

POL 345: Quantitative Analysis and Politics

Fall 2011

Kosuke Imai

Department of Politics, Princeton University

What accounts for who votes and their choice of candidate? Would universal health insurance improve the health of the poor? Researchers and policy makers use statistics to answer these questions. However, the validity of their conclusions depends upon underlying assumptions and correct application of statistical methods. The course will introduce basic principles of statistical inference and programming skills for data analysis. The goal is to provide students with the foundation necessary to analyze data in their independent research at Princeton and to become a critical consumer of news articles and academic studies that use statistics.

1 Who Should and Should Not 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 college-level statistics course should not enroll.
- I am willing to spend considerable time *outside* of classroom each week in order to keep up with the course materials.
- I would like to *use* statistics in my junior paper, senior thesis, and/or job in the future.

Please note that there are a number of alternative course offerings available for satisfaction of the QR/analytical requirements, and you should evaluate carefully whether this course is appropriate for your interests. In particular, the analytical requirement of the politics department can be fulfilled by taking any of the following courses: POL 346, POL 347, POL 350, ECO 202, ECO 302, ECO 312, ORF 245, PSY 251, SOC 301, SOC 404, WWS 332.

2 POL 345 and Beyond

POL 345 is an introductory statistics and data analysis course, and I encourage students who take this course to continue using statistics. Alumni of this course have used statistics in their senior thesis and won best thesis prizes, and many of them sent me an email about their encounter with statistics in their summer internships and post-graduate career. In today's information world, data are available everywhere and the role of statistics is rapidly increasing in academia, business, medicine, public policy, and many other parts of the society. I echo the message of *The New York Times* which published an article entitled "For Today's Graduate, Just One Word: Statistics."

3 Contact Information

Name	Kosuke Imai	Alex Acs	Carlos Velasco
Office	Corwin Hall 036	Robertson Hall 324	Corwin 026
Email	kimai@princeton.edu	aacs@princeton.edu	cvelasco@princeton.edu
Office Hours	Fridays 1:30–3:30pm or stop by any time	Mondays 2:00–3:30pm (Corwin 023) Thursdays 10:30–12:00pm (Bendheim 215)	Thursdays 2:00–3:30pm (Bendheim 317)

In addition to office hours, each of us is also available by appointment. 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.

4 Logistics

- **Lectures:** Mondays and Wednesdays 3:30pm–4:20pm, Robertson 002
Lecture slides will be posted by 9am before lecture. Students should bring their own copy to the lectures.
- **Precepts:** All precepts will be held in Friend Center 005. This is a computer lab, and so there is no need to bring your own laptop although it is fine for you to do so. Precept assignment will be done through Blackboard during the first week of the semester.

Day	Tuesday	Tuesday	Tuesday	Wednesday	Wednesday
Time	9:00-9:50am	11:00-11:50am	2:30-3:20pm	10:00-10:50am	11:00-11:50am

5 Course Requirements

- **Precept participation (10%):** There will be ten precepts during the semester. Each precept will be equally weighted. Students are required to turn in their answers to a short list of programming questions *24 hours* prior to each precept via the precept folder at Blackboard. The answers will not be graded but will be used to facilitate the discussion during the precept.
- **Problem sets (20%):** There will be four problem sets during the semester. The problem sets provide an opportunity for students to conduct data analysis and learn important statistical concepts. Each problem set will be equally weighted, and the computer code should also be submitted via the problem set folder at Blackboard.
- **In-class quizzes (30%):** There will be three closed-book, in-class quizzes. Each quiz will be equally weighted. The quizzes will serve as one way for the course staff and students to assess how well students are understanding the key concepts covered in the class.
- **Take-home final exam (40%):** There will be an open-book, take-home final exam. Students should not discuss its contents with anyone before submission. The exam questions will be posted at noon on January 18 and your writeup will be due noon in the Corwin department office on January 23. The computer code should be submitted via the final folder at Blackboard.

Neither problem sets nor the final exam may be submitted electronically. Your answers should contain annotated computer code, an electronic copy of which should also be submitted through the appropriate folder at Blackboard. 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.

6 Submission of the Computer Code via Blackboard Folders

For precepts, problem sets, and final exam, students are required to turn in the computer code used via assignment folders at Blackboard. For each assignment, submit your code to the appropriate folder as a *single* file named as `xxxHandoutX.R` or `xxxPSetX.R` where `xxx` is your NetID and `X` is the handout/problem set number. For example, it would be `kimaiHandout2.R` for me if I am submitting my code for the questions in Handout 2.

7 Precept Format and Policy

- **Precept 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 final exam. We will post the precept handout at Blackboard by Friday 9am of the previous week.
- **What you need to do before coming to the precept:** Before coming to each precept, you should go over the handout and answer the small set of programming questions, called “Precept Questions” that are included in the handout. You are allowed to collaborate with others, but must comply with the collaboration policy of this course. These questions are designed to check whether you understood the materials covered in the handout and facilitate the discussion during precepts. While you must *turn in your answers to the appropriate precept folder at Blackboard 24 hours before the precept*, they will not be graded.
- **What you will do during the precept:** The precept will begin by briefly reviewing the handout. Come to the precept with questions. We will give special attention to issues that seemed problematic for students in the Precept Questions. The remainder of the precept time will then be spent on practice problems and additional materials, which we will prepare and distribute during the precept. The answers to these practice problems will be posted on Blackboard by Thursday 9am. Students are strongly encouraged to solve all practice problems including the ones we do not have time to solve during the precept in order to prepare for problem sets and final exam.
- **Participation:** Precept participation 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 should notify your preceptor at least 2 days before the precept so that we can make an appropriate arrangement. In either case, letters from doctors or coaches may be requested.

8 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 collaboration is allowed). At the same time, as with any class that includes analytical exercises and 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. *Please take this guideline seriously.* In the past, plagiarism cases typically resulted in one-year suspension from 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 *sometimes* 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. For example, there exists computer software that can detect plagiarism.

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.

9 Textbooks

The main textbook for this course is:

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

You are required to read most chapters of this book and so we recommend that you purchase it. In addition, parts of the following book may be helpful, but the purchase of these books is not required:

- Agresti, Alan and Barbara Finlay. (2008). *Statistical Methods for the Social Sciences*. 4th eds. Prentice Hall.
- Verzani, John. (2005). *Using R for Introductory Statistics*. Chapman & Hall.

Both books are available at Labyrinth and on reserve at Firestone. Some parts of Agresti and Finlay (2008) are available electronically at Blackboard (through E-reserve).

10 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.

11 Getting Help

Because POL 345 is a challenging course for some of you, 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 (September 21) where we will go over the handout step by step. Attendance at 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 5:00-6:00pm in Corwin 127.
- **Study Hall:** 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, Cicely Jones and Jason Weinreb, who did very well in POL 345 last year. Unless otherwise announced, the tutoring session, a.k.a. “Study Hall,” will be available in the Frist Campus Center from 7:30pm to 10:30pm on Sundays (Cicely) and Mondays (Jason), starting from the second week of the semester.
- **Peer Tutoring Program:** This program is designed to help those students who are having difficulties in a course because of gaps in their preparation, a particularly challenging subject matter, loss of time due to illness, and other factors. Like the Study Hall tutors, tutors of this program are undergraduates who have performed well in POL 345. Unlike Study Hall, however, this individualized tutoring service is not free. For the details of the program, see <http://www.princeton.edu/odoc/services/tutoring/>

Please note that Princeton undergraduates are permitted to use only those tutoring services authorized by the Office of the Dean of the College.

- **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 available. You may also e-mail to set up an appointment with either 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 answers to those questions. You should also feel free to respond to questions that you can answer.

12 Weekly Schedule

To sum up, the following is the schedule for a typical week of POL 345.

Monday	Tuesday	Wednesday	Thursday	Friday	Sunday
9:00am Slides posted		9:00am Slides posted	9:00am Answers posted	9:00am Handout posted	
2:00–3:30pm Office hour		Precept (submit your answers)	10:30am-12:00pm Office hour	1:30-3:30pm Office hour	
3:30–4:20pm Lecture		3:30-4:20pm Lecture	2:00–3:30pm Office hour		
7:30–10:00pm Study hall					7:30–10:00pm Study hall

13 Course Plan

We will make the assigned chapters of A&F available through Blackboard. We also list relevant parts of Verzani for each precept for your information, but in general the handouts we distribute should be sufficient.

Introduction

WEEK 1: SEPTEMBER 19–23

- **LECTURES:** Overview of the course, Descriptive and statistical inference, Research design
- **READINGS:** FPP Chapters 1 and 2
- **R INFORMATION SESSION:** Introduction to **R** (Verzani 1.2.1–1.2.3)
Wednesday, September 21: 5:00-6:00pm in Corwin 127

Descriptive Statistics: Summarizing Data

WEEK 2: SEPTEMBER 26–30

- LECTURES: Data measurement, Numerical and graphical summary
- READINGS: FPP Chapters 3–7
- PRECEPT 1: Data manipulation (Verzani 1.2.4–1.4.3)

WEEK 3: OCTOBER 3–7

- LECTURES: Correlation
- READINGS: FPP Chapters 8–9
- PRECEPT 2: Summarizing univariate data (Verzani 2.1–2.3)
- PROBLEM SET 1: Posted on Wednesday, Oct. 5

WEEK 4: OCTOBER 10–14

- LECTURE: Regression
- READINGS: FPP Chapters 10–12
- PRECEPT 3: Summarizing bivariate data (Verzani 3.1–3.4)
- PROBLEM SET 1: Due on Wednesday, Oct. 12

Probability: Modeling Uncertainty

WEEK 5: OCTOBER 17–21

- LECTURES: Conditional probability, Independence, Random variables, Expectation
- READINGS: FPP Chapters 13–14
- PRECEPT 4: Loop and conditional statements (Verzani 5.1, 6.2, and 6.4)
- IN-CLASS QUIZ 1: Monday, Oct. 17
- PROBLEM SET 2: Posted on Wednesday, Oct. 19

WEEK 6: OCTOBER 24–28

- LECTURES: Binomial and normal distributions, Law of large numbers
- READINGS: FPP Chapters 15–17; A&F Chapter 4 (4.1 – 4.3)
- PRECEPT 5: Review of the materials
- PROBLEM SET 2: Due on Friday, Oct. 28

FALL BREAK: OCTOBER 29–NOVEMBER 6

Statistical Inference: Learning from Data

WEEK 7: NOVEMBER 7–11

- LECTURES: Central limit theorem, sample surveys, point estimation
- READINGS: FFP Chapters 18–21; A&F Chapters 4 (4.4–4.5) and 5 (5.1)
- PRECEPT 6: Probability and simulations (Verzani 5.2, 5.3, and 6.3)

WEEK 8: NOVEMBER 14–18

- LECTURES: Standard error, confidence intervals
- READINGS: FFP Chapter 23; A&F Chapter 5 (5.2–5.4)
- PRECEPT 7: Point estimation and confidence intervals (Verzani 7.1–7.3)

WEEK 9: NOVEMBER 21–23

- IN-CLASS QUIZ 2: Monday, Nov. 21
- WEDNESDAY LECTURE: Review of the materials
- NO PRECEPT: Extra Office Hours
- PROBLEM SET 3: Posted on Monday, Nov. 21

THANKSGIVING BREAK: NOVEMBER 24–27

WEEK 10: NOVEMBER 28–DECEMBER 2

- LECTURES: Hypothesis tests
- READINGS: FFP Chapters 26 and 29; A&F Chapter 6 (6.1–6.6)
- PRECEPT 8: Statistical test (Verzani 8.1–8.5)
- PROBLEM SET 3: Due on Friday, Dec. 2

WEEK 11: DECEMBER 5–9

- LECTURE: Randomized experiments, Regression
- READINGS: FFP Chapter 27; A&F Chapter 7 (7.1–7.3)
- PRECEPT 9: Analysis of experimental data (Verzani 7.5, 8.6.1–8.6.2)
- PROBLEM SET 4: Posted on Wednesday, Dec. 7

WEEK 12: DECEMBER 12–16

- WEDNESDAY LECTURE: Regression
- READINGS: A&F Chapter 9 (9.5–9.6)
- PRECEPT 10: Regression in \mathbf{R} (Verzani 10.1–10.2)
- IN-CLASS QUIZ 3: Monday, Dec. 12
- PROBLEM SET 4: Due on Friday, Dec. 16