POL345/SOC305: Introduction to Quantitative Social Science

Fall 2016

Margaret Frye (Sociology)   Kosuke Imai (Politics)

Princeton University

Would universal health insurance improve the health of the poor? Do patterns of arrests in US cities show evidence of racial profiling? What accounts for who votes and their choice of candidates? This course will teach students how to address these and other social science questions by analyzing quantitative data. The course introduces 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 own research and to become critical consumers of statistical claims made in the news media, in policy reports, and in academic research.

1   Who Should and Should Not Take This Course

Here is a checklist to consider when deciding whether to take POL345/SOC305:

- I am a sociology or politics concentrator (other students who are interested in quantitative social science are welcome too).

- In addition to developing my knowledge of statistical concepts, I want to learn the computational skills needed to manipulate and analyze data.

- 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 and politics analytical requirements, and you should evaluate carefully whether this course is appropriate for your interests. Other introductory statistics courses include ECO202, ORF245, PSY251, and WWS200.

2   POL345/SOC305 and Beyond

POL345/SOC305 is an introductory statistics and data analysis course, and we encourage students who take this course to continue using statistics. Alumni of previous statistics courses that we have taught have used statistics in their senior thesis and won best thesis prizes, and many have sent us emails about their encounters 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 society. We echo the message of *The New York Times* which published an article entitled "For Today's Graduate, Just One Word: Statistics."

3   Contact Information

Course Instructors:

NAME	Margaret Frye	Kosuke Imai
OFFICE	Wallace Hall 147	Corwin Hall 036
OFFICE HOURS	Wednesdays 1:15–3:00 (sign up at WASS)	Tuesdays 2:00–3:30 (sign up at WASS)
EMAIL	mfrye@princeton.edu	kimai@princeton.edu

Preceptors:

NAME	Herrissa Lamothe	Mariana Campos Horta
OFFICE	Wallace Hall 127	Wallace Hall 225
OFFICE HOURS	Fridays 2:00–3:30	Mondays 11:00–12:30
EMAIL	hlamothe@princeton.edu	mcampos@princeton.edu

NAME	Adeline Lo	Alexander Kustov
OFFICE	Corwin Hall 037	Corwin Hall 027
OFFICE HOURS	Thursdays 9:30–11	Mondays 4:30–6:00
EMAIL	adelinel@princeton.edu	akustov@princeton.edu

NAME	Diana Stanescu	Daniela Urbina Julio
OFFICE	Corwin Hall 235A	Wallace Hall 284
OFFICE HOURS	Tuesdays 4:30–6:00	Thursdays 11:00–12:30
EMAIL	stanescu@princeton.edu	djulio@princeton.edu

NAME	Yo-Yo Chen	Hannah Korevaar
OFFICE	Wallace Hall 284	Wallace Hall 227
OFFICE HOURS	Wednesdays 16:30–18:00	Tuesdays 10:00–11:30
EMAIL	sc37@princeton.edu	hannahmk@princeton.edu

Workshop Instructors:

NAME	Yunkyu Sohn	Ethan Fosse
OFFICE	Corwin Hall 037	Wallace Hall 108
OFFICE HOURS	Wednesdays 4:30–6:00	Fridays 2:00–3:00
EMAIL	ysohn@princeton.edu	efosse@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 on 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, Friend Center 101
  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 a computer lab. There is no need to bring your own laptop, although it is fine for you to do so. Precepts will start meeting on Wednesday, September 21 and will meet at the following times and locations:

Precept:	P01	P02A; P02B	P03A; P03B	P04A; P04B
Day:	Wednesday	Thursday	Thursday	Thursday
Time:	7:30-8:50pm	11:00-12:20am	1:30-2:50pm	3:00-4:20pm
Location:	Robertson 011	Friend 005; 007	Friend 005; 007	Friend 005; 007

Precept:	P05A; P05B	P06A; P06B	P07A; P07B
Day:	Friday	Friday	Thursday
Time:	11:00-12:20am	1:30-2:50pm	7:30-8:50pm
Location:	Friend 005; 007	Friend 005; 007	Friend 005; 007

- Precept Assignments: You will be able to select your preferred precept on Tiger Hub from Wednesday (9/14) at 9am to Friday (9/16) at 7pm. If you do not choose a precept by then, we will assign you one. We will also change some assignments to ensure that precept sizes are balanced. You will be notified of your final precept assignment on Monday (9/19).

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 Piazza.

- **Online programming assignments (10%):** There will be weekly online programming assignments. They will not be graded but students are expected to complete them on time. **Collaboration is permitted**. These assignments are directly based on the textbook and are designed to check whether you understood the materials covered in the textbook. If, for some reason, you have a trouble installing the required R packages, you may use the RStudio sever we have set up. In a web browser (Chrome recommended), type the following url, `https://pol-rstudio.princeton.edu` Note that you must be on the campus network to access this server.

- **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 21. 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 take-home exam is equally weighted.

- **Final group project (20%)**: There will be a final group project due on "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 at least 24 hours 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 submit a print out of their writeup produced using an Rmarkdown file. Students must also upload an electronic copy of the Rmarkdown file to Blackboard using `xxx.Rmd` as a file name where `xxx` is your NetID.

7   Precept Policy

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

This course uses a manuscript of the following textbook.

> Imai, Kosuke (2017). *Quantitative Social Science: Introduction.* Princeton University Press, Forthcoming.

The textbook, aka "QSS," is made freely available to the students via Blackboard. Due to the copyright issues, this file should not be distributed to those who are not taking this class.

10 Statistical Software

In this course we use the open-source statistical software R (http://www.r-project.org). R can be more powerful than other statistical software such as SPSS, STATA and SAS, but it can also be more difficult to learn. A variety of resources will be made available for POL345/SOC305 students in order to learn R as efficiently as possible. To help make using R easier, we'll be using RStudio (http://www.rstudio.com/)—a user interface that simplifies many common operations.

11 Getting Help

Because POL345/SOC305 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 COMPASS Workshops** You may find R challenging, especially at the beginning of the semester. To help you to master the computational demands of the course, we have created a new resource: a series of optional weekly computing workshops designed around the curriculum of the course. The workshops will be held on Tuesdays from 7:30–9:00pm in Room 307 at the McGraw Teaching Center in the Frist Campus Center. Some of the workshops will provide reinforcement of basic concepts, while others will teach additional skills that go beyond the content of the course but will be very helpful for your own research and career. The workshops are open to all students, staff, and affiliates at Princeton, so please tell your friends about them! Attendance is voluntary but strongly recommended, as this resource will really help you to master the programming element of the course. To fully participate in the workshops you should bring a laptop, although there will be a few laptops available at the workshops.

- **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 25. The head tutor of this Study Hall is Vilma Jimenez (vjimenez@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 POL345/SOC305. 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. 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 12–16

- Topic: Overview of the course, Introduction to R
- Textbook: Chapter 1 (Section 1.3)

Causality

Week 1: September 19–23

- Topic: Randomized experiments
- Textbook: Chapter 2 (Sections 2.1–2.4)
- Workshop: Introduction to R and RStudio
- Problem Set 1: Posted on Thursday, September 22

Week 2: September 26–30

- Topic: Observational studies

- Textbook: Chapter 2 (Sections 2.5–2.7)

- Workshop: Data Wrangling in R

- Problem Set 1: Due on Wednesday, September 28

Measurement

Week 3: October 3–7

- Topic: Survey sampling

- Textbook: Chapter 3 (Sections 3.1–3.4)

- Workshop: Base Graphics in R

- Problem Set 2: Posted on Thursday, Oct. 6

Week 4: October 10–14

- Topic: Clustering

- Textbook: Chapter 3 (Sections 3.5–3.7)

- Workshop: Data Visualization in R with ggplot2

- Problem Set 2: Due on Wednesday, Oct. 12

Prediction

Week 5: October 17–21

- Topic: Prediction and Loop

- Textbook: Chapter 4 (Section 4.1)

- Workshop: Programming Loops in R

- Take-home Exam 1: Posted on Friday, Oct. 21

Week 6: October 24–28

- Topic: Regression

- Textbook: Chapter 4 (Sections 4.2 and 4.3)

- Take-home Exam 1: Due on Friday, Oct. 28

FALL BREAK: OCTOBER 29–NOVEMBER 6

Probability

WEEK 7: NOVEMBER 7–11

- TOPIC: Probability and conditional probability

- TEXTBOOK: Chapter 6 (Sections 6.1–6.3)

- WORKSHOP: Probability and Simulations in R

- PROBLEM SET 3: Posted on Thursday, November 10

WEEK 8: NOVEMBER 14–18

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

- TEXTBOOK: Chapter 6 (Sections 6.4–6.5)

- WORKSHOP: Monte Carlo Simulations in R

- PROBLEM SET 3: Due Wednesday, November 16

WEEK 9: NOVEMBER 21

IN-CLASS QUIZ

THANKSGIVING BREAK: NOVEMBER 23–27

Uncertainty

WEEK 10: NOVEMBER 28–DECEMBER 2

- TOPIC: Estimation

- TEXTBOOK: Chapter 7 (Section 7.1)

- WORKSHOP: Text Analysis in R

- PROBLEM SET 4: Posted on Thursday, December 1

WEEK 11: DECEMBER 5–9

- TOPIC: Hypothesis tests

- TEXTBOOK: Chapter 7 (Section 7.2)

- PROBLEM SET 4: Due on Wednesday, December 7

- WORKSHOP: Hypothesis Testing in R

- TAKE-HOME EXAM 2: Posted on Thursday, December 8

Week 12: December 12–16

- Topic: Regression with uncertainty

- Textbook: Chapter 7 (Section 7.3)

- Workshop: Regression Analysis in R

- Take-home Exam 2: Due on Friday, December 16