The linear regression model is the most commonly used statistical method in political science. One is likely to come across some type of regression analysis in just about any recent issue of a top political science journal. As a result, it is important for political scientists to understand the theory and practice of regression analysis.

This seminar is designed to introduce students to multivariate regression analysis. We will focus most of our energy on the interpretation and use of single equation, linear regression models. Toward the end of the semester, we will cover some more advanced topics (including common nonlinear models). We will cover some mathematical and statistical principles used in regression analysis. The seminar will prepare students to use regression analysis appropriately in original research, as well as understand and critically evaluate published research that relies on regression results. Though we begin with a review of basic regression, this course presumes that students are familiar with research design and basic statistics.

REQUIRED BOOKS


There will be some journal articles included in the assigned readings (on MyGateway).
RECOMMENDED READING


We will use STATA (version 10) as the main software program for the course, for both class examples and homework problems. For those of you with earlier versions of Stata, this will pose a bit of a challenge. The commands are pretty much the same in Stata 10, but data files are sometimes not compatible between different versions of the program. The Hamilton book contains several chapters on commands used for regression analysis, as well as general tips for Stata users.

SUPPLEMENTARY READINGS

For those who want to read more, the following books provide a basic introduction to several topics covered in the course. The Achen and Berry and Sanders monographs cover the OLS regression model, while the two Long books cover maximum likelihood models for limited dependent variables. The Wooldridge book is a popular econometrics text, somewhat similar to Fox. Gill’s book covers many mathematical concepts in more detail.


COURSE REQUIREMENTS

Class Performance: Scheduled class meetings will largely follow a lecture format. However, from time to time I will call on students for answers and explanations of topics in the assigned readings. I expect students to complete the required readings before class. I will make available data files I use for classroom examples, in case you want to reanalyze the data and practice statistical techniques on your own.

Problem Sets: Throughout the semester, I will distribute problem sets in class. Each assignment may involve explaining regression concepts, computation by calculator or spreadsheet, the use of a computer software program, or all of the above. The computer assignments can be completed using Stata, which is available on all classroom and lab computers on campus, as well as the Econ Lab (452 SSB).

Term Paper: Students are required to write a 15 page paper on a social science topic that involves estimating at least one regression model. You may use any data set and statistical software in completing the analysis. You should think of the paper as an expanded version of the methodology section in a typical article. In organizing the paper, you should:

1. Briefly introduce the substance of the topic
2. Justify the use of regression analysis
3. Describe your regression diagnostic efforts in detail
4. Interpret the results of your analysis
5. Discuss the generalizeability of the results

A one-page proposal for the paper is due at the beginning of class during the fifth week of the semester (by Feb. 18). Earlier proposals are strongly encouraged. Proposals related to on-going or previous original research (such as for another seminar or conference paper) are also encouraged. The term paper is due Wednesday, May 6. You will present your completed research project to the class (a fifteen minute presentation) during the last week of classes.

Midterm Exam: A 48-hour take-home midterm exam will be held during the 6th week of classes (February 23).

Grade Distribution
Problem Sets: 40%
Term Paper: 30%
Presentation: 10%
Midterm Exam: 20%
COURSE OUTLINE

I. Introduction to OLS Regression

Week 1 (January 21)
Introduction
Review of Bivariate Regression

Read:
Fox, chapter 1 and Appendix A
Hamilton, chapter 1

Applications:


Week 2 (January 26-28)
Mathematics and Statistics Review
Ordinary Least Squares and Its Assumptions
Introduction to Stata

Read:
Fox, chapters 2 and 3, Appendix D pp. 65-73, Appendix B pp. 5-15 (on web)
Berry, chapters 2-4
Hamilton, chapters 1-2

Week 3 (February 2 – 4)
Ordinary Least Squares, Statistical Inference, and Hypothesis Testing

Read:
   Fox, chapters 5 and 6, and Appendix D pp. 89-91 (on web)
   Hamilton, chapter 6

Supplementary Readings:

   With Coincidence, Improbability, and Chance.” *American Journal of Political
   Science* 36:1023-1046.

   *Political Research Quarterly* 52(September):647-674.

Application:

II. Scalar Introduction to Multiple Regression

Week 4 (February 9-11)
Estimation of Multiple Regression & Functional Form

Read:
   Fox, chapters 4 and 5
   Berry, chapter 2-5

Applications:

   Review* 102:107-123.

   (non-linear functional form)


   Countries: A Cross Sectional and Longitudinal Analysis.” *International
   Organization* 37:441-467.
Paper proposal due Feb. 18

Week 5 (February 16-18)
Multiple Regression and Statistical Inference

Read:
Fox, chapters 5 and 6

Supplementary Readings:
Articles by Green, King, and Lewis-Beck & Skalaban on the use of $R^2$ in *The Political Methodologist*


Applications

Campbell, James E. 1997. “The Presidential Pulse and the 1994 Midterm Congressional Election.” *Journal of Politics* 59:830-857. (Note the use of $R^2$ in Figure 1)


Monday, Feb. 23
Distribute take-home midterm exam (due in class Wednesday, Feb. 25)
III. Matrix Introduction to Multiple Regression

Week 6 (February 23-25)
The Linear Model in Matrix Form

Read:
Fox, chapter 9

IV. Dummy and Interaction Effects in OLS Regression

Weeks 7 and 8 (March 2-11)
Dummy Variables, Interactions, and Presenting Results

Read:
Fox, chapter 7
Berry, pp. 60-67
Hamilton, pp. 192-202


Applications:


Supplementary Reading:


V. **Regression Diagnostics: Finding and Treating Estimation Problems**

**Week 9 (March 16-18)**

**Heteroskedasticity**

**Read:**

Fox, chapter 12  
Berry, pp. 67, 72-81  
Hamilton, chapter 7 (pp. 209-214)


**Applications:**


**No classes the week of March 23 (Spring Break)**

**Week 10 (March 30 – April 1)**

**Analyzing Residuals: Outliers, Leverage, Non-Linearity, and Non-Normality**

**Read:**

Berry, pp. 81-82  
Fox, chapter 11 and parts of chapter 12  
Hamilton, chapter 7 (pp. 214-224), chapter 9, and pp. 141-145

Supplementary Reading:

Applications:


Joshua Benton and Holly K. Hacker. 2004. “Cheating may be Pervasive.” *Dallas Morning News*, December 18, 1A, 14A.


Week 11 (April 6 – 8)
Multicollinearity, Specification Error, and Measurement Error

Read:
Berry, pp. 24-27, 30-45, 49-60
Fox, chapter 13
Hamilton, pp. 224-228, chapter 12

Applications:


Using Highly Correlated Data Sets.” *Political Analysis* 11:196-203. (Importance of measurement of the dependent variable)


**Week 12 (April 13-15)**  
**Time Series Data and Autocorrelation**

**Read:**  
Fox, chapter 16  
Berry, pp. 67-72, 78-81  
Hamilton, chapter 13

**Applications:**  


**VI. Advanced Topics in Regression**

**Week 13 (April 20-22)**  
**Limited Dependent Variables and Maximum Likelihood Estimation: Basics**

**Read:**  
Fox, chapters 14 and 15

**Applications:**  
**Supplementary Reading:**


**Week 14 (April 27-29)**

Maximum Likelihood Estimation: Applications

**Read:**

Hamilton, chapter 10


**Supplementary Reading:**


**Applications:**


**Week 15 (May 4-6)**

Present Research Papers

**Read:**

Wooldridge, chapter 19