

**PA 552: Quantitative Analysis of Political Data  
Spring 2003**

Dave Weimer  
262-5713 or 263-2325  
weimer@lafollette.wisc.edu

Mondays/Wednesdays, 9:00 to 10:45 a.m.  
225 Ingram

Office Hours: Wednesdays 2 to 4 p.m., 201 La Follette  
Mondays and Wednesdays 11 a.m. to noon, 215 North Hall  
Other times welcome by appointment.

Effective participation in the social sciences requires familiarity with the basic elements of multivariate statistics. As social scientists rarely have the opportunity to study phenomena or behavior through controlled experiments, empirical tests of hypotheses derived from theory must often be coaxed either from data collected without the benefit of random assignment or from data that "happens" to be available as a byproduct of some non-research process. It is usually necessary, therefore, to use multivariate techniques to control statistically for those factors that cannot be controlled by random assignment. Absent familiarity with these basic techniques, social scientists cannot critically evaluate empirical results in their substantive areas of interest. Without some facility for actually using the techniques, they are less likely to be able to contribute in an important way to the testing of theory or even to the description of complicated phenomena.

Our objective is to prepare for the roles of consumer and producer of multivariate statistical analysis. Because it is commonly used, intuitively appealing, and fairly flexible, we focus on the basic linear regression model. It also provides a frame of reference for considering other techniques. We try to develop appropriate practical use and intuitive understanding rather than an ability to prove theorems. At the same time, however, we must be careful to develop an adequate theoretical base to allow continued learning beyond the course. Consequently, although we will cover relatively few formal proofs in class, we will go through a number of derivations to convey key points and increase your capability for continued learning after the course.

***Math Camp***

Applying some basic concepts and techniques drawn from calculus and linear algebra enable us to develop a deeper understanding of multivariate estimation and inference. As some of you have not been previously exposed to these concepts and techniques, or were exposed but desire a refresher, we will use some class time for math review. We will use the entire **second and third classes** to review basic differential calculus. In subsequent classes, we will hold a "math camp" prior to each class from **9 a.m. to 9:30 a.m.** to continue to review the basic math we will be using in the course. The pace will be as slow as necessary to bring everyone along. Once we have covered the topics we require for the course, we will either stop meeting at this time or use the time for going over problem sets.

### *Statistical Computing*

A number of course assignments will require you to use the STATA statistical package. Enough guidance will be provided for the assignments. I highly recommend that you concurrently take PA 553 (1 credit), which will develop your statistical computing skills in more depth.

### *Course Requirements*

*Examinations:* Midterm (20 percent) on **March 12**; final (50 percent) **as scheduled**.

*Assignments:* Approximately weekly assignments will be in a variety of formats: problem sets, computing exercises, Monte Carlo experiments, and memoranda tied to data analysis (20 percent).

*Project:* Attempt to answer a disciplinary or policy question by applying techniques learned in course to data that you have assembled (10 percent). Due **May 5**.

### *Texts*

The following texts are available in the University Bookstore and on reserve at the College Library:

Damodar N. Gujarati, *Basic Econometrics* 4<sup>th</sup> (New York: McGraw-Hill, 2003).

William H. Greene, *Econometric Analysis* 4<sup>th</sup> (New York: Macmillan Publishing Company, 2000).

Gudmund R. Iversen, *Calculus* (Thousand Oaks, CA: Sage Publications, 1996).

The text by Gujarati provides clear and accessible coverage of course topics. Greene provides a much more comprehensive survey of the theory underlying the commonly used basic techniques. If you are planning on doing methods as a field, then I highly recommend Greene. Otherwise, I recommend Gujarati.

Iverson serves as a text for our calculus review. If you already have a calculus text, then no need to purchase Iverson.

## ***Outline of Topics***

### ***I. Introduction***

Overview

### ***II. Calculus Review***

Derivatives  
Optimization

Iversen, 1, 2, 4

### ***III. Bivariate Regression***

History; multiple regression and hypothesis testing  
Fitting curves to data  
Correlation and regression  
Ordinary least squares (OLS)  
Hypothesis testing, power, confidence intervals  
Properties of least squares estimators  
Maximum likelihood estimators (MLEs)

Gujarati, 1 to 6

### ***IV. Multivariate Regression***

Review of matrix notation  
Gauss-Markov theorem and BLUE estimators  
Properties of estimators  
Statistical inference

Gujarati, Appendix B, C, 7 and 8; Greene, 2 and 6

David Weimer and Aidan Vining, *Policy Analysis: Concepts and Practice* (Englewood Cliffs, N.J.: Prentice-Hall, 1999), Chapter 15: "Revising the Lead Standard for Gasoline." (On reserve)

### ***V. Model Specification***

Non-linear models, Cobb-Douglas models, interaction terms Indicator variables  
Analysis of residuals  
Specification error

Gujarati, 9; Greene, 8

## **VI. Pathologies and Treatments**

Multicollinearity  
 Heteroscedasticity and generalized least squares (GLS)  
 Feasible GLS  
 Autocorrelation  
 Aggregation bias  
 Measurement error

Gujarati, 10 to 13, 17; Greene, 11 to 13

## **VII. Models with Discrete Dependent Variables**

Contingency table analysis  
 Linear probability models, logit, and probit

Gujarati, 15; Greene, 19

R. Michael Alvarez and Jonathan Nagler, "When Politics and Models Collide: Estimating Models of Multiparty Elections," *American Journal of Political Science* 42:1 (1998), 55-96. (J-Store at [www.jstor.org](http://www.jstor.org))

## **VIII. Simultaneous Equation Models**

Identification  
 Estimation: instrumental variables; two-stage least squares

Gujarati, 18 to 20; Greene, 15 and 16

David Figlio and Jens Ludwig, "The Effects of Private Schooling on Academic and Non-Academic Outcomes," National Bureau of Economic Research Working Paper, October 2001. (NBER at [nber.org](http://nber.org))

Robert S. Erikson and Thomas R. Palfrey, "Campaign Spending and Incumbency: An Alternative Simultaneous Equations Approach," *Journal of Politics* 60:2 (1998), 355-373. ((J-Store at [www.jstor.org](http://www.jstor.org)))

Larry M. Bartels, "Instrumental and 'Quasi-Instrumental' Variables," *American Journal of Political Science* 35:3 (1991), 777-800. (J-Store at [www.jstor.org](http://www.jstor.org))

**IX. *Additional Topics as Time Permits***

Pooled Time Series/Cross Sections  
Censored data  
Selection models

Gujarati, 16; Greene, 20

Nathaniel Beck and Jonathan N. Katz, "What to Do (and Not to Do) with Time-Series Cross-Section Data," *American Political Science Review* 89:3 (1995), 634-647. (J-Store at [www.jstor.org](http://www.jstor.org))

\*\* \*\* \*