Course Overview

In this course, we are going to build on the foundations in statistical theory covered in the first semester to explore in detail the statistics associated with estimating relationships between variables. The primary focus of this course is on the use of regression analysis, a statistical technique for quantifying and making inferences about relationships between variables. Regression analysis is increasingly used to evaluate and provide quantitative assessments of existing public policies and programs. For example, these empirical techniques have been employed to assess the effects of head start on the long term socioeconomic outcomes of children, to test whether lenders racial discriminate against African-Americans seeking mortgages, and to assess the effect of the prison incarceration rate on crime. As policy analysts, you will need to understand these research methods, how to apply them, and how to assess the validity of social scientific research based on these empirical techniques.

While the principal objective of the course is to introduce the statistical models used to measure association between random variables (specifically, linear and non-linear regression analysis), an important sub-theme that will be salient in most of our discussions this semester concerns the issue of actually determining causality. In a nutshell, regression analysis uses data samples to estimate the relationship between dependent and explanatory variables. An important point that should always be kept in mind is that for most empirical studies in the social sciences correlation does not imply causation no matter how fine the correlation is measured. When analyzing observational or non-experimental data, there are a number of factors that may lead one to conclude that a given variable causally affects another variable, when in fact, there is no effect. For example, the explanatory variable of interest may not have an effect but may be highly correlated with variables that do. Alternatively, the direction of causation may run in the opposite direction of that which is being hypothesized. Throughout the course, we will be highlighting such problems and presenting statistical techniques designed to address these issues.

This course has several goals. The first is to familiarize you with the vocabulary, methods, and results of econometric theory. We will be mainly concerned with using, rather than proving, the theory. Over the course of the semester, we will study how to address the most important issues that arise when doing an econometric study. These issues include (1) endogeneity of regressors due to omitted variables, simultaneity, or selection, (2) functional form, (3) choice of control variables, (4) measurement error in regressors and handling of outliers, (5) efficient estimation in the presence of heteroskedasticity, (6) appropriate estimation of standard errors and test statistics, and (7) presentation of results.

Second, you should strive to become an educated/critical consumer of empirical research. We will read and critique a number of empirical papers and program evaluations. The usual approach will be to discuss 1) the economic, political, and policy issues that motivate the application, 2) the econometric techniques and issues related to the application (e.g., data, specification, estimation techniques), and 3) the results of the empirical analysis and what we can learn from them.

Your work in this course will be multi-faceted. You will be assigned 5 (depending on how quickly we move through the material) “hands on” applied problem sets. These frequently will involve replicating some results from published work. Replicating what others have done is a good way to understand the choices involved in applied work. We will use STATA, the leading computer program for statistical analysis in the social sciences, and by the end of the course you should have acquired a variety of skills that will be useful for doing public policy research. To gain skill as an “educated/critical consumer” you will be required to write a summary and critique of an applied policy research paper. Detailed guidelines on this writing assignment will be given to you early in the semester. There will be three exams.
**Meeting Times**

Class will meet Mondays and Wednesdays from 8:30 to 10:00 am. In addition, there are discussion sections on Fridays where material presented during the week will be reviewed.

**Grades**

Grades will be based on 3 exams (25 percent each), 5 problems sets (worth a cumulative 15 percent) and a writing assignment (10 percent). Problem sets will be handed out in class and are due a week later (we will then distribute answer keys). Late problem sets are docked 20 percentage points per day.

The writing assignment (on an assigned paper/topic yet to be determined) will involve writing a short (3 to 5 pages) summary and critique of an applied policy research paper. You are to identify the general premise of the article, the hypothesis being tested, the data used in the study, the econometric methods employed, and the sources of variation/identification that the researcher employs to estimate parameters. Close attention should be paid to how the parameter estimates are/can be interpreted. Lastly, you should briefly discuss how believable you find the estimates.

**Readings**

The required textbook for this class is *Introductory Econometrics*, Jeffrey M. Wooldridge (Southwestern, 2nd edition, 2002). More than others, this text has lots of examples and is written for students who are or will be doing applied work. The textbook is available at the campus bookstore. There is also a reader for this course. The reader will be available at Vic’s Copies, 2843 Hearst Avenue, Berkeley. The reader is a collection of scholarly articles that we will be reading throughout the semester.

**STATA**

STATA version 8.0 is available in the computer lab. However, you can buy a one-year license to use STATA at home if you wish, through the STATA website (at [http://www.stata.com/order/new/edu/gradplans/gp2-order.html](http://www.stata.com/order/new/edu/gradplans/gp2-order.html)). The student rate is $39 for “small” STATA (a scaled-back version) and $89 for the standard Intercooled STATA. We will distribute a STATA tutorial guide. (If you prefer SPSS or SAS that is fine too).

**Office Hours**

Johnson’s office is on the third floor of the new building, room #347. His office hours for this course are Tuesdays, 3-5. Johnson can also be reached by phone at 643-0169 and Email at: ruckerj@berkeley.edu
Week 1: Introduction

January 19

**TOPIC:** Course Intro • Review of Syllabus • Overview of Quantitative Program Evaluation
• Inferring Causality: experimental vs. non-experimental data analysis

**READING:**
• Wooldridge Ch 1

Week 2: Review of Statistical Inference & Introduction to Linear Regression Analysis

January 24

**TOPIC:** Review of Probability & Distribution (expectation and moments) • Review of Statistical Inference (Point and Interval Estimation—confidence intervals; Hypothesis Testing—testing for differences in means; sampling distributions and inference; approximate asymptotic distribution of the sample mean)

**READING:**
• W Appendices A, B
• W Appendix C
• (mathematics, probability, and statistics review for refreshing memories)

* **ASSIGNMENT:** Hand out first problem set (due Feb. 2nd)

January 26

**TOPIC:** Introduction to Regression Analysis—Why & How? (Bivariate Regression)

Week 3: Introduction to Linear Regression Analysis & the Two-Variable Model

January 31

**TOPIC:** Introduction to Regression Analysis—Why & How? (Bivariate Regression—Structure of data; Mechanics of OLS, Assumptions required for Unbiasedness, Relationship to Correlation)

**READING:**
• W Ch 2

February 2

**TOPIC:** The Two Variable Regression Model (cont’d)—Sampling Distribution of Regression Estimates; Residuals, Fitted Values & Goodness of Fit; Hypothesis Tests; Confidence Intervals

**READING:**
• W Ch 2

* **ASSIGNMENT DUE:** First Problem set due
## Week 4: Multiple Regression Analysis: Estimation

**February 7**

**TOPIC:** Motivation for Multiple Regression Analysis • Interpretation of Coefficients • Omitted Variable Bias  
**READING:**  
- W Ch 3  
* **ASSIGNMENT:** Hand out second problem set (due Feb. 14th)

**February 9**

**TOPIC:** Multiple Regression Analysis—Inference & Hypothesis Tests: t-Test, p-value, confidence intervals  
**READING:**  
- W Ch 3, Sections 4.1-4.3  

## Week 5: Multiple Regression Analysis: Inference & Hypothesis Testing

**February 14**

**TOPIC:** Multiple Regression Analysis—Inference & Hypothesis Testing (cont’d) • Multicollinearity  
**READING:**  
- W Ch 3, Sections 4.1-4.3  
- W Ch 3 (p.97-100)  
* **ASSIGNMENT DUE:** Second problem set due

**February 16**

**TOPIC:** Model Specification—Dummy Variables, Interaction terms, & Quadratics • Tests of Joint Hypotheses—Goodness of Fit Measures, Tests involving more than 1 regression coefficient  
**READING:**  
- W Ch 7.1-7.4, 7.6  
- W Sections 4.4-4.6

## Week 6: First Midterm

**February 21**

NO CLASS

**February 23**

*Review for Midterm*

**February 25**

**MIDTERM DURING SECTION**
Week 7: Application of Multiple Regression Models: Decomposing Group Differences, Specification Choices

February 28

**Application#1:** The Black/White Wealth Gap

**Application#2:** Racial Disparities in Federal Sentencing

March 2

**TOPIC:** Model Specification—Functional Form and Variable Selection

**READING:**
- W Ch 6.1-6.3 (Also p.685-691)

**ASSIGNMENT:** Hand out third problem set (due March 9th)

Week 8: “Real World” Issues—Violation of the Basic Assumptions

March 7

**TOPIC:** Heteroscedasticity

**READING:**
- W, Ch 8.1-8.4

March 9

**TOPIC:** Measurement Error, Missing Data, and Sensitivity of Results to Particular Observations

**READING:**
- W Sections 9.3,9.4

**ASSIGNMENT DUE:** Third Problem Set Due

Week 9: Study Designs to Distinguish Correlation from Causality—Accounting for Omitted Variable Bias

March 14

**TOPIC:** Using Proxy Variables for Unobserved Explanatory Variables to address Omitted Variable Bias

**READING:**
- W Section 9.2

March 16

**TOPIC:** Adding Controls to Account for Omitted Variable Bias

**Application:** Explaining Black-White Wage Differences


**ASSIGNMENT:** Hand out fourth problem set (due March 28th)

March 21

**NO CLASS:** Spring Recess

March 23

**NO CLASS:** Spring Recess
Week 10: Panel Data Techniques

March 28

**TOPIC:** Introduction to Panel Data Techniques (Sources of variation/identification, Difference Estimators)

**READING:**
- W Ch 13.3-13.5

March 30

**REVIEW FOR MIDTERM**

**ASSIGNMENT DUE:** Fourth problem set due.

April 1

**MIDTERM EXAM DURING SECTION**

Week 11: Panel Data Applications and Natural Experiments

April 4

**TEXT READING:** W Ch 13.3-13.5, 14.3

**Application#1:** The Effects of Teenage Pregnancy on Socioeconomic Outcomes


**Application#2:** The Effects of Head Start on Socioeconomic Outcomes


April 6

**TOPIC:** Natural/quasi Experimental Designs: pre-/post-intervention observation with a comparison group (Difference-in-Difference Estimators)

**TEXT READING:** W Ch 13.1-13.2

**Application:** The Effect of the Minimum Wage

**READING:**

Week 12: Experimental Methods continued

April 11

**TOPIC:** Empirical Methods for Program Evaluation

**Application#1:** The Effects of 1996 welfare reform


**Application#2:** Evaluating the Effect of Tuition and Financial Aid Policies


* **ASSIGNMENT:** Hand out writing assignment (due April 18th)

April 13

**TOPIC:** Advantage & Disadvantages of Experiments

**Application#1:** Effect of Price of Health Care on Health Care Utilization

**READING:** Manning, Willard G. et al. (1987). Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. American Economic Review,
Application#2: **Neighborhood Effects**

**READING:**

**Exec Summary:** [LINK]

---

**Week 13: Randomized Experiments and Instrumental Variables Approaches**

**April 18**

**Application:** **Labor Market Discrimination**

**READING:**

* **ASSIGNMENT DUE:** Writing assignment due
April 20  
**TOPIC:** Instrumental Variables Approach/ 2 Stage Least Squares for Omitted-Variables Problem  
**TEXT READING:** W Ch 15

**Application#1:** Effect of Class Size on Achievement  

**Application#2:** Effect of School Vouchers on Achievement  
**READING:**  

---

**Week 14: Instrumental Variables Continued and Introduction to Binary Dependent Variables**

April 25  
**TOPIC:** Endogeneity, Instrumental Variables Approach/ 2 Stage Least Squares for Omitted-Variables Problem  
**READING:**  
- W Ch 15  

**Application:** Labor-Supply Consequences of Childbearing  

April 27  
**TOPIC:** Binary Dependent Variables—Logit and Probit Model  
**READING:** W Sections 7.5, 8.5, 17.1  
* **ASSIGNMENT:** Hand out fifth problem set (due May 4th)

---

**Week 15: Models with Categorical Dependent Variables: Binary Dependent Variable**

May 2  
**TOPIC:** Binary Dependent Variables—Logit and Probit Model  
**READING:** W Sections 7.5, 8.5, 17.1  
**Application:** Racial discrimination in the markets for mortgage capital and insurance  

May 4  
**REVIEW FOR THE FINAL**  
* **ASSIGNMENT DUE:** Fifth problem set due