Sociology 542
Analysis of Sociological Data II
Fall 2010
Wednesday 4:10 – 6:50 p.m.; Davison Hall 128 (Seminar Room)

Professor: Deborah Carr
Davison Hall 047
E-mail: carrds@sociology.rutgers.edu
Office Hours: Tuesday 1:30-3:30 p.m. and by appointment

Lab Assistant: Neha Gondal
Davison Computer Lab
Email: ngondal@sociology.rutgers.edu
Office Hours: Thursday 2-3 p.m. and by appointment

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 (recently renamed PASW Statistics), 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 Douglass College Bookstore. The required articles and chapters are available at the Soc 542 Sakai website. The required readings will be supplemented by recommended readings throughout the semester, based on students’ requests for further information on particular topics or methods.
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. Given the vast amount of complex and often technical material that we must cover during the semester, we will NOT have time to engage in theoretical debates in class. I strongly encourage you to carry out these discussions via the Sakai website, however.

(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 PASW (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, at least one week prior to deadline. I anticipate that we will have seven or eight assignments with each accounting for ~7.5 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, conference paper, or journal article. Completion of the paper will proceed in five steps.

1. Each student will submit via email or via the Sakai drop box 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 3. 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 10 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

W 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:


Recommended Reading:

W September 8: Monday classes are held

W 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:


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


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

Recommended Reading:

**W 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:


Norusis, Marija. 2010. Chapter 4: Counting Responses; Chapter 5: Computing Descriptive Statistics; Appendix A: Obtaining Charts in SPSS; and Appendix B: Transforming and Selecting Data.


Recommended Reading:
W 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.


W 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.


W 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.


**W 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.


**W 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.


FRIDAY October 29 (9:30-12:10): 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.

W 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.

References:


**W 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.


**MONDAY November 22: Class Cancelled**

**W November 24:** All University classes cancelled

**W 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:**


Marriage and Family 67: 791-798.

Recommended:

Friday December 3  9:30 a.m. – 12:10
Optional ‘bonus’ class session: 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.


EHA


SEM


W December 8:  Student paper presentations

HAVE A WONDERFUL WINTER VACATION!