*(2), 269–282.

Lave, C., & March, J. G. (1978). *An introduction to models in the social sciences.* New York: Harper & Row.

Lavrakas, P. J. (1987). *Telephone survey methods.* Newbury Park, CA: Sage.

Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection, and supervision* (2nd ed.). Newbury Park, CA: Sage.

Lavrakas, P. J. (1998). Methods for sampling and interviewing in telephone surveys. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 429–472). Thousand Oaks, CA: Sage.

Lawlis, G. F., & Lu, E. (1972). Judgment of counseling process: Reliability, agreement, and error. *Psychological Bulletin, 78,* 17–20.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28,* 563–575.

Lax, D. A. (1985). Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association, 80,* 736–741.

Layder, D. (1993). *New strategies in social research.* Cambridge, UK: Polity.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis & the interpretation and mathematical foundation of latent structure analysis. In S. A. Stoufffer.(Ed.), *Measurement and prediction* (pp. 362–472). Princeton, NJ: Princeton University Press.

Lazarsfeld, P. F. (1955). Interpretation of statistical relations as a research operation. In P. F. Lazarsfeld & M. Rosenberg (Eds.), *The language of social research* (pp. 111–125). Glencoe, IL: Free Press.

Lazarsfeld, P. F. (1958). Evidence and inference in social research. *Daedalus, 87,* 120–121.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

Leach, C., Freshwater, K., Aldridge, J., & Sunderland, J. (2001). Analysis of repertory grids in clinical practice. *British Journal of Clinical Psychology, 40,* 225–248.

Leamer, E. (1978). *Specification searches: Ad hoc inference with nonexperimental data.* New York: Wiley.

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data.* New York: John Wiley.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review, 73*(1), 31–43.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis.* New York: Wiley.

Lebreton, J.-D., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs, 62,* 67–118.

Lee, A. M. (1978). *Sociology for whom.* New York: Oxford University Press.

Lee, P. M. (1997). *Bayesian statistics: An introduction* (2nd ed.). London: Arnold.

Lee, R. (2000). *Unobtrusive methods in social research.* Buckingham, UK: Open University Press.

Lee, R. B. (1993). *The Dobe !Kung.* New York: Holt, Rinehart and Winston.

Lee, R. M. (1993). *Doing research on sensitive topics.* London: Sage.

Lee, R. M. (1995). *Dangerous fieldwork.* London: Sage.

Lehtonen, M. (2000). *The cultural analysis of texts.* London: Sage.

Leisering, L., & Leibfried, S. (1999). *Time and poverty in Western welfare states.* New York: Cambridge University Press.

Lemaitre, G. (1992). *Dealing with the seam problem* (Survey of Labour and Income Dynamics Research Papers 92–05). Ottawa: Statistics Canada.

Lenski, G. (1966). *Power and privilege.* New York: McGraw-Hill.

Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 259–272). New York: John Wiley.

Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika, 33,* 183–212.

Lett, J. (1987). The importance of the emic/etic distinction. In *The human enterprise: A critical introduction to anthropological theory.* Boulder, CO: Westview.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics* (pp. 278–292). Stanford, CA: Stanford University Press.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory, 10,*: 707–710. (Original work published 1965)

Levin, A., Liukkonen, J., & Levine, D. W. (1996). Equivalent inference using transformations. *Communications in Statistics, Theory and Methods, 25*(5), 1059–1072.

Levin, I. P. (1999). *Relating statistics and experimental design: An introduction* (Sage University Paper Series on

Quantitative Applications in the Social Sciences, 07–125). Thousand Oaks, CA: Sage.

Levin, J. R., & Subkoviak, M. J. (1977). Planning an experiment in the company of measurement error. *Applied Psychological Measurement, 1*(3), 331–338.

Levinas, E. (1998). *On thinking-of-the-Other: Entre nous.* New York: Columbia University Press.

Levine, E. K. (1982). Old people are not all alike: Social class, ethnicity/race, and sex are bases for important differences. In J. E. Sieber (Ed.), *The ethics of social research: Surveys and experiments* (pp. 127–144). New York: Springer-Verlag.

Lévi-Strauss, C. (1963). *Structural anthropology.* New York: Basic Books.

Lévi-Strauss, C. (1969). *The elementary structures of kinship.* London: Eyre and Spottiswoode.

Levy, P. S., & Lemeshow, S. (1999). *Sampling of populations: Methods and applications.* New York: Wiley.

Lewin, K. (1943). Defining the "field at a given time." *Psychological Review, 50,* 292–310.

Lewin, K. (1959). *A dynamic theory of personality: Selected papers* (D. K. Adams & K. E. Zener, Trans.). New York: McGraw-Hill. (Original work published 1935)

Lewin, K. (1997). Experiments in social space. In G. W. Lewin (Ed.), *Resolving social conflicts: Field theory in social science* (pp. 59–67). Washington, DC: American Psychological Association. (Originally published in 1939)

Lewis, D. K. (1973). *Counterfactuals.* Cambridge, MA: Harvard University Press.

Lewis-Beck, M. S. (1980). *Applied regression: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07–022). Beverly Hills, CA: Sage.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–103). Thousand Oaks, CA: Sage.

Lewis-Beck, M. S. (Ed.). (1993). *Regression analysis.* London: Sage/Toppan.

Lewis-Beck, M. S., & Skalaban, A. (1990). The R-squared: Some straight talk. *Political Analysis, 2,* 153–171.

Li, H., Rosenthal, R., & Rubin, D. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods, 1*(1), 98–107.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73,* 13–22.

Liao, T. F. (1989). A flexible approach for the decomposition of rate differences. *Demography, 26,* 717–726.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models.* Thousand Oaks, CA: Sage.

Liao, T. F. (2001). How responsive is U.S. population growth to immigration? A situational sensitivity analysis. *Mathematical Population Studies, 9,* 217–229.

Liao, T. F. (2002). *Statistical group comparison.* New York: Wiley.

Lichfield, N. (1996). *Community impact evaluation.* London: University College Press.

Lieberson, S. (1987). *Making it count: The improvement of social research and theory.* Berkeley: University of California Press.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140,* 44–53.

Lincoln, Y. S., & Guba, E. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Lincoln, Y. S. & Guba, E. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 163–188). Thousand Oaks, CA: Sage.

Lind, J. (1753). *A treatise of the scurvy: Of three parts containing an inquiry into the nature, causes and cure of that disease.* Edinburgh, UK: Sands, Murray and Cochran.

Lindahl, L. (2002). *Do birth order and family size matter for intergenerational income mobility? Evidence from Sweden* (Working Paper No. 5). Stockholm: Swedish Institute for Social Research.

Lindell, M. K., & Brandt, C. J. (1999). Assessing interrater agreement on the job relevance of a test: A comparison of the CVI, T, $r_{WG(J)}$, and $r^*_{WG(J)}$ indexes. *Journal of Applied Psychology, 84,* 640–647.

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23,* 127–135.

Lindesmith, A. (1968). *Addiction and opiates.* Chicago: Aldine.

Lindley, D. V. (2001). Thomas Bayes. In C. C. Heyde & E. Seneta (Eds.), *Statisticians of the centuries* (pp. 68–71). New York: Springer-Verlag.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, 34,* 1–41.

Lindquist, E. F. (1953). *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin.

Lindsay, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related models, including a simple latent class model for item analysis. *Journal of the American Statistical Association, 86,* 96–107.

Lingis, A. (1994). *Abuses.* Berkeley: The University of California Press.

Lingis, A. (1994). *The community of those who have nothing in common.* Bloomington: Indiana University Press.

Lingis, A. (1998). *The imperative.* Bloomington: Indiana University Press.

Linn, R. L. (Ed.). (1989). *Educational measurement* (3rd ed.). New York: Macmillan.

Linstone, H. A. (1978). The Delphi technique In R. B. Fowles (Ed.), *Handbook of futures research.* Westport, CT: Greenwood.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger. R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute, Inc.

Little, D. (1991). *Varieties of social explanation: An introduction to the philosophy of social science.* Boulder, CO: Westview.

Little, D. (1998). *Microfoundations, method and causation: On the philosophy of the social sciences.* New Brunswick, NJ: Transaction Publishers.

Little, R. J. A. (1988). Missing data adjustments in large surveys. *Journal of Business and Economic Statistics, 6,* 287–301.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: John Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: John Wiley.

Litwin, M. S. (1995). *How to measure survey reliability and validity.* Thousand Oaks, CA: Sage.

Livingston, G. (1999). Beyond watching over established ways: A review as recasting the literature, recasting the lived. *Review of Educational Research, 69*(1), 9–19.

Loader, C. (1999). *Local regression and likelihood.* New York: Springer.

Locke, K. D. (2000). *Using grounded theory in management research.* London: Sage.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports,* 3(Suppl. 9), 635–694.

Loewenthal, L. (1989). Sociology of literature in retrospect. In P. Desan et al. (Eds.), *Literature and social practice.* Chicago and London: University of Chicago Press.

Lofland, J., & Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis.* Belmont, CA: Wadsworth.

Loman, L. A., & Larkin, W. E. (1976). Rejection of the mentally ill: An experiment in labeling. *Sociological Quarterly, 17,* 555–560.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

Long, J. S., & Cheng, S. (2003). Regression models for categorical outcomes. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis.* London: Sage.

Long, J. S., & Ervin, L. H. (2000) Using heteroskedasticity consistent standard errors in the linear regression model. *American Statistician, 54,* 217–224.

Longford, N. T. (1993). *Random coefficient models.* New York: Oxford University Press.

Looney, C. G. (1997). *Pattern recognition using neural networks.* Oxford, UK: Oxford University Press.

Lord, F. M. (1956). Sampling error due to choice of split in split-half reliability coefficients. *Journal of Experimental Education, 24,* 245–249.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Lorencz, B. (1991). Becoming ordinary: Leaving the psychiatric hospital. In J. M. Morse & J. Johnson (Eds.), *The illness experience: Dimensions of suffering* (pp. 140–200). Newbury Park, CA: Sage.

Lorenz, E. N. (1963). Deterministic non-periodic flow. *Journal of Atmospheric Science, 20,* 130–141.

Lorenz, K. (1981). *The foundations of ethology.* New York: Simon and Schuster.

Lotka, A. J. (1998). *Analytical theory of biological populations: Part II: Demographic analysis with particular application to human populations* (D. P. Smith & H. Rossert, Trans.). New York: Plenum. (Original work published 1939)

Luger, G. F. (2002). *Artificial intelligence: Structures and strategies for complex problem solving* (4th ed.). Reading, MA: Addison-Wesley.

Lukes, S. (1994). Methodological individualism reconsidered. In M. Martin & L. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 451–459). Cambridge: MIT Press.

Lundberg, G. (1939). *Foundations of sociology.* New York: Macmillan.

Lupia, A., McCubbins, M. D., & Popkin, S. L. (Eds.). (1998). *Elements of reason.* Cambridge, UK: Cambridge University Press.

Luttrell, W. (2003). *Pregnant bodies, fertile minds: Gender, race, and the schooling of pregnant teens.* New York: Routledge.

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology, 55,* 6–10.

Lykken, D. T. (1995). *The antisocial personalities.* Mahwah, NJ: Lawrence Erlbaum.

Lynch, M. (2000). Against reflexivity as an academic virtue and source of privileged knowledge. *Theory, Culture and Society, 17*(3), 26–54.

Lyotard, J.-F. (1984). *The postmodern condition: A report on knowledge.* Minneapolis: University of Minnesota Press.

Macaulay, A. C., Paradis, G., Potvin, L., Cross, E. J., Saad-Haddad, C., McComber, A., et al. (1997). The Kahnawake Schools Diabetes Prevention Project: Intervention, evaluation, and baseline results of a diabetes primary prevention program with a native community in Canada. *Preventive Medicine, 26*(6), 79–90.

MacDonald, I. L., & Zucchini, W. (1997). *Hidden Markov models and other types of models for discrete-valued time series.* London: Chapman & Hall.

MacDonald, R. R. (2002). The incompleteness of probability models and the resultant implications for theories of statistical inference. *Understanding Statistics, 1*(3), 167–189.

MacDougall, D. (2001). Renewing ethnographic film: Is digital video changing the genre? *Anthropology Today, 17*(3), 15–21.

Machlin, S. R., & Taylor, A. K. (2000). *Design, methods, and field results of the 1996 Medical Expenditure Panel Survey Medical Provider Component* (MEPS Methodology Report No. 9, AHRQ Pub. No. 00–0028). Rockville, MD: Agency for Healthcare Research and Quality.

Mackay, A. L. (1977). *Scientific quotations.* New York: Crane, Russak.

Mackie, J. L. (1974). *The cement of the universe: A study of causation.* Oxford, UK: Oxford University Press.

Mackintosh, N. J. (1995). *Cyril Burt: Fraud or framed?* Oxford, UK: Oxford University Press.

Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology, 28,* 143–166.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics.* Cambridge, UK: Cambridge University Press.

Maddala, G. S. (1992). *Introduction to econometrics.* Englewood Cliffs, NJ: Prentice Hall.

Madden, D. (2000). Towards a broader explanation of male-female wage differences. *Applied Economics Letters, 7,* 765–770.

Madjar, I. (2001). The lived experience of pain in the context of clinical practice. In S. K. Toombs (Ed.), *Handbook of phenomenology and medicine* (pp. 263–277). Boston: Kluwer Academic.

Magidson, J. (1993). The CHAID approach to segmentation modeling: CHi-squared Automatic Interaction Detection. In R. Bagozzi (Ed.), *Handbook of marketing research* (pp. 118–159). London: Blackwell.

Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology, 31,* 223–264.

Magnusson, D. (1967). *Test theory.* Reading, MA: Addison-Wesley.

Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society, 29,* 507–548.

Maines, D. R. (1992). Theorizing movement in an urban transportation system by use of the constant comparative method in field research. *Social Science Journal, 29,* 283–292.

Maines, D. R. (2001). *The faultline of consciousness: A view of interactionism in sociology.* New York: Aldine De Gruyter.

Makridakis, S., & Hibbon, M. (2000). The M3-competition: Results, conclusions, and implications. *International Journal of Forecasting, 16,* 451–476.

Malinoswki, B. (1989). *A diary in the strict sense of the term.* Stanford, CA: Stanford University Press.

Malinowski, B. (1922). *Argonauts of the Western Pacific.* London: Routledge & Kegan Paul.

Manski, C. F. (1995). *Identification problems in the social sciences.* Cambridge, MA: Harvard University Press.

Manton, K. G., & XiLiang, G. (2003). *Variation in disability decline and Medicare expenditures.* Working Monograph, Center for Demographic Studies.

Manton, K. G., & Stallard, E. (1988). *Chronic disease modelling: Vol. 2. Mathematics in medicine.* New York: Oxford University Press.

Manton, K. G., Singer, B., & Woodbury, M. A. (1992). Some issues in the quantitative characterization of heterogeneous populations. In J. Trussell, R. Hankinson, & J. Tilton (Eds.), *Demographic application of event history analysis* (pp. 9–37). Oxford, UK: Clarendon.

Manton, K. G., Woodbury, M. A., & Tolley, H. D. (1994). *Statistical applications using fuzzy sets.* Wiley Series in Probability and Mathematical Statistics. New York: John Wiley.

Marcus, G., & Fischer, M. (1986). *Anthropology as cultural critique.* Chicago: University of Chicago Press.

Marcuse, H. (1964). *One dimensional man.* Boston: Beacon.

Marcuse, H. (1968). *One dimensional man: The ideology of industrial society.* London: Sphere.

Mare, R. D. (1994). Discrete-time bivariate hazards with unobserved heterogeneity: A partially observed contingency table approach. In P. V. Marsden (Ed.), *Sociological methodology 1984* (pp. 341–383). Cambridge, MA: Blackwell.

Markham, A. N. (1998). *Life online: Researching real experience in virtual space.* Walnut Creek, CA: AltaMira.

Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics, 12,* 591–612.

Marquis, K. (1984). Record checks for sample surveys. In T. Jabine, E. Loftus, M. Straf et al. (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines.* Washington, DC: National Academy Press.

Marsh, C. (1982). *The survey method: The contribution of surveys to sociological explanation.* London: Allen & Unwin.

Marsh, C. (1988). *Exploring data: An introduction to data analysis for social scientists.* Cambridge, UK: Polity.

Marsh, L. C., & Cormier, D. R. (2001). *Spline regression models.* Thousand Oaks, CA: Sage.

Marshall, P. A. (1992). Research ethics in applied anthropology. *IRB: A Review of Human Subjects Research, 14*(6), 1–5.

Martens, H., & Naes, T. (1989). *Multivariate calibration.* London: Wiley.

Martin, J. A., Hamilton, B. E., Ventura, S. J., Menacker, F., & Park, M. M. (2002, February 12). *Births: Final data for 2000* (National Vital Statistics Report, Vol. 50, No. 5). Hyattsville, MD: National Center for Health Statistics.

Martin, K. (1998). *When a baby dies of SIDS: The parents' grief and searching for a reason.* Edmonton, Canada: Qual Institute Press.

Mason, J. (1996). *Qualitative researching.* London: Sage.

Mason, J. (2002). *Qualitative researching* (2nd ed.). London: Sage.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, W. K. (1973). Some methodological issues in the cohort analysis of archival data. *American Sociological Review, 38,* 242–258.

Mathiowetz, N., & McGonagle, K. (2000). An assessment of the current state of dependent interviewing in household surveys. *Journal of Official Statistics, 16*(4), 401–418.

Matza, D. (1969). *Becoming deviant.* Englewood Cliffs, NJ: Prentice-Hall.

Maxim, P. S. (1999). *Quantitative research methods in the social sciences.* New York: Oxford University Press.

Maxwell, J. A. (1996). *Qualitative research design.* Thousand Oaks, CA: Sage.

Maxwell, S. E., & Delaney, H. D. (1990). *Design experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature, 26,* 459–467.

Maynard, D. W. (1984). *Inside plea bargaining: The language of negotiation.* New York: Plenum.

McAdams, D. P., & Constantian, C. A. (1983). Intimacy and affiliation motives in daily living: An experience sampling analysis. *Journal of Personality and Social Psychology, 45,* 851–861.

McCleary, R., & Hay, R. A., Jr. (1980). *Applied time series analysis for the social sciences.* London: Sage.

McCloskey, D. N. (1986). *The rhetoric of economics.* Madison: University of Wisconsin Press.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B, 42,* 109–142.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear and mixed models.* New York: John Wiley.

McCullough, B. D., & Vinod, H. D. (1997). The numerical reliability of econometric software. *Journal of Economic Literature, 37*(2), 633–665.

McDonald, J. F., & Moffitt, R. A. (1980). The uses of Tobit analysis. *Review of Economics and Statistics, 62,* 318–321.

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A., Jr. (1980). *Interrupted time series analysis.* Beverly Hills, CA: Sage.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of econometrics* (pp. 105–142). New York: Academic Press.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30–46.

McGrew, W. (1972). *An ethological study of children's behavior.* New York: Academic Press.

McGuigan, J. (1992). *Cultural populism.* London: Routledge.

McGuinness, M., & Sapsford, R. (2002). *Middlesbrough Citizens' Advice Bureau: Report for 2001.* Middlesbrough, UK: University of Teesside, School of Social Sciences.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage, 3,* 143–157.

McIver, J. P., & Carmines, E. G. (1980). *Unidimensional scaling.* Beverly Hills, CA: Sage.

McKean, K. (1987, January). The orderly pursuit of pure disorder. *Discover,* pp. 72–81.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology, 4,* 103–120.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and application to clustering.* New York: Marcel Dekker.

McSweeny, A. J. (1978). Effects of response cost on the behavior of a million persons: Charging for directory assistance in Cincinnati. *Journal of Applied Behavior Analysis, 11,* 47–51.

Meacham, S. J. (1998). Threads of a new language: A response to Eisenhart's "On the subject of interpretive reviews." *Review of Educational Research, 68*(4), 401–407.

Mead, M. (1928). *Coming of age in Samoa.* New York: William Morrow.

Meadows, L. M., & Dodendorf, D. M. (1999). Data management and interpretation: Using computers to assist. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (pp. 195–218). Thousand Oaks, CA: Sage.

Melton, G. B., Levine, R. J., Koocher, G. P., Rosenthal, R., & Thompson, W. C. (1988). Community consultation in socially sensitive research: Lessons from clinical trials on treatments for AIDS. *American Psychologist, 43,* 573–581.

Menard, S. (1995). *Applied logistic regression analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–106). Thousand Oaks, CA: Sage.

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–106). Thousand Oaks, CA: Sage.

Menard, S. (2002). *Longitudinal research* (2nd ed.). Thousand Oaks, CA: Sage.

Menard, S. (in press). *Logistic regression.* Thousand Oaks, CA: Sage.

Menard, S., & Elliott, D. S. (1990). Longitudinal and cross-sectional data collection and analysis in the study of crime and delinquency. *Justice Quarterly, 7,* 11–55.

Menchú, R. (1984). *I, Rigoberta Menchú: An Indian woman in Guatemala.* (Ed. and intro. Elisabeth Burgos-Debray, Trans. Ann Wright.) London: Verso.

Mendel, G. (1866). Versuche über Pflanzen-Hybriden [Experiments in plant hybridization]. In *Verhandlungen des naturforschenden Vereins* [Proceedings of the Natural History Society]. Available in both the original German and the English translation at www.mendelweb.org

Merkle, D. M., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 243–258). New York: Wiley.

Merleau-Ponty, M. (1962). *Phenomenology of perception.* London: Routledge Kegan Paul.

Merton, R. K. (1949). *Social theory and social structure.* New York: Free Press.

Merton, R. K. (1968). *Social theory and social structure* (3rd ed.). New York: Free Press.

Merton, R. K. (1987). The focused interview and focus groups: Continuities and discontinuities. *Public Opinion Quarterly, 51,* 550–566.

Merton, R. K., & Kendall, P. L. (1946). The focused interview. *American Journal of Sociology, 51*(6), 514–557.

Merton, R. K., Fiske, M., & Kendall, P. L. (1990). *The focused interview: A manual of problems and procedures* (2nd ed.). New York: Free Press.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5–11.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., p. 1487–1495). New York: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculation by fast computing machines. *Journal of Chemical Physics, 21,* 1087–1092.

Meulman, J. J. (1998). Book review of W. J. Krzanowski, & F. H. C. Marriott, *Multivariate Analysis. Part I. Distributions, Ordinations, and Inference,* London: Edward Arnold, 1994. *Journal of Classification, 15,* 287–293

Meulman, J. J., & Heiser, W. J. (2000). *Categories.* Chicago: SPSS.

Michel, Y., & Haight, B. K. (1996). Using the Solomon four design. *Nursing Research, 45*(6), 367–369.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67,* 371–378.

Milgram, S. (1974). *Obedience to authority.* London: Tavistock.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: Harper & Row.

Mill, J. S. (1947). *A system of logic.* London: Longman, Green. (Originally published in 1879)

Mill, J. S. (1956). *A system of logic, ratiocinative and inductive.* London: Longmans, Green & Company. (Original work published 1843)

Mill, J. S. (1973). *A system of logic.* Toronto: University of Toronto Press. (Original work published 1865)

Mill, J. S. (1974). *Collected works of John Stuart Mill: Vol. 8. A system of logic* (Books 4–6). Toronto: University of Toronto Press.

Millar, A., Simeone, R. S., & Carnevale, J. T. (2001). Logic models: A system tool for performance management. *Evaluation and Program Planning, 24,* 73–81.

Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science.* New York: Praeger.

Miller, A. G. (1995). Constructions of the obedience experiments: A focus upon domains of relevance. *Journal of Social Issues, 51,* 33–53.

Miller, A. J. (1990). *Subset selection in regression.* New York: Chapman Hall.

Miller, D. (1998). Writing and retelling multiple ethnographic tales of a soup kitchen for the homeless. *Qualitative Inquiry, 4,* 469–492.

Miller, L., Rustin, M., Rustin, M., & Shuttleworth, J. (1989). *Closely observed infants.* London: Duckworth.

Miller, R. G. (1981). *Simultaneous statistical inference* (2nd ed.). New York: Springer-Verlag.

Miller, R. W. (1987). *Fact and method: Explanation, confirmation and reality in the natural and the social sciences.* Princeton, NJ: Princeton University Press.

Miller, W. L., & Crabtree, B. F. (1999). Clinical research: A multimethod typology and qualitative roadmap. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (2nd ed., pp. 3–32). Thousand Oaks, CA: Sage.

Millett, K. (1970). *Sexual politics.* Garden City, NY: Doubleday.

Milligan, G. W. (1979). Ultrametric hierarchical clustering algorithms. *Psychometrika, 44*(3), 343–346.

Milliken, G. A., & Johnson, D. A. (1984). *Analysis of messy data: Vol. 1. Designed experiments.* New York: Van Nostrand Reinhold.

Milliken, G. A., & Johnson, D. E. (1992). *Analysis of messy data: Vol. 1. Designed experiments.* London: Chapman & Hall.

Mills, C. W. (1959). *The sociological imagination.* New York: Oxford University Press.

Mills, T. C. (1990). *Time series techniques for economists.* Cambridge, UK: Cambridge University Press.

Millward, L. J. (2000). Focus groups. In G. M. Breakwell, C. Fife-Schaw, & S. Hammond (Eds.), *Research methods in psychology* (2nd ed., pp. 303–324). London: Sage.

Milner, A. (1996). *Literature, culture and society.* London: UCL Press.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy, 66,* 281–302.

Mincer, J. (1974). *Schooling, experience and earnings.* New York: Columbia University Press.

Mishler, E. G. (1986). *Research interviewing: Context and narrative.* Cambridge, MA: Harvard University Press.

Mishler, E. G. (1991). Representing discourse: The rhetoric of transcription. *Journal of Narrative and Life History, 1*(4), 255–280.

Mishler, E. G. (1995). Models of narrative analysis: A typology. *Journal of Narrative and Life History, 5*(2), 87–123.

Mitchell, J. (1983). Case and situation analysis. *Sociological Review, 31*(2), 186–211.

Mitchell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin, 100,* 398–407.

Mittelhammer, R. C., Judge, G. G, & Miller, D. J. (2000). *Econometric foundations.* New York: Cambridge University Press.

Mohr, D. C., Likosky, W., Bertagnolli, A., Goodkin, D. E., Van der Wende, J., Dwyer, P., et al. (2000). Telephone administered cognitive-behavioral therapy for the treatment of depressive symptoms in multiple sclerosis. *Journal of Consulting and Clinical Psychology, 68*(2), 356–361.

Mohr, L. B. (1990). *Understanding significance testing* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–073). Newbury Park, CA: Sage.

Monge, P., & Contractor, N. (2003). *Theories of communication networks.* New York: Oxford University Press.

Monk-Turner, E., & Turner, C. G. (2001). Sex differentials in the South Korean labour market. *Feminist Economics, 7*(1), 63–78.

Mood, A. M., & Graybill, F. A. (1963). *Introduction to the theory of statistics* (2nd ed.). New York: McGraw-Hill.

Mooney, C. Z. (1997). *Monte Carlo simulation.* Thousand Oaks, CA: Sage.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference.* Thousand Oaks, CA: Sage.

Moore, D. S. (1997). *Statistics: Concepts and controversies* (4th ed.). New York: W. H. Freeman.

Moore, D. S., & McCabe, G. P. (1998). *Introduction to the practice of statistics.* New York: W. H. Freeman.

Moore, D. W. (1999, June/July). Daily tracking polls: Too much "noise" or revealed insights? *Public Perspective,* pp. 27–31.

Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology, 65,* 131–149.

Moreno, L. (1994). Frailty selection in bivariate survival models: A cautionary note. *Mathematical Population Studies, 4,* 225–233.

Morgan, D. L. (1988). *Focus groups as qualitative research.* Newbury Park, CA: Sage.

Morgan, D. L. (1993). *Successful focus groups: Advancing the state of the art.* London: Sage.

Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications for health research. *Qualitative Health Research, 8,* 362–376.

Morgan, D. L., & Krueger R. A. (Eds.). (1998). *Focus group kit.* Thousand Oaks, CA: Sage.

Morley, D. (1980). *The "Nationwide" audience.* London: British Film Institute.

Morley, D. (1986). *Family television—Cultural power and domestic leisure.* London: British Film Institute.

Morris, M., & Western, B. (1999). Inequality in earnings at the close of the 20th century. *Annual Review of Sociology, 25,* 623–657.

Morrison, D. E., & Henkel, R. (Eds.). (1970). *The significance test controversy.* Chicago: Aldine.

Morrow, R. A. (with Brown, D. D.). (1994). *Critical theory and methodology.* London: Sage.

Morse, J. M. (1989). Strategies for sampling. In J. Morse (Ed.), *Qualitative nursing research: A contemporary dialogue* (pp. 117–131). Rockville, MD: Aspen.

Morse, J. M. (1997). Considering theory derived from qualitative research. In J. Morse (Ed.), *Completing a qualitative project: Details and dialogue* (pp. 2163–2188). Thousand Oaks, CA: Sage.

Morse, J. M., & Richards, L. (2002). *Read me first for a user's guide to qualitative methods.* Thousand Oaks, CA: Sage.

Morse, J. M., Swanson, J. M., & Kuzel, A. J. (Eds.). (2001). *The nature of qualitative evidence.* Thousand Oaks, CA: Sage.

Morton, Rebecca B. (2000). *Methods and models: A guide to the empirical analysis of formal models in political science.* New York: Cambridge University Press.

Moser, C. A., & Kalton, G. (1971). *Survey methods in social investigation* (2nd ed.). Aldershot, UK: Gower.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics.* Reading, MA: Addison-Wesley.

Mounce, H. O. (1997). *The two pragmatisms: From Peirce to Rorty.* London: Routledge and Kegan Paul.

Moustakas, C. (1990). *Heuristic research: Design, methodology, and applications.* Newbury Park, CA: Sage.

Moustakas, C. (1994). *Phenomenological research methods.* Thousand Oaks, CA: Sage.

Moyé, L. A. (2000). *Statistical reasoning in medicine.* New York Springer-Verlag.

Mruck, K., Corti, L., Kluge, S. & Opitz, D. (Eds.). (2000, December). Text archive: Re-analysis. *Forum Qualitative Sozialforschung* [Forum: Qualitative Social Research] [Online Journal], *1*(3). Available: http://qualitative-research.net/fqs/fqs-eng.htm

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56,* 63–75.

Mueller, D. J. (1986). *Measuring social attitudes: A handbook for researchers and practitioners.* New York: Teachers College, Columbia University.

Mueller, R. O. (1996). *Basic principles of structural equation modeling: An introduction to LISREL and EQS.* New York: Springer.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw-Hill.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research, 22,* 267–305.

Murphy, M. (1990). Minimising attrition in longitudinal studies: Means or end? In D. Magnusson & L. R. Bergman (Eds.), *Data quality in longitudinal research* (pp. 148–156). Cambridge, UK: Cambridge University Press.

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F. B., Ashkam, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development [Entire issue]. *Health Technology Assessment, 2*(3).

Musch, J., & Reips, U. (2000). A brief history of Web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61–88). San Diego, CA: Academic Press.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115–132.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*(1), 81–117.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 51,* 431–462.

Myers, J. L., & Well, A. D. (1995). *Research design and statistical analysis.* Hillsdale, NJ: Lawrence Erlbaum.

Naes, T., & Risvik, E. (Eds.). (1996). *Multivariate analysis of data in sensory science.* New York: Elsevier.

Nagel, E. (1979). *The structure of science: Problems in the logic of scientific explanation.* Indianapolis, IN: Hackett.

Nagler, J. (1995). Coding style and good computing practices. *Political Methodologist, 6,* 2–8.

Namboodiri, K. (1984). *Matrix algebra: An introduction.* Beverly Hills, CA: Sage.

Narayan, K. (1993). How native is the "native" anthropologist? *American Anthropologist, 95,* 671–686.

National Center for Health Statistics. (2003). *United States life tables, 2000.* Retrieved from www.cdc.gov/nchs/data/lt2000.pdf.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research* (DHEW Publication No. (OS) 78–0012). Washington, DC: Government Printing Office.

National Institutes of Health. (2003, July 22). *Certificates of confidentiality.* Retrieved from http://grants1.nih.gov/grants/policy/coc/index.htm

National Research Council. (2001). *The 2000 census: Interim assessment.* Washington, DC: National Academy Press.

Naughton, J. (1984). *Soft systems analysis: An introductory guide.* Milton Keynes, UK: Open University Press.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco: W. H. Freeman.

Nelder, J. A., & Lane, P. W. (1995). The computer analysis of factorial experiments: In memoriam—Frank Yates. *The American Statistician, 49,* 382–385.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A, 135,* 370–384.

Nelson, J. S., Megill, A., & McCloskey, D. N. (Eds.). (1987). *The rhetoric of the human sciences.* Madison: University of Wisconsin Press.

Nesbary, D. K. (2000). *Survey research and the World Wide Web.* Boston, MA: Allyn and Bacon.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models* (3rd ed.). Chicago: Irwin.

Neter, J., Wasseman, W., & Kutner, M. H. (1989). *Applied linear regression models* (2nd ed.). Boston: Irwin.

Neter, J., Wasserman, W., & Kutner, M. (1990). *Applied linear statistical models* (3rd ed.). Homewood, IL: Irwin.

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement, 22,* 287–293.

New, C. (1996). *Agency, health and social survival.* London: Taylor & Francis.

Newbold, P., & Bos, T. (1985). *Stochastic parameter regression models* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–051). Beverly Hills, CA: Sage.

Newbold, P., & Bos, T. (1994). *Introductory business forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Newby, H. (1977). In the field: Reflections on a study of Suffolk farm workers. In C. Bell & H. Newby (Eds.), *Doing sociological research* (pp. 108–129). London: Allen and Unwin.

Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science, 134,* 2011–2017.

Newell, A., & Simon, H. A. (1971). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Newell, R. W. (1986). *Objectivity, empiricism and truth.* London: Routledge and Kegan Paul.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55,* 703–708.

Newton, I. (1952). *Mathematical principles of natural philosophy.* Chicago: Britannica. (Originally published in 1686)

Newton, R. R., & Rudestam, K. E. (1999). *Your statistical consultant: Answers to your data analysis questions.* Thousand Oaks, CA: Sage.

NHS Centre for Reviews and Dissemination. (2001). *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews* (Report number 4, 2nd ed.). York, UK: CRD.

Nicholson, L. (Ed.). (1990). *Feminism/postmodernism.* New York: Routledge Kegan Paul.

Nida, E. A. (1975). *Componential analysis of meaning: An introduction to semantic structures.* The Hague, The Netherlands: Mouton.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Nishisato, S. (1994). *Elements of dual scaling: An introduction to practical data analysis.* Hillsdale, N.J.: Lawrence Erlbaum.

Nishisato, S. (1996). Gleaning in the field of dual scaling. *Psychometrika, 61,* 559–599.

Nishisato, S., & Clavel, J. G. (2003). A note on between-set distances in dual scaling and correspondence analysis. *Behaviormetrika, 30,* 87–98.

NIST/SEMATECH. (2003). *E-handbook of statistical methods.* Retrieved from http://www.itl.nist.gov/div898/handbook/eda

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies.* Newbury Park, CA: Sage.

Nocedal, J., & Wright, S. (1999). *Numerical optimization.* New York: Springer-Verlag.

Noldus, L. P. J. J., Trienes, R. J. H., Hendriksen, A. H. M., Jansen, H., & Jansen, R. G. (2000). The Observer

Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments & Computers, 32,* 197–206.

Norpoth, H. (1995). Is Clinton doomed? An early forecast for 1996. *Political Science & Politics, 27,* 201–206.

Norton, H. W. (1939). The 7 × 7 squares. *Annals of Eugenics, 9,* 269–307.

Norusis, M. (1990). *SPSS/PC+ Statistics 4.0.* Chicago: SPSS.

Norusis, M. J. (1990). *SPSS base system user's guide.* Chicago: SPSS.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues, 58,* 161–176.

Nowak, A., & Latané, B. (1994). Simulating the emergence of social order from individual behaviour. In N. Gilbert & J. Doran (Eds.), *Simulating societies: The computer simulation of social phenomena* (pp. 63–84). London: UCL.

Noymer, A. (2001). The transmission and persistence of "urban legends": Sociological application of age-structured epidemic models. *Journal of Mathematical Sociology, 25,* 299–323.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Nutley, S., & Webb, J. (2000). Evidence and the policy process. In H. Davies, S. Nutley, & P. Smith (Eds.), *What works? Evidence-based policy and practice in public services* (pp. 13–41). Bristol, UK: Policy Press.

O'Brien, R. M. (2000). Age-period-cohort-characteristic models. *Social Science Research, 29,* 123–139.

O'Hear, A. (1989). *An introduction to the philosophy of science.* Oxford, UK: Clarendon.

O'Hear, A. (Ed.). (1996). *Verstehen and humane understanding.* Supplement to *Philosophy,* Royal Institute of Philosophy Supplement 41. Cambridge, UK: Cambridge University Press.

O'Rourke, D., & Blair, J. (1983). Improving random respondent selection in telephone surveys. *Journal of Marketing Research, 20,* 428–432.

Oakley, A. (2000). *Experiments in knowing: Gender and method in the social sciences.* Cambridge, UK: Polity.

Ohnuki-Tierney, E. (1984). "Native" anthropologists. *American Ethnologist, 11,* 584–586.

Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics,* 7(3), 349–365.

Olmsted, P., & Weikart, D. P. (Eds.). (1994). *Families speak: Early childhood care and education in 11 countries.* Ypsilanti, MI: High/Scope Press.

Olsen W. K. (1993) Random samples and repeat surveys in South India. In S. Devereux & J. Hoddinott (Eds.) *Fieldwork in Developing Countries,* (pp. 57–72). Boulder, CO: Lynne Rienner.

Olsen, W. K. (1999). *Path analysis for the study of farming and micro-enterprise* (Bradford Development Paper No. 3). Bradford, UK:University of Bradford, Development and Project Planning Centre.

Olsen, W. K., Warde, A., & Martens, L. (2000). Social differentiation and the market for eating out in the UK. *International Journal of Hospitality Management, 19*(2), 173–190.

Oneal, J. R., & Russett, B. (1997). The classical liberals were right: Democracy, interdependence, and conflict, 1950–1985. *International Studies Quarterly, 41,* 267–294.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 776–783.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143–179). New York: Academic Press.

Orr, L. L. (1999). *Social experiments: Evaluating public programs with experimental methods.* Thousand Oaks, CA: Sage.

Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157–159.

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5,* 343–355.

Osgood, C. E. (1952). The nature of and measurement of meaning. *Psychological Bulletin, 49,* 197–237.

Osgood, C. E., & Tzeng, O. C. S. (Eds.). (1990). *Language, meaning, and culture: The selected papers of C. E. Osgood.* New York: Praeger.

Osgood, C. E., Suci, C. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Ostrom, C. W., Jr. (1990). *Time series analysis.* Thousand Oaks, CA: Sage.

Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs, 62,* 1–135.

Outhwaite, W. (1975). *Understanding social life: The method called verstehen.* London: Allen & Unwin.

Outhwaite, W. (1987). *New philosophies of social science: Realism, hermeneutics and critical theory.* London: Macmillan.

Paccagnella, L. (1997). Getting the seats of your pants dirty: Strategies for ethnographic research on virtual communities. *Journal of Computer Mediated Communication, 3*(1). Retrieved from http://www.ascusc.org/jcmc/vol3/issue1/paccagnella.html

Paechter, C. (1996). Power, knowledge and the confessional in qualitative research. *Discourse: Studies in the Politics of Education, 17*(1), 75–84.

Page, B. I., & Shapiro, R. Y. (1992). *The rational public.* Chicago: University of Chicago Press.

Page, S. (2000). The lost art of unobtrusive measures. *Journal of Applied Social Psychology, 30*(10), 2126–2128.

Pagès, J., & Tenenhaus, M. (2001). Multiple factor analysis combined with PLS path modeling: Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgments. *Chemometrics and Intelligent Laboratory Systems, 58,* 261–273.

Palmer, R. E. (1969). *Hermeneutics: Interpretation theory of Schleiermacher, Dilthey, Heidegger, and Gadamer.* Evanston, IL: Northwestern University Press.

Pampel, F. C. (2000). *Logistic regression: A primer.* Thousand Oaks, CA: Sage.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology, 47,* 409–410.

Papineau, D. (1993). *Philosophical naturalism.* Oxford, UK: Basil Blackwell.

Park, A., & Jowell, R. (1997). *Consistencies and differences in a cross-national survey.* London: SCPR.

Parker, I. (1988). Deconstructing accounts. In C. Antaki (Ed.), *Analysing everyday explanation: A casebook of methods* (pp. 184–198). London: Sage.

Parkin, F. (1973). *Class inequality and political order.* London: Paladin.

Parkinson, B., & Manstead, A. S. R. (1993). Making sense of emotion in stories and social life. *Cognition and Emotion, 7,* 295–323.

Parsons, T. (1968). Émile Durkheim. In D. L. Sills (Ed.), *International encyclopedia of the social sciences, Vol. 4.* New York: Macmillan.

Parsons, T. (1968). *The structure of social action* (3rd ed.). New York: Free Press.

Passmore, J. (1967). Logical positivism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 5, pp. 52–57). New York: Macmillan.

Paterson, B. L., Thorne, S. E., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research: A practical guide to meta-analysis and meta-synthesis.* Thousand Oaks, CA: Sage.

Patrick, C. J. (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology, 31,* 415–428.

Pattison, P. E. (1993). *Algebraic models for social networks.* New York: Cambridge University Press.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (1999). *Grand Canyon celebration: A father-son journey of discovery.* Amherst, NY: Prometheus.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation.* London: Sage.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 6*(2), 559–572.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal, 3,* 1243–1246.

Peat, J., Mellis, C., Williams, K., & Xuan W. (2002). *Health science research: A handbook of quantitative methods.* London: Sage.

Pedhazer, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart and Winston.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). New York: Wadsworth.

Peirce, C. S. (1931–1958). *Collected papers of Charles Sanders Peirce.* Cambridge, MA: Harvard University Press.

Pelusi, J. (1997). The lived experience of surviving breast cancer. *Oncology Nursing Forum, 24*(8), 1343–1353.

Perks, R., & Thomson, A. (Eds.). (1998). *The oral history reader.* London: Routledge.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research, 21,* 381–391.

Pettofrezzo, A. J. (1978). *Matrices and transformations.* New York: Dover.

Pfungst, O. (1965). *Clever Hans: The horse of Mr. von Osten.* New York: Holt, Rinehart, & Winston. (Original work published 1911)

Phillips, D. (1976). *Holistic thought in social science.* Stanford, CA: Stanford University Press.

Phillips, N., & Brown, J. (1993). Analyzing communication in and around organizations: A critical hermeneutic approach. *Academy of Management Journal, 36*(6), 1547–1576.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica, 59,* 283–306.

Piaget, J. (1929). *The child's conception of the world.* New York: Harcourt, Brace Jovanovich.

Piaget, J. (1955). *The construction of reality in the child.* New York: Routledge Kegan Paul. (Original work published 1937)

Pickering, M. (2001). *Stereotyping: The politics of representation.* Basingstoke, UK: Palgrave.

Pierce, J. L., Gardner, D. G., Cummings, L. L., & Dunham, R. B. (1989). Organization-based self-esteem: Construct definition, measurement, and validation. *Academy of Management Journal, 32,* 622–648.

Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review, 94,* 251–267.

Pike, K. L. (1967). *Language in relation to a unified theory of the structure of human behavior* (2nd ed.). The Hague: Mouton. (Original work published 1954)

Pindyck, R. S., & Rubinfeld, D. L. (1991). *Econometric models and economic forecasts* (3rd ed.). New York: McGraw-Hill.

Pindyck, R. S., & Rubinfeld, D. L. (1998). *Econometric models and economic forecasts* (4th ed.). New York: McGraw-Hill.

Pink, S. (2001). *Doing visual ethnography: Images, media and representation in research.* London: Sage.

Plackett, R. L. (1972). The discovery of the method of least squares. *Biometrika, 59,* 239–251.

Plake, B. S., & Impara, J. C. (Eds.). (2001). *The fourteenth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

Platt, J. (2002). The history of the interview. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 33–54). Thousand Oaks, CA: Sage.

Plessy v. Ferguson, 163 U.S. 537 (1896).

Plewis, I. (1985). *Analysing change: Measurement and exploration using longitudinal data.* Chichester, UK: Wiley.

Plummer, K. (2001). *Documents of life 2: An invitation to a critical humanism.* London: Sage.

Poland, B. (2001). Transcription quality. In J. Gubrium & J. Holstein (Eds.), *Handbook of interview research.* Thousand Oaks, CA: Sage.

Poland, B. D. (1995). Transcript quality as an aspect of rigor in qualitative research. *Qualitative Inquiry, 1*(3), 290–310.

Pollock, K. H., Nichols, J. D., Brownie, C., & Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs, 107,* 1–97.

Pope, C., & Mayes, N. (1995). Qualitative research: Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *British Medical Journal, 311,* 42–45.

Popkin, S. L. (1991). *The reasoning voter.* Chicago: University of Chicago Press.

Popper, K. (1959). *The logic of scientific discovery.* London: Hutchinson.

Popper, K. (1965). *Conjectures and refutations: The growth of scientific knowledge* (2nd ed.). New York: Basic Books.

Popper, K. (1966). *The open society and its enemies* (5th ed., Vol. 1). London: Routledge. (Original work published 1945)

Popper, K. (1972). *Objective knowledge: An evolutionary approach.* Oxford, UK: Clarendon.

Popper, K. R. (1957). *The poverty of historicism.* Boston: Beacon.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Popper, K. R. (1961). *The poverty of historicism.* London: Routledge & Kegan Paul.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge.* New York: Basic Books.

Popper, K. R. (1965). *The logic of scientific discovery.* New York: Harper & Row.

Popper, K. R. (1972). *Conjectures and refutation.* London: Routledge & Kegan Paul.

Popper, K. E. (1986). *The poverty of historicism.* London: ARK.

Porter, S. (1993). Critical realist ethnography: The case of racism and professionalism in a medical setting. *Sociology, 27,* 591–609.

Potter, J. (1996). *Representing reality: Discourse, rhetoric and social construction.* London: Sage.

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour.* London: Sage.

Poundstone, W. (1988). *Labyrinths of reason: Paradoxes, puzzles and the frailty of knowledge.* New York: Anchor/Doubleday.

Poundstone, W. (1992). *Prisoner's dilemma.* New York: Doubleday.

Power, M. (1990). Modernism, postmodernism and organisation. In J. Hassard & D. Pym (Eds.), *The theory and philosophy of organisations.* London: Routledge Kegan Paul.

Powers, D. A., & Xie, Y. (2000). *Statistical methods for categorical data analysis.* San Diego: Academic Press.

Prais, S. J., & Winsten, C. B. (1954). *Trend estimators and serial correlation.* Chicago: Cowles Commission.

Prendergast, C. (1995). *Cultural materialism: On Raymond Williams.* Minneapolis: University of Minnesota Press.

Prentice, R. L., & Zhao, L. P. (1991). Estimating equations for parameters in mean and covariances of multivariate discrete and continuous responses. *Biometrics, 47,* 825–839.

Preston, S. H., & Coale, A. J. (1982). Age structure, growth, attrition and accession: A new synthesis. *Population Index, 48*(2), 217–259.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes.* Oxford, UK: Basil Blackwell.

Price, L. J. (1989). In the shadow of biomedicine: Self-medication in two Ecuadorian pharmacies. *Social Science and Medicine, 28,* 905–915.

Price, L. J. (2002). Carrying out a structured observation. In M. V. Angrosino (Ed.), *Doing cultural anthropology* (pp. 107–114). Prospect Heights, IL: Waveland.

Price, V. (1992). *Public opinion.* Newbury Park, CA: Sage.

Proctor, C. H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika, 35,* 73–78.

Przeworski, A., & Teune, H. (1970). *The logic of comparative social inquiry.* New York: John Wiley.

Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B, 11,* 18–84.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). *GLLAMM manual technical report 2001/01.* London: King's College, University of London, Department of Biostatistics and Computing, Institute of Psychiatry. Retrieved from www.iop.kcl.ac.uk/IoP/Departments/BioComp/programs/manual.pdf

Rachlin, H. (2000). *The science of self-control.* Cambridge, MA: Harvard University Press.

Radest, H. B. (1996). *Humanism with a human face: Intimacy and the enlightenment.* London: Praeger.

Radley, A., & Taylor, D. (2003). Remembering one's stay in hospital: A study in recovery, photography and forgetting.

Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine, 7(2), 129–159.

Rae, D. W., & Yates, D. (1981). *Equalities.* Cambridge, MA: Harvard University Press.

Raftery, A. E. (1996). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (Vol. 25, pp. 111–163). Oxford, UK: Basil Blackwell.

Raftery, A. E. (1996). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 163–187). London: Chapman and Hall.

Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association, 90,* 54–63.

Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies.* Berkeley: University of California Press.

Ragin, C. C. (1994). *Constructing social research.* Thousand Oaks, CA: Pine Forge Press.

Ragin, C. C. (2000). *Fuzzy-set social science.* Chicago: University of Chicago Press.

Ragin, C. C., & Becker, H. S. (1992). *What is a case? Exploring the foundations of social inquiry.* New York: Cambridge University Press.

Ramazanoglu, C., with Holland, J. (2002). *Feminist methodology.* London: Sage.

Ramsay, J. O. (1977). Maximum likelihood estimation in MDS. *Psychometrika, 42,* 241–266.

Ramsey, F. P. (1960). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays.* New York: Harcourt Brace.

Random House. (1997). *Random House Webster's college dictionary.* New York: Author.

Rao, P., & Griliches, Z. (1969). Small-sample properties of several two-stage regression methods in the context of auto-correlated errors. *Journal of the American Statistical Association, 64,* 253–272.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma.* Ann Arbor: University of Michigan Press.

Rässler, S. (2002). Statistical matching: A frequentist theory, practical applications, and alternative Bayesian approaches. *Lecture Notes in Statistics, 168.* New York: Springer.

Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education, 15,* 423–476.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1999) *Applied regression analysis: A research tool.* New York: Springer.

Reason, R., & Brabury, H. (Eds.). (2001). *Handbook of action research: Participative inquiry and practice.* London: Sage.

Redington, M., & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in the Cognitive Sciences, 1*(7), 273–281.

Reed-Danahay, D. (2001). Autobiography, intimacy and ethnography. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography* (pp. 405–425). London: Sage.

Reed-Danahay, D. E. (Ed.). (1997). *Auto/ethnography: Rewriting the self and the social.* Oxford, UK: Berg.

Rees, P., Martin, D., & Williamson, P. (Eds.). (2002). *The census data system.* Chichester, UK: Wiley.

Reichenbach, H. (1951). *The rise of scientific philosophy.* Berkeley: University of California Press.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). New York: Cambridge University Press.

Rengger, J. (1996). *Retreat from the modern: Humanism, postmodernism and the flight from modernist culture.* Exeter, UK: Bowerdean.

Rescher, N. (1997). *Objectivity: The obligations of impersonal reason.* Notre Dame, IN: University of Notre Dame Press.

Retherford, R. D., & Choe, M. K. (1993). *Statistical models for causal analysis.* New York: John Wiley.

Rex, J. (1974). *Sociology and the demystification of the modern world.* London: Routledge & Kegan Paul.

Reyment, R. A., & Jöreskog, K. G. (1993). *Applied factor analysis in the natural sciences.* Cambridge, UK: Cambridge University Press.

Reynolds, H. T. (1984). *Analysis of nominal data* (2nd ed.). Beverly Hills, CA: Sage.

Rhee, J. W. (1996). How polls drive campaign coverage. *Political Communication, 13,* 213–229.

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury.

Richards, T. J., & Richards, L. (1994). Using computers in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 445–462). Thousand Oaks, CA: Sage.

Richardson, L. (1985). *The new other woman: Contemporary single women in affairs with married men.* London: Collier-Macmillan.

Richardson, L. (1997). *Fields of play: Constructing an academic life.* New Brunswick, NJ: Rutgers University Press.

Richardson, L. (2000). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 923–948). Thousand Oaks, CA: Sage.

Richardson, L., & Lockridge, E. (in press). *Travels with Ernest: Crossing the Literary-Ethnographic Divide.* Walnut Creek, CA: AltaMira.

Ricoeur, P. (1981). *Hermeneutics and the human sciences: Essays on language, action and interpretation.* New York: Cambridge University Press.

Ricoeur, P. (1992). *Oneself as another.* Chicago: The University of Chicago Press.

Ridgeway, C. L. (2001). Inequality, status, and the construction of status beliefs. In J. H. Turner (Ed.), *Handbook* *of sociological theory* (pp. 323–340). New York: Kluwer Academic/Plenum.

Rieger, J. (1996). Photographing social change. *Visual Sociology, 11*(1), 5–49.

Riessman, C. K. (2002). Doing justice: Positioning the interpreter in narrative work. In W. Patterson (Ed.), *Strategic narrative: New perspectives on the power of personal and cultural storytelling* (pp. 195?216). Lanham, MA: Lexington Books.

Riessman, C. K. (2002). Positioning gender identity in narratives of infertility: South Indian women's lives in context. In M. C. Inhorn & F. van Balen (Eds.), *Infertility around the globe: New thinking on childlessness, gender, and reproductive technologies.* Berkeley: University of California Press.

Riessman, C. K. (2003). Performing identities in illness narrative: Masculinity and multiple sclerosis. *Qualitative Research,* 3(1), 5–33.

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research, 30,* 359–383.

Riggs, P. J. (1992). *Whys and ways of science.* Melbourne, Australia: Melbourne University Press.

Riker, W. H. (1980). Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review, 74,* 432–446.

Ripley, B. D. (1996). *Pattern recognition and neural networks.* Cambridge, UK: Cambridge University Press.

Ritchie, J., & Lewis, J. (2003). *Qualitative research practice.* London: Sage.

Ritzer, G. (1980). *Sociology: A multiple paradigm science.* Boston: Allyn & Bacon.

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics, 39,* 347–366.

Robert, C. P. (2001). *The Bayesian choice: A decision theoretic motivation* (2nd ed.). New York: Springer-Verlag.

Robert, C. P., & Casella, G. (1999). *Monte Carlo statistical methods.* New York: Springer-Verlag.

Roberts, B. (2002). *Biographical research.* Buckingham, UK: Open University Press.

Roberts, N., Andersen, D. F., Deal, R. M., Grant, M. S., & Shaffer, W. A. (1983). *Introduction to computer simulation: A system dynamics modeling approach.* Reading, MA: Addison-Wesley.

Robey, B., Rutstein, S. O., & Morris, L. (1992). The reproductive revolution: New survey findings (Technical Report M-11). *Population Reports.* Baltimore, MD: Johns Hopkins University Center for Communication Programs.

Robins, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine, 17,* 269–302.

Robinson, J. (1994). White women researching/representing "others": From anti-apartheid to postcolonialism. In A. Blunt & G. Rose (Eds.), *Writing, women and space* (pp. 97–226). New York: Guilford.

Robinson, J. P., & Godbey, G. (1997). *Time for life.* University Park: Pennsylvania State University Press.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15,* 351–357.

Robinson, W. S. (1951). The logical structure of analytic induction. *American Sociological Review, 16,* 812–818.

Robson, C. (1993). *Real world research.* Oxford, UK: Blackwell.

Rochford, E. B., Jr. (1992). On the politics of member validation: Taking findings back to Hare Krishna. In G. Miller & J. A. Holstein (Eds.), *Perspectives on social problems* (Vol. 3, pp. 99–116). Greenwich, CT: JAI.

Rodgers, J. L. (1999). The bootstrap, the jackknife, and the randomization test: A sampling taxonomy. *Multivariate Behavioral Research, 34,* 441–456.

Rodríguez, G. (1994). Statistical issues in the analysis of reproductive histories using hazard models. In K. L. Campbell & J. W. Wood (Eds.), *Human reproductive ecology: Interaction of environment, fertility and behavior* (pp. 266–279). New York: New York Academy of Sciences.

Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker.* Cambridge, MA: Harvard University Press.

Rogers, A. (1995). *Multiregional demography.* Chichester, UK: Wiley.

Rohrer, L. G. (1965). The great response-style myth. *Psychological Bulletin, 63,* 129–156.

Romney, A. K., Weller, S. C., & Batchelder, W. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist, 88,* 313–338.

Ronan, W. W., & Latham, G. P. (1974). The reliability and validity of the critical incident technique: A closer look. *Studies in Personnel Psychology, 6*(1), 33–64.

Roncek, D. W. (1992). Learning more from Tobit coefficients. *American Sociological Review, 57,* 503–507.

Rorty, R. (1991). *Objectivity, relativism, and truth: Philosophical papers volume 1.* Cambridge, UK: Cambridge University Press.

Rosaldo, R. (1989). *Culture & truth.* Boston: Beacon.

Rose, G. (2001). *Visual methodologies.* London: Sage.

Rosenau, P. M. (1992). *Post-modernism and the social sciences: Insights, inroads and intrusions.* Princeton, NJ: Princeton University Press.

Rosenbaum, P. R. (1995). *Observational studies.* New York: Springer-Verlag.

Rosenbaum, P. R. (1998). Multivariate matching methods. *Encyclopedia of Statistical Sciences, 2,* 435–438.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55.

Rosenberg, A. (1988). *Philosophy of social science.* Oxford, UK: Clarendon.

Rosenberg, M. (1968). *The logic of survey analysis.* New York: Basic Books.

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal &

R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279–349). New York: Academic Press.

Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering method in multivariate research. *Multivariate Behavioral Research, 10,* 489–502.

Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis.* New York: Wiley.

Rosenhan, D. (1973). On being sane in insane places. *Science, 179,* 250–258.

Rosenthal, G. (Ed.). (1998). *The Holocaust in three generations: Families of victims and persecutors of the Nazi regime.* London: Cassell.

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts.

Rosenthal, R. (1976). *Experimenter effects in behavioral research* (rev. ed.). New York: Irvington.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart & Winston.

Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom: Expanded edition.* New York: Irvington.

Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject.* New York: Wiley.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Rosenthal, R., & Rosnow, R. L. (Eds.). (1969). *Artifact in behavioral research.* New York: Academic Press.

Rosenthal, R., & Rubin, D. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences, 3,* 377–415.

Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural "natural experiments" in economics. *Journal of Economic Literature, 38,* 827–874.

Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review, 40,* 337–354.

Rosnow, R. L., & Rosenthal, R. (1970). Volunteer effects in behavioral research. In K. H. Craik, B. Kleinmuntz, R. L. Rosnow, R. Rosenthal, J. A. Cheyne, & R. H. Walters (Eds.), *New directions in psychology* (pp. 213–277). New York: Holt, Rinehart and Winston.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research.* New York: W. H. Freeman.

Ross, S. (1999). *A first course in probability.* Upper Saddle River, NJ: Prentice Hall.

Rossi, P. H. (1951). *The application of latent structure analysis to the study of social stratification.* Unpublished doctoral dissertation, Columbia University.

Rossi, P. H. (1979). Vignette analysis: Uncovering the normative structure of complex judgments. In R. K. Merton, J. S. Coleman, & P. H. Rossi (Eds.), *Qualitative and quantitative social research: Papers in honor of Paul F. Lazarsfeld* (pp. 176–186). New York: Free Press.

Rossi, P. H., & Anderson, A. B. (1982). The factorial survey approach: An introduction. In P. H. Rossi & S. L. Nock (Eds.), *Measuring social judgments: The factorial survey approach* (pp. 15–67). Beverly Hills, CA: Sage.

Rossi, P. H., & Berk, R. A. (1985). Varieties of normative consensus. *American Sociological Review, 50,* 333–347.

Rossi, P. H., Freeman, H. E., & Lipsey, M. (1998). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.

Rossi, P. H., Sampson, W. A., Bose, C. E., Jasso, G., & Passel, J. (1974). Measuring household social standing. *Social Science Research, 3,* 169–190.

Rothenberg, J., & Rothenberg, D. (Eds.). (1983). *Symposium of the whole: A range of discourse toward an ethnopoetics.* Berkeley: University of California Press.

Rothman, K. J. (1977). Epidemiologic methods in clinical trials. *Cancer, 39,* 1771–1775.

Rothman, K. J. (1986). *Modern epidemiology.* Boston: Little, Brown.

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott.

Rousseau, J.-J. (1953) *The confessions.* Harmondsworth, UK: Penguin. (Original work published 1781)

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression & outlier detection.* New York: Wiley.

Routledge, P. (1996). The third space as critical engagement. *Antipode, 28*(4), 399–419.

Rowell, R. K. (1996). Partitioning predicted variance into constituent parts: How to conduct commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 33–44). Greenwich, CT: JAI.

Royston, J. P. (1993). A toolkit for testing for non-normality in complete and censored samples. *The Statistician, 42,* 37–43.

Rubenstein, A. (1998). *Modeling bounded rationality.* Cambridge: MIT Press.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics, 4,* 87–95.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology, 2,* 169–188.

Rubin, D. B., & Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association, 81,* 366–374.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo method.* New York: John Wiley.

Rudas, T. (1997). *Odds ratios in the analysis of contingency tables.* Thousand Oaks, CA: Sage.

Rummel, R. J. (1970). *Applied factor analysis.* Evanston, IL: Northwestern University Press.

Ruskey, F. (2001). A survey of Venn diagrams. *The Electronic Journal of Combinatorics,* Dynamic Survey #5 [Online]. Retrieved from http://www.combinatorics.org/Surveys/ds5/VennEJC.html

Russell, B. (1912). *Mysticism and logic.* London: Allen & Unwin.

Russett, B. (1969). Inequality and instability: The relation of land tenure to politics. In D. Rowney & J. Graham (Eds.), *Quantitative history: Selected readings in the quantitative analysis of historical data* (pp. 356–367). Homewood, IL: Dorsey.

Ryle, G. (1971). Thinking and reflecting. In G. Ryle, *Collected papers, Volume 2.* London: Hutchinson.

Ryssevik, J., & Musgrave, S. (2001). The social science dream machine: Resource discovery, analysis, and delivery on the Web. *Social Science Computing Review, 19*(2), 163–174.

Sackett, G. P. (1979). The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In J. D. Osofsky (Ed.), *Handbook of infant development* (1st ed., pp. 623–649). New York: Wiley.

Sacks, H. (1992). *Lectures on conversation* (2 vols.). Oxford, UK: Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696–735.

Sade, A. (1951). An omission in Norton's list of 7 × 7 squares. *Annals of Mathematical Statistics, 22,* 306–307.

Sale, J. E. M., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and Quantity, 36,* 43–53.

Salem, A., & Mount, T. (1974). A convenient descriptive model of income distribution: The gamma density. *Econometrica, 42,* 1115–1127.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century.* New York: Holt.

Sandelowski, M., Docherty, S., & Emden, C. (1997). Qualitative metasynthesis: Issues and techniques. *Research in Nursing & Health, 20,* 365–371.

Sankoff, D., & Kruskal, J. B. (1983). *Time warps, string edits, and macromolecules.* Reading, MA: Addison-Wesley.

Sapir, E. (1951). Culture, genuine and spurious. In D. G. Mandelbaum (Ed.), *Selected writings of Edward Sapir in language, culture and personality* (pp. 308–331). Berkeley: University of California Press.

Saris, W. E. (1991). *Computer-assisted interviewing.* Newbury Park, CA: Sage.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology 1987* (pp. 105–129). San Francisco: Jossey-Bass.

Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling.* New York: Springer-Verlag.

Saunders, M. N. K., & Cooper, S. A. (1993). *Understanding business statistics: An active learning approach.* London: DP Publications.

Saunders, P. (1979). *Urban politics: A sociological interpretation.* London: Hutchinson.

Saunders, P. T. (1980). *An introduction to catastrophe theory.* New York: Cambridge University Press.

Saussure, F. (1966). *Course in general linguistics.* New York: McGraw-Hill. (Originally published in 1915)

Saussure, F. de (1974). *Course in general linguistics.* London: Fontana/Collins.

Sayer, A. (1992). *Method in social science: A realist approach* (2nd ed.). London: Routledge.

Sayer, A. (1997). Essentialism, social constructionism and beyond. *Sociological Review, 45,* 453–487.

Sayer, A. (2000). *Realism and social science.* London: Sage.

Sayer, A. (2000). System, lifeworld and gender: Associational versus counterfactual thinking. *Sociology, 34,* 707–725.

Sayer, A. (2001). Reply to Holmwood. *Sociology, 35,* 967–984.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Scheaffer, R. L., Mendenhall, W., & Ott, L. (1990) *Elementary survey sampling.* Boston: PWS-Kent.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika, 40,* 87–104.

Scheffé, H. (1959). *The analysis of variance.* New York: Wiley.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist, 70,* 1075–1095.

Schegloff, E. A. (1989). Harvey Sacks—lectures 1964–1965: An introduction/memoir. *Human Studies, 12,* 185–209.

Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defence of intersubjectivity in conversation. *American Journal of Sociology, 97*(5), 1295–1345.

Schegloff, E.A. (1996). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology, 2,* 161–216.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1,* 143–186.

Schelling, T. C. (1978). *Micromotives and macrobehavior.* New York: Norton.

Schensul, J. J., & LeCompte, M. D. (Eds.). (1999). *Ethnographer's toolkit.* Walnut Creek, CA: AltaMira.

Schensul, S. L., Schensul, J. J., & LeCompte, M. D. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires.* Walnut Creek, CA: AltaMira.

Schervish, M. J. (1996). *P*-values: What they are and what are they not. *American Statistician, 50,* 203–206.

Schlesselman, J. J. (1982). *Case-control studies: Design, conduct, analysis.* New York: Oxford University Press.

Schmidt, P. (1976). *Econometrics.* New York: Marcel Dekker.

Schmidt, P., & Witte, A. D. (1988). *Predicting recidivism using survival models.* New York: Springer-Verlag.

Schneider, H. (1986). *Truncated and censored samples from normal populations.* New York: Marcel Dekker.

Schober, M., & Conrad, F. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61,* 576–602.

Schoenberg, R. (1977). Dynamic models and cross-sectional data: The consequences of dynamic misspecification. *Social Science Research, 6,* 133–144.

Schölkopf, B., & Smola, A. J. (2003). *Learning with kernel.* Cambridge: MIT Press.

Schrodt, P. A. (2002). Mathematical modeling. In J. B. Manheim, R. C. Rich, & L. Willnat (Eds.), *Empirical political analysis: Research methods in political science* (5th ed.). New York: Longman.

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. . (1986). *Understanding regression analysis: An introductory guide.* Beverly Hills, CA: Sage.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys.* New York: Academic Press.

Schuman, H., & Presser, S. (1996). *Questions & answers in attitude surveys.* Thousand Oaks, CA: Sage.

Schütz, A. (1963). Concept and theory formation in the social sciences. In M. A. Natanson (Ed.), *Philosophy of the social sciences* (pp. 231–249). New York: Random House.

Schutz, A. (1964). Don Quixote and the problem of reality. In *Collected papers, Vol. 2.* The Hague: Martinus Nijhoff.

Schütz, A. (1967). *Der sinnhafte Aufbau der sozialen Welt* [Phenomenology of the social world]. Evanston, IL: Northwestern University Press. (Originally published in 1932)

Schutz, A. (1973). *Collected papers* (4 vols.). Dordrecht, The Netherlands: Kluwer.

Schütz, A. (1976). Equality and the meaning structure of the social world. In *Collected Papers II* (pp. 226–273). The Hague: Martinus Nijhoff. (Originally published in 1955)

Schütz, A. (1976). *The phenomenology of the social world* (G. Walsh & F. Lehnert, Trans.). London: Heinemann.

Schwandt, T. (1996). Farewell to criteriology. *Qualitative Inquiry, 2,* 58–72.

Schwandt, T. A. (1998). The interpretive review of educational matters: Is there any other kind? *Review of Educational Research, 68*(4), 409–412.

Schwarz, N., & Sudman, S. (Eds.). (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Francisco: Jossey-Bass.

Schwarz, N., Hippler, H.-J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly, 49*(3), 388–395.

Scott, J. (1990). *A matter of record: Documentary sources in social research.* Cambridge, UK: Polity.

Scott, J. (1992). *Social network analysis.* London: Sage.

Scott, J. W. (1988). *Gender and the politics of history.* New York: Columbia University Press.

Scott, M. B., & Lyman, S. (1968). Accounts. *American Sociological Review, 33,* 46–62.

Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly, 19,* 321–325.

Scott, W. A. (1968). Attitude measurement. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Reading, MA: Addison-Wesley.

Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science, 130,* 477–482.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Scriven, M. (2001). Evaluation: Future tense. *The American Journal of Evaluation, 22,* 301–307.

Seal, H. L. (1967). The historical development of the Gauss linear model. *Biometrika, 54,* 1–23.

Seale, C. F. (1999). *The quality of qualitative research.* London: Sage.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components.* New York: Wiley.

Sechrest, L. (Ed.). (1979). *Unobtrusive measures today.* San Francisco: Jossey-Bass.

Segrè, E. (1970). *Enrico Fermi: Physicist.* Chicago: University of Chicago Press.

Sen, A. (1973). *On economic inequality.* New York: Norton.

Sewell, W. H. (1942). The development of a sociometric scale. *Sooциometry, 5*(3), 279–297.

Shadish, W. R. (2000). The empirical program of quasi-experimentation. In L. Bickman (Ed.), *Validity and social experimentation: Donald Campbell's legacy* (pp. 13–35). Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice.* Newbury Park, CA: Sage.

Shadish, W. R., Jr., Newman, D. L., Scheirer, M. A., & Wye, C. (1995). *Guiding principles for evaluators* (New Directions for Program Evaluation, No. 66). San Francisco: Jossey-Bass.

Shapiro, G., & Markoff, J. (1998). *Revolutionary demands: A content analysis of the Cahier de Doléances of 1789.* Stanford, CA: Stanford University Press.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52,* 591–611.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Newbury Park, CA: Sage.

Shaw, C. (1966). *The jack roller.* Chicago: University of Chicago Press.

Shaw, I., Bloor, M., Cormack, R., & Williamson, H. (1996). Estimating the prevalence of hard-to-reach populations: The illustration of mark-recapture methods in the study of homelessness. *Social Policy and Administration, 30*(1), 69–85.

Sheskin, D. J. (1997). *Handbook of parametric and nonparametric statistical procedures.* Boca Raton, FL: CRC Press.

Shneiderman, B. (1992). Tree visualization with treemaps: A 2-D space-filling approach. *ACM Transactions on Graphics, 11*(1), 92–99.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Shryock, H. S., Siegel, J. S., & Associates. (1973). *The methods and materials of demography.* New York: Academic Press. (Condensed edition by Edward G. Stockwell.)

Si, M., Neufeld, R. R., & Dunbar, J. (1999). Removal of bedrails on a short-term nursing home rehabilitation unit. *The Gerontologist, 39,* 611–614.

Sieber, J. E. (1992). *Planning ethically responsible research.* Newbury Park, CA: Sage.

Sieber, J. E. (1996). Typically unexamined communication processes in research. In B. H. Stanley, J. E. Sieber, & G. B. Melton (Eds.), *Research ethics: A psychological approach.* Lincoln: University of Nebraska Press.

Sieber, J. E. (Ed.). (1991). *Sharing social science data: Advantages and challenges.* Newbury Park, CA: Sage.

Siegel, J. S. (2002). *Applied demography.* San Diego, CA: Academic Press.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory.* Thousand Oaks, CA: Sage.

Silverman, D. (1993). *Interpreting qualitative data: Methods for analysing talk, text, and interaction.* London: Sage.

Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association, 49,* 467–479.

Simon, H. A. (1957). *Models of man.* New York: Wiley.

Simon, H. A. (1976). Discussion: Cognition and social behavior. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, NJ: Erlbaum.

Simons, H. (1996). The paradox of case study. *Cambridge Journal of Education, 26*(2), 225–240.

Sims, C. A. (1972). Money income and causality. *American Economic Review, 62,* 540–552.

Singer, E. (1978). Informed consent: Consequences for response rate and response quality in social surveys. *American Sociological Review, 43,* 144–162.

Singer, E., & Frankel, M. R. (1982). Informed consent procedures in telephone interviews. *American Sociological Review, 47,* 416–427.

Singer, E., vonThurn, D., & Miller, E. (1995). Confidentiality assurances and response. *Public Opinion Quarterly, 59,* 66–77.

Singleton, A. (1999). Measuring international migration: A case study of European cross-national comparisons. In D. Dorling & S. Simpson (Eds.), *Statistics in society: The arithmetic of politics* (pp. 148–158). London: Arnold.

Skinner, C. J., Holt, D., & Smith, T. M. F. (Eds.). (1989). *Analysis of complex surveys.* Chichester, UK: Wiley.

Sklodowska, E. (1992). *Testimonio hispanoamericano: Historia, teoría, poética.* New York: Peter Lang.

Skoudis, E. (2002). *Counter Hack: A step-by-step guide to computer attacks and effective defenses.* Upper Saddle, NJ: Prentice Hall.

Skvoretz, J., & Faust, K. (1999). Logit models for affiliation networks. *Sociological Methodology, 29,* 253–280.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5–11.

Smart, B. (2000). Postmodern social theory. In B. Turner (Ed.), *The Blackwell companion to social theory* (pp. 447–480). Oxford, UK: Blackwell.

Smelser, N. (1997). *Problems of sociology.* Berkeley: University of California Press.

Smith, B. H. (1997). *Belief and resistance: Dynamics of contemporary intellectual controversy.* Cambridge, MA: Harvard University Press.

Smith, D. (1997). Comment on Heckman's "Truth and method: Feminist standpoint theory revisited." *Signs, 22*(21): 392–397.

Smith, D. P. (1992). *Formal demography.* New York: Plenum.

Smith, G. (1997, Winter). Do statistics test scores regress toward the mean? *Chance,* pp. 42–45.

Smith, J. (1993). *After the demise of empiricism: The problem of judging social and educational inquiry.* Norwood, NJ: Ablex.

Smith, J. K. (1989). *The nature of social and educational inquiry: Empiricism versus interpretation.* Norwood, NJ: Ablex.

Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational enquirers. *Educational Researcher, 15,* 4–12.

Smith, J., & Deemer, D. (2000). The problem of criteria in the age of relativism. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 877–922). Thousand Oaks, CA: Sage.

Smith, L. M. (1978). An evolving logic of participant observation, educational ethnography and other case studies. *Review of Research in Education, 6,* 316–377.

Smith, M. (1998). *Social science in question.* London: Sage.

Smith, N. W. (2001). *Current systems in psychology.* Belmont, CA: Wadsworth.

Smith, S. K., Tayman, J., & Swanson, D. A. (2001). *State and local population projections: Methodology and analysis.* New York: Kluwer Academic/Plenum Publishers.

Smith, S., & Watson, J. (2001). *Reading autobiography: A guide for interpreting life narratives.* Minneapolis: University of Minnesota Press.

Smith, T. W. (1992). Thoughts on the nature of context effects. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 163–184). New York: Springer-Verlag.

Smithson, M. (1987). *Fuzzy set analysis for behavioral and social sciences.* New York: Springer-Verlag.

Smithson, M. (2003). *Confidence intervals* (Sage University Papers on Quantitative Applications in the Social Sciences, 07–140). Thousand Oaks, CA: Sage.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification.* San Francisco: Freeman.

Snedecor, G. W. (1937). *Statistical methods.* Ames: Iowa State University Press.

Snedecor, G. W. (1956). *Statistical methods* (5th ed.). Ames: Iowa State College Press.

Snider, J. G., & Osgood, C. E. (Eds.). (1969). *Semantic differential technique: A sourcebook.* Chicago: Aldine.

Snijders, T. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology, 31,* 361–395.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Snow, D. (1980). The disengagement process: A neglected problem in participant observation research. *Qualitative Sociology, 3,* 100–122.

Sobel, M. E. (1995). Causal inference in the social and behavioral sciences. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 1–38). New York: Plenum.

Sobel, M., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology 1985* (pp. 152–178). San Francisco: Jossey-Bass.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin, 38,* 1409–1438.

Sokal, R. R., & Rohlf, F. J. (1981). *Biometry* (2nd ed.). San Francisco: W.H. Freeman.

Sokal, R., & Sneath, P. (1963). *Principles of taxonomy.* San Francisco: Freeman.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin, 46,* 137–150.

Solomon, W. D. (1978). Ethics: Rules and principles. In W. T. Reich (Ed.), *Encyclopedia of bioethics.* New York: Free Press.

Sørensen, A. B. (1979). A model and a metric for the analysis of the intragenerational status attainment process. *American Journal of Sociology, 85,* 361–384.

Sørensen, J. B. (2002). The use and misuse of the coefficient of variation in organizational demography research. *Sociological Methods & Research, 30,* 475–491.

Soulliere, D., Britt, D. W., & Maines, D. R. (2001). Conceptual modeling as a toolbox for grounded theorists. *Sociological Quarterly, 42,* 233–251.

Soyland, A. J. (1994). *Psychology as metaphor.* London: Sage.

Spanos, A. (1999). *Probability and statistical inference: Econometric modeling with observational data.* Cambridge, UK: Cambridge University Press.

Spearman, C. (1904). General intelligence,objectively determined and measured. *American Journal of Psychology, 15,* 366–374.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15,* 72–101.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 12,* 271–295.

Special section: Speaking in the name of the real: Freeman and Mead on Samoa. (1983). *American Anthropologist, 85,* 908–947.

Spector, P. E. (1987). Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology, 72,* 438–443.

Spector, P. E. (1992). *Summated rating scale construction: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–082). Newbury Park, CA: Sage.

Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior, 15,* 385–392.

Spector, P. E., & Brannick, M. T. 91995). The nature and effects of method variance in organizational research. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology: 1995* (pp. 249–274). West Sussex, England: Wiley.

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management, 23,* 659–677.

Spiegel, M. R., Schiller, J., & Srinivasan, R. A. (2000). *Schaum's outline of probability and statistics* (2nd ed.). New York: McGraw-Hill.

Spradley, J. P. (1979). *The ethnographic interview.* New York: Holt, Rinehart and Winston.

SPSS base 10.0 user's guide. (1999). Chicago: SPSS.

SPSS. (1990). *SPSS reference guide.* Chicago: SPSS.

SPSS. (2001). *Reliability* (Technical section). Retrieved from www.spss.com/tech/stat/algorithms/11.0/reliability.pdf

SPSS. (2002, August). Crosstabs. In *SPSS 11.0 statistical algorithms* [Online]. Available: http://www.spss.com/tech/stat/algorithms/11.0/crosstabs.pdf

Stanley, L. (1992). *The auto/biographical I: Theory and practice of feminist auto/biography.* Manchester, UK: Manchester University Press.

Stanley, L., & Wise, S. (1993). *Breaking out again: Feminist ontology and epistemology.* London: Routledge Kegan Paul.

Stanton, A. L., Burker, E. J., & Kershaw, D. (1991). Effects of researcher follow-up of distressed subjects: Tradeoff between validity and ethical responsibility? *Ethics & Behavior, 1*(2), 105–112.

Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing.* New York: Wiley.

Stem, D. E., Jr., & Lamb, C. W., Jr. (1981). The marble-drop technique: A procedure for gathering sensitive information. *Decision Sciences, 12,* 702–707.

Stern, J., Stackowiack, R., & Greenwald, R. (2001). *Oracle essentials: Oracle9i, Oracle8i and Oracle 8.* Sebastopol, CA: O'Reilly.

Stevens, J. (2001). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 677–680.

Stevens, S. S. (1951). *Handbook of experimental psychology.* New York: Wiley.

Stevens, S. S. (1968). Measurement, statistics, and the schemapiric view. *Science, 161,* 849–861.

Stewart, D. K., & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin, 70,* 160–163.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice* (Applied Social Research Methods Series, Vol. 20). Newbury Park, CA: Sage.

Stigler, S. M. (1986). *The history of statistics:The measurement of uncertainty before 1900.* Cambridge, MA: Belknap Press of Harvard University Press.

Stigler, S. (1986). Laplace's 1774 memoir on inverse probability. *Statistical Science, 1,* 359–378.

Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods.* Cambridge, MA: Harvard University Press.

Stoll, D. (1999). *Rigoberta Menchú and the story of all poor Guatemalans.* Boulder, CO: Westview.

Stone, A. A., Shiffman, S., & DeVries, M. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 27–38). New York: Russell Sage.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (2000). *The science of self-report: Implications for research and practice.* Mahwah, NJ: Lawrence Erlbaum.

Stouffer, S. A. (1950). Some observations on study design. *American Journal of Sociology, 55,* 355–361.

Strauss, A. (1987). *Qualitative analysis for social scientists.* New York: Cambridge University Press

Strauss, A. L., Corbin, J., Fagerhaugh, S., Glaser, B., Maines, D., Suczek, B., & Weiner, C. (1984). *Chronic illness and the quality of life.* St. Louis, MO: Mosby.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Grounded theory procedures and techniques* (2nd ed.). Thousand Oaks, CA: Sage.

Stringer, E. T. (1996). *Action Research: A handbook for practitioners.* Thousand Oaks, CA: Sage.

Strohmetz, D. B., & Rosnow, R. L. (1994). A mediational model of research artifacts. In J. Brzezinski (Ed.), *Probability in theory-building: Experimental and nonexperimental approaches to scientific research in psychology* (pp. 177–196). Amsterdam: Editions Rodopi.

Stuart, A. (1984). *The ideas of sampling.* London: Griffin.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics* (Vol. 1). London: Edward Arnold.

Stuart, A., Ord, J. K., & Arnold, S. (1999). *Kendall's advanced theory of statistics: Vol. 2A. Classical inference & the linear model* (6th ed.). New York: Oxford University Press.

Studenmund, A. H. (1992). *Using econometrics: A practical guide* (2nd ed.). New York: HarperCollins.

Student. (1908). The probable error of a mean. In E. S. Pearson & J. Wishart (Eds.), *Student's collected papers.* London: University College.

Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs.* New York: Russell Sage.

Sudman, S. (1976). *Applied sampling.* New York: Academic Press.

Sudman, S., & Bradburn, N. (1982). *Asking questions.* San Francisco: Jossey-Bass.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Sudnow, D. (1978). *Ways of the hand.* Cambridge, MA: Harvard University Press.

Sudnow, D. (2001). *Ways of the hand* (Rev. ed.). Cambridge, MA: MIT Press.

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioural observation data.* London: Lawrence Erlbaum.

Suppes, P. (1970). *A probabilistic theory of causality.* Amsterdam: North Holland.

Swokowski, E. W. (1979). *Calculus with analytic geometry* (2nd ed.). Boston: Prindle, Weber & Schmidt.

Symon, G. J., & Clegg, C. W. (1991). Technology-led change: A study of the implementation of CADCAM. *Journal of Occupational Psychology, 64,* 273–290.

Szalai, A. (Ed.). (1972). *The use of time.* The Hague: Mouton.

Sztompka, P. (1990). *Robert Merton: An intellectual profile.* London: Macmillan.

Tabachnick, B. G., & Fidell, L. S. (2001). *Computer-assisted research design and analysis.* Boston: Allyn and Bacon.

Tacq, J. J. A. (1984). *Causaliteit in Sociologisch Onderzoek. Een Beoordeling van Causale Analysetechnieken in het Licht van Wijsgerige Opvattingen over Causaliteit* [Causality in sociological research: An evaluation of causal techniques of analysis in the light of philosophical theories of causality]. Deventer, The Netherlands: Van Loghum Slaterus.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis.* London: Sage.

Takane, Y. (1981). Multidimensional scaling of sorting data. In Y. P. Chaubey & T. D. Dwivedi (Eds.), *Topics in applied statistics.* Montreal: Concordia University Press.

Takane, Y., Young, F. W., & DeLeeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika, 42,* 7–67.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association, 82,* 528–550.

Taris, T. W. (2000). *A primer in longitudinal data analysis.* London: Sage.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Taylor, C. (1994). The politics of recognition. In C. Taylor & A. Gutmann (Eds.), *Multiculturalism and the politics of recognition* (pp. 25–73). Princeton, NJ: Princeton University Press.

Taylor, M. F., with Brice, J., Buck, N., & Prentice-Lane, E. (Eds.). (2000). *British Household Panel Survey user manual volume B9: Codebook.* Colchester, UK: University of Essex.

Taylor, S. J., & Bogdan, R. (1998). *Introduction to qualitative research methods: A guide & resource.* New York: Wiley.

Tedlock, D. (1999). Poetry and ethnography: A dialogical approach. *Anthropology and Humanism, 24*(2), 155–167.

ten Have, P. (1999). *Doing conversation analysis.* Thousand Oaks, CA: Sage.

Tenenhaus, M. (1998). *La régression PLS* [PLS regression]. Paris: Technip.

The 1980 U.S. census [Special issue]. (1992). *Survey Methodology, 18.*

The 1990 U.S. census [Special issue]. (1993). *Journal of the American Statistical Association, 88.*

The 1990 U.S. census [Special issue]. (1994). *Statistical Science, 9.*

The 2000 U.S. census. (2001). *Society, 39,* 2–53.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model.* New York: Springer-Verlag.

Theus, M. (1997). Visualization of categorical data. In W. Bandilla & F. Faulbaum (Eds.), *SoftStat '97: Advances in statistical software* (Vol. 6, pp. 47–55). Stuttgart, Germany: Lucius & Lucius.

Thom, R. (1975). *Structural stability and morphogenesis.* Reading, MA: Benjamin.

Thomas, H. (1997). Dancing: Representation and difference. In J. McGuigan (Ed.), *Cultural methodologies.* London: Sage.

Thomas, J. (1992). *Doing critical ethnography.* Newbury Park, CA: Sage.

Thomas, W. I., & Znaniecki, F. (1918–1920). *The Polish peasant in Europe and America* (5 vols.). New York: Dover.

Thomas, W. I., & Znaniecki, F. (1958). *The Polish peasant in Europe and America* (2 vols.). New York: Dover. (Original five-volume work published in 1918–1920).

Thompson, B. (1985). Alternate methods for analyzing data from experiments. *Journal of Experimental Education, 54,* 50–55.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement, 57,* 5–19.

Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 285–316). Washington, DC: American Psychological Association.

Thompson, J. (1981). *Critical hermeneutics: A study in the thought of Paul Ricoeur and Jürgen Habermas.* Cambridge, UK: Cambridge University Press.

Thompson, J. (1990). *Ideology and modern culture.* Stanford, CA: Stanford University Press.

Thompson, P. (2000). *The voice of the past* (3rd ed.). Oxford, UK: Oxford University Press.

Thorndike, R. L. (1967). The analysis and selection of test items. In D. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 201–216). New York: McGraw-Hill.

Thorne, S. (1994). Secondary analysis in qualitative research: Issues and implications. In J. Morse (Ed.), *Critical issues in qualitative research methods* (pp. 263–279). Thousand Oaks, CA: Sage.

Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology, 16,* 433–451.

Thurstone, L. L. (1926). The scoring of individual performance. *Journal of Educational Psychology, 17,* 446–457.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–286.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 23,* 529–554.

Thurstone, L. L. (1929). Fechner's law and the method of equal-appearing intervals. *Journal of Experimental Psychology, 12,* 214–224.

Thurstone, L. L. (1931). Measurement of social attitudes. *Journal of Abnormal and Social Psychology, 26,* 249–269.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitudes.* Chicago: University of Chicago Press.

Timm, N. H. (1975). *Multivariate analysis with applications in education and psychology.* Monterey, CA: Brooks/Cole.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20,* 410–433.

Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology, 22,* 358–376.

Tintner, G. (1968). *Methodology and mathematical economics and econometrics.* Chicago: University of Chicago Press.

Titon, J. (1980). The life story. *Journal of American Folklore, 93*(369), 276–292.

Titscher, S., Meyer, M., Wodak, R., & Vetter, R. (2000). *Methods of text and discourse analysis.* London: Sage.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica, 26,* 24–36.

Tolley, H. D., & Manton, K. G. (1992). Large sample properties of estimates of discrete grade of membership model. *Annals of the Institute of Statistical Mathematics, 41,* 85–95.

Tolley, H. D., Kovtun, M., Manton, K. G., & Akushevich, I. (2003). Statistical properties of grade of membership models for categorical data. Manuscript to be submitted to *Proceedings of National Academy of Sciences*, 2003.

Tommerup, P. (1993). *Adhocratic traditions, experience narratives and personal transformation: An ethnographic study of the organizational culture and folklore of the Evergreen State College, an innovative liberal arts college.* Unpublished doctoral dissertation, University of California, Los Angeles.

Toombs, S. K. (2001). Reflections on bodily change: The lived experience of disability. In S. K. Toombs (Ed.), *Handbook of phenomenology and medicine* (pp. 247–261). Boston: Kluwer Academic.

Toothaker, L. E. (1991). *Multiple comparisons for researchers.* Newbury Park, CA: Sage.

Toothaker, L. E. (1993). *Multiple comparison procedures.* Newbury Park, CA: Sage.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Toulmin, S. E. (1978). Science, philosophy of. In *Encyclopaedia Britannica* (15th ed., Vol. 16, pp. 375–393). Chicago: Britannica.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge, UK: Cambridge University Press.

Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin, 96,* 394–401.

Townsend, P., Davidson, N., & Whitehead, M. (1988). *Inequalities in health: The Black report and the health divide.* Harmondsworth: Penguin.

Tracy, K. (2002). *Everyday talk: Building and reflecting identities.* New York: Guilford.

Tracy, P. E., & Fox, J. A. (1981). The validity of sensitive measurements: A comparison of two measurement strategies. *American Sociological Review, 46,* 187–200.

Train, K. (1986). *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand.* Cambridge: MIT Press.

Tranfield, D., & Starkey, K. (1998). The nature, organization and promotion of management research: Towards policy. *British Journal of Management, 9,* 341–353.

Traugott, M. W., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly, 43*(3), 359–377.

Trice, A. D. (1987). Informed consent: VII: Biasing of sensitive self-report data by both consent and information. *Journal of Social Behavior and Personality, 2,* 369–374.

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology, 9,* 507–533.

Troyna, B. (1994). Reforms, research and being reflexive about being reflective. In D. Halpin & B. Troyna (Eds.), *Researching education policy.* London: Falmer.

Trussell, J., & Rodríguez, G. (1990). Heterogeneity in demographic research. In J. Adams, D. A. Lam, A. I. Hermalin, & P. E. Smouse (Eds.), *Convergent issues in genetics and demography* (pp. 111–132). Oxford, UK: Oxford University Press.

Truzzi, M. (1974). *Verstehen: Subjective understanding in the social sciences.* Reading, MA: Addison-Wesley.

Tsay, R. S. (2002). *Analysis of financial time series.* New York: Wiley.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences, 72,* 20–22.

Tsuji, R. (1999). Trusting behavior in prisoner's dilemma: The effects of social networks. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 60*(5-A), 1774. (UMI No. 9929113)

Tuchman, G. (1994). Historical social science: Methodologies, methods, and meanings. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 306–323). Thousand Oaks, CA: Sage.

Tucker, L. R. (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychometric scaling: Theory and applications* (pp. 155–167). New York: Wiley.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Tukey, J. W. (1958). Bias and confidence in not quite large samples [abstract]. *Annals of Mathematical Statistics, 29,* 614.

Tukey, J. W. (1960). A survey of sampling from contaminated normal distributions. In I. Olkin et al. (Eds.), *Contributions to probability and statistics.* Stanford, CA: Stanford University Press.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods.* Orlando, FL: Academic Press.

Turner, A. C. (1982). What subjects of survey research believe about confidentiality. In J. E. Sieber (Ed.), *The ethics of social research: Surveys and experiments* (pp. 151–165). New York: Springer-Verlag.

Turner, R. H. (1951). The quest for universals. *American Sociological Review, 18*(6), 604–611.

Turner, S. P., & Factor, R. A. (1981). Objective possibility and adequate causality in Weber's methodological writings. *Sociological Review, 29*(1), 5–28.

Turner, V., & Bruner, E. (Eds.). (1986). *The anthropology of experience.* Urbana: University of Illinois Press.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* 251–278.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin, 109,* 524–536.

U.S. Census Bureau. (2001). *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy* *on Adjustment for Non-Redistricting Uses* (With supporting documentation, Reps. 1–24). Washington, DC: Government Printing Office.

U.S. Federal Regulations of Human Research 45 CFR 46. Retrieved from http://ohrp.osophy.dhhs.gov.

UK Data Archive (n.d.). [Online]. Available: http://www.data-archive.ac.uk/.

United Nations Statistics Division. (2001). *Handbook on census management for population and housing censuses–Series F* (No. 83, Revision 1). New York: United Nations.

United Nations. (1995). *World population prospects: The 1994 revision.* New York: Author.

United Nations. (1996). *Demographic yearbook 1994.* New York: Author.

University of Mississippi. (n.d.). PsychExperiments. Retrieved from http://psychexps.olemiss.edu

Utts, J. M. (1996). *Seeing through statistics.* Belmont, CA: Duxbury.

Valero, P., Young, F., & Friendly, M. (2003). Visual categorical analysis in ViSta. *Computational Statistics and Data Analysis, 43*(4), 495–508.

Vallet, L.-A. (2001). Forty years of social mobility in France. *Revue Française de Sociologie: An Annual English Selection, 42*(Suppl.), 5–64.

Valsiner, J., & Benigni, L. (1986). Naturalistic research and ecological thinking in the study of child development. *Developmental Review, 6*(3), 203–223.

Van de Geer, J. (1971). *Introduction to multivariate analysis for the social sciences.* San Francisco: Freeman.

Van den Ende, H. W. (1971). *Beschrijvende Statistiek voor Gedragswetenschappen.* Amsterdam/Brussel: Agon Elsevier.

van der Ark, L. A., & van der Heijden, P. G. M. (1998). Graphical display of latent budget analysis and latent class analysis, with special reference to correspondence analysis. In M. Greenacre & J. Blasius (Eds.), *Visualization of categorical data* (pp. 489–508). San Diego: Academic Press.

van der Ark, L. A., van der Heijden, P. G. M., & Sikkel, D. (1999). On the identifiability of the latent model. *Journal of Classification, 16,* 117–137.

van der Heijden, P. G. M., Mooijaart, A., & de Leeuw, J. (1992). Constrained latent budget analysis. In P. Marsden (Ed.), *Sociological methodology* (pp. 279–320). Cambridge, UK: Basil Blackwell.

van der Heijden, P. G. M., van der Ark, L. A., & Mooijaart, A. (2002). Some examples of latent budget analysis and its extensions. In J. A. Hagenaars & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 107–136). Cambridge, UK: Cambridge University Press.

Van der Heijden, P. G. M., van Gils, G., Bouts, J., & Hox, J. J. (2000). A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning: Eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods & Research, 28,* 505–537.

Van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer.

Van Dijk, T. A. (Ed.). (1985). *Handbook of discourse analysis: Vols. 1–4.* London: Academic Press.

Van Maanen, J. (1979). The fact and fiction in organizational ethnography. *Administrative Science Quarterly, 24,* 539–550.

Van Maanen, J. (1988). *Tales of the field: On writing ethnography.* Chicago: University of Chicago Press.

van Manen, M. (1997). *Researching lived experience: Human science for an action sensitive pedagogy.* London, Ontario: Althouse.

van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard, 16*(40), 33–36.

Vapnik, V. N. (1999). *Statistical learning theory.* New York: Wiley.

Vaupel, J. W. (1990). Relatives' risks: Frailty models of life history data. *Theoretical Population Biology, 37,* 220–234.

Vaupel, J. W., & Yashin, A. I. (1985). The deviant dynamics of death in heterogenous populations. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 179–211). San Francisco: Jossey-Bass.

Vaupel, J. W., & Yashin, A. I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *American Statistician, 39,* 176–185.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography, 16,* 439–454.

Velleman, P. F. (1988). *DataDesk version 6.0 statistics guide.* Ithaca, NY: Data Description, Inc.

Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval and ratio typologies are misleading. *American Statistician, 47,* 65–72.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer-Verlag.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer.

Verbrugge, L. (1980). Health diaries. *Medical Care, 18,* 73–95.

Vermunt, J. K. (1997). *Log-linear models for event histories.* Thousand Oaks, CA: Sage.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge, UK: Cambridge University Press.

Vermunt, J. K., Rodrigo, M. F., & Ato-Garcia, M. (2001). Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods & Research, 30,* 170–196.

Vogt, P. W. (1993). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences.* Newbury Park, CA: Sage.

Vogt, W. P. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (2nd ed.). Thousand Oaks, CA: Sage.

Von Collani, E., & Drager, K. (2001). *Binomial distribution handbook.* Basel, Switzerland: Birkhauser.

von Glasersfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality.* New York: Norton.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behaviour.* Princeton, NJ: Princeton University Press.

von Wright, G. H. (1971). *Explanation and understanding.* Ithaca, NY: Cornell University Press.

Wachter, K. W., & Trussell, J. (1982). Estimating historical heights. *Journal of the American Statistical Association,* 77(378), 279–293.

Wackerly, D. D., Mendelhall, W., III, & Scheaffer, R. L. (1996). *Mathematical statistics with applications* (5th ed.). Belmont, CA: Duxbury.

Wahba, G. (1990). *Spline models for observational data.* Philadelphia: SIAM.

Walby, S., & Olsen, W. (2002, November). *The impact of women's position in the labour market on pay and implications for UK productivity.* Cabinet Office Women and Equality Unit. Retrieved from www.womenandequalityunit.gov.uk.

Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population. *Annals of Mathematical Statistics, 11,* 147–162.

Walker, E., & Lev, J. (1953). *Statistical inference.* New York: Henry Holt.

Walker, R. (Ed.). (1985). *Applied qualitative research.* Aldershot, UK: Gower.

Walkerdine, V., Lucey, H., & Melody, J. (2001). *Growing up girl: Psycho-social explorations of gender and class.* Basingstoke, UK: Palgrave.

Wallgren, A. et al. (1996). *Graphing statistics and data: Creating better data.* Thousand Oaks, CA: Sage.

Walsh, A., & Ollenburger, J. C. (2001). *Essential statistics for the social and behavioral sciences: A conceptual approach.* Upper Saddle River, NJ: Prentice Hall.

Walt, S. M. (1987). *The origins of alliances.* Ithaca, NY: Cornell University Press.

Walton Braver, M., & Braver, S. L. (1996). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Psychological Bulletin, 104*(1), 150–154.

Walton, J. (1992). Making the theoretical case. In C. C. Ragin & H. S. Becker (Eds.), *What is a case? Exploring the foundations of social inquiry* (pp. 121–137). New York: Cambridge University Press.

Ward, A. (2002). *Oral history and copyright.* Retrieved June 7, 2002 from http://www.oralhistory.org.uk

Warde, A., Martens, L., & Olsen, W. K. (1999). Consumption and the problem of variety: Cultural omnivorousness, social distinction and dining out. *Sociology, 30*(1), 105–128.

Warner, R. M. (1998). *Spectral analysis of time-series data.* New York: Guilford.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60,* 63–69.

Warren, C. A. B. (1987). *Madwives: Schizophrenic women in the 1950s.* New Brunswick, NJ: Rutgers University Press.

Warren, C. A. B. (2002). Qualitative interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 83–101). Thousand Oaks, CA: Sage.

Warren, C. A. B., & Hackney, J. K. (2000). *Gender issues in ethnography* (Qualitative Research Methods Series vol. 9, 2nd ed.). Thousand Oaks, CA: Sage.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* New York: Cambridge University Press.

Wasserman, S., & Galaskiewicz, J. (Eds.). (1994). *Advances in social network analysis.* Thousand Oaks, CA: Sage.

Watson, J. B. (1959). *Behaviorism.* Chicago: University of Chicago Press.

Watson, M. W. (1994). Vector autoregression and cointegration. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, pp. 2843–2915). Amsterdam: Elsevier.

Watts, D. (1999). *Small worlds: The dynamics of networks between order and randomness.* Princeton, NJ: Princeton University Press.

Weaver, L., & Cousins, B. (2001, November). *Unpacking the participatory process.* Paper presented at the annual meeting of the American Evaluation Association, St. Louis, MO.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive measures in the social sciences.* Chicago: Rand McNally.

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences* (2nd ed.). Boston: Houghton Mifflin.

Webb, J. F., Khazen, R. S., Hanley, W. B., Partington, M. S., Percy, W. J. L., & Rathborn, J. C. (1973). PKU screening—is it worth it? *Canadian Medical Association Journal, 108,* 328–329.

Weber, M. (1949). Die "Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis ["Objectivity" in social science and social policy]. In *The methodology of the social sciences: Max Weber* (pp. 50–112). New York: Free Press. (Originally published in 1904)

Weber, M. (1964). *The theory of social and economic organization* (A. M. Henderson & T. Parsons, Trans.). New York: Free Press.

Weber, M. (1975). *Roscher and Knies.* New York: Free Press.

Weber, R. P. (1990). *Basic content analysis.* Newbury Park, CA: Sage.

Wedel, M., & DeSarbo, W. S. (1994). A review of recent developments in latent class regression models. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 352–388). Cambridge, UK: Basil Blackwell.

Weisberg, H. F. (1974). Dimensionland: An excursion into spaces. *American Journal of Political Science, 18,* 743–776.

Weisberg, H. F. (1974). Models of statistical relationship. *American Political Science Review, 68,* 1638–1655.

Weisberg, H. F. (1992). *Central tendency and variability* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–083). Newbury Park, CA: Sage.

Weisberg, S. (1985). *Applied linear regression.* New York: Wiley.

Weiss, C. (1977). *Use of social research in public policy.* Lexington, MA: D. C. Heath.

Weiss, C. H. (1972). A treeful of owls. In C. H. Weiss (Ed.), *Evaluating action programs* (pp. 3–27). Boston: Allyn & Bacon.

Weiss, C. H., & Bucuvalas, M. (1980). Truth test and utility test: Decision makers' frame of reference for social science research. *American Sociological Review, 45,* 302–313.

Weiss, H. M. (2002). Deconstructing job satisfaction: Separating evaluations, beliefs and affective experiences. *Human Resources Management Review, 12,* 173–194.

Weiss, R. (1995). *Learning from strangers.* New York: Free Press.

Weitzman, E. A., & Miles, M. B. (1995). *Computer programs for qualitative data analysis: A software source book.* Thousand Oaks, CA: Sage.

Welch, B. L. (1947). The generalization of student's problem when several different population variances are involved. *Biometrika, 34,* 28–35.

Weller, S. C., & Romney, A. K. (1988). *Systematic data collection* (Qualitative Research Methods, Vol. 10). Newbury Park, CA: Sage.

Weller, S. C., & Romney, A. K. (1990). *Metric scaling: Correspondence analysis.* Newbury Park, CA: Sage.

Wellman, B., & Berkowitz, S. D. (Eds.). (1997). *Social structures: A network approach* (updated ed.). Greenwich, CT: JAI.

Wengraf, T. (2001). *Qualitative research interviewing: Biographic-narrative and semi-structured method.* London: Sage.

Werner, O., & Schoepfle, G. M. (1987). *Systematic fieldwork* (2 vols.). Newbury Park, CA: Sage.

West, S. G., Hepworth, J. T., McCall, M. A., & Reich, J. W. (1989). An evaluation of Arizona's July 1982 drunk driving law: Effects on the City of Phoenix. *Journal of Applied Social Psychology, 19,* 1212–1237.

Wetherell, M., & Potter, J. (1992). *Mapping the language of racism: Discourse and the legitimation of exploitation.* Brighton, UK: Harvester Wheatsheaf.

Wetherell, M., Taylor, S., & Yates, S. J. (Eds.). (2001). *Discourse theory and practice.* London: Sage.

Wheldall, K., & Merrett, F. (1989). *Positive teaching: The behavioural approach.* Birmingham, UK: Positive Products.

Wherry, R. J., Sr. (1984). *Contributions to correlational analysis.* New York: Academic Press.

Whewell, W. (1847). *The philosophy of the inductive sciences* (Vols. 1 & 2). London: Parker.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48,* 817–838.

White, H. (1984). *Asymptotic theory for econometricians.* Orlando, FL: Academic Press.

White, L. (1999). *Political analysis: Technique and practice* (4th ed.). Orlando, FL: Harcourt Brace.

Whiting, J. W., Child, I. L., & Lambert, W. W. (1966). *Field guide for the study of socialization.* New York: Wiley.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics.* Chichester, UK: Wiley.

Whyte, W. F. (1955). *Street corner society: The social structure of an Italian slum* (2nd ed.). Chicago: University of Chicago Press. (Original work published 1943)

Whyte, W. F. (1984). *Learning from the field: A guide from experience.* Beverly Hills, CA: Sage.

Whyte, W. F. (Ed.). (1991). *Participatory action research.* Newbury Park, CA: Sage.

Wiggershaus, R. (1995). *The Frankfurt School: Its history, theories and political significance* (M. Robertson, Trans.). Cambridge, UK: Polity.

Wiggins, L. M. (1973). *Panel analysis.* Amsterdam: Elsevier.

Wilcox, R. (2001). *Fundamentals of modern statistical methods: Substantially improving power and accuracy.* New York: Springer.

Wilcox, R. R. (1996). *Statistics for the social sciences.* San Diego, CA: Academic Press.

Wilcox, R. R. (2001). *Fundamentals of modern statistical methods.* New York: Springer.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques.* San Diego, CA: Academic Press.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics, 1,* 80–83.

Wildt, A. R., & Ahtola, O. T. (1978). *Analysis of covariance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–012). Beverly Hills, CA: Sage.

Willer, D. (1967). *Scientific sociology: Theory and method.* Englewood Cliffs, NJ: Prentice Hall.

Williams, G. (1984). The genesis of chronic illness: Narrative re-construction. *Sociology of Health & Illness, 6*(2), 175–200.

Williams, L. J., & Brown, B. K. (1994). Method variance in organizational behavior and human resources research: Effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes, 57,* 185–209.

Williams, M. (1998). The social world as knowable. In T. May & M. (Eds.), *Knowing the social world* (pp. 5–21). Buckingham, UK: Open University Press.

Williams, M. (2000). Interpretivism and generalisation. *Sociology, 34*(2), 209–224.

Williams, M. (2000). *Science and social science: An introduction.* London: Routledge.

Williams, M., & May, T. (1996). *Introduction to the philosophy of social research.* London: UCL Press.

Williams, R. (1977). *Marxism and literature.* Oxford, UK: Oxford University Press.

Willson, V. L., & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal, 19,* 249–258.

Wilson, P. J. (1974). *Oscar: An inquiry into the nature of sanity.* New York: Vintage.

Wilson, T. P. (1969). A proportional reduction in error interpretation for Kendall's tau-b. *Social Forces, 47,* 340–342.

Winch, P. (1990). *The idea of a social science and its relation to philosophy.* London: Routledge. (Originally published in 1958)

Winer, B. J. (1962). *Statistical principles in experimental designs.* New York: McGraw-Hill.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology, 18,* 327–350.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology, 25,* 659–706.

Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research, 23*(2), 230–257.

Wiseman, F., Moriarty, M., & Schafer, M. (1975–1976). Estimating public opinion with the randomized response model. *Public Opinion Quarterly, 39*(4), 507–513.

Wishart, J. (1938). Growth-rate determinations in nutritional studies with the Bacon pig, and their analysis. *Biometrika, 30,* 16–28.

Wittgenstein, L. (1974). *On certainty* (G. E. M. Anscombe & G. H. Von Wright, Eds.). Oxford, UK: Blackwell.

Wodak, R., & Meyer, M. (2001). *Methods in critical discourse analysis.* Thousand Oaks, CA: Sage.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiaah (Ed.), *Multivariate analysis* (pp. 391–420). New York: Academic Press.

Wolf, M. (1992). *A thrice told tale: Feminism, postmodernism and ethnographic responsibility.* Stanford, CA: Stanford University Press.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixtures. *Multivariate Behavioral Research, 5,* 329–350.

Wonnacott, T. H., & Wonnacott, R. J. (1990). *Introductory statistics for business and economics* (4th ed.). San Francisco: Jossey-Bass.

Wood, B. D. (1992). Modeling federal implementation as a system: The clean air case. *American Journal of Political Science, 1,* 40–67.

Wood, J. W., & Weinstein, M. (1990). Heterogeneity in fecundability: The effect of fetal loss. In J. Adams, D. A. Lam, A. I. Hermalin, & P. E. Smouse (Eds.), *Convergent issues in genetics and demography* (pp. 171–188). Oxford, UK: Oxford University Press.

Wood, J. W., Holman, D. J., Yasin, A., Peterson, R. J., Weinstein, M., & Chang, M.-C. (1994). A multistate model of fecundability and sterility. *Demography, 31,* 403–426.

Wood, S. (2001). mgcv: GAMs and generalized ridge regression for R. *R News, 1,* 20–25.

Woodbury, M. A., & Clive, J. (1974). Clinical pure types as a fuzzy partition. *Journal of Cybernetics, 4*(3), 111–121.

Woodward, J. (1995). Causation and explanation in econometrics. In D. Little (Ed.), *On the reliability of economic models: Essays in the philosophy of economics* (pp. 9–61). Boston: Kluwer Academic.

Wooldridge, J. M. (2000). *Introductory econometrics: A modern approach.* Cincinnati, OH: South-Western.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data.* Cambridge: MIT Press.

Wooldridge, J. M. (2003). *Introductory econometrics: A modern approach* (2nd ed.). Cincinnati, OH: South-Western College Publishing.

Woolgar, S. (Ed.). (1988). *Knowledge and reflexivity.* London: Sage.

World Bank, International Economics Department. (1997). *World tables of economic and social indicators, 1950–1992* [Computer file]. Ann Arbor, MI: Interuniversity Consortium for Political and Social Science Research.

Wright, B. D. (1997). S. S. Stevens revisited. *Rasch Measurement Transactions, 11,* 552–553.

Wright, S. (1918). On the nature of size factors. *Genetics, 3,* 367–374.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5,* 161–215.

Wright, W. (1975). *Six guns and society.* Berkeley: University of California Press.

Wu, L. L., & Martinson, B. C. (1993). Family structure and the risk of a premarital birth. *American Sociological Review, 58*(2), 210–232.

Wu, L. L., & Tuma, N. B. (1990). Local hazard models. In C. C. Clogg (Ed.), *Sociological methodology 1990* (pp. 141–180). Oxford, UK: Basil Blackwell.

Wylie, R. (1974). *The self concept: A review of methodological considerations and measuring instruments.* Lincoln: University of Nebraska Press.

Xie, Y. (1989). An alternative purging method: Controlling the composition-dependent interaction in an analysis of rates. *Demography, 26,* 711–716.

Xie, Y. (1992). The log-multiplicative layer effect model for comparing mobility tables. *American Sociological Review, 57,* 380–395.

Xie, Y., & Pimentel, E. E. (1992). Age patterns of marital fertility: Revisingthe Coale-Trussell method. *Journal of the American Statistical Association, 87,* 977–984.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different classes. *Journal of the American Statistical Association, 29,* 51–66.

Yeo, I.-K., & Johnson, R. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika, 87,* 954–959.

Yin, R. K. (1989). *Case study research: Design and methods.* Newbury Park, CA: Sage.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage.

Yin, R. K. (1998). The abridged version of case study research. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 229–259). Thousand Oaks, CA: Sage.

Young, F. W., & Hamer, R. M. (1987). *Multidimensional scaling: History, theory, and applications.* Hillsdale, NJ: Lawrence Erlbaum.

Young, K., Ashby, D., Boaz, A., & Grayson, L. (2002). Social science and the evidence-based policy movement. *Social Policy and Society, 1*(3), 215–224.

Yuen, K. K. (1974). The two-sample trimmed *t* for unequal populations variances. *Biometrika, 61,* 165–170.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika, 2,* 121–134.

Yule, G. U. (1971). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. In A. Stuart & M. Kendall (Eds.), *Statistical papers of George Udny Yule* (pp. 389–420). New York: Hafner. (Original work published 1927)

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Griffin.

Yule, G. U., & Kendall, M. G. (1968). *An introduction to the theory of statistics.* New York: Hafner.

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8,* 338–353.

Zaller, J. R. (1992). *The nature and origins of mass opinion.* Cambridge, UK: Cambridge University Press.

Zaman, A. (1996). *Statistical foundations for econometric techniques.* New York: Academic Press.

Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology, 29,* 703–709.

Zeeman, E. C. (1972). Differential equations for the heartbeat and nerve impulse. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 8–67). Edinburgh, UK: Edinburgh University Press.

Zeisel, H., & Kaye, D. H. (1997). *Prove it with figures.* New York: Springer.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association, 57,* 348–368.

Zellner, A., & Theil, H. (1962). Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica, 30,* 54–78.

Zetenyi, T. (Ed.). (1988). *Fuzzy sets in psychology.* Amsterdam: North-Holland.

Zimmerman, D. H., & Wieder, D. (1977). The diary-interview method. *Urban Life, 5,* 479–498.

Zimmerman, M. (2001–2002). Rigoberta Menchú, David Stoll, subaltern narrative and testimonial truth: A personal testimony. *Antípodas: Journal of Hispanic and Galician Studies, 13/14,* 103–124.

Znaniecki, F. (1934). *The method of sociology.* New York: Farrar and Rhienhart.

Znaniecki, F. (1969). *On humanistic sociology: Selected papers* (R. Bierstadt, Ed.). Chicago: University of Chicago Press.

Zorn, C. J. W. (2001). Generalized estimating equation models for correlated data: A review with applications. *American Journal of Political Science, 45,* 470–490.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.

Zwillinger, D. (1995). *CRC standard mathematical tables and formulae* (30th ed.). Boca Raton, FL: CRC Press.