

Problems of Data Analysis
Sociology 509
Spring 2009
Case Western Reserve University

Instructor: Dr. Jessica Kelley-Moore

Office: 230 Mather Memorial Hall

Phone: 216-368-8879

E-mail: jak119@case.edu

Class Meeting: Tuesday/Thursday 8:30am – 9:45am; Clark Hall 210

Course Summary

The purpose of this course is to teach students advanced techniques commonly used in the analysis of social science data. At the end of this course, it is expected that students will be able to:

1. Critically evaluate the general design and statistics in existing studies.
2. Select and estimate correct models based on the research question and measurement of the dependent variable.
3. Apply current techniques used in complex social science data, such as handling missing data, weighting the sample, adjusting for clustering, and measuring change.
4. Use STATA for analyses and data manipulation, including the Help functions, Do-Files, and graphics functions.

Text

Rabe-Hesketh, Sophia and Brian S. Everitt. 2007. *A Handbook of Statistical Analyses Using STATA*. Fourth Edition. Boca Raton: Chapman and Hall/CRC.

NOTE: It is strongly recommended that you purchase the fourth edition of this text, as it provides the most updated instructions for using STATA. In addition, not all chapters are presented in previous editions.

All other assigned readings are available through Blackboard.

Data

All of the examples and exercises for this class will be drawn from the study *Americans' Changing Lives*. This is a four-wave panel study of more than 3,600 adults ages 25 to 96. Focusing especially on differences between Black and White Americans in middle and late life, these data constitute the first, second, third, and fourth waves in a national longitudinal panel survey covering a wide range of sociological, psychological, mental, and physical health items. The description of the data and the codebooks are available on Blackboard.

Students will use these data for their individual projects, as well. Using such rich data, students should be able to develop these final projects into a conference research presentation, a publication, research proposal, or possible dissertation. *Please Note: If you intend to publish or present this project, you must obtain IRB Approval. You are eligible for the #4 Exemption, Use of Secondary Data that has been de-identified.*

Grade Distribution

Item	Points	Date
Homework 1	50	February 2
Homework 2	50	February 19
Homework 3	50	February 27
Homework 4	50	March 25
Homework 5	50	April 1
Homework 6	50	April 10
Unit 1 Project	100	March 6
Unit 2 Project	100	May 1
Individual Project		
Topic Paragraph & Variable List	25	February 10
Lit Review and Analytic Plan	50	February 24
Final Project	100	May 7
Total Points	675	

90 – 100% = A; 80 – 89% = B; 70 – 79% = C; 60 – 69% = D; Below 59% = F

Blackboard

This course requires the use of Blackboard, which is an online course management program.

Blackboard will be my primary tool for announcements, changes in readings or due dates, web site links, online discussions, and posting grades. You are responsible for keeping up with the material on Blackboard. The ACL data, documentation, and help sheets will also be posted on Blackboard.

Software

This course will use STATA, a statistical processing package. Part of the curriculum of this course is to teach students how to use STATA. No prior experience with this package is necessary. As a note, most techniques taught in this course can easily be estimated in SAS or SPSS but only output produced in STATA will be accepted for homework assignments and final papers.

Unit Projects

In lieu of exams, you will complete two Unit Projects. While the homework assignments focus on the mechanics of data analysis and STATA, the Unit projects will focus on the conceptual issues related to data analysis. The Unit questions will be applied and require critical thinking and clear writing. You will receive the assigned questions in class one week before they are due. You have one week to complete and submit to me. You may use your course notes, readings, and any material posted on Blackboard. **YOU MAY NOT WORK TOGETHER.** You may at any time contact me for questions. I will provide verbal feedback, but will not review drafts of your answers.

Homework

All homework assignments will be either posted on Blackboard or distributed in class. The assignment will lose 5 points for each day it is late. All homework assignments should be typed in 12pt. font.

Output should be cut and pasted into the Word document for the homework. All output should include titles, annotated notes, and clearly relate to the question being answered. *As a note: STATA output will appear best in Word documents when you change the font to Courier New, Font Size 9.* In some assignments, you will be asked to provide your syntax. This should also be labeled clearly.

Individual Project

In this course, students will have the opportunity to conduct their own statistical analysis. The *ACL* data are freely available from ICPSR at the University of Michigan. Thus, the original work produced by the students is the foundation for a single-authored presentation or publication. This class will focus on the analysis portion of the paper so it only requires an abbreviated literature review.

For the final project, students will submit a research paper with the abbreviated literature review, methods section (including description of the analysis), results section [with tables – NOT OUTPUT], and brief discussion section. Students will also attach all relevant output for the presented analysis. Further details, including grading criteria, will be posted on Blackboard.

UNIT 1: Principles and Practices of Good Data Management

January 13: Introduction to Principles of Quantitative Data Analysis

Elder, Glen, Eliza Pavalko, and Elizabeth C. Clipp. 1993. Pages 1 – 49 in *Working with Archival Data: Studying Lives*. Thousand Oaks: Sage Publications.

Campbell, Richard T. 1994. “A Data-Based Revolution in the Social Sciences.” *ICPSR Bulletin* 14 (3): 1 – 4.

Rabe-Hesketh & Everitt. Chapters 1 and 2.

January 15: Review of the Practice of Quantitative Data Analysis

Complete Online Data Exercises

Racial Disparities in Mental Health: A Data-Driven Learning Guide

<http://www.icpsr.umich.edu/cocoon/OLC/racementalhealth.xml?token=a01>

Exploring the Second Shift: A Data-Driven Learning Guide

<http://www.icpsr.umich.edu/cocoon/OLC/secondshift.xml?token=a01>

Please Note: You will need to create an ICPSR account to complete the exercises.

January 20: A Review of Regression and its Assumptions

Rabe-Hesketh & Everitt. Chapter 3.

Kennedy, Peter. 1998. “The Classical Linear Regression Model.” Pages 42-53 in *Econometrics*. Cambridge, MA: MIT Press.

January 22: The Use and Application of Dummy Variables

January 27: Interaction Terms

January 29: Nested Models & Model Comparison Across Groups

Computer Lab

February 2: Homework 1 Due: Linear Regression with Interactions and Dummy Variables

February 3: Structures and Shapes of Data – Implications for Design

Long, J. Scott and Jeremy Freese. 2001. First Part of Chapter 2 (Pages 13 – 36) of *Regression Models for Categorical Dependent Variables Using STATA*. College Station: STATA Press.

February 5: Measurement and Recoding: Selecting the Best Distribution

Computer Lab

Long, J. Scott and Jeremy Freese. 2001. Remainder of Chapter 2 (Pages 36 - 61) of *Regression Models for Categorical Dependent Variables Using STATA*. College Station: STATA Press.

February 10: Missing Data Part 1: Where Did My N Go?

Little, Roderick J. and Nathaniel Schenker. 1995. "Missing Data." Pp. 39-75 in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel. New York: Plenum Press.

****TOPIC PARAGRAPH & VARIABLE LIST DUE****

February 12: Missing Data Part 2: Getting My N Back

Review slides from Paul Allison course on Missing Data.

February 17: There Are No Fixed Rules for Managing Missing Data!

Computer Lab

February 19: Homework 2 Due (Missing Data)

February 19: Complex Sampling Design: Promises and Pitfalls

****LIT REVIEW & ANALYTIC PLAN DUE****

Winship, Christopher and Larry Radbill. 1994. "Sampling Weights and Regression Analysis." *Sociological Methods and Research* 23:230-257.

Korn, Edward L. and Barry I Graubard. 1991. "Epidemiologic Studies Utilizing Surveys: Accounting for the Sampling Design." *American Journal of Public Health* 81(9):1166-1173.

February 24: Nested Data Designs

Luke, Douglas. 2004. "The Need for Multilevel Modeling." Pages 1 – 9 of *Multilevel Modeling*. Thousand Oaks: Sage Publications.

February 26: Weights, Clustering, and Nested Data

Computer Lab

February 27: Homework 3 Due (Weighting and Clustering)

March 3: NO CLASS

March 6 [NOTE: FRIDAY]: Unit 1 Project Due

March 10: SPRING BREAK

March 12: SPRING BREAK

UNIT 2: Making Reasoned Decisions about Analysis and Design

March 17: Limited and Non-Normal Dependent Variables: Logistic Regression

Rabe-Hesketh and Everitt. 2007. Chapter 6.

Long, J. Scott and Jeremy Freese. 2001. Chapter 4 (Pages 100 - 136) of *Regression Models for Categorical Dependent Variables Using STATA*. College Station: STATA Press.

Kennedy, Peter. "Qualitative Dependent Variables." Pages 233-248 in *A Guide to Econometrics*. Cambridge, MA: MIT Press.

March 19: Logistic Regression, cont.

Computer Lab

March 24: Logistic Regression

March 25: Homework 4 Due (Logistic Regression)

March 26: Limited and Non-Normal Dependent Variables: Other Models

Long, J. Scott and Jeremy Freese. 2001. Chapter 7 (Pages 223 - 262) of *Regression Models for Categorical Dependent Variables Using STATA*. College Station: STATA Press.

March 31: Ordered Logistic & Multinomial Logistic Regression

Computer Lab

April 1: Homework 5 Due (Models for Limited Dependent Variables) [No Fooling!]

April 2: Longitudinal Designs and Analysis

Kelley-Moore, Jessica. 2007. "Longitudinal Research." *Encyclopedia of Health and Aging*. Sage Books.

Menard, Scott. 1991. Chapters 4 & 5 in *Longitudinal Research*. Newbury Park: Sage.

April 7: Measuring and Modeling Change

Rabe-Hesketh and Everitt. 2007. Chapter 8.

Allison, Paul D. 1990. "Change Scores as Dependent Variables in Regression Analysis." Pp. 93 – 114 in *Sociological Methodology, 1990*, edited by Clifford D. Clogg. Washington, D.C.: American Sociological Association.

April 9: Attrition and Selection Bias

Berk, Richard A. and Subhash C. Ray. 1982. "Selection Biases in Sociological Data." *Social Science Research* 11: 352-398.

Winship, Christopher and Robert D. Mare. 1992. "Models for Sample Selection Bias." *Annual Review of Sociology* 18: 327-350.

Stolzenberg, Ross M. and Daniel A. Relles. 1997. "Tools for Intuition about Sample Selection Bias and Its Correction." *American Sociological Review* 62: 494-507.

Optional: Heckman, James J. 1979. "Sample Selection Bias as Specification Error." *Econometrica* 47: 153-161.

April 10: Homework 6 Due (Measuring Change)

April 14: Event History and Survival Analysis

Rabe-Hesketh and Everitt. 2007. Chapter 12.

April 16: Measuring Change With Three or More Time Points

Rabe-Hesketh and Everitt. 2007. Chapter 9.

April 21: Latent Growth Curves

Computer Lab

April 23: A Brief Introduction to Maximum Likelihood

Rabe-Hesketh and Everitt. 2007. Chapter 13.

May 1: Unit 2 Projects Due

May 7: Final Paper Due