Sociology 2020:
Multivariate Statistical Methods II
(AKA Regression Models for Categorical
Dependent Variables)

Class Location and Times: Maxcy 108 Tuesdays 9-11:50
Class Instructor: Professor Leah VanWey
Office Location and Hours: BERT 305, Wednesday 9-10:30
TA: Heather Randell
TA Office Location and Hours: Maxcy 409, Tuesday 3-5

Course Goals:
    Sociology is becoming both more sophisticated in the application of quantitative data analysis techniques, including regression models, and more mixed-method. While some of you may become committed single-method scholars, it is more likely that you will mix qualitative and quantitative methods in single projects or over the course of your career. Regardless of the method(s) you use, you can expect to be asked to review papers or proposals, or to supervise students, using quantitative methods. To prepare you for these roles, this course provides you the foundation for analyzing data in which you wish to estimate a regression model with a categorical dependent variable, and for evaluating the work of others who are doing so. This includes not only having a rudimentary grasp of the assumptions and statistical foundations of the models, but also knowing things like how results relate to underlying causal models you might posit in your work. In short, my goal is that you will leave this course able to complete and write up a quantitative analysis without doing anything wrong (that's a higher bar than you might think), and with doing the additional sensitivity tests that make the difference between a good article and a great article.
    I firmly believe that the only way to learn statistics and skills in programming and data management, is by doing. In this course, you will complete assignments that are designed to lead you step-by-step through the various regression models that we will cover, and will then take those step-by-step models and implement your own analysis in a final paper. We will also read and discuss recently published articles using the models we are learning, in order to give you practice in the most important part of your job as a quantitative analyst: translating theory to analysis and results back to theory.

Organization of the Course:
    We will cover the topics covered in the first textbook listed below (Long 1997). These include the binary regression model (logit or probit), the ordered logit model, the multinomial logit model, and the Poisson and negative binomial models. Along the way, we will talk about the myriad tests that one must do to address common criticisms (in Sociology, not in some other social sciences). For each model (or general type of model), the in-class portion of the course will cover the basic statistical foundations, estimation, hypothesis testing, etc., and the interpretation of results. Outside of this time, you will actually estimate the models and tests using Stata and complete assignments that write up your results.

Requirements:
    This course requires full participation and engagement by students. I expect you to grapple with the textbook and example articles before we come to class, ask questions in class when lectures are

unclear, and to locate, prepare, and analyze data for a final project. Data for the majority of labs will be made available to you, but **will require cleaning and variable creation**. This is an important part of the research process, and you should not underestimate the amount of time that it will take. In addition, assignments for the course, including homework exercises and your final project, should be prepared in a professional format. **Tables should be formatted as if for journal publication, variables and data should be described as you would for the submission of a manuscript for publication (i.e. assuming the reader is not familiar with the data and variables), and all data manipulations, tests, and sample restrictions should be presented.**

Graded written and data analysis work will include four homework assignments and the final paper project. The homework assignments will involve the manipulation and analysis of data, and a writeup and formal presentation of results. In each, you will turn in the **formal written assignment with tables and text**, and an accompanying **stata do-file** and **log file**. Please do not print the log file; email it. The paper project will involve you turning in short written assignments periodically to ensure that you have started and are on the right track before the end of the semester.

The final project will involve your presenting a research question and an abbreviated literature review to motivate the analysis, your acquisition and preparation of data, analysis of these data, and a paper in a manuscript format (abbreviated theoretical section) reporting results and conclusions. The final paper that you turn in should be accompanied in printed form by an appendix reporting (in formal tables) the additional tests that you conducted that are only footnoted in the main text, and in electronic form by your stata data file, do-file and log file.

Books:

SUGGESTED: Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications

This book introduces the intuition and motivation behind the models, the estimation and interpretation, and other related topics in a clear and readable fashion. I find it a very useful reference even now.

REQUIRED: Long, J. Scott, and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press.

This book provides a very brief overview of each of the topics covered in the first book and covers the practical aspects of implementing the models in Stata and interpreting the output. It provides all the information that you need to make sure that you do not do anything wrong.

Articles:

For some meetings, we will read and discuss recent articles in top Sociology journals that use the methods that we are learning. Please read the articles before coming to class on the days indicated.

The full citations for articles are:

Gauchat, Gordon. 2012. "Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010" *American Sociological Review*, April 2012; vol. 77, 2: pp. 167-187.

Lin, Ken-Hou, and Jennifer Lundquist. 2013. "Mate Selection in Cyberspace: The Intersection of Race, Gender, and Education" *American Journal of Sociology* Vol. 119, No. 1, pp. 183-215. Stable URL: http://www.jstor.org/stable/10.1086/673129

Wang, Dan, and Sarah Soule. 2012. "Social Movement Organizational Collaboration: Networks of Learning and the Diffusion of Protest Tactics, 1960–1995" *American Journal of Sociology*, Vol. 117, No. 6, pp. 1674-1722. Stable URL: http://www.jstor.org/stable/10.1086/664685

COURSE OUTLINE

DATE	TOPIC	CORRESPONDING CHAPTER(S)	Articles
Tuesday January 28	Introduction Assessment / Review Matrix Notation Motivating the Binary Model	Long Ch. 1 Read Long and Freese Ch. 1-3 , and verify Stata components are installed in prep for homework; Ch. 9 is also a useful reference	
Tuesday February 4	Maximum Likelihood Estimation Estimation of the Binary Model	Long Ch. 2, 3	
Tuesday February 11	Interpretation of Coefficients, and Hypothesis Testing in the Binary Model	Long Ch. 4 Long and Freese Ch. 4	Gauchat 2012
Tuesday February 18	NO CLASS – "LONG WEEKEND"		
Tuesday February 27	Binary Model continued… Goodness of Fit		
Tuesday March 4	Interpretation of the Binary Model: Two Approaches to Predicted Probabilities	Long and Freese Ch. 3	Lin and Lundquist 2013
Tuesday March 11	Ordered Logit Models	Long Ch. 5 Long and Freese Ch. 5	
Tuesday March 18	Multinomial Models	Long Ch. 6 Long and Freese Ch. 6	
Tuesday March 25	NO CLASS – SPRING BREAK		
Tuesday April 1	Interpretation and Testing in Multinomial Models		
Tuesday April 8	Motivating Count Models Estimation of Count Models (and the difference between Poisson and negative binomial)	Long Ch. 8 Long and Freese Ch. 8	
Tuesday April 15	Interpretation of Coefficients in Count Models Advanced Testing in Count Models		Wang and Soule 2012
Tuesday April 22	Calculation of Predicted Values in Count Models		
Tuesday April 29	Fancy Stuff	Long Ch. 7	
Tuesday May 6	Wrap Up / Catch Up		
Friday May 16, 5pm	**FINAL PAPER DUE**		