POL 451: Statistical Methods in Political Science

Fall 2007

Kosuke Imai

Department of Politics, Princeton University

1 Contact Information

Office: Corwin Hall 041
Phone: 609–258–6601
Fax: 973–556–1929
Email: kimai@Princeton.Edu
URL: http://imai.princeton.edu
Office Hours: Mondays 3:00pm – 4:00pm or by appointment

2 Catalog Description

Mondays and Wednesdays, 1:30pm – 2:50pm.
Friend Center 202.

In this course, students will learn basic research design and data analysis methodology in empirical social science research. The main goal is to learn how statistical theory can be used to make causal inferences in experimental and observational studies. The course satisfies the analytical methods requirement for politics majors. The materials of this course are particularly useful for those who plan to use quantitative analysis in their junior papers and senior thesis as well as for those who wish to apply for graduate programs in the social sciences. Familiarity with elementary probability theory is helpful, but is not required.

3 What This Course Is Really About

This course introduces the basics of statistics and probability to students with little or no mathematical background. It is intended as a first course on probability and statistical methods in the social sciences. One main focus of the course is the key question of how to use statistics to make causal inferences, which are the main goals of most social science research. Descriptive inferences and survey sample surveys are also covered. Students will learn basic data analysis techniques as well as elementary statistical and probability concepts.

4 Course Requirements

The final course grade is based on in-class participation (20%), problem sets (30%), and the choice of a project or exams (50%).

- **In-class participation:** This is a seminar course, and regular attendance is required. Students are expected to read the assigned materials before classes and actively participate in discussion.

- **Problem sets:** There will be several problem sets during the semester. Students are allowed to discuss the questions with each other, but they must write up their answers to the questions on their own. Late submission will be penalized (for each day, 30% reduction will be applied).

- **Research project:** If you choose this option, your research project should contain an original analysis of a data set addressing substantive questions of interest. Some potential projects are:

  - analysis of a data set that you collected, an experiment that you conducted, or a survey that you administered.
  - the empirical analysis portion of your senior thesis or junior paper.
  - your reanalysis of a data set that has been analyzed in a published scholarly paper.

  By the end of the semester, you must complete a research report that describes a substantive hypothesis of interest and explains the methods and results of the empirical analysis. This report is due at 5pm on January 21, in my departmental mailbox (Corwin Hall). Late submission will not be accepted.

  To support continued progress for the project, students are required to:

  1. meet with me to discuss their project during September.
  2. give a 15 minute in-class presentation about research design right after the fall break.
  3. give 15 minute in-class presentation about preliminary results in December.

  The presentations will be 20% of the course grade, whereas the final project report will be 30%. In addition, I will be available for individual meetings anytime during the semester to discuss projects and comment on drafts.

- **Exams**: If you choose this option, you are required to take a mid-term exam (October 24) and a final exam (the date specified by the registrar). Both exams will be in-class and closed-book. The mid-term exam will be 20% of the course grade, and the final exam will be 30%.

5 Textbooks

The main textbook we use in this class is:

- Freedman, David, Robert Pisani, and Roger Purves. (2007). *Statistics* 4th eds. Norton.

This book is available for purchase at the U-store and is also reserved at the library. The course also relies on journal articles and book chapters.

In addition to the required textbook, you may find some materials from the following textbooks useful.

- Achen, Christopher. (1982). *Interpreting and Using Regression* Sage.

- Fox, John. (2002) *An R and S-PLUS Companion to Applied Regression* Sage.

- DeGroot, Morris H. and Mark J. Schervish (2002) *Probability and Statistics*, 3rd Edition, Addison Wesley.

The first is an introductory textbook to regression, and the second one is a reference book for the statistical software **R**, which will be the (free!) software for data analysis in this course; this book is available for purchase at U-store. The third book is a more advanced (i.e., mathematical) textbook on probability and statistics.

6 Tentative Course Plan

6.1 Introduction

1. (September 17). OVERVIEW

    (a) What is statistical inference?

    (b) When are statistics useful?

    - Salsburg, David. (2002). *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century* Owl Books.
    - Taleb, Nassim Nicholas. (2007). *The Black Swan: The Impact of the Highly Improbable* Random House.

6.2 Causal Inference and Research Design

1. (September 19). RANDOMIZED EXPERIMENTS VS. OBSERVATIONAL STUDIES

    (a) What is causal inference?

    (b) Statistics and causal inference

    (c) Advantages and disadvantages of randomized experiments and observational studies

    - Freedman *et al.* Chapters 1 and 2.
    - Holland, P. W. (1986) "Statistics and causal inference" (with discussions) *Journal of the American Statistical Association* **81**, 945–970.
    - Taubes, Gary. (2007). "Unhealthy Science" *The New York Times Magazine* September 16, Section 6.

2. (September 24 and 26). EXPERIMENTS IN THE REAL WORLD

(a) Pros and cons of randomized experiments

(b) Applications of field, social, and natural experiments

- Burtless, Gary. (1995). "The Case for Randomized Field Trials in Economic and Policy Research." *Journal of Economic Perspectives* **9**, 63–84.

- Heckman, James J. and Smith, Jeffrey A. (1995). "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* **9**, 85–110.

- Bertrand, Marianne and Mullainathan, Sendhil. (2004). "Are Emily and Greg More Employable than Lakisha and Jamal?: A Field Experiment on Labor Market Discrimination." *American Economic Review* **94**, 991–1013.

- Wantchekon, L. (2003). "Clientelism and voting behavior: Evidence from a field experiment in Benin." *World Politics* **55**, 399–422.

- Chattopadhyay, Raghabendra and Esther Duflo. (2004). "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, **72**, 1409–1443.

3. (October 1). STRATEGIES FOR DESIGNING OBSERVATIONAL STUDIES

(a) How can we use observational data to make credible causal inferences?

(b) What are the advantages and limitations of observational studies?

- Card, D. and Krueger, A. B. (1994) "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* **84**, 772–793.

- Angrist, Joshua and Victory Lavy. (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Student Achievement" *Quarterly Journal of Economics*, **114**, 533–575.

- Lee, David, Enrico Moretti, and Matthew J. Butler. (2004). "Do Voters Affect or Elect Policies? Evidence from the U.S. House." *Quarterly Journal of Economics*, **119**, 807–859.

6.3 Descriptive Statistics

1. (October 3 and 8). SUMMARIZING THE DATA

(a) How do we numerically summarize the data?

(b) How do we graphically summarize the data?

(c) Introduction to **R**

- Freedman *et al.*, Chapters 3, 4, 5 and 7.
- Fox, Chapters 1.1, 2 (Skip 2.4), and 3 (Skip 3.4).

2. (October 10). DATA MEASUREMENT PROBLEMS

(a) What happens when the data are measured with error?

- Freedman *et al.*, Chapter 6.
- McDonald, M. P. and Popkin, S. L. (2001) "The Myth of the Vanishing Voter." *American Political Science Review* **95**, 963–974.
- Uggen, C. and J. Manza (2002) "Democratic Contraction? Political Consequences of Felon Disenfranchisement in the United States." *American Sociological Review* **67**, 777–803.

6.4 Correlation and Regression

1. (October 15). CORRELATION

   (a) Correlation is not causation.
   (b) What is ecological correlation?

   - Freedman *et al.* Chapters 8 and 9.
   - Robinson, W. S. (1950). "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* **15**, 351–357.

2. (October 17, and 22). REGRESSION

   (a) What is the regression line?
   (b) How do we graphically display regression lines?
   (c) How do we calculate error for regression?
   (d) Things may not be linear.

   - Freedman *et al.* Chapters 10, 11 and, 12.
   - Fox Chapters 1.2 and 4.1
   - Freedman, D. A., S. P. Klein, J. Sacks, C. A. Smyth, and C. G. Everett (1991) "Ecological Regression and Voting Rights" *Evaluation Review*, **15**, 673–711.

6.5 Midterm Exam and Student Presentations

1. (October 24). MIDTERM EXAM

   (a) In-class midterm exam for those who choose the exam option.
   (b) Individual meetings for those who choose the project option.

2. (November 5). STUDENT PRESENTATIONS

   (a) Short presentations about research design for those who choose the project option.
   (b) Those who choose the exam option should also participate in the discussion.

6.6 Probability

1. (November 7). INTRODUCTION TO PROBABILITY

   (a) What is probability?
   (b) Conditional probability and independence
   (c) Who is Bayes?

   - Freedman *et al.* Chapters 13 and 14.
   - King, G. and Lowe, W. (2003) "An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design." *International Organization* **57**, 617–642.

2. (November 12). PROBABILITY DISTRIBUTIONS AND RANDOM VARIABLES

   (a) Binomial and normal distributions
   (b) How can we draw from a probability distribution?

   - Freedman *et al.* Chapters 5 and 15.

3. (November 14). LAW OF LARGE NUMBERS AND CENTRAL LIMIT THEOREM

   (a) Expected values.
   (b) What happens to the sum of random variables when the sample size increases?
   (c) Why is the normal distribution so special?
   (d) What is standard error?

   - Freedman *et al.* Chapters 16, 17, and 18.

6.7 Sampling

1. (November 19 and 21). *Basic Mechanisms of Sample Surveys*

   (a) Why sample rather than enumerate?
   (b) How accurate is survey sampling?

   - Freedman *et al.* Chapters 19, 20, 21, and 23.

2. (November 26). *Survey and Polls in the Social Sciences*

   - Freedman *et al.* Chapter 22.
   - Gelman, A. and G. King. (1993) "Why Are American Presidential Election Campaign Polls So Variable When Votes are So Predictable?" *British Journal of Political Science*, **23**, 409–451.

6.8 Statistical Inference

1. (November 28, December 3 and 5). STATISTICAL ANALYSIS OF RANDOMIZED EXPERIMENTS

   (a) What is the statistical test?

   (b) Null hypothesis, test statistic, and p-value.

   (c) Statistical significance is not scientific significance.

   (d) What is the relationship between statistical tests and confidence intervals?

   - Freedman *et al.* Chapters 26, 27, 28 and 29.
   - Fisher, R. A. (1935) *The Design of Experiments.* Oliver and Boyd, Ch. II.

2. (December 10 and 12). STATISTICAL INFERENCE WITH REGRESSION

   (a) Statistical tests in the regression analysis

   (b) How do we construct confidence intervals around the regression line?

   - Achen.

6.9 Final Exam and Student Presentations

1. (December 17). STUDENT PRESENTATIONS

   (a) Short presentations about preliminary results for those who choose the project option.

   (b) Those who choose the exam option should also participate in the discussion.

2. (January TBD). FINAL EXAM

   (a) In-class final exam for those who choose the exam option.