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 from data that "happens" to be available. It is usually necessary, therefore, to use multivariate techniques to control statistically those factors that cannot be controlled by experiment. 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.

**Course Requirements**

*Examinations:* Midterm (20 percent) on **March 2**; final (50 percent) as scheduled.

*Assignments:* Four data analysis exercises (10 percent) must be completed on time. A number of additional problem sets and Monte Carlo exercises, which are important for developing theoretical understanding, should also be completed (10 percent).

*Project:* Attempt to answer an empirical question by applying techniques learned in course to data that you have assembled (10 percent). Due **April 25**.

*Texts*

You may wish to purchase one or more of the following texts as support for lectures:


The text by Greene provides a comprehensive survey of the theory underlying the commonly used basic techniques. It would be the best investment for those planning to take PSC 505. Gujarati’s text is less comprehensive and less technical, though it does provide quite clear discussions of topics. As both texts are on reserve, and my lectures are generally self-contained, it would be possible to avoid buying either. I recommend, however, that you purchase one or the other.

**Outline of Topics**

I. Introduction

Overview; history; multiple regression and hypothesis testing

II. Bivariate Regression

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

III. Multivariate Regression

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

Gujarati, 7 to 9; Greene, 2, 6, and 7

IV. Case Study: Determining Causality


IV. Model Specification

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

Gujarati, 13 to 15; Greene, 8

V. Pathologies and Treatments

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

Gujarati, 10 to 12, 21 Greene, 11, 12, 13, 17, and 18


VI. Models with Discrete Dependent Variables

Contingency table analysis
Linear probability models, logit, and probit

Gujarati, 16; Greene, 19


VII. Simultaneous Equation Models

Identification
Estimation: instrumental variables; two-stage least squares

Gujarati, 18 to 20; Greene, 15 and 16

VIII. Additional Topics as Time Permits

Pooled Time Series/Cross Sections
Censored data
Selection models

Green, 14 and 20


*** *** ***

Assignments

**Memorandum Exercises** (Tentative List):

- Hypothesis testing (type I error)
- Refuse collection (residuals)
- Discrimination (dummy variables)
- Pre-sentence reports (logit)

**Monte Carlo Mini-Projects** (Teams):

- Correlation between independent variable and disturbance
- Errors-in-variable problem and instrumental variables

**Problem Sets** (Tentative List):

- Review of expectations
- Summations
- MLE
- Non-linear regression (team project)
- Review of matrix operations
- Outliers
- Heteroscedasticity

**Project**

Frame hypothesis; collect relevant data; test hypothesis; diagnose pathologies.