

Sociology 542
Analysis of Sociological Data II
Fall 2009
Tuesday 9:50-12:30; Lucy Stone Hall A256

Professor: Deborah Carr
Lucy Stone Hall (LSH) A332
Office Phone: 932-4068
E-mail: carrds@sociology.rutgers.edu
Office Hours: Tuesday 1:30-3:30 p.m.

Lab/SPSS Assistant: Neha Gondal
Email: ngondal@sociology.rutgers.edu

Course Description: This is the second in a two-course series of graduate-level statistics and data analysis. The course begins with an in-depth examination of data sources, content, and limitations. We will also review basic statistical concepts and will then move onto multiple regression, the main focus of the course. In the final weeks, we will cover the basic conceptual underpinnings and purposes of more advanced statistical methods, including binary and multinomial logistic regression, event history analysis, and structural equation modeling.

There is only one way to become a proficient quantitative sociologist and data analyst: by diving into the process and working one's way through the intricacies of preparing, coding, analyzing, presenting, and interpreting the data. This process of discovery is only possible, however, with a strong grasp of the statistical concepts that underlie one's analyses, and a basic knowledge of the computer software packages that generate those analyses.

This is not a course on how to use SPSS, although we will devote much of the course to interpreting the output generated from the computer. This also is not a math course, although we will learn and apply core concepts of basic algebra. Finally, this is not a course in a substantive area of sociology, although we will be reading about a variety of important substantive topics, and will be using statistical tools to test hypotheses and further our understanding of substantive questions. Rather, this is a course on how to conduct theoretically-guided methodologically rigorous social science research, using quantitative research methods and high-quality survey (or administrative) data.

Prerequisites: Sociology 541 (Analysis of Sociological Data I) or its equivalent. You should be familiar with basic descriptive statistics, inferential statistics, and simple bivariate regression.

Readings: Required reading and supplemental recommended readings are noted below.. Students must complete each week's readings prior to the start of class, and should be prepared to both ask and answer questions on the readings during the class session. The four required books are available at the Livingston College Bookstore. The required articles and chapters are available at the Soc 542 Sakai website.

Allison, Paul D. 1999. *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge Press. ISBN: 978-0761985334.

Kiecolt, K. Jill and Laura E. Nathan. 1985. *Secondary Analysis of Survey Data*. Newbury Park, CA: Sage Publications. ISBN: 0803923023.

Miller, Jane E. 2005. *The Chicago Guide to Writing About Multivariate Analysis*. Chicago: University of Chicago Press. ISBN: 978-0226527833.

Norusis, Marija J. 2009. *SPSS 17.0 Guide to Data Analysis*. Upper Saddle River, NJ: Prentice-Hall. ISBN: 978-0321621436.

Recommended Reference Book: This book is a must-own for anyone who reads sociological journals, conducts sociological research, or simply wants to be able understand what quantitative sociologists do.

Vogt, W. Paul. 2005. *Dictionary of Statistics and Methodology: A Non-technical Guide for the Social Sciences*, 3rd edition. Newbury Park, CA: Sage Publications. ISBN: 978-0761988557.

Overview of Course Requirements and Grading: The course is organized as a combination of a lecture class and a “working group” meeting. I will provide a 1.5 to 2 hour lecture on each week’s topics, followed by a discussion where research challenges are puzzled out and, hopefully, resolved. During the discussion, students are encouraged to raise questions about the readings and lecture content, and should talk freely about their challenges with the past week’s data analysis activities. Course grades will be based on three major requirements: participation in discussion, weekly homework assignments, and the final course paper (see details below). I will not give grades of “incomplete.” In the “real world,” journal and book editors, publishers, and grant proposal reviewers do not give breaks to those who cannot meet deadlines.

(1) Participation in weekly discussion (5 percent of class grade).

This grade reflects participation in discussion and the quality of both the questions and answers offered during class. Students should read carefully and complete all homework assignments so that they come to class with insightful comments and questions. Students who read materials that are not on the syllabus but that are germane to the discussion should feel free to contribute this knowledge to the discussion; the class as a whole will benefit from the specialized knowledge of all participants.

(2) Completion of weekly homework assignments (55 percent of class grade).

Each week, students are required to complete a homework assignment. These assignments may include simple algebraic calculations, presentation of descriptive data, and analysis of one’s data set. Analyses are to be conducted using SPSS, unless a student has a persuasive reason for choosing another software package. It is critically important that each homework assignment is completed carefully and on deadline. Each of these homework assignments builds on the past assignment, and will comprise a critical component of the final research paper. Assignments will

be distributed and discussed in class, one week prior to deadline. I anticipate that we will have 11 assignments with each accounting for 5 percent of the course grade. However, the content, frequency, and timing of the homework assignments is contingent upon students' comprehension of and comfort with the course material, as demonstrated in the preceding week's class.

(3) Final research paper or proposal (40 percent of class grade).

The main course requirement is the preparation of an original empirical research paper (roughly 25 pages, double spaced). Each student will present a brief synopsis of his or her research project during the final class session. Paper topic, methodology, and data source will be chosen in consultation with the instructor. The paper should define a research question, review relevant prior studies, analyze suitable data, and present the findings and their implications. The centerpiece of the analysis **MUST** be a multivariate analysis (i.e., OLS regression, binary logistic regression, or multinomial logistic regression). Students are encouraged to use this class project to develop a qualifying paper or journal article. Completion of the paper will proceed in five steps.

1. Each student will submit a brief (2-3 page) research proposal **by the start of class on September 22**. This brief proposal should describe your research question, the data to be used, and the analyses to be undertaken. You are encouraged to meet with me early in the semester to discuss your research plan.
2. An annotated bibliography summarizing at least five seminal sources related to your research question. This will help to ensure that you are deriving hypotheses and models with reference to the existing literature. The annotated bibliography is due **by the start of class on October 20**.
3. **OPTIONAL**: A first draft of the paper may be submitted **by 5 p.m. on Friday December 4**. Detailed feedback will be given to students by Monday December 7.
4. Each student will make a 10-15 minute presentation of his or her research to the class on **December 8**.
5. The final draft of the paper is due **no later than Friday December 12 at 5 p.m.**

Data Sources: All students are **required** to conduct an original analysis of secondary data. Many data sets are available through University of Michigan's Inter-University Consortium for Political and Social Research (<http://www.icpsr.umich.edu>). Information on accessing data sets will be made available during the first two weeks of the course. Among these data sets are:

Adolescent Health Study (AD-Health)
Americans' Changing Lives (ACL)
Changing Lives of Older Couples (CLOC)
General Social Survey (GSS)
Health and Retirement Survey (HRS)
Longitudinal Study of Aging (LSOA)
Mexican Health and Aging Study (MHAS)
Midlife in the United States (MIDUS)

Monitoring the Future (MTF)
National Longitudinal Study (NLS)
National Survey of Black Americans (NSBA)
National Survey of Families and Households (NSFH)
Panel Study of Income Dynamics (PSID)
Wisconsin Longitudinal Study (WLS).

READING SCHEDULE

September 1: Introduction: Linking Data and Theory

Objectives: (1) To derive testable hypotheses from sociological theories, particularly midrange theories; (2) To develop operationalized measures from theoretical concepts; (3) To understand the process through which broad sociological ideas are translated into conceptual models, and eventually into statistical models; (4) To understand the role of the Institutional Review Board (IRB) and prepare requests for IRB approval/exemption.

Required Reading:

- (SK) Bryson, Bethany. 1996. "‘Anything but Heavy Metal’: Symbolic Exclusion and Musical Dislikes." *American Sociological Review* 61: 994-899.
- (SK). Carr, Deborah et al. 2009. "Research Methods." Pp. 465-481 in *Encyclopedia of the Life Course and Human Development: Volume 3*, edited by Deborah Carr, Robert Crosnoe, M. E. Hughes, and Amy Pienta). Farmington Hills, MI: Gale/CENGAGE Learning.
- (SK) Cox, D. R. 1990. "Role of Models in Statistical Analysis." *Statistical Science* 5(2): 169-174.
- (SK) Gross, Neil and Solon Simmons. 2002. "Intimacy as a Double-Edged Phenomenon? An Empirical Test of Giddens." *Social Forces* 81(2): 531-555.

Recommended Reading:

- (SK) Oakes, J. Michael Oakes. 2002. "Risks and Wrongs in Social Science Research: An Evaluator’s Guide to the IRB." *Evaluation Research* 26: 443-479.

September 15: Data Options

Objectives: (1) To choose a data set for your course paper; (2) To understand the basic principles of collecting survey data, including sampling frame, response rate, and attrition; (3) To understand the strengths and weaknesses of cross-sectional, repeated cross-section, longitudinal, and cohort studies; (4) To begin preparing secondary data for your own empirical analysis.

Required Reading:

- (SK) Carr, Deborah. 2006. "Methodological Issues in Studying Bereavement." Pp. 19-47 in *Spousal Bereavement in Late Life*, edited by Deborah Carr, Randolph Nesse, and Camille B. Wortman. New York: Springer.
- Kiecolt, K. Jill and Laura E. Nathan. 1985. *Secondary Analysis of Survey Data*. Newbury Park, CA: Sage Publications. (Pp. 9-14, and 47-76).
- (SK) Wienke, Chris and Gretchen J. Hill, G. J. 2009. "Does the ‘Marriage Benefit’ Extend to Partners in Gay and Lesbian Relationships? Evidence from a Random Sample of Sexually Active Adults." *Journal of Family Issues* 30: 259-289.

Miller, Jane. 2005. Chapter 12: Writing About Data and Methods (Pp. 272-300). [We will revisit this chapter throughout the semester]

Norusis, Marija. 2009. Chapter 3: Sources of Data. Pp. 31-46.

Website for Inter-University Consortium for Political Research at University of Michigan:
<http://www.icpsr.umich.edu/>

Recommended Reading:

(SK) Smith, Emma. 2008. "Pitfalls and Promises: The Use of Secondary Data Analysis in Education Research." *British Journal of Educational Sciences* 56: 323-339.

September 22: Understanding Variables: What They Are, What They Do, What They Look Like, and What to Do When They're Missing.

Objectives: (1) To understand the differences and similarities among independent, dependent, mediating, confounding, and control variables; (2) To identify and recognize the appropriate use of continuous, ordinal, and categorical variables; (3) To develop an understanding of scale construction, use, and replicability; (4) To understand the diverse types of missing data, and basic strategies for handling missing data.

Required Reading:

(SK) Horwitz, Allan. 2002. "Outcomes in the Sociology of Mental Health and Illness: Where Have We Been and Where Are We Going?" *Journal of Health and Social Behavior* 43(2): 125-42.

(SK) deWit, Leonore, M., Annemieke van Straten, Marieke van Herten, Brenda WJH Penninx, and Pim Cuijpers. 2009. "Depression and Body Mass Index, a U-Shaped Association." *BMC Public Health* 9:1-6.

Miller, Jane. 2005. Chapter 3: Causality, Significance, and Substantive Significance (Pp. 34-41).

Norusis, Marija. 2009. Chapter 4: Counting Responses (Pp. 47-68); Chapter 5: Computing Descriptive Statistics (Pp. 79-96); Appendix A: Obtaining Charts in SPSS (Pp.579-595); and Appendix B: Transforming and Selecting Data (Pp. 597-621).

(SK) Reidpath, Daniel D., David Crawford, Linda Tilgner, and Carl Gibbons. 2002. "Relationship between Body Mass Index and the Use of Healthcare Services in Australia." *Obesity Research* 10: 526-31.

(SK) Ryff, Carol D. 1989. "Happiness is Everything, or is it? Explorations on the Meaning of Psychological Well-Being." *Journal of Personality and Social Psychology* 57: 1069-1081.

Recommended Reading:

(SK) Link, Bruce G., Jo C. Phelan, Richard Miech, and Emily Leckman Westin. 2008. "The Resources that Matter: Fundamental Social Causes of Health Disparities and the Challenge of Intelligence." *Journal of Health & Social Behavior* 49: 72-91.

(SK) Sabatelli, Ronald M. 1988. "Measurement Issues in Marital Research: A Review and Critique of Contemporary Survey Instruments." *Journal of Marriage and Family* 50: 891-915. [Scan only].

September 29: Simple Regression

Objectives: (1) To understand the basic assumptions of ordinary least squares (OLS) regression; (2) To interpret regression coefficients; (3) To understand the relationship between multiple regression and correlation; (4) To recognize the importance of statistically significant and not statistically significant predictor variables.

Allison, Paul. D. 1999. Chapter 4: How Do I Run a Multiple Regression (Pp. 71-96); Chapter 5: How Does Bivariate Regression Work? (Pp. 97-117); Chapter 9: How is Multiple Regression Related to Other Statistical Techniques (Pp. 175-177).

(SK) Colwell, John and Makiko Kato. 2005. "Video Game Play in British and Japanese Adolescents." *Simulation & Gaming* 36: 518-530.

Miller, Jane. 2005. Chapter 3: Causality, Significance, and Substantive Significance (Pp. 41-49); Chapter 4: Five More Technical Principles (Pp. 50-67); Chapter 9: Quantitative Comparisons for Multivariate Models (Pp. 208-215).

Norusis, Marija J. 2009. Chapter 20: Linear Regression and Correlation. (Pp. 443-476).

October 6: Multiple Regression

Objectives: (1) To understand the purposes and form of multiple regression; (2) To understand the process of model building and interpreting results; (3) To recognize and address problems including multicollinearity.

Allison, Paul. 1999. Chapter 1: What is Multiple Regression (Pp. 1-23); Chapter 2: How Do I Interpret Multiple Regression Results (Pp. 25-49); Chapter 7: What Can Be Done about Multicollinearity (Pp. 137-151).

(SK) Freedman, David A. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology* 21: 291-313.

Norusis, Marija. 2009. Chapter 21: Testing Regression Hypotheses (Pp. 477-498) and Chapter 23: Building Multiple Regression Models (Pp. 523-558).

October 13: Multiple Regression (cont'd)

Objectives: (1) To understand and remedy issues including sample bias; (2) To understand the use of sampling weights; and (3) To recognize the problems posed by outliers.

Allison, Paul. 1999. Chapter 6: What are the Assumptions of Multiple Regression (Pp. 119-135); Chapter 3: What Can Go Wrong with Multiple Regression (Pp. 57-60).

(SK) Berk, Richard A. 1983. "An Introduction of Sample Selection Bias in Sociological Data." *American Sociological Review* 48: 386-398. [Scan only]

(SK) Fomby, Paula and Andrew J. Cherlin. 2007. "Family Instability and Child Well-Being." *American Sociological Review* 72: 181-204.

Norusis, Marija 2009. Chapter 9: Plotting Data (Pp. 171-198); Chapter 11: The Normal Distribution (Pp. 217-228).

October 20: Causal Ordering and Causal Inference

Objectives: (1) To understand the purpose of nested models; (2) To develop an understanding of mediation effects; (3) To recognize the threats to causal inferences in OLS regression; (4) To understand omitted variable bias and its threats to causal inference.

(SK) Ainsworth, James W. 2002. "Why Does it Take a Village? The Mediation of Neighborhood Effects on Educational Attainment." *Social Forces* 81(1): 117-52.

Allison, Paul. 1999. Chapter 3: What Can Go Wrong with Multiple Regression (Pp. 50-57; 60-62).

(SK) Baron, Reuben M. and David A. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6): 1173-1182.

(SK) Cagney, Kathleen A. and Diane S. Lauderdale. 2002. "Education, Wealth, and Cognitive Function in Later Life." *Journal of Gerontology, Social Sciences*: 57:P163-P172

(SK) Fomby, Paula and Andrew J. Cherlin. 2007. "Family Instability and Child Well-Being." *American Sociological Review* 72: 181-204. [reread tables and results].

Miller, Jane. 2005. Chapter 14: Writing About Multivariate Models (Pp. 342-344).

(SK) Springer, Kristen. 2009. "Childhood Physical Abuse and Midlife Physical Health: Testing a Multi-Pathway Life Course Model." *Social Science & Medicine* 69: 138-146.

October 27: Model Form, Testing and Fit

Objectives: (1) To identify the meaning of model fit statistics; (2) To develop strategies for presenting and discussing regression model fit; (3) To devise strategies for transforming data to handle non-linear relationships; and (4) To develop a general understanding of sensitivity analyses.

Allison, Paul. 1999. Chapter 8: How Can Multiple Regression Handle Nonlinear Relationships (Pp. 153-166).

(SK) Flegal, Katherine M., Barry I. Graubard, David F. Williamson, and Mitchell H. Gail. 2005. "Excess Deaths Associated with Underweight, Overweight and Obesity." *Journal of American Medical Association* 293(15): 1861-1867.

Miller, Jane. 2005. Chapter 9: Quantitative Comparisons for Multivariate Models (Pp. 213-215); Chapter 10: Choosing How to Present Statistical Test Results (Pp. 231-254)

(SK) Morgan, Laurie A. and Michelle M. Arthur. 2005. "Methodological Considerations for Estimating the Gender Pay Gap for Employed Professionals." *Sociological Methods and Research* 33: 383-403.

Norusis, Marija. 2009. Chapter 24: Multiple Regression Diagnostics (Pp. 559-572).

November 3: Interaction Terms

Objectives: (1) To define "main effect" and "interaction effect"; (2) To recognize when interaction terms should be used (versus separate models for subpopulations); (3) To calculate and interpret the main and interactive effects of both continuous and dichotomous variables.

Allison, Paul. 1999. Chapter 8: How Can Multiple Regression Handle Nonlinear Relationships ? (Pp. 166-174)

(SK) Carr, Deborah. 2004. "Gender, Pre-Loss Marital Dependence and Older Adults' Adjustment to Widowhood." *Journal of Marriage and Family* 66: 220-35.

Miller, Jane. 2005. Chapter 9: Quantitative Comparisons for Multivariate Models (Pp. 215-218); Chapter 13: Writing About Distributions and Associations (Pp. 311-315); Appendix D: Using a Spreadsheet for Calculations (Pp. 433-437).

Norusis, Marija. 2009. Chapter 8: Counting Responses for Combinations of Variables. Pp. 142-161.

(SK) Mouw, Ted and Yu Xie. 1999. "Bilingualism and the Academic Achievement of First- and Second-Generation Asian Americans: Accommodation with or without Assimilation?" *American Sociological Review* 64: 232-252.

November 10: Logistic Regression

Objectives: (1) To recognize when logistic regression should be used; (2) To interpret the effects of continuous and categorical variables in logistic regression models; (3) To interpret main and interaction effects in logistic regression models.

(SK) Carr, Deborah. 2004. "The Desire to Date and Remarry Among Older Widows and Widowers." *Journal of Marriage and Family* 66: 1051-68.

(SK) DeMaris, Alfred. 1995. "A Tutorial in Logistic Regression." *Journal of Marriage and Family* 57: 956-968. [Focus on pp. 956-965]

(SK) Hunt, Matthew O. 2007. "African American, Hispanic, and White Beliefs about Black/White Inequality, 1977-2004." *American Sociological Review* 72: 39-415.

(SK) Kreager, Derek A. 2007. "Unnecessary Roughness? School Sports, Peer Networks, and Male Adolescent Violence." *American Sociological Review* 72: 705-24.

(SK) Morgan, S. Philip and Jay D. Teachman. 1988. "Logistic Regression: Description, Examples, and Comparisons." *Journal of Marriage and Family* 50: 929-36.

(SK) Wuensch, Karl L. 2007. "Binary Logistic Regression with SPSS." Handout.

November 17: Multinomial Logistic Regression

Objectives: (1) To recognize when multinomial logistic regression should be used; (2) To interpret the effects of continuous and categorical variables in multinomial logistic regression models; (3) To interpret main and interaction effects in multinomial logistic regression models; (4) To develop a conceptual understanding of the differences between multinomial logistic regression and ordered probit and ordinal regression models.

(SK) DeMaris, Alfred. 1995. "A Tutorial in Logistic Regression." *Journal of Marriage and Family* 57: 956-968. [Focus on pp. 965-968].

(SK) Lundquist, Jennifer Hicke. 2008. "Ethnic and Gender Satisfaction in the Military: The Effect of a Meritocratic Institution." *American Sociological Review* 73: 477-96. [ordered probit]

(SK) Sarkisian, Natalia and Naomi Gerstel. 2004. "Kin Support among Blacks and Whites: Race and Family Organization." *American Sociological Review* 69: 812-837. [multinomial logistic regression].

(SK) Vogt Yuan, Anastasia S. 2007. "Perceived Age Discrimination and Mental Health." *Social Forces* 86: 291-311. [multinomial logistic regression]

November 24: Conceptual Overview of Advanced Methods: Event History Analysis (EHA) and Structural Equation Modeling (SEM)

Objectives: (1) To recognize when event history analyses and structural equation modeling should be used; (2) To interpret the effects of continuous and categorical variables in these models; (3) To develop a conceptual understanding of the differences between OLS regression and structural equation models; and (4) To develop a conceptual understanding of the differences between logistic regression and event history models..

Allison, Paul. D. 1999. Chapter 9: How is Multiple Regression Related to Other Statistical Techniques (Pp. 177-187).

EHA

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. Newbury Park, CA: Sage Publications. (Pp. 9-42). [EHA]

- (SK) Burton, Russell P. D., Johnson, Robert J., Ritter, Christian, and Clayton, Richard R. 1996. "The Effects of Role Socialization on the Initiation of Cocaine Use: An Event History Analysis from Adolescence into Middle Adulthood." *Journal of Health and Social Behavior* 37(1): 75-90. [EHA]
- (SK) Heaton, Tim B. and Vaughn R. A. Call. 1995. "Modeling Family Dynamics with Event History Techniques." *Journal of Marriage and Family* 57: 1078-1090. [EHA]
- (SK) Hallinan, Maureen and Richard Williams. 1987. "The Stability of Students' Interracial Friendships." *American Sociological Review* 52(5): 653-664. [EHA]

SEM

- (SK) Godwin, Deborah D. 1988. "Causal Modeling in Family Research." *Journal of Marriage and the Family* 50: 917-927. [SEM]
- (SK) Sewell, William H., Archibald Haller, and Alejandro Portes. 1969. "The Educational and Early Occupational Attainment Process." *American Sociological Review* 34(1): 82-92. [SEM]

December 1: Pulling it Together: Writing a Quantitative Research Paper

Objectives: (1) To produce a theoretically-informed, methodologically rigorous research paper, drawing together this semester's lessons!

Required:

- (SK) Cherlin, Andrew J., Linda M. Burton, Tera R. Hurt, and Diane M. Purvin. 2004. "The Influence of Physical and Sexual Abuse on Marriage and Cohabitation." *American Sociological Review* 69: 768-789.
- Miller, Jane. 2005. Chapter 5: Creating Effective Tables (Pp. 81-119); Chapter 14: Writing About Multivariate Models (Pp. 317-348); Appendix C: Terminology for Ordinary Least Squares (OLS) and Logistic Models (Pp. 423-422).
- (SK) Van Dyke, Nella, Marc Dixon, Helen Carlon. 2007. "Manufacturing Dissent: Labor Revitalization, Union Summer and Student Protest." *Social Forces* 86: 193-214.
- (SK) White, Lynn. 2005. "Writes of Passage: Writing an Empirical Journal Article." *Journal of Marriage and Family* 67: 791-798.

Recommended:

- (SK) Miller, Jane E., and Yana van der Meulen Rodgers. 2008. "Economic Importance and Statistical Significance: Guidelines for Communicating Empirical Research." *Feminist Economics* 14: 117-49.

December 8: Student paper presentations

HAVE A WONDERFUL WINTER VACATION!