

---

Washington University in St. Louis  
Department of Political Science

**Pol Sci 572. Quantitative Methods in Political Analysis II: Linear Models**

Fall 2003 Semester

---

**Instructor**

Andrew D. Martin, Ph.D.  
Office: Eliot Hall 326  
Email: [admartin@artsci.wustl.edu](mailto:admartin@artsci.wustl.edu)  
WWW: <http://adm.wustl.edu>  
Phone: (314) 935-5863  
Hours: By appointment only (see website for availability)

**Class Schedule**

Tuesday 6:30-9:00 PM  
Eliot Hall 314

## Introduction

Political Science 572 is the second of a sequence of three graduate methodology courses in Political Science at Washington University. This course builds on the foundations of probability theory, distribution theory, and hypothesis testing developed in Political Science 571. All students are expected to have an understanding of these topics, as well as a basic understanding of differential calculus, integral calculus, and matrix algebra. This course will focus on a particular statistical model used throughout the social and natural sciences—the linear regression model—through the lens of probability modeling. We develop this model, both in scalar and matrix form, and discuss issues of estimation, inference, and specification. We also study various diagnostic tools one can use to test for the violation of the assumptions of the classical linear regression model, and remedies for these problems.

In addition to these statistical issues, we also cover a series of topics that are important to applied political science. We begin by discussing model specification, and the difficulties of measuring theoretically important quantities. We also explore the substantive interpretation of statistical models by reading applied work in the field, discussing the coding of variables, learning data management skills, and writing essays about statistical analyses. Finally, there will be instruction in two commonly used statistical packages: Stata and R.

## Requirements and Evaluation

The requirements for this course are simple—attend class and complete the assignments on time. This is a lecture-based workshop, which will proceed as quickly or slowly as is necessary. Since this is a graduate level course, it need not be said that class attendance is mandatory. Late assignments will not be accepted. Students will be required to complete weekly, graded homework assignments. I will discuss these assignments in class. Some assignments will be computational or

analytical in nature, similar to those from last semester. Others will involve writing short essays or research notes. Many will come from the textbook. Clarity of presentation and argument are of utmost importance when preparing these assignments. There will be approximately twelve (12) homework assignments. Homeworks count for 40% of the final course grade. Students are to complete assignments on their own.

There will also be two written examinations: a mid-term and a final. The midterm exam is worth 25% of the final grade, and the final exam is worth 35%. Grades will be assigned based on the written assignments and the examinations. In graduate courses assigning grades is an artificial, and oftentimes meaningless, process. Each assignment and exam will be assigned a letter grade (based on a curve if appropriate). The letter grades represent: A = mastery of the material, B = satisfactory work (passing), C = unsatisfactory work for graduate school (below passing), D or F = completely unsatisfactory (failing). Because this course is a prerequisite for future courses, no incompletes will be assigned for any reason. Failure to meet the requirements of the course will result in a failing grade.

## Readings and Software

There is one required text for this course. This text is available at the bookstore or through your favorite internet bookseller:

John Fox. 1997. *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage. ISBN: 080394540X.

I will supplement the material in Fox with two texts on statistics. One is a nice introduction to mathematical statistics:

Morris H. DeGroot and Mark J. Schervish. 2002. *Probability and Statistics*, Third Edition. Reading, MA: Addison-Wesley. ISBN: 0201524880.

The other is a good undergraduate textbook which includes a nice introduction to linear models:

Alan Agresti and Barbara Finlay. 1997. *Statistical Methods for the Social Sciences*, Third Edition. Upper Saddle River, NJ: Prentice Hall. ISBN: 0135265266.

I will also assign some political science articles throughout the semester; some will be methodological in nature, and others will be applications.

Practical training in the use of modern statistical software is an important component of the course. For the most part I will take a “learn-by-doing” approach, although at times we will meet in the computer classroom for detailed instruction. In years past I have taught this course using the commercial Stata package (<http://www.stata.com>). Stata is a powerful package, and has become standard fare in political science. Everything we will do in class this semester can be done in Stata. John Fox has posted example datasets and Stata code at his homepage:

<http://socserv.socsci.mcmaster.ca/jfox/>

Students can purchase academic versions of Stata for most platforms at (somewhat) reasonable prices.

In this iteration of the course I will spend most of my time talking about the free, open source R package (<http://www.r-project.org>). As I will argue throughout the semester, R has a number of nice features that make it ideally suited for applied data analysis. I have assigned an optional textbook about linear models in the R package:

John Fox. 2002. *A R and S-Plus Companion to Applied Regression*. Thousand Oaks, CA: Sage. ISBN: 0761922806.

All of the code discussed in this book is available on the web. Students can choose either package for course assignments. Ideally students will learn both packages, as each has advantages and disadvantages. I will be available during office hours for software assistance.

To provide a forum for communication outside of class, not only between instructor and student, but also between students, I have established an email mailing list. It is necessary for each of you to sign up for the list, and I encourage you to contribute to it. Oftentimes if you have a question about the text, lecture, or software, many of your colleagues will as well. It is best to use the listserv as a forum to pose the questions you have, and to answer the questions that are posed. Of course you can ask the instructor these questions, but the common good will be served by using the mailing list.

## Tentative Course and Topics Schedule

### September 2

Fox (1997) Chapters 1-2. Fox (2002) Chapters 1-2.

### September 9

Fox (1997) Chapters 3, 5 (skim Chapter 4). Fox (2002) Chapter 3.

### September 16

Fox (1997) Chapter 6. Fox (2002) Chapter 4.

### September 23

Fox (1997) Chapter 7.

Friedrich, Robert J. 1982. "In Defense of Multiplicative Terms in Multiple Regression Equations." *American Journal of Political Science*. 26: 787-833.

Wright, Jr., Gerald C. 1976. "Linear Models for Evaluating Conditional Relationships." *American Journal of Political Science*. 20: 349-373.

## September 30

Fox (1997) Chapter 8.

## October 7

Catch-up and review.

King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science*. 30: 666-687.

## October 14

Midterm Exam.

## October 21

Fox (1997) Chapter 9, Appendices C and D (review linear algebra).

*Political Analysis* Symposium on  $R^2$  (1990).

## October 28

Fox (1997) Chapter 10, Appendix B.

## November 4

Fox (1997) Chapter 11. Fox (2002) Chapter 6.

## November 11

Fox (1997) Chapter 12, 13.

Farrar, Donal E., and Robert R. Glauber. 1967. "Multicollinearity in Regression Analysis: The Problem Revisited." *The Review of Economics and Statistics*. 49: 92-107.

## November 18

Fox (1997) Chapter 14.

## December 2

Fox (1997) Chapter 15.

**TBA**

Final Exam