PS 552: Multivariate Statistical Inference
Spring 2010

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

Mondays/Wednesdays, 8:30 to 10:15 a.m.
Van Vleck B223

Office Hours: Mondays 3:30 p.m. to 5 p.m., 201 La Follette
Mondays and Wednesdays 10:30 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 "happen" 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 primarily on the basic linear regression model. It also provides a frame of reference for considering other techniques that we will consider. 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 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 **8:30 a.m. to 9:00 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. However, I highly recommend that you concurrently take PS 553 (1 credit), which will develop your statistical computing skills in more depth. It will also help you develop effective data handling skills that will be useful as you begin your own research projects.

Course Requirements

Examinations: Midterm (20 percent) on **March 10**; final (50 percent) **as scheduled** (currently May 11 at 2:45 p.m.).

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 they are on reserve at the College Library:

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

William H. Greene, *Econometric Analysis* 5th (New York: Macmillan Publishing Company, 2002).

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 and you already have some mathematical confidence, then I recommend Greene. Otherwise, I recommend Gujarati. In any event, I attempt to make lectures self-contained so the primary use of either text is to get a second view. Therefore, if you already have a comparable text, then you do not necessarily have to purchase either of these texts.

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

Readings and exercises are available on learn@uw.

Outline of Topics

I. *Introduction*

Overview

II. *Calculus Review*

Derivatives
Optimization

Iversen, 1, 2, 4

III. *Bivariate Regression*

History
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, 8

David Weimer and Aidan Vining (2005) *Policy Analysis: Concepts and Practice* 4th Ed. (Englewood Cliffs, N.J.: Prentice-Hall,), Chapter 18: "Revising the Lead Standard for Gasoline," 452-476. (Available at learn@uw.)

Phil Cook's lessons on presenting statistical analysis. (Available at learn@uw)

V. *Model Specification*

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

Gujarati, 9

Thomas Brambor, William Roberts Clark, and Matt Golder (2006) Understanding Interaction Models: Improving Empirical Analysis. *Political Analysis* 14(1), 63-82. (http://pan.oxfordjournals.org/cgi/reprint/14/1/63)

VI. *Pathologies and Treatments*

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

Gujarati, 10 to13, 17

VII. *Models with Discrete Dependent Variables*

Contingency table analysis
Linear probability models, logit, and probit

Gujarati, 15

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

Chunrong Ai and Edward C. Norton (2003) Interaction Terms in Logit and Probit Models. *Economic Letters* 80(1), 123-129. (Available at learn@uw.)

Tue Tjur (2009) Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination. *American Statistician* 63(4), 366-372. (http://pubs.amstat.org/doi/pdfplus/10.1198/tast.2009.08210)

VIII. *Simultaneous Equation Models*

Identification

Estimation: instrumental variables; two-stage least squares

Gujarati, 18 to 20

Joshua D. Angrist and Alan B. Kruger (2001) Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives* 15(4), 69-85. (J-Store at www.jstor.org)

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

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

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

Panel data
Censored data
Seemingly unrelated regressions
Selection models
Hierarchical models
Regression discontinuity

Gujarati, 16, 17

Charles H. Franklin (1989) Estimation Across Data Sets: Two-Stage Auxiliary Instrumental Variables Estimation (2SAIV). *Political Analysis* 1 (1989), 1-24. (Available at learn@uw.)

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

Curtis S. Signorino (1999) Strategic Interaction and the Statistical Analysis of International Conflict. *American Political Science Review* 93(2), 279-297. (J-Store at www.jstor.org)

Michael Tomz, Joshua A. Tucker, and Jason Wittenberg (2002) An Easy and Accurate Regression Model for Multiparty Elections. *Political Analysis* 10(1), 66-83. (http://intl-pan.oupjournals.org)