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 attempt 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 will 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.

**Mathematics**

Applying some basic concepts and techniques drawn from calculus and linear algebra enables us to develop a deeper understanding of multivariate estimation and inference. I assume that you have a familiarity with basic differential calculus but not necessarily with matrix algebra. We will spend several classes covering the latter after we have completed our introductory tour of bivariate regression. Later in the course I will introduce some integral calculus as needed.
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 11; 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 8.

Texts

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


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.

Readings and exercises are available on learn@uw.
Outline of Topics

I. Introduction (class 1)

   Overview

II. Calculus Review (classes 2 and 3)

   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


   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
VI. Pathologies and Treatments

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

VII. Models with Discrete Dependent Variables

Contingency table analysis
Linear probability models, logit, and probit
Ordered probit, multinomial and conditional logit

VIII. Simultaneous Equation Models

Identification
Estimation: instrumental variables; two-stage least squares


**IX. Additional Topics as Time Permits**

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

Gujarati, 16, 17


W. Robert Reed and Rachel Webb (2010) The PCSE Estimator is Good—Just Not as Good as You Think. *Journal of Time Series Econometrics* 2(1), Article 8, 1–24. (Available at learn@uw)
