

Sociology G4075: Introduction to Social Data Analysis II

December 21, 2005

Tuesday and Thursday, 4:10pm-5:25pm
311 Fayerweather

Lab: Wednesday 9-10:50am
Mathematics 407

	Instructor	TA
	Aaron Gullickson	Lucas Dauter
office:	412 Fayerweather Hall	
email:	ag2319@columbia.edu	
office hours:	Wednesday, 2-4pm	

Course Objectives

This course is a continuation of the introductory statistics course for sociology graduate students. This semester, we will focus on extensions of the OLS regression model to incorporate various kinds of dependent variables, using the framework of the generalized linear model. Additional topics on common sociological methods which fall outside this framework will also be discussed.

Readings

There is only one required textbook for the class:

Powers, Daniel A. and Yu Xie. 2000. *Statistical Methods for Categorical Data Analysis*. New York: Academic Press.

This book should be available at the Columbia bookstore. There will be additional articles made available to students over the courseworks website or on reserve.

Course Organization

This course will consist of biweekly lectures (Tuesdays and Thursdays), and a lab section on Wednesdays. The course is organized around a series of topic modules which generally last about two weeks each. Each module will focus on one or two specific articles which make use of the method under study. There will be a homework project for mosts module which will be due at the conclusion of the module. The lab sections will provide a chance for students to work collectively on projects and to received guidance from the instructor. In addition to the homework, there will be a final take-home exam and a final paper.

The Paper

At the conclusion of the class, each student will turn in an original research paper which makes use of the statistical methods learned in the course. This paper should address a research question of substantive interest. Students will be responsible for finding and managing data for their project, although the instructor can usually suggest appropriate data sets for the research question. Students who are writing papers for other courses are welcome to consider “doubling up” of papers, as long as the research question can be adequately addressed by the methods learned in this course.

The paper should be written as a full-fledged empirical research article including an introduction/statement of the problem, state of current knowledge, data and methods description, findings, and conclusion. At the end of the semester each student will present their research to the rest of the class. The papers will be due on the last day of class.

In order to encourage a healthy research process, there will be several deadlines prior to the turning of the final paper. These are as follows.

- At the end of the 3rd week, students should turn in a brief one-page summary of their research question in broad terms.
- At the end of the 6th week, students should have identified the data set they will be using.
- At the end of the 8th week, students should turn in the introduction and literature review for their paper.
- The final paper will be due the last day of class.

Grading

A HW assignment will be due for each module. At the end of the semester, each student will turn in an original research paper which makes use of the statistical methods used in the class. A take-home exam will also be given out on the last day of class and will be due during finals week.

Homework will count for 35% of the grade, the take-home exam will count for 25% of the grade, and the final paper will count for 40% of the grade.

Course Modules

OLS Regression Techniques and Issues (week 1-3) Chapters 1 and 2, Appendix A

Assignment reading: Ellison, Christopher G. and Marc A. Musick. 1993. "Southern Intolerance: A Fundamentalist Effect?" *Social Forces* 72(2): 379-398.

1. Review of the Linear Regression Model
2. Model selection
3. Missing data/Weighting
4. Advanced methods
 - (a) Multilevel models
 - (b) Factor analysis and structural equation modeling
 - (c) Addressing "endogeneity"
 - i. Fixed-effects models
 - ii. Instrumental variables

Generalized Linear Models (week 4) Appendix B

No additional reading or assignment

1. The linear probability model and generalized least squares
2. Generalized linear models
3. Maximum likelihood estimation

Logistic Regression (week 5-6) Chapter 3

Assignment reading: Mare, Robert D. 1980. "Social Background and School Continuation Decisions." *Journal of the American Statistical Association* 75(370): 295-305.

1. Odds and odds ratios
2. Logistic regression
3. Goodness of fit and other types of link functions

Log-linear Models (week 7-8) Chapter 4

No additional reading or assignment

1. Two-way tables and odds ratios
2. Log-linear models for two-way tables
3. Log-linear models for multiway tables
4. Model Selection

Multinomial and Ordered Logit Models (week 9-10) Chapter 6 and 7

Brooks, Clem. 2000. "Civil Rights Liberalism and the Suppression of a Republican Political Realignment in the United States, 1972 to 1996." *American Sociological Review* 65(4): 483-505.

1. Multinomial logit models
2. The conditional logit model
3. The ordered logit model
4. Comparison of the logistic, multinomial, ordered logit, and log-linear models

Event History Models (week 11-12) Chapter 5

Assignment reading: Oppenheimer, Valerie Kincade, Matthijs Kalmijn, and Nelson Kim. "Men's Career Development and Marriage Timing During a Period of Rising Inequality." *Demography* 34(3): 311-330.

1. Events, rates, and relative risk
2. Parametric survival models
3. Semi-parametric survival models
4. Discrete time approximation
5. Event count poisson regression

Week off for research (week 13)

Class presentations (week 14-15)