

POL 345: Quantitative Analysis and Politics

Fall 2009

Kosuke Imai

Department of Politics, Princeton University

What accounts for who votes and their choice of candidate? Do politicians make policy based on constituency interests or their own ideologies? Would universal health insurance improve the health of the poor? Policy makers and academic researchers use statistics to answer these questions. However, the validity of their conclusions depends upon underlying assumptions and correct application of statistical methods. This course introduces the basics of applied statistics to students who have had little previous exposure to the subject. Topics will include causal inference, descriptive inference, survey analysis, and probability theory.

1 Should I Take This Course?

Here is a checklist to consider when deciding whether to take POL 345:

- This is my first statistics course at Princeton. Those who have already taken a statistics course should not enroll.
- I am willing to spend time on problem sets and learn statistical programming for data analysis.
- Please note that there are several alternative course offerings available for satisfaction of the QR/analytical requirements, and you should evaluate carefully whether this course is appropriate for your interests.

2 Contact Information

Name	Kosuke Imai	Will Bullock	Jaquilyn Waddell Boie
Office	Corwin Hall 036	Corwin Hall 235A	Corwin Hall 235A
Phone	609-258-6601	609-258-0092	609-258-9560
Email	kimai@princeton.edu	wbullock@princeton.edu	jwaddell@princeton.edu
Office Hours	Fridays 3-5pm	Wednesdays 1:30-3:30pm	Thursdays 2-4pm

Questions about lectures, readings, problem sets, and exams should be posted at the Discussion Board on BlackBoard so that other students in the class can benefit from them. For other matters, the best way to reach us is via e-mail. You can usually expect a response within 24 hours.

3 Logistics

- Lectures: Mondays and Wednesdays 3:30pm–4:20pm, Robertson 016
Lecture slides will be posted by 9am. Students should bring their own copy to the lectures.

- Precepts: All precepts will be held in Friend Center 005

Day	Tuesday	Tuesday	Tuesday	Wednesday	Wednesday
Time	11:00-11:50am	1:30-2:20pm	2:30-3:20pm	12:30-1:20pm	1:30-2:20pm

4 Course Requirements

- **Precept attendance (10%):** There will be ten precepts during the semester. Each precept attendance will be equally weighted. All precepts will be held in a computer lab at the Friend Center. There is no need to bring your own laptop although it is fine for you to do so.
- **Problem sets (40%):** There will be four problem sets during the semester. Each problem set will be equally weighted. See the problem set collaboration policy below.
- **In-class quizzes (20%):** There will be two closed-book, in-class quizzes. Each quiz will be equally weighted.
- **Take-home final exam (30%):** There will be an open-book final exam. Students should not discuss its contents with anyone before submission. The exam questions will be posted at noon on January 13 and will be due noon in the Corwin department office on January 15.

Neither problem sets nor the final exam may be submitted electronically. Answers are posted shortly after the deadline, and so late submission is not allowed without prior notice. More detailed instructions will be given for each assignment.

5 Precept Format and Policies

- **Handouts:** We will prepare a handout for each precept, which contains all the materials we expect you to learn for that week. The mastery of these materials is required for problem sets, quizzes, and exams. We will post the precept handout by Sunday 9am of each week. *We expect you to have printed the handout and read it over before coming to the precept.*
- **Practice problems:** We will also provide practice problems each week as an optional study guide that can help you make sure that you have mastered the materials covered in the precept and in the lectures. *You do not need to turn in your answers*, but we will be happy to discuss them with you should you have questions. Practice problems will be handed out at the beginning of the precept so that those who have already understood the handouts can start working on them during the precepts. The answers will be posted by Saturday 9pm.
- **Attendance:** Precept attendance is mandatory, and you must show up on time to your assigned precept. If medical illness or a family emergency arises, please let your preceptor know as soon as possible. In cases not as serious as those (e.g., a conflict with an extracurricular activity), you may be able to attend one of the other precepts offered during the week. In either case, letters from doctors or coaches may be requested and you should discuss the situation with your preceptor in advance.

6 Problem Set Collaboration Policy

Some problem sets for this course present opportunities for students to discuss questions and collaborate to find a solution together (For each problem set, please first check whether the collaboration is allowed). At the same time, as with any class that includes computer programming, there is a clear distinction between permissible collaboration and unacceptable plagiarism. This course will follow a modified version of the guidelines used for computer science classes here at Princeton.

Programming necessitates that you reach your own understanding of the problem and discover a path to its solution. During this time, discussions with other people (whether via the Internet or in person) are permitted and encouraged. However, when the time comes to write code that solves the problem, such discussions (except with course staff members) are no longer appropriate: the code must be your own work. For each assignment, please list the names of any individuals with whom you collaborated.

Do not, under any circumstances, copy another person's code. Incorporating someone else's code into your program in any form is a violation of academic regulations. Abetting plagiarism or unauthorized collaboration by sharing your code is also prohibited. Sharing code in digital form is an especially egregious violation: do not e-mail your code to anyone.

Novices often have the misconception that copying and mechanically transforming a program (by rearranging independent code, renaming variables, or similar operations) makes it something different. Actually, identifying plagiarized source code is easier than you might think.

This policy supplements the University's academic regulations, making explicit what constitutes a violation for this course. Princeton Rights, Rules, Responsibilities handbook asserts:

The only adequate defense for a student accused of an academic violation is that the work in question does not, in fact, constitute a violation. Neither the defense that the student was ignorant of the regulations concerning academic violations nor the defense that the student was under pressure at the time the violation was committed is considered an adequate defense.

If you have any questions about these matters, please consult a course staff member.

7 Textbooks

The main textbook for this course is:

- Agresti, Alan and Barbara Finlay. (2008). *Statistical Methods for the Social Sciences*. 4th eds. Prentice Hall.

In addition, parts of the following two books may be helpful:

- Freedman, David, Robert Pisani, and Roger Purves. (2007). *Statistics*. 4th eds. Norton.
- Verzani, John. (2005). *Using R for Introductory Statistics*. Chapman & Hall.

All of these books are available at Labyrinth and on reserve at Firestone.

8 Statistical Software

In this course, we use the open-source statistical software **R** (<http://www.r-project.org>). Recently, a *New York Times* article (“Data Analysts Captivated by **R**'s Power” January 6, 2009) featured **R** as,

a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca [...] whether being used to set ad prices, find new drugs more quickly or fine-tune financial models. Companies as diverse as Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group and Shell use it. [...] “The great beauty of R is that you can modify it to do all sorts of things,” said Hal Varian, chief economist at Google. “And you have a lot of prepackaged stuff that’s already available, so you’re standing on the shoulders of giants.”

R is much more powerful than other statistical software such as SPSS, STATA and SAS, but it’s a bit more difficult to learn. A variety of resources will be made available for POL 345 students in order to learn R as efficiently as possible.

9 How Do I Get Help?

Because POL 345 is a challenging course, we have made the following resources available to you in order to facilitate efficient learning about statistics and data analysis. We encourage you to take advantage of them whenever you have questions about the course materials and are struggling with problem sets.

- **R information session:** You may find R challenging especially at the beginning of the semester. To help you get started, we will hold an optional R information session during the first week (Thursday, September 24th) where the preceptors will go over the handout step by step. The attendance to this information session is voluntary, but we assume that you understand the materials in the handout before the precepts begin in the second week. The session will be held from 7:30-8:30pm in Robertson 016.
- **Tutors:** In addition to the instructor and the preceptors, we will have tutors available at the McGraw Center who can help you learn the course materials. The tutors are your fellow undergraduate students who did very well in POL 345 last year. Unless otherwise announced, the tutoring session, Aka “Study Hall,” will be available in the Frist Campus Center from 7:30pm to 10:30pm on Sundays (Harrison Frye) and Mondays (Catherine Che).
- **Office hours:** Each of us will hold weekly office hours, starting the first week. In addition to these office hours, I have an open door policy – you can stop by my Corwin office anytime without an appointment and I will answer your questions so long as I am in the office. You may also e-mail to set up an appointment with any of us outside of our office hours.
- **Discussion board:** In addition to office hours and individual appointments, we will be available online to answer any questions you may have about the course materials and the problem sets. We use the Discussion Board functionality of the BlackBoard system. Before posting your question, please review previous posts to make sure that a similar question has not been answered. *In accordance with the problem set collaboration policy described above, you should not directly post your code for a problem set.* You should frame your questions in general terms rather than trying to have us debug your code directly. You may subscribe to the Discussion Board so that you receive your fellow students’ questions and our answers to those questions. You should also feel free to respond to questions that you can answer.
- **Individual advising:** In order to make sure that nobody is falling behind, I will meet briefly with each of you after the first in-class quiz and before the fall break. This meeting will give

me an opportunity to get to know each of you and receive feedback from you about the course, and I will provide study tips for anyone who is having difficulties. Please sign up through the web appointment system <https://wass.princeton.edu/pages/login.page.php>.

10 Course Plan

Introduction

Week 1: September 21–25

- LECTURES: Overview of the course, Descriptive vs. inferential statistics, Forecasting
- READINGS: A&F Chapter 1
- INFORMATION SESSION: Introduction to **R**
- SUPPLEMENTAL READINGS: Verzani 1.2.1-1.2.3

Week 2: September 28–October 2

- LECTURES: Survey sampling, Causal inference
- READINGS: A&F Chapter 2 (2.2–2.4)
- PRECEPT 1: Data manipulation
- PROBLEM SET 1: Posted on Thursday, Oct 1
- SUPPLEMENTAL READINGS: FPP Chapters 1, 2, and 19; Verzani 1.2.4-1.4.3

Descriptive Statistics

Week 3: October 5–9

- LECTURES: Data measurement, Numerical and graphical summary
- READINGS: A&F Chapters 2 (2.1) and 3 (3.1–3.4, 3.6–3.7);
- PRECEPT 2: Summarizing univariate data
- PROBLEM SET 1: Due on Wednesday, Oct. 7
- SUPPLEMENTAL READINGS: FPP Chapters 3–7; Verzani 2.1-2.3

Week 4: October 12–16

- LECTURE: Correlation
- READINGS: A&F Chapter 3 (3.5); FPP Chapters 8–9
- PRECEPT 3: Summarizing bivariate data
- IN-CLASS QUIZ 1: Given on Monday, Oct. 12
- SUPPLEMENTAL READINGS: Verzani 3.1-3.4

Week 5: October 19–23

- LECTURES: Regression
- READINGS: A&F Chapter 9 (9.1–9.4)
- PRECEPT 4: Loop and conditional statements
- PROBLEM SET 2: Posted on Wednesday, Oct. 21
- SUPPLEMENTAL READINGS: FPP Chapters 10–12; Verzani 5.1, 6.2, 6.4

Probability

Week 6: October 26–30

- LECTURES: Conditional probability, Independence, Random variables, Expectation
- READINGS: A&F Chapter 4 (4.1–4.2)
- NO PRECEPT
- PROBLEM SET 2: Due on Wednesday, Oct. 28
- SUPPLEMENTAL READINGS: FPP Chapters 13–14

Fall Break: October 31–November 8

Week 7: November 9–13

- LECTURES: Binomial and normal distributions, Law of large numbers, Central limit theorem
- READINGS: A&F Chapter 4 (4.3–4.7)
- PRECEPT 5: Probability and simulations
- SUPPLEMENTAL READINGS: FPP Chapters 15–18; Verzani 5.2-5.3, 6.3

Statistical Inference

Week 8: November 16–20

- LECTURES: Statistical analysis of sample surveys
- READINGS: A&F Chapter 5
- PRECEPT 6: Unbiasedness of estimator and confidence intervals
- PROBLEM SET 3: Posted on Monday, Nov. 16
- SUPPLEMENTAL READINGS: FPP Chapters 20–23; Verzani 7.1-7.3

Week 9: November 23–27

- LECTURES: Statistical hypothesis tests
- READINGS: A&F Chapter 6
- PRECEPT 7: Statistical test
- PROBLEM SET 3: Due on Monday, Nov. 23
- SUPPLEMENTAL READINGS: FFP Chapters 26 and 29; Verzani 8.1-8.5

Week 10: November 30–December 4

- LECTURE: Statistical analysis of randomized experiments
- READINGS: A&F Chapter 7
- PRECEPT 8: Analysis of experimental data
- IN-CLASS QUIZ 2: Given on Wednesday, Dec. 2
- SUPPLEMENTAL READINGS: FFP Chapter 27; Verzani 7.5, 8.6.1-8.6.2

Week 11: December 7–11

- LECTURES: Statistical inference with simple regression
- READINGS: A&F Chapters 9 (9.5–9.6)
- PRECEPT 9: Simple regression in **R**
- PROBLEM SET 4: Posted on Thursday, Dec. 10
- SUPPLEMENTAL READINGS: Verzani 10.1-10.2

Week 12: December 14–19

- LECTURES: Statistical inference with multiple regression
- READINGS: A&F Chapters 10 and 11
- PRECEPT 10: Multiple regression in **R**
- PROBLEM SET 4: Due on Wednesday, Dec. 16
- SUPPLEMENTAL READINGS: Verzani 10.3