SOCY 7704: REGRESSION MODELS FOR CATEGORICAL DATA

Instructor: Natasha Sarkisian
Email: natasha@sarkisian.net
Office phone: (617) 552-0495
Home phone: (617) 755-3178
Mailbox: McGuinn 426

Office: McGuinn 417
Office hours: By appointment
Class time: Wednesdays 2-4:20 PM
Class location: Campion 131
Webpage: http://www.sarkisian.net/socy7704/

COURSE DESCRIPTION[*]

This applied course is designed for graduate students with a prior background in statistics at the level of SC703: Multivariate Statistics (or its equivalent). This means that students should have considerable experience with ordinary-least-squares (OLS) regression: I assume you have an understanding of multiple OLS regression and an ability to conduct such analyses using some statistical software (e.g., SPSS, SAS, Stata, etc.). Major topics of the course will include OLS regression diagnostics, binary, ordered, and multinomial logistic regression, models for the analysis of count data (e.g., Poisson and negative binomial regression), treatment of missing data, and the analysis of clustered and stratified samples.

Dependent variable		Type of analysis
Nominal	Binary	Logistic regression (or probit)
	Multi-category	Multinomial logistic regression
Ordinal		Ordered logistic regression (or ordered probit)
Interval	Continuous	OLS regression
	Count	Poisson and negative binomial regression

We will be using Stata for all the analyses throughout the course. No previous Stata experience is necessary: One of the goals of the course is to familiarize you with this statistical software package. I will provide an introduction to Stata in the beginning of the course and guide you throughout the course. Your main textbook (Long and Freese 2014) also provides an introduction to Stata and a step-by-step guide for many analyses. For the topics not covered in this book, I will provide handouts covering relevant Stata commands. For your assignment, you can use Stata on Citrix: see http://apps.bc.edu .

The goals of the course are to develop the skills necessary to critically evaluate contemporary social research using advanced quantitative methods and to identify an appropriate technique, estimate models, and interpret results for independent research. The course will be applied in the sense that we will focus on estimating models and interpreting the results, rather than on understanding in detail the mathematics behind the techniques. I hope that the course will provide you with a solid foundation in advanced quantitative methods, which is in high demand in many fields, both in and out of academia. For those of you in the Sociology Department, the course can also provide a foundation for the "Advanced Quantitative Methods" area examination.

COURSE POLICIES

For each topic in the course, I will give a lecture focusing on the reasoning behind the technique, and provide a review of the syntax used to do analyses and the output generated by Stata. You will then get a chance to practice conducting the analyses and interpreting the results. We will also discuss and critically evaluate published research based on the various techniques. Make sure that you carefully read these examples of published research before class and be prepared to discuss them. The course is based on an interactive relationship between the instructor and students, as well as on collaboration among the

---

[*] This syllabus draws upon ideas presented in syllabi by a number of people, including Robert Kunovich, John Williamson, Joya Misra, and Doug Anderton.

students. You are strongly encouraged to ask questions and discuss the material in class. I also encourage collaboration among the students. Please feel free to help each other when running analyses for assignments. However, everyone must turn in their own report and statistical output. You will also check each other's work and provide feedback on drafts of your two major assignments.

I also would like to stress that you are always welcome to ask any additional questions. Email is usually the best way to get in touch with me in order to get a quick question answered or to set up an appointment to discuss something at length. You are also welcome to call me either in my office or at home (any time between 9 AM and 10 PM); however, be prepared to leave your name and number if I am not available to pick up the phone. Also, please check our course website regularly: various course materials (assignments, notes, etc.) will be posted there regularly. And make sure to check your email, too – from time to time I may send some announcements.

Finally, a note on feedback. I would like to know how I could make this course experience as useful and interesting as possible. Therefore, every class in the end of class I will ask you to submit a sheet of paper with the date and at least one sentence of reaction to that class meeting, indicating what you learned, or something you liked or did not like, found interesting or controversial, found clear or too simplistic, or found confusing and in need of further (or better) explanation. You may also submit comments on the course in general.

REQUIRED MATERIALS

Required books:
The following books should be available for purchase at the BC bookstore. They are also placed on reserve at the library.
1. Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata,3rd Edition.* College Station, TX: Stata Press.
2. Berry, William D. 1993. *Understanding Regression Assumptions.* (Quantitative Applications in the Social Sciences, 07-092). Thousand Oaks, CA: Sage Publications.
3. Fox, John. R. 1991. *Regression Diagnostics.* (Quantitative Applications in the Social Sciences, 07-079). Thousand Oaks, CA: Sage Publications.
4. Jaccard, James J., and Robert Turrisi. 2003. *Interaction Effects in Multiple Regression. 2nd edition.* (Quantitative Applications in the Social Sciences, 07-072). Thousand Oaks, CA: Sage Publications.
5. Pampel, Fred C. 2000. *Logistic Regression: A Primer.* (Quantitative Applications in the Social Sciences, 07-132). Thousand Oaks, CA: Sage Publications
6. Borooah, Vani K. 2002. *Logit and Probit: Ordered and Multinomial Models*. (Quantitative Applications in the Social Sciences, 07-138). Thousand Oaks, CA: Sage Publications.
7. Enders, Craig K. 2010. Applied Missing Data Analysis. Guilford Press.
8. Lee, Eun Sul, and Ronald N. Forthofer. 2006. *Analyzing Complex Survey Data. 2nd edition.* (Quantitative Applications in the Social Sciences, 07-071). Thousand Oaks, CA: Sage Publications.

Other required readings:
Other required readings (listed below in the course outline) will be available on electronic reserve in the library: see http://www.bc.edu/reserves

Additional resources for learning Stata:
Resources to Help You Learn and Use Stata. http://www.ats.ucla.edu/stat/stata/
Stata Reference Manuals. College Station, TX: Stata Press.
Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.   http://nbecerra.files.wordpress.com/2011/03/rmcv_long.pdf

COURSE REQUIREMENTS AND GRADING

There will be a number of assignments in this course --three larger assignments, one smaller assignment, and responses to articles. The due dates can be found in the schedule below.

Assignment	% of grade
Assignment 1 draft 1	5
Assignment 1 peer feedback	5
Assignment 1 final draft	20
Assignment 2 draft	5
Assignment 2 peer feedback	5
Assignment 2 final draft	20
Count data mini-assignment	10
Answers to article questions	10
Article write-up	20

The larger assignments will involve selecting a research question and variables, running analyses, conducting diagnostics and applying remedies, and interpreting the results. You will create a running log to document all the steps in this process and then edit it, keeping only the relevant steps and adding comments, explanations, and interpreting the results. I will distribute more guidelines later in the semester.

In addition, for one of these larger assignments, you will write up the results like you would for a journal publication. This will include an Introduction that will provide a short substantive description of your theoretical argument, your research questions, and hypotheses (1 page max.). Second, your write-up will include a brief Data and Methods section (1-2 pages) describing the variables and the methodology. You will include any discussion of diagnostics and modifications in this section. Also, here you will include a table with descriptive statistics for your sample. Third, you will provide a 3-4 page description of the results including tables (in journal format) and graphs assisting in the interpretation of results. Finally, you will include a brief conclusion summarizing your findings (1 page max.).

For all of the assignments, I will provide data, although you can use your own data if they are appropriate for the technique (see me in advance). My intention is that these assignments will assist in the completion of the advanced quantitative methods area exam in sociology and/or will facilitate your own independent research projects.

Each assignment will consist of two drafts, to be submitted electronically. Small files can be sent by email; any large files should be submitted using MyFiles or another file sharing website. You will get peer feedback as well as feedback from me (along with a temporary grade) on the first draft. At that point, we will also discuss the common problems and mistakes. You will then get a chance to submit a revised draft. If you are satisfied with your temporary grade, you do not need to revise the assignment – just let me know. The letter grades for the assignments will be determined as follows:

93-100	A
90-92	A-
87-89	B+
83-86	B
80-82	B-
60-79	C
0-59	F

COURSE OUTLINE

September 3. Introduction to the course and to Stata.

September 10. Introduction to data management using Stata
Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stat, 3rd edition.* Chapter 2. College Station, TX: Stata Press.

September 17. OLS regression assumptions and diagnostics
Berry, William D. 1993. *Understanding Regression Assumptions.* (Quantitative Applications in the Social Sciences, 07-092). Thousand Oaks, CA: Sage Publications.

September 24. OLS regression assumptions and diagnostics (continued)
Fox, John. R. 1991. *Regression Diagnostics.* (Quantitative Applications in the Social Sciences, 07-079). Thousand Oaks, CA: Sage Publications.

October 1. OLS regression assumptions and diagnostics (continued)
Answers to article questions due at 9 AM
Jaccard, James J., and Robert Turrisi. 2003. *Interaction Effects in Multiple Regression. 2nd edition.* (Quantitative Applications in the Social Sciences, 07-072). Thousand Oaks, CA: Sage Publications.
Kenworthy, Lane, and Melissa Malami. 1999. "Gender Inequality in Political Representation: A Worldwide Comparative Analysis." *Social Forces, 78*: 235-268. RESERVE.

October 8. Binary logistic regression.
Pampel, Fred C. 2000. *Logistic Regression: A Primer.* (Quantitative Applications in the Social Sciences, 07-132). Thousand Oaks, CA: Sage Publications
Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata, 3rd edition.* Chapters 3 and 5. College Station, TX: Stata Press.

October 15. Binary logistic regression (continued)
Assignment 1 (OLS Regression) first draft due for peer feedback
Answers to article questions due at 9 AM
Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata.* Chapters 4 and 6. College Station, TX: Stata Press.
Allison, Paul D. 1999. "Comparing Logit and Probit Coefficients Across Groups." *Sociological Methods and Research,* 28: 186-208.
Alba, Richard, John Logan, Amy Lutz, and Brian Stults. 2002. "Only English by the Third Generation? Loss and Preservation of the Mother Tongue among the Grandchildren of Contemporary Immigrants." *Demography, 39*: 467-484. RESERVE.

October 22. Ordered logistic regression
Assignment 1 (OLS Regression) first draft peer feedback due to student and instructor
Answers to article questions due at 9 AM
Borooah, Vani K. 2002. *Logit and Probit: Ordered and Multinomial Models.* (Quantitative Applications in the Social Sciences, 07-138). Chapters 1 and 2. Thousand Oaks, CA: Sage Publications.
Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata.* Chapter 7. College Station, TX: Stata Press.
Michelson, Melissa R. 2003. The Corrosive Effect of Acculturation: How Mexican Americans Lose Political Trust. *Social Science Quarterly, 84(4)*: 918-933. RESERVE

October 29. Multinomial logistic regression
Answers to article questions due at 9 AM
Borooah, Vani K. 2002. *Logit and Probit: Ordered and Multinomial Models*. (Quantitative Applications in the Social Sciences, 07-138). Chapter 3. Thousand Oaks, CA: Sage Publications.
Long, J. Scott and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. Chapter 8. College Station, TX: Stata Press.
Reynolds, Jeremy. 2004. "When Too Much Is Not Enough: Actual and Preferred Work Hours in the United States and Abroad." *Sociological Forum, 19*: 89-120. RESERVE.

November 5. Count data models
***Assignment 1 (OLS Regression) final draft due ***
Answers to article questions due at 9 AM
Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Chapter 8, pp. 217-239. Thousand Oaks, CA: Sage Publications. RESERVE.
Long, J. Scott and Jeremy Freese. 2014. Chapter 9: 9.1 to 9.3. *Regression Models for Categorical Dependent Variables Using Stata, 3rd edition*. College Station, TX: Stata Press.
Van der Burg, Brigitte, Jacques Siegers, and Rudolf Winter-Ebmer. 1998. "Gender and Promotion in the Academic Labour Market." *Labour,* 12: 701-713. RESERVE.

November 12. Zero-inflated and zero-truncated count data models
Assignment 2 (Logistic Regression) first draft due for peer feedback
Answers to article questions due at 9 AM
Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Chapter 8, pp. 239-249. Thousand Oaks, CA: Sage Publications. RESERVE.
Long, J. Scott and Jeremy Freese. 2014. Chapter 9: 9.4 to 9.8. *Regression Models for Categorical Dependent Variables Using Stata, 3rd edition*. College Station, TX: Stata Press.
Sarkisian, Natalia and Naomi Gerstel. 2004. "Explaining the Gender Gap in Help to Parents: The Importance of Employment." *Journal of Marriage and the Family, 66*: 431-451. RESERVE

November 19. Missing data.
Assignment 2 (Logistic Regression) first draft peer feedback due to student and instructor
Enders, Craig K. 2010. Chapters 1-5. Applied Missing Data Analysis. Guilford Press.

November 26: No class

December 3. Missing data (continued).
***Assignment 2 (Logistic Regression) final draft due ***
Enders, Craig K. 2010. Chapters 6-10. Applied Missing Data Analysis. Guilford Press.

December 10. Complex survey data.
***Mini-assignment (Count Data) due ***
Lee, Eun Sul, and Ronald N. Forthofer. 2006. *Analyzing Complex Survey Data. 2nd edition.* (Quantitative Applications in the Social Sciences, 07-071). Thousand Oaks, CA: Sage Publications.

Article write-up due December 15