

Professor Ted Gerber

Office Hours: Weds. 10-12, or by appointment
4450 Social Science Building
tgerber@ssc.wisc.edu

TA: Jungmyung Kim

O/H: Friday 12-2, OBA
2413 Social Science Bldg.
jungmyung.kim@wisc.edu

INTRODUCTION

This course covers selected expansions of basic (linear) multiple regression analysis—the topic of the pre-requisite course, Sociology 361—that are common in social science applications. It consists of three broad units. First, we closely consider core assumptions that underlie the use of Ordinary Least Squares (OLS) to estimate regression models for interval-level dependent variables, learn how to “diagnose” common violations of those assumptions in specific applications (multicollinearity, heteroskedasticity, and overly influential observations), and examine steps that can be taken to address these violations once they are diagnosed. Next, we take up forms of regression that are appropriate for analyzing dependent variables that are not interval-level, such as dichotomous variables, ordinal variables, categorical variables (with more than two categories), count variables, and interval variables that are truncated or censored. These forms of regressions in some ways resemble standard linear regression, but they require a different approach to estimating model parameters, “maximum likelihood estimation,” which we will learn about in a non-technical manner. Third, we examine forms of regression that are appropriate in applications where residuals are likely to be correlated, focusing on three common situations: 1) when the “cases” consist of observations on a single unit taken at different points in time (time series regression); 2) when they consist of a sample of units such as persons or countries sampled at different points in time (panel data regression); and 3) when they consist of individual units such as persons that are “nested” or “clustered” within higher-level units such as schools, neighborhoods, or states whose characteristics exert common effects on all those nested within them (hierarchical linear models or multilevel models.)

These extensions of regression beyond the core linear model make it more flexible and useful tool in social science research, thus helping practitioners and consumers of research realize the potential of regression for understanding relationships between variables based on sample data and avoiding common pitfalls that beset many applications. Thus, the larger objective of the course is to advance students’ familiarity with regression beyond the basic design and function of linear multiple regression models, to see how regression can be used to test a much wider array research hypotheses in the social sciences, while avoiding common pitfalls that can undermine conclusions if they are not dealt with. Throughout, the course emphasizes concrete applications to specific research problems using real data sets analyzed with statistical software, as well as how to interpret results.

By the end of the course, students should have a broader and firmer understanding of multiple regression models, including the assumptions on which they are based and the appropriate forms of regression to use depending on the nature of the data and problem at hand. Students will acquire the ability to perform a wider range of multiple regression analyses using the statistical software package, Stata, as well as another package (HLM) designed specifically for estimating

hierarchical linear models. In addition, students will build on their initial experiences from Sociology 361 (or an equivalent course) locating and obtaining large, publicly available data sets, defining a research question, setting up the right form regression analysis to test it, checking for problems or issues in the application of that specific regression technique, and presenting the findings in an intuitively appealing and professional fashion. They will also gain the ability to interpret and methodologically assess articles in sociological and other social science journals that use regression-based quantitative analysis techniques. Their familiarity with aspects of regressions models covered in Sociology 361—such as the inclusion of categorical variables on the righthand side of regression equations as sets of dummy variables, specifications of curvilinear relationships, interactions among righthand side variables, and basic concepts of mediation and confounding—will be reinforced and enhanced. The skills covered in the course will provide a foundation for further advanced work in social statistics or for original analyses of social science data for research purposes in an academic or professional context.

FORMAT

The course consists of lectures, lab sessions, and methodological discussions of published sociological articles (during lectures). Lectures will focus on conceptual material, while labs will emphasize software, computer work, practical tools for completing assignments, and review of problem sets. Note: all lecture and lab sessions are *mandatory*. Most or all labs will all involve use of Stata or HLM on laptops: if you do not have a laptop, one will be provided from the SSCC's mobile lab. All lab sessions will be run by the teaching assistant, Jungmyung Kim.

READINGS

The **required textbook** for the course is: *Regression Models for Categorical and Limited Dependent Variables*, by J. Scott Long (Sage Publications, 1997). It is highly recommended that students purchase the book, as all chapters will be assigned during the course.

Other books (selected chapters assigned);

Gordon, Rachel A. 2015. *Regression Analysis for the Social Sciences*. 2nd Edition. Routledge.
Ostrom, Charles W. 1990. *Time Series Analysis: Regression Techniques*. 2nd Edition. Sage.

Articles:

Bollen, Kenneth A. and Robert W. Jackman. 1985. "Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases." *Sociological Methods and Research*. 13: 510-42.

Bollen, Kenneth A., and Robert W. Jackman. 1995. "Income inequality and democratization revisited: Comment on Muller." *American Sociological Review* 60: 983-989.

Budig, Michelle J. and Paula England. 2001. "The wage penalty for motherhood." *American Sociological Review*. 66: 204-225.

DiPrete, Thomas and Jerry D. Forristal. 1994. "Multilevel Models: Methods and Substance." *Annual Review of Sociology*. 20: 331-57.

Gerber, Theodore P., Jane Zavisca, and Jia Wang. 2022. "Market and Nonmarket Pathways to Home Ownership and Social Stratification in Hybrid Housing Regimes: Evidence from Four Post-Soviet Countries." *American Journal of Sociology*. 128: 866-913.

Halaby, Charles. 2004. "Panel Models in Sociological Research: Theory into Practice." *Annual Review of Sociology*. 30: 507-54.

- Monk, Ellis P., Jr. 2015. "The Cost of Color: Skin Color, Discrimination, and Health among African-Americans." *American Journal of Sociology*. 121: 396-444.
- Muller, Edward N. 1995a. "Economic Determinants of Democracy." *American Sociological Review*. 60: 966-82.
- Muller, Edward N. 1995b. "Income inequality and democratization: reply to Bollen and Jackman." *American Sociological Review* 60: 990-996.
- Xie, Yu and Kimberlee A. Shauman. 1998. "Sex Differences in Research Productivity: New Evidence About an Old Puzzle." *American Sociological Review*. 63: 847-70.

REQUIREMENTS

- 1) Students must complete 6 *problem sets* (two problems in the first five, four in the last one), graded for accuracy. Assignments will be posted on the course's Canvas website on the Monday night preceding their Friday due day. They must be submitted *by 12 noon* on the due dates (all Fridays). Each problem set is worth 5% of your final grade, aside from the last one, which is worth 10%. Late problem sets will be penalized, in fairness to those who submit it on time and in fairness to the TA, and no problem sets will be accepted after the following Monday (when the prior Friday's problem sets will be reviewed during lab sessions). Lab sessions on the Mondays starting the weeks when problem sets are due will include in-class practice problems that will prepare students to complete the problem sets due later that week. Overall, problem sets will be worth 35% of your final grade.
- 2) Students must complete 5 *methodological analyses* of sociology articles that use one of the techniques covered in the course. These must be 1-2 pages, single spaced, and must focus on methodological aspects (as opposed to theoretical or substantive arguments) of the assigned articles. They will be due on Tuesdays or Thursdays at 2 pm (before the start of lecture), because some lecture time will be devoted to discussions of the articles. They will be graded for accuracy, depth, insight, clarity, and presentation. More detailed instructions will be distributed during the second week of class. They are each worth 3% of your final grade (thus 15% combined).
- 3) A *seminar paper* presenting an original data analysis using one of techniques covered during the semester. The paper should be no more than 20 pages (double spaced) in length. I will provide a handout with more detailed instructions early in the semester. The paper offers students the chance to execute an original research project using regression analyses. It will be due at 5 pm on the designated final exam day for the course, May 8. It will count for 50% of your final grade

LEARNING OBJECTIVES

Students completing the course will be ready to conduct their own research applying extended forms of regression analysis (beyond the basic linear model) at a level suitable for research assistants on faculty research projects and, potentially, publication in the form of articles in social science journals or professional reports. They will be able to find, download, and format social science data for analysis of concrete research questions. They will know how to diagnose common problems in the applications of regression and take actions to address these problems. They will acquire the skills to identify and use the appropriate form of regression for different types of dependent variables and data structures, and to critically evaluate the methodological

strengths and weaknesses of published sociological research using the methods covered. They will be able to write an article-length description of an independent research project that uses these techniques. The course will provide a solid foundation for taking advanced courses in social statistics that assume a broad understanding of regression in social science applications.

SCHEDULE

Week 1 (Jan 23-27): Introduction and Review of Linear Regression

Jan 26: Read Long, chs 1-2

Week 2 (Jan 30/Feb 3) Regression Diagnostics

Jan 30 Lab: Lab introduction, Stata review, initial diagnostic commands using Stata

Read: Gordon, ch. 11; Bollen and Jackman (1985)

Week 3 (Feb 6/10) Regression Diagnostics (cont.)

Feb 6 Lab: More diagnostic commands using Stata, practice problems for Problem Set 1

Read: Muller (1995a), Bollen and Jackman (1995), Muller (1995b)

Feb 10: Problem Set 1 due at 12 pm

Week 4 (Feb 13/17) Logit/Probit and Maximum Likelihood Estimation

Feb 13 Lab: Review Problem Set 1 solutions and solidify comprehension of diagnostics

Read: Long, chs. 3-4

Feb 14: Methodological Analysis 1 due by 2:00 pm: Muller (1995a, 1995b), Bollen and Jackman (1995)

Methodological Discussion: Muller (1995a, 1995b); Bollen and Jackman (1995)

Week 5 (Feb 20/24) Logit/Probit and Maximum Likelihood Estimation (cont.)

Feb 20 Lab: Stata estimation and postestimation commands, Excel application, and practice problems for Problem Set 2

Read: (Monk 2015)

Feb 24: Problem Set 2 due at 12 pm

Week 6 (Feb 27/Mar 3) Ordinal Logit/Probit and Multinomial Logistic Regression

Feb 27 Lab: Review Problem Set 2 solutions and solidify comprehension of logit/probit models

Read: Long, chs. 5-6

Mar 2: Methodological Analysis 2 due at 2:00 pm (Monk 2015)

Methodological Discussion: Monk (2015)

Week 7 (Mar 6/10) Ordinal Logit/Probit and Multinomial Logistic Regression (cont.)

Mar 6 Lab: Stata estimation and postestimation commands, Excel application, and practice problems for Problem Set 3

Mar 10: Problem Set 3 due at 12 pm

Spring Break (Mar 11-19)

Week 8 (Mar 20/24) Tobit, Poisson, and Negative Binomial Regression

Mar 20 Lab: Review Problem Set 3 solutions and solidify comprehension of ordinal and multinomial models

Read: Long ch. 7-8; Read: Gerber, Zavisca, and Wang (2022)

Mar 21: Methodological Analysis 3 due at 2:00 pm (Gerber, Zavisca, and Wang 2022)

Methodological Discussion: Gerber, Zavisca, and Wang (2022)

Week 9 (Mar 27/31) Tobit, Poisson, and Negative Binomial Regression (cont.)

Mar 27 Lab: Stata estimation and postestimation commands, Excel application, and practice problems for Problem Set 4.

Read: Xie and Shauman (1998)

Mar 31: Problem Set 4 due at 12 pm

Week 10 (Apr 3/7) Basics of Time Series Analysis

Apr 3 Lab: Review Problem Set 4 Solutions and solidify comprehension of Tobit, Poisson, and Negative Binomial Regression

Read: Ostrom (1990), chs. 1-3

Apr 4: Methodological Analysis of Xie and Shauman (1998) due by 2:00 pm

Methodological Discussion: Xie and Shauman (1998)

Apr 6: Virtual lecture session (Zoom link to be distributed)

Week 11 (Apr 10/14) Panel Data Models

Apr 10 Lab: Stata Commands and practice problems for Problem Set 5

Read: Halaby (2004)

Apr 14 Problem Set 5 due at 12 pm

Week 12 (Apr 17/21) Panel Data Models (cont.)

Apr 17 Lab: Review Problem Set 5 solutions, solidify comprehension of Basic Time Series

Read: Budig and England (2001)

April 21: Methodological Analysis of Budig and England (2001) due by 2:00 pm

Methodological Discussion: Budig and England (2001)

Week 13 (Apr 24/28) Hierarchical Linear Models

Apr 24 Lab: Stata estimation commands for panel data models, learn HLM

Read: DiPrete and Forristal (1994)

Week 14 (May 1/5) Hierarchical Linear Models (cont.)

May 1 Lab: Practice problems for Problem Set 6

May 5 Problem Set 6 (worth 10 points) due at 12 pm

May 8, 5:00 pm: Final Paper Due!!!