POL 572: Quantitative Analysis II

Spring 2009

Kosuke Imai

1 Contact Information

Office: Corwin Hall 041
Office Phone: 258–6601
Email: kimai@Princeton.Edu
URL: http://imai.princeton.edu

2 Logistics

- Lectures: Wednesdays 9:00–10:20am in McCosh 62, and Thursdays 10:30am–11:50am in Frist 307
- Precepts (taught by Teppei Yamamoto tyamamot@Princeton.EDU): Monday 6pm – 7:20pm in Friend 004
- Kosuke's Office hours: Fridays 10am – noon or stop by anytime
- Teppei's Office hours: Mondays & Tuesdays 2pm – 4pm in 235A Corwin

3 Questions about the Course Materials

In addition to precepts and office hours, please use the *discussion board* at Blackboard when asking questions about lectures, problem sets, and other course materials. This allows all students to benefit from the discussion and help each other understand the materials. Teppei will be primarily responsible for handling questions about precepts and problem sets, while I will answer the questions about the lectures. But, students are also encouraged to participate in discussions and answer any quesions that are posted.

4 Course Description

Catalogue Description. This is the second course in the quantitative methods sequence. It will emphasize the flexibility of the maximum likelihood framework in the context of regression models, models that mix qualitative and continuous endogenous variables, hazard models, and scaling models.

What this course is about. This course is the first course in applied statistical methods for social scientists. Students will learn a variety of basic *cross-section* regression models (as time permits!) including linear regression model, discrete choice models, duration (or hazard) models, event count models, structural equation models, and others. Unlike traditional courses on applied regression modeling, I will emphasize the connections between these methods and causal inference, which is the primary goal of social science research.

5 Prerequisites

There are three prerequisites for this course.

1. Mathematics covered in POL 502
2. Probability and statistics covered in POL 571
3. Statistical computing and programming covered in the statistical software workshop held at the end of January

6 Course Requirements

The final grades are based on the following two items:

1. Problem sets (40%): Although you are allowed to discuss the problem sets with others, you should not copy someone else's computer code or answers. In particular, sharing a paper or electronic copy of your code and answers with other students is strictly prohibited.
2. Midterm exam (20%): The midterm exam will be held around the spring break.
3. Final exam (40%): It is also possible to conduct a final project in stead of taking the final exam if I view your proposed project appropriate for this course. If you prefer this option, you must obtain my permission by the end of February.

7 Computing

In this course, we use a statistical computing environment, called R. R is available for any platform and without charge at http://www.r-project.org/. In a recent *New York Times* article, R is described as software that "allows statisticians to do very intricate and complicated analyses without knowing the blood and guts of computing systems."

8 Books

1. Much of the following book will be assigned as required reading prior to the lectures:

   David A. Freedman. *Statistical Models: Theory and Practice.* Cambridge University Press, Cambridge, 2005.

2. In addition, you may find the following monographs useful:

   (a) Political Methodology

Gary King. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. University of Michigan Press, Ann Arbor, 1998.

(b) Probability and Statistics

Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York, 2005.

(c) Econometrics

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA, 2002.

(d) R

John Fox. *An R and S-plus Companion to Applied Regression*. Sage Publications, Thousand Oaks, CA, 2002.

9 Course Outline

Each topic is followed by the list of required readings. I assume that students read these materials *prior to* each lecture.

1. Review, Introduction, and Overview

   (a) Causal inference, experimentation, and statistics

   Chapter 1 of Freedman

   (b) The potential outcomes framework of causal inference

   Paul W. Holland. Statistics and causal inference (with discussion). *Journal of the American Statistical Association*, 81:945–960, 1986.

   Junni L. Zhang and Donald B. Rubin. Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*, 28(4):353–368, 2003.

   **For Zhan and Rubin (2003), skip Sections 4 – 7.*

   (c) Randomized inference for classical randomized experiments

   Chapter 2 of Paul Rosenbaum *Observational Studies*, 2nd ed. 2002.

   Daniel E. Ho and Kosuke Imai. Randomization inference with natural experiments: An analysis of ballot effects in the 2003 California recall election. *Journal of the American Statistical Association*, 101(475):888–900, September 2006.

   **For Rosenbaum (2002), skip Sections 2.4.4, 2.5.4., 2.8, 2.9, and 2.10.*

   (d) Classical randomized experiments and simple regression

   Freedman, Chapter 2

David A. Freedman. On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40(2):180–193, February 2008.

**For Freedman (2008), skip Sections 5 – 8.*

(e) Observational studies, parametric adjustment, and matching

Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3):199–236, Summer 2007.

2. Multiple regression

(a) Least squares

Chapter 4 of Freedman

**If you are not comfortable with matrix algebra, you may refer to Chapter 3 of Freedman.*

(b) Maximum likelihood

Section 6.1 of Freedman

3. Structural equation modeling, and instrumental variables

(a) Selection problem, instrumental variables, and causality

Chapter 8 of Freedman

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, 91(434):444–455, 1996.

(b) Direct and indirect effects

Chapter 5 of Freedman

Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51–71, February 2010.

4. Parametric regression modeling with various data types

(a) Likelihood theory

Chapter 4 of Gary King *Unifying Political Methodology* University of Michigan Press, 1998.

(b) Parametric and nonparametric Bootstrap

Chapter 7 of Freedman

Gary King, Michael Tomz, and Jason Wittenberg. Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, 44:341–355, 2000.

(c) Binary data

Chapter 6 of Freedman

(d) Other data types (as time permits)

i. Ordered and multinomial data
ii. Truncated and censored data
iii. Survival data
iv. Event count data
v. Bivariate and multivariate data
vi. Data with nonrandom sample selection