POL 345: Quantitative Analysis and Politics

Fall 2015

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. Other introductory statistics courses include ECO202, ORF245, PSY251, SOC301, and WWS200.

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

INSTRUCTOR:

Name	Kosuke Imai
Office	Corwin Hall 036
Email	kimai@princeton.edu

PRECEPTORS:

Name:	Brandon de la Cuesta	Ted Enamorado
Office:	Roberston 008	Robertson 006
Office Hours:	Mondays 10:00 - 11:30am	Tuesdays 4:30 - 6:00pm
Email:	bjmiller@princeton.edu	tede@princeton.edu

Name:	Tolya Levshin	Bella Wang
Office:	Robertson 015	Fisher B04
Office Hours:	Fridays 10:00 - 11:30am	Thursdays 3:30 - 5:00 pm
Email:	alevshin@princeton.edu	bxw@princeton.edu

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 Piazza 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, McCosh 28
  Lectures will often contain in-class data analysis exercises. Students should bring their own laptop to each lecture. Lecture slides will be posted shortly after lecture at Blackboard.

- 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:	Thursday	Thursday	Thursday	Thursday	Thursday
Time:	10:00-10:50am	11:00-11:50am	12:30-1:20pm	1:30-2:20pm	2:30-3:20pm

	Thursday	Friday	Friday
	3:30pm-4:20pm	1:30-2:20pm	2:30-3:20pm

5 Course Requirements

The course requirements consist of the following four components:

- **Participation (10%):** Students should actively participate in all aspects of the course. Class participation will be judged based on questions asked/answered during the lectures, the precepts, and on the online discussion board.

- **Problem sets (10%):** 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 graded as $\checkmark-$ (unsatisfactory), $\checkmark$(satisfactory), or $\checkmark+$ (excellent), and counts equally towards the final course grade. Collaboration is permitted, but students must write up the code and answers on their own.

- **In-class quiz (10%)**: There will be one closed-book, in-class quiz held on November 23. The quiz 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 up to that point in the semester.

- **Take-home exams (40%)**: There will be two open-book, take-home exams, one during the midterm week and the other during the final week of the semester. No collaboration is allowed, and students should not discuss their contents with anyone before submission. Each problem set is equally weighted.

- **Final group project (30%)**: There will be a final group project due "Dean's Date." Students will find a data set of interest, analyze it, and report findings in a short memo. The details will be announced later during the semester.

For all assignments, late submission is not allowed without prior notice.

6 Submission of the Computer Code via Blackboard Folders

For the problem sets, take-home exam, and final project, students are required to electronically submit their computer code as well as their write-up to the assignment folders at Blackboard. For each assignment, submit your code to the appropriate folder as a *single* file named as `xxxExamX.R` or `xxxPSetX.R` where `xxx` is your NetID and `X` is the exam/problem set number. For example, it would be `kimaiExam2.R` for me if I am submitting my code for the questions in the second take-home exam.

7 Lecture and Precept Format and Policy

- **Small assignments:** There will be small assignments throughout the semester. These assignments are directly based on the textbook and are designed to check whether you understood the materials covered in the textbook. They are not graded but the lectures and precepts assume that students have completed them. You are allowed to obtain any help you would need in order to complete these assignments.

- **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

Problem sets for this course present opportunities for students to discuss questions and collaborate to find a solution together. 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 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.

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 Textbook

The main textbook for this course is:

> Imai, Kosuke. *A First Course in Quantitative Social Science.* Princeton University Press, forthcoming.

The textbook, a.k.a. "QSS," is made freely available to the students via Perusall at `https://perusall.com`. Students must register first at this site in order to view the textbook. The instruction is available at Blackboard. Due to the copyright issues, this file should not be distributed to those who are not taking this class. Perusall enables students to comment on and ask questions about the textbook directly. We welcome any feedback including typos, mistakes, and clarifications. Students who provide useful comments and questions will receive **extra credits** (up to 10% of the course grade).

10 Statistical Software

In this course, we use the open-source statistical software R (http://www.r-project.org). 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 many 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 introduction session:** You may find R challenging especially at the beginning of the semester. To help you get started, we will hold an optional R introduction session during the first week. Attendance at this introduction session is voluntary, but we assume that you understand the materials in Chapter 1 of QSS before the second week begins. The time and location of the session is TBD.

- **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 know R but may not be familiar with the course materials. Unless otherwise announced, the tutoring session, a.k.a. "Study Hall," will be available at the McGraw Center in the Frist Campus Center from 7:30pm to 10:30pm from Sundays till Wednesday, starting on September 27. The head tutor of this Study Hall is Seth Morokoff (morokoff@Princeton.edu).

- **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. 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 preceptor 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.

- **Piazza discussion forum:** 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 Piazza discussion forum that will be linked on Blackboard course page or accessible directly at `http://piazza.com`. You should also feel free to respond to questions that you can answer. Piazza also has a free smartphone application if you are interested.

12   Course Plan

We have set up a Google Calendar with events, due dates, and sessions. The URL for this calendar is `https://goo.gl/MtKhyH`.

Introduction

WEEK 0: SEPTEMBER 16–21

- TOPIC: Overview of the course, Introduction to R

- TEXTBOOK: Chapter 1 (Section 1.3)

- R INTRODUCTION SESSION: TBD

Causality

WEEK 1: SEPTEMBER 21–25

- TOPIC: Randomized experiments

- TEXTBOOK: Chapter 2 (Sections 2.1–2.4)

- PROBLEM SET 1: Posted on Thursday, September 24

WEEK 2: SEPTEMBER 28–OCTOBER 2

- TOPIC: Observational studies

- TEXTBOOK: Chapter 2 (Sections 2.5–2.7)

- PROBLEM SET 1: Due on Wednesday, September 30

Measurement

WEEK 3: OCTOBER 5–9

- TOPIC: Survey sampling

- TEXTBOOK: Chapter 3 (Sections 3.1–3.4)

- PROBLEM SET 2: Posted on Thursday, Oct. 8

WEEK 4: OCTOBER 12–16

- TOPIC: Clustering

- TEXTBOOK: Chapter 3 (Sections 3.5–3.7)

- PROBLEM SET 2: Due on Wednesday, Oct. 14

Prediction

WEEK 5: OCTOBER 19–23

- TOPIC: Prediction and Loop

- TEXTBOOK: Chapter 4 (Section 4.1)

- TAKE-HOME EXAM 1: Posted on Friday, Oct. 23

WEEK 6: OCTOBER 26–30

- TOPIC: Regression

- TEXTBOOK: Chapter 4 (Sections 4.2 and 4.3)

- TAKE-HOME EXAM 1: Due on Friday, Oct. 30

FALL BREAK: OCTOBER 31–NOVEMBER 8

Probability

WEEK 7: NOVEMBER 9–13

- TOPIC: Probability and conditional probability

- TEXTBOOK: Chapter 6 (Sections 6.1–6.3)

- PROBLEM SET 3: Posted on Thursday, November 12

WEEK 8: NOVEMBER 16–20

- TOPIC: Random variables and their distributions, Large sample theorems

- TEXTBOOK: Chapter 6 (Sections 6.4–6.5)

- PROBLEM SET 3: Due Wednesday, November 18

WEEK 9: NOVEMBER 23

IN-CLASS QUIZ

THANKSGIVING BREAK: NOVEMBER 24–29

Uncertainty

WEEK 10: NOVEMBER 30–DECEMBER 4

- TOPIC: Estimation

- TEXTBOOK: Chapter 7 (Section 7.1)

- PROBLEM SET 4: Posted on Thursday, December 3

WEEK 11: DECEMBER 7–11

- TOPIC: Hypothesis tests

- TEXTBOOK: Chapter 7 (Section 7.2)

- PROBLEM SET 4: Due on Wednesday, December 9

- TAKE-HOME EXAM 2: Posted on Friday, December 11

WEEK 12: DECEMBER 14–18

- TOPIC: Regression with uncertainty

- TEXTBOOK: Chapter 7 (Section 7.3)

- TAKE-HOME EXAM 2: Due on Wednesday, December 16