

Sociology 401

Statistical Analysis of Social Data: Applied Regression Methods I

Instructor: Professor Juan Onésimo Sandoval
Office: Sociology Department, 1810 Chicago, Room 314
Office Hours: Tuesday 2:00 p.m. to 4:00 p.m. and by Appointment
Phone: 847-467-4139
Email: j-sandoval@northwestern.edu
TA: Nehal Patel (n-patel@northwestern.edu)

Course Description

This course is part of the quantitative methods sequence for graduate students in sociology. Students in other disciplines are welcome to take this course. After a brief review of basic statistics, the course will move quickly to the theory, methods, and practice of linear regression analysis. We will cover ordinary least squares (OLS) methods, multiple regression, practical regression techniques, OLS assumptions, regression diagnostics, and some bi-variate non-linear regression techniques. We will use statistical software in this class. STATA, SAS, and SPSS will be used heavily for numerical computation and data analysis. However, you also be required to manually calculate several homework problems. You can use Excel if you wish for the written homework. Students completing this class are expected to: (1) read, understand, and evaluate sociological research that uses OLS regression; (2) use OLS regression techniques and statistical software to conduct quantitative research; and (3) become prepared for more advanced statistical methods such as Sociology 401-2 and 402.

Unlike similar courses that are either too technical or too interpretative, this course adopts a middle ground and emphasizes the linkages between mathematics and real data through calculation. Throughout the quarter, we will be using STATA, SAS, and SPSS as a calculator to derive statistical output from real data. The philosophy underlying this computational approach is that knowing how to calculate statistical results helps to better understand both statistical foundations and practical techniques of the method.

Prerequisites

This course is built upon Sociology 400, and thus students are assumed to have working knowledge of elementary statistics (up to analysis of variance, i.e., ANOVA) and one of the following, STATA, SAS, and SPSS software. Those who do not meet either one of these two requirements should not take the course. Under special circumstances, students who have adequate statistical experience with other software packages (e.g., S-PLUS or R-PLUS) may be allowed to enroll. If you belong to this category, you need to realize that other software will not be supported and that it is your responsibility to develop adequate experience in STATA, SPSS or SAS.

Requirements, Coursework, and Grading

The course materials will be tightly organized and students are expected to attend every class and lab session. Before each lecture, students are required to read the assigned readings and exercises. Students are assumed to have access to STATA, SAS and SPSS software and a working data set that will be used for most of the coursework (those who have not had their own data ready should consult the TA and acquire the data in the first two weeks).

Assignments: There will be 6 assignments. Students can work in teams of two depending on class size for all the assignments except the final assignment. Students are required to submit a final research paper. The research paper is not expected to be fully polished, but it should have all the makings of a journal article, such as a clear research question, an appropriate data set, sufficient literature review, justified hypotheses, insightful data analysis, as well as proper interpretation and presentation of the major findings. The final paper is due by 5 p.m., March 12 (Monday).

Class Participation: Class participation and lab attendance is required.

About Deadlines: All assignments should be turned in by the deadline. Late assignments will not be graded except under very special circumstances (in which permission of an extended deadline must be obtained in advance).

Personal Meeting

During the second half of the quarter, I would like to have a 30-minute personal meeting with each of you to discuss your progress and problems. I hope to get to know each of you on a personal level and help you advance your research topic.

Evaluation will be based on the following:

Assignment #1	15%	Due January 22, 2007
Assignment #2	15%	Due January 29, 2007
Assignment #3	10%	Due February 5, 2007
Assignment #4	10%	Due February 12, 2007
Assignment #5	15%	Due March 5, 2007
Assignment #6	25%	Due March 12, 2007
Class & Lab participation	10%	
Total	100%	

Course Management System (CMS)

We will use the Course Management System for many of the administrative tasks associated with this class. I will put materials for in-class exercises, as well as the data used in the exercises and problems on the CMS site. In addition, I will encourage the students to use this system to discuss course topics and questions as well as to collaborate on the problems sets. Finally, this will be a useful medium for us to directly and indirectly receive ongoing feedback as to how the course is going. Of course, for this to work we need you to use the system to let us know your questions and concerns. It should also be noted that we see this system not as a replacement, but as a supplement to traditional, face-to-face student/faculty and student/student interaction.

Textbooks

Allison, P. D. (1999), *Multiple Regression: A Primer*, Pine Forge Press, Thousand Oaks.

Chatterjee, Samprit, Ali Hadi, and Bertram Price. 1999. *Regression Analysis by Example*. New York: Wiley-Interscience Publication.

Recommended Textbooks

Achen, Christopher H. 1982. *Interpreting and Using Regression*. New York: Sage.

Hardy, Melissa A. 1993. *Regression with Dummy Variables*. New York: Sage.

Lewis-Beck, Michael S. 1980. *Applied Regression: An Introduction*. New York: Sage.

Weisberg, Sanford. 1985. *Applied Linear Regression*. New York: Wiley-Interscience Publication.

Class Schedule

Week 1: Review of Basic Statistics Part I and Part II (January 8)

Frankfort-Nachmias and Leon-Guerrero - Chapter 4 – Measures of Central Tendency
Frankfort-Nachmias and Leon-Guerrero - Chapter 13 – Testing Hypotheses

Week 2: NO CLASS (January 15)

Week 3: Simple OLS Regression Part (January 22)

*Allison – Chapter 5
**Chatterjee, Hadi, and Price – Chapter 1-2

Week 4: Multiple OLS Regression I (January 29)

**Chatterjee, Hadi, and Price – Chapter 3

Week 5: Multiple OLS Regression II (February 5)

**Chatterjee, Hadi, and Price – Chapter 5

Week 6: OLS Assumptions and Regression Diagnostics I (February 12)

*Allison – Chapter 1-3
**Chatterjee, Hadi, and Price – Chapter 4

Week 7: OLS Assumptions and Regression Diagnostics II (February 19)

*Allison – Chapter 7
**Chatterjee, Hadi, and Price – Chapter 5-6

Week 8: Practical Regression Techniques I (February 26)

*Allison – Chapter 8
**Chatterjee, Hadi, and Price – Chapter 7-8

Week 9: Practical Regression Techniques II (March 5)

*Allison – Chapter 6 & 9
**Chatterjee, Hadi, and Price – Chapter 9-11