

POL 572: Quantitative Analysis II

Spring 2016

Kosuke Imai

Ted Enamorado

Department of Politics
Princeton University

This course is the first course in applied statistical methods for social scientists. We begin by studying the fundamental principles of statistical inference. Students will then learn a variety of basic *cross-section* regression models including linear regression model, structural equation and instrumental variables models, discrete choice models, and models for missing data and sample selection. Unlike traditional courses on applied regression modeling, we will emphasize the connections between these methods and causal inference, which is a primary goal of social science research.

1 Contact Information

	Kosuke Imai (Instructor)	Ted Enamorado (Preceptor)
Office:	Corwin Hall 036	Corwin Hall 021
Phone:	258-6601	615-752-8093
Email:	kimai@Princeton.Edu	tede@princeton.edu
URL:	http://imai.princeton.edu	

2 Logistics

- Lectures: Tuesdays and Thursdays, 10:30am – 11:50am, 127 Corwin Hall
- Precepts: Fridays, 3:00 PM – 4:30 PM, TBA
- Kosuke’s office hours: stop by anytime or make an appointment
- Ted’s office hours: Mondays and Wednesdays, 4:30pm – 6:00pm, Robertson 006
- Google calendar at: <https://goo.gl/BnjTKk> (iCAL) and <https://goo.gl/GbVe3T> (HTML)

3 Questions and Announcements

In addition to precepts and office hours, please use the *Piazza Discussion Board* at <https://piazza.com/> when asking questions about lectures, problem sets, and other course materials. This allows all students to benefit from the discussion and to help each other understand the materials. Both students and instructors are encouraged to participate in discussions and answer any questions that are posted.

To join the POL 572 Piazza site, click on “Search Your Classes” from the Piazza homepage. After specifying Princeton University as your school, search for “POL 572: Quantitative Analysis

II.” You will then be prompted to enter your `princeton.edu` email address to confirm your registration. Piazza can also be accessed from within Blackboard by going to the POL 572 course page and clicking on the link to “Piazza Messageboard.” In addition, all class announcements will be made through Piazza. Blackboard will still be used for hosting all class materials.

4 Prerequisites

There are two prerequisites for this course:

- Probability and statistics covered in POL 571: at the level of Blitzstein and Hwang (2015) and DeGroot and Schervish (2002).
- Statistical programming covered in the statistical programming camp held at the end of January. The camp materials are posted at the Program for Quantitative and Analytical Political Science (Q-APS) website <http://q-aps.princeton.edu>.

In addition, the following is recommended but not required:

- Mathematics covered in POL 502 : Basic real analysis, calculus, and linear algebra.

5 Course Requirements

The final grades are based on the following items:

- **Participation** (10%): The level of engagement in lectures, precepts, and Piazza discussions.
- **Problem sets** (30%): Approximately bi-weekly problem sets will be given throughout the semester. Each problem set will equally contribute to the final grade and contain both analytical and data analysis questions. For this class, problem sets provide opportunities for individual learning and group collaboration. The following instructions will apply to all problem sets:
 - *Group Collaboration.* Each student will be assigned to a group. Groups are required to work together and produce a single set of solutions to each problem set. To facilitate individual learning, all group members should contribute to all problem set questions rather than dividing them up. You must work through the whole problem set together with your group, rather than working separately on different portions of the problem sets. There is to be no collaboration between groups, aside from public posts on Piazza. Groups will be reassigned for each problem set.
 - *Submission policy.* Students hand in problem sets, as a group, with the name of each group member on the assignment. All answers should be typed via **Rmarkdown** and the source code should also be submitted. Please ensure your code adheres to the Google’s **R** Style Guide rules (see below for URL), as style errors will be penalized. Neither late submission nor electronic submission will be accepted unless you obtain a prior approval from the instructor.
- **Quizzes** (30%): Two closed-book quizzes will be given on March 25 and May 6 (both from 3:00pm to 4:30pm, location TBA), covering the first and second half of the course materials, respectively. Each quiz is equally weighted.
- **Take-home final exam** (30%): The take-home open-book final exam will be given on May 9 and be due on May 16. The exam consists of data analysis questions alone.

6 How to Get the Most out of this Course

To get the most out of this course, students must keep up with the new materials that will be introduced each week. Do not fall behind by leaving your questions unanswered! Because the materials in the later part of the course (and POL573) build upon those covered earlier in the semester, falling behind will mean that you will be lost for the rest of the quantitative methods sequence. ☺ Here are some recommendations we give to students:

- Go over each lecture slide carefully and try to fill in and understand every detail. Forming a small study group for this purpose is a good idea. Help others understand the materials through study group meetings and Piazza.
- Attend precepts and use office hours and Piazza discussion board to clarify any questions you may have about the course materials.
- Use the assigned readings to supplement the lectures.
- Start working on the problem sets as soon as you receive them. Try to solve questions on your own first before meeting with others to discuss them.
- Carefully go over the graded problem sets and exams as well as their solutions so that you do not repeat your mistakes.

Finally, POL572 is a “statistics bootcamp” where you are introduced to the fundamentals of applied statistics and data analysis. In POL573, you will begin to learn how to conduct original research using statistics and data analysis. Thus, at the minimum, you need to take POL572 in order to get the most out of POL573. This means that you should carefully consider whether POL572 (and POL573) offers the kind of methodological training you want to receive.

7 Statistical Computing

A major emphasis of this course is to have students learn how to better present and communicate the results of their statistical analysis in a manner that can be easily understood by the general audience who has little statistical training. To achieve this goal, we use a statistical computing environment, called **R**. **R** is available for any platform and without charge at <http://www.r-project.org/>. In the past, we noticed that some students ended up spending an unnecessarily large amount of time debugging their **R** code for problem sets. To avoid this unpleasant experience, here are some tips you should keep in mind:

- Ask for help during office hours and at Piazza! The whole point of taking statistics courses (as opposed to learning statistics on your own) is to help you understand materials efficiently. You should take full advantage of available resources.
- First lay out the structure of code and then figure out how to implement it in **R**. If you cannot figure out what **R** command is necessary for implementing your ideas, just ask! There are many online **R** forums and so Google should be able to help you too.
- Build your code step by step. It is often difficult to locate a bug in a large number of code lines. It is more efficient if you check the accuracy of each small code chunk as you write it. For example, when writing a loop, it would be a good idea to first write a code that works for one iteration and make sure that it works appropriately. Then, you can put it inside of a loop.

- Good coding style makes it easy for you to understand your own code. Follow the Google’s R Style Guide available at <https://google.github.io/styleguide/Rguide.xml>

8 Course Outline

Each topic is followed by the list of suggested readings. As you can see, the list of topics is quite ambitious. My current plan is to spend two to three weeks on each topic but the plan is subject to change, depending on how students are keeping up with the course materials.

Basic Principles of Statistical Inference

1. Descriptive, Predictive, and Causal inference

Lecture notes “Statistical Framework of Causal Inference”

Freedman (2009) Chapter 1

Imbens and Rubin (2015) Part I

2. Identification, Estimation, Confidence Interval, and Hypothesis Testing

Lecture notes: “Classical Approaches to Statistical Analysis of Randomized Experiments”

Imbens and Rubin (2015) Chapters 4–6 and 21–22

Manski (2007) Introduction and Chapter 7

DeGroot and Schervish (2002) Sections 7.1–7.5, 8.1, 8.5–8.7, 9.3–9.5

3. Problem Sets: Sample surveys, Randomized experiments

Linear Regression

1. Simple Regression

Freedman (2009) Chapter 2

Imbens and Rubin (2015), Chapter 7

DeGroot and Schervish (2002) Sections 10.1–10.3

Angrist and Pischke (2009), Section 6.1

Guido W. Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, February 2008.

2. Multiple Regression

Freedman (2009) Chapter 3

(Easier) Freedman (2009), Chapters 4 and 5; or (Harder) Hayashi (2000), Chapters 1 and 2

Imbens and Rubin (2015), Chapter 8

3. Matching and Regression

Lecture notes: “Selection Bias in Observational Studies”

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.

Elizabeth A. Stuart. Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1):1–21, 2010.

Imbens and Rubin (2015), Part III

4. Fixed Effects, First Differences, and Difference-in-Differences

Lecture notes: “Causal Inference with Repeated Measures in Observational Studies”

Angrist and Pischke (2009), Chapter 5

5. Problem Sets: Sharp regression discontinuity design, Ecological inference

Structural Equation Modeling

1. Instrumental Variables

Lecture notes: “Randomized Experiments with Noncompliance”

Angrist and Pischke Chapters 4 and 6

Imbens and Rubin (2015), Part VI

2. Direct and Indirect Effects

Kosuke Imai, Luke Keele, Dustin Tingley, and Teppei Yamamoto. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(4):765–789, November 2011.

3. Problem Sets: Fuzzy regression discontinuity design, Causal mediation analysis

Maximum Likelihood and Regression Models

1. Likelihood Theory

(Shorter) Freedman (2009) Section 7.1; or (Longer) King (1998) Chapter 4.

David A. Freedman. On the so-called “Huber sandwich estimator” and “robust standard errors”. *American Statistician*, 60(4):299–302, 2006.

2. Bootstrap and Monte Carlo Approximation

Freedman (2009) Chapter 8.

Gary King, Michael Tomz, and Jason Wittenberg. Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, 44(2): 341–355, 2000.

3. Discrete Choice Models

(Shorter) Freedman (2009) Sections 7.2–7.3; or (Longer) King (1998) Sections 5.1–5.4.

4. Sample Selection Models and Missing Data

James J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1): 153–161, January 1979.

Gary King, James Honaker, Anne Joseph, and Kenneth Scheve. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95(1):49–69, March 2001.

5. Problem Sets: Retrospective sampling design, Multiple imputation

9 Books

There is no single textbook for this course. However, you may find the following books (listed below in the alphabetical order) useful and some of them are used for this course. They are also available for purchase at the Labyrinth bookstore and on reserve at the library.

- Data analysis and **R**

John Fox. *An R and S-plus Companion to Applied Regression*. Sage Publications, Thousand Oaks, CA, 2nd edition, 2011.

Kosuke Imai. *A First Course in Quantitative Social Science*. A book manuscript, 2016.

- Causal inference

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, 2009.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2nd edition, 2009.

- Econometrics, Applied Statistics, and Political Methodology

David A. Freedman. *Statistical Models: Theory and Practice*. Cambridge University Press, Cambridge, 2nd edition, 2009.

Fumio Hayashi. *Econometrics*. Princeton University Press, Princeton, 2000.

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

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA, 2nd edition, 2010.

- Probability and Statistics

Joseph K. Blitzstein and Jessica Hwang. *Introduction to Probability*. CRC Press, Boca Raton, 2015.

Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*. Addison Wesley, Boston, 3rd edition, 2002.