

USP/SOC 982:
Methods of Research and Analysis for Urban Social Institutions II
Fall, 2006
Monday, 4:30 p.m. - 7:10 p.m.
Room 293 Bolton Hall

Professor Nancy Mathiowetz

e-mail: nancym2@uwm.edu

Phone: 414-229-2216

Office hours: Mondays, 1:00 - 3:00 p.m. and by appointment

Course Description and Goals: The focus of this course is on the application of statistical techniques to analyze social science data. Students will choose a data set and prepare a research paper based upon the data; the final product for the course will be an original research paper suitable for presentation at a professional conference or for publication in a scholarly journal. This is not a statistics course, *per se*, but rather a course in which statistical theories and techniques will be reviewed (and some new techniques introduced). The emphasis in this course will be on the appropriate use and interpretation of statistical techniques in the analysis of social science data. The material covered in this course will enable students to intelligently and critically read professional publications as well as prepare students to successfully complete the steps in the empirical research process, moving from a research idea through data analysis and presentation of findings.

Prerequisites: At least one graduate statistics course (that included multivariate methods) and a passing grade of 85% or higher on the diagnostic exam for this course. The course work assumes that students understand principles of statistical inference, including confidence intervals and hypothesis testing, measures of central tendency and variation, multiple linear regression, analysis of variance, regression with limited dependent variables (i.e., multivariate logistic regression) and the analysis of categorical data, including multivariate contingency tables (and associated statistical tests of association and relationship strength).

Required Texts: There are two required texts for this class and a coursepack (available at Clark Graphics, 2915 N. Oakland Ave. Telephone number: 414-962-4633). The required texts are:

Carol Aneshensel (2002) *Theory-Based Data Analysis for the Social Sciences*. Thousand Oaks: Pine Forge Press.

J. Scott Long and Jeremy Freese (2006) *Regression Models for Categorical Dependent Variables Using Stata* (2nd Edition) College Station, TX: Stata Corporation

In addition, I recommend the following textbooks, as needed (depending upon your statistical training and the analysis you will be conducting for your paper):

Agresti, A. (1996) *An Introduction to Categorical Data Analysis*. New York: John Wiley and Sons. A very thorough treatment of categorical data analysis. Chapters 1-3 provide an excellent introduction.

Hosmer, D. And Lemeshow, S. (2000) *Applied Logistic Regression*. Second Edition. New York: John Wiley and Sons. Chapters 1-3 provide an excellent introduction to logistic regression analysis.

Knoke, D., Bohrnstedt, G., and Mee, A. (2002) *Statistics for Social Data Analysis*. Fourth Edition. F.E. Peacock Publishers

Korn, E. And Graubard, B. (1999) *Analysis of Health Surveys*. New York: John Wiley and Sons. An excellent text concerning the analysis of survey data, not just health surveys.

Long, S. (1997) *Regression Models for Categorical and Limited Dependent Variables*, Vol. 7 of *Advanced Quantitative Techniques in the Social Sciences*. Thousand Oaks, CA: Sage.

Sage Series in Quantitative Applications in the Social Sciences (Thousand Oaks, CA). There are a number of inexpensive volumes in this series that provide a useful introduction to particular methods. The entire series can be found at www.sagepublications.com .

Computing

This course will make extensive use of statistical software for the analysis of complex survey data. The course will be taught using Stata. Many of the statistical tests and models we will be examining are not available in SPSS nor can one use the standard SPSS package for the analysis of complex survey data. Stata is available in the computer labs throughout campus. In addition, I have set up a Grad Plan with the Stata corporation. I will forward an email from Stata to class members with the direct link. You will be able to order either of the following:

Intercooled Stata 9

One-year license, *Getting Started* manual **\$89**

Intercooled Stata 9

Perpetual license, *Getting Started* manual **\$145**

In addition, you might wish to order:

Base Reference Manual (3 volumes), Data Management Reference Manual, User's Guide, and Quick Reference and Index, \$165

Grading. Grading for the course will be based on the following:

1. **Homework assignments (200 points).** There will be 8 homework assignments, throughout the semester, each worth 25 points. The schedule of assignments can be found in the attached schedule. The homework is intended to reinforce the lecture materials and readings and allow you the opportunity to apply your skills to real datasets. I encourage collaboration among students with respect to mastering the material and debugging statistical software runs. **However, you are required to write up your own answers to the homework, independent of other students. You are also required to produce and attach your own computer printout, whenever printout is needed.** No photocopies or second printings of other students' printouts are allowed. Violations of this policy will result in serious penalties.

Late assignments will be downgraded by 10 points per day that they are late.

2. **Midterm Exam (100 points):** There will be a midterm exam scheduled for October 23rd. The exam will emphasize the selection of appropriate models, interpretation of models, and ability to succinctly write up interpretations of findings. The exam will be open note, open text and will reiterate homework assignments and lectures.
2. **Final project–written paper (250 points).** The final project will require you to select a dataset of interest to you, draft a hypothesis or hypotheses you wish to examine using these data (theory or empirically based hypothesis), motivate the research with a literature review, complete the statistical analysis, and write up the results. The final paper is expected to be of the quality one would find in a leading social science peer-reviewed journal. Many of the assignments throughout the semester are directed toward making progress toward the final paper. This is a great opportunity to find the data for your MA or Ph.D. thesis and begin to conduct preliminary analysis! Students will peer-review and critique papers by fellow students. **Work on the final paper should begin the first week of class!**
3. **Final paper: presentation (50 points).** In addition to writing a final paper, you will present a profession-level presentation to the class, 20-30 minutes of your findings. This presentation will be similar to a presentation you would present at a professional meeting. All students will grade and critique all of the presentations.

Class website: Lecture notes (to the extent there are any), exercises, and any supplemental readings not included in the texts or coursepack can be found on the course website, using D2L. I encourage you to check the D2L website prior to class so that you can have copies of the material during the lecture.

Lecture and Lab format: The course runs from 4:30 to 7:10 p.m. on Monday; the course will be divided between lecture and lab, depending upon the nature of the material to be covered. For example, during the weeks that we are focusing on statistical techniques, lab time will be spent applying (that is trying out) the various techniques. I encourage you to have either a floppy disk/CD/flash drive with you for all classes, so that you can save your work.

As you can see, many of the weeks in which we discuss a statistical technique include an application paper. This paper is intended to illustrate use and exposition of the particular statistical technique. We will discuss these papers in class, so it is imperative that the paper is read prior to the lecture.

You are expected to have read the material prior to the lecture. The lectures only make sense when you have prepared adequately for the class. Since this is a small, Ph.D.-level course, all students will be expected to contribute to the discussion and explication of material during the class. Attendance at both the lecture and lab are required. Office hours are intended to be used for clarification of material and assistance with assignments, readings, and the final project and not to reiterate material that was covered in lecture and/or lab.

Schedule of Topics

NOTE: FP denotes assignment related to final paper

Date	Topic	Readings	Exercise
9.11	Introduction; Theory Based Data Analysis	S. Levitt & S. Dubner, <i>Freakonomics</i> (coursepack) <i>Theory-Based Data Analysis</i> pp. 1-94	
9.18	Modeling, Testing Theories; Causal Models	<i>Theory-Based Data Analysis</i> pp. 95-245 Klem “Path Analysis” (coursepack) Application: Chesley “Blurring Boundaries? Linking Technology Use, Spillover, Individual Distress, and Family Satisfaction” (Coursepack)	Exercise 1: Draft hypothesis for final paper and identification of data for analysis (Due 9.25)
9.25	Linear Regression Review; Linear Regression Application Introduction to Stata	<i>Getting Started with Stata</i> Long and Freese, pp. 3-72; 415-427 Application: Klebanov, Brooks-Gunn, and Duncan “Does Neighborhood and Family Poverty Affect Mothers’ Parenting, Mental Health and Social Support” (coursepack)	Exercise 2: Multivariate linear regression model with Stata: analysis and short write-up (Due 10.2)

10.2	Binary Outcomes: Logits and Probits	Long and Freese, pp. 75-181 DeMaris “A Tutorial in Logistic Regression” (coursepack) Application: McLaughlin and Lichter “Poverty and Marital Behavior of Young Women” (coursepack)	Exercise 3: Multivariate logistic regression model with Stata: analysis and short write up (Due 10.9)
10.9	Multicategory Logit Models	Long and Freese, pp. 183-291 Application: Chesley “Families in a High Tech Age” (coursepack)	Exercise 4: Categorical Outcome Models(Due 10.16)
10.16	Count Data	Long and Freese, pp. 349-414 Application: Sakinsian and Gerstel “Explaining the Gender Gap in Help to Parents: The Importance of Employment” (coursepack)	Exercise 5: Zero-inflated negative binomial models (Due 10.23)
10.23	Midterm Exam		Exercise 6: Preliminary Data Analysis for Final Paper (Due 10.30)
10.30	Data Reduction, Scales, Factor Analysis	Bryant and Yarnold “Principal Components Analysis and Exploratory and Confirmatory Factor Analysis” (coursepack)	Exercise 7: Data reduction techniques (Due 11.6)

11.6	Dealing with Missing Data (Imputation); Outliers	Kalton and Kasprzyk “The Treatment of Missing Survey Data” (coursepack) Little “Regression with Missing X’s: A Review (coursepack)	
11.13	Analyzing Complex Survey Data: Weights and Variance Estimates	Korn, E. And Graubard, B. <i>Analysis of Health Surveys</i> pp. 8-21 and 159-177 (coursepack)	Exercise 8: Complex variance estimation for survey data (Due 11.20)
11.20	Other issues in the analysis of survey data (e.g., time series, longitudinal)	TBD	
11.27	Consultations, Lab Time (Buffer week in case we fall behind!)		
12.4	Final Presentations		critiques and comments of presentations
12.11	Final Presentations		critiques and comments of presentations
12.18	Final Papers Due		