POL 345/SOC 305: Introduction to Quantitative Social Science

Fall 2017

Marc Ratkovic (Politcs) Matthew Salganik (Sociology)

Princeton University

Why do people vote the way they do? Can universal health insurance lead to a longer lifespan? What countries are more or less likely to erupt in civil conflict? Assessing these questions requires the ability to think analytically about data and statistics. This course will provide an introduction to causal inference, probability theory, and estimation. The focus of this course will be on hands-on data analysis and the practical application of basic statistical methods to real-world, relevant problems.

1 What Are the Broad Goals of the Course?

This course is a course on statistical theory, reasoning, and argument in the social sciences. The course has three goals, in terms of increasing difficulty. At the end of this course, you should

1. Be an intelligent and critical consumer of statistics, in the academic and popular literature.
2. Be able to implement standard statistical methods and interpret their output.
3. Be able to tailor these methods to a question of interest in your own research.

The course will have a focus on preparing you for a Junior Paper, Senior Thesis, or graduate-level work. Statistical theory is the language we will be using, but we will be emphasizing quality of communication, thought, and argument.

2 Who Should Take This Course?

POL 345/SOC 305 is one of many courses satisfying several Departments' requirements for quantitative reasoning. You should take this course if

1. You have not taken any other college-level Statistics courses.
2. You have an interest in political science, policy-making, economics, or another social science.
3. You expect your Junior Paper or Senior Thesis to involve some form of data analysis.
4. You are willing to spend considerable time outside of class in order to keep up with the material.

3 How Does This Course Compare to Other Statistics Courses?

POL 345/SOC 305 will be a very "hands-on" exploration of statistics and the concepts underlying data analysis. We will not spend much time answering mathematical questions. Instead, we will focus on analyzing data from different social sciences. Please do not mistake the light emphasis on technical rigor with an easy class. We will be learning how to analyze data through the use of a free statistical package, **R** . The problem sets will involve you analyzing and reporting on an actual data set, and will require you to integrate communicating, analytic thought, and your knowledge of the class material.

POL 345/SOC 305 is the first in a two-course, undergraduate sequence in applied statistics for the social sciences. The second course is POL 346. Upon completing POL 345/SOC 305 and POL 346, you may be interested in pursuing the University's Certificate in Statistics and Machine Learning (link)..

4 What Requirements are Satisfied by this Course? What Other Courses are Available?

Politics Majors: POL 345/SOC 305 satisfies the analytical requirement for the Politics Department (link). Other courses taught at the same level as POL 345/SOC 305 that satisfy the analytical requirement are POL 341, POL 346, POL 347, POL 451, ANT 300A, ANT 301A, ECO 202, ECO 302, ECO 312, ORF 245, PHI 201, SOC 404, WWS 200, and WWS 332. POL 245 and POL 250 do NOT satisfy the requirement. Among quantitative courses, ECO 302, ECO 312, and ORF 245 are offered this semester. Of the others, POL 350 and WWS 200 are generally offered in the Spring, and both are comparable but less intensive than POL 345/SOC 305.

Woodrow Wilson Majors: POL 345/SOC 305 satisfies the statistics requirement for admission to the Woodrow Wilson School. Other courses that satisfy this requirement are WWS 200 and POL 346. WWS 200 is generally offered in the Spring, and is comparable but less intensive than POL 345/SOC 305.

Sociology Majors: Juniors majoring in Sociology are required to take POL 345/SOC 305.

Distribution Requirements: POL 345/SOC 305 satisfies the quantitative reasoning requirement. ORF245/EGR 245, PSY 251, WWS 200 can be used to replace this course.

Certificates: POL 345/SOC 305 can be used towards the **Statistics and Machine Learning** certificate. POL 345/SOC 305 counts as a course from the *Foundations of Statistics* category for the certificate. POL 346 counts towards the *Applied Statistics* category. Please see the Center's website for details (link).

POL 345/SOC 305 may also be used towards the **Program in Political Economy** certificate. Please see the Program's website for details (link).

5 Some Tips for Success

Below are some tips for succeeding in this course:

1. Attend lecture and precepts.
2. Complete the precept practice assignments.
3. *Do not fall behind.* This course is cumulative, and we will regularly build on previous weeks' material. If you start falling behind, see your preceptor *immediately*, in order to plan a course for catching up.
4. Start the problem sets the day we send them out.
5. Speak up in precept with any questions you have.

6 Data Analysis Using R

This course will make regular use of the open source statistical package, **R** (http://www.r-project.org). The package is free, and is quickly becoming an industry and academic standard. From the New York Times, "Data Analysts Captivated by R's Power," NYT, 01/07/2009:

`To some people R is just the 18th letter of the alphabet. To others, it's the rating on racy movies, a measure of an attic insulation or what pirates in movies say.`

`R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca partly because data mining has entered a golden age, 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.`

7 Statistics Now and Later in Your Life

You may find immediate use for Statistics in other assignments, including your Junior Papers and Senior Theses. As our world grows more and more data-intensive, the demand for people with a working knowledge of statistics has only increased. From "The Age of Big Data," NYT, 02/11/2012:

`GOOD with numbers? Fascinated by data? The sound you hear is opportunity knocking... A report last year by the McKinsey Global Institute, the research arm of the consulting firm, projected that the United States needs 140,000 to 190,000 more workers with ''deep analytical'' expertise and 1.5 million more data-literate managers, whether retrained or hired.`

8 Textbook

Required:

- Imai, Kosuke. *Quantitative Social Science: Introduction.* Princeton University Press.

Optional:

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

9 Course Components

9.1 Precept Material (10 %)

There will be ten precept assignments. These assignments will introduce you to commands in **R** . These assignments will have two components. The first point is for a preliminary exercise, to be done before precept. The second point will be for your precept assignment and will be due on Blackboard at the end of precept.

When submitting precept assignments, please adhere to the following guidelines:

1. Title your **R** script `xxxPreceptX.R` where `xxx` is your Net ID and `X` is the precept number. For example, if Marc were submitting the script from Precept 3 to Blackboard, he would name my **R** script `ratkovicPrecept3.R`.

2. Head your script as follows:

```
## Marc Ratkovic
## Precept 3
## Preceptor Name
## September 11, 2017
```

9.2 Problem Sets (30 %)

You will have three problem sets due through the semester. You will have roughly one week to finish each assignment, which will be distributed via Blackboard. You will be required to produce the document using Rmarkdown. An electronic copy of your Rmarkdown file and the resulting PDF must be submitted to Blackboard.

Please use the same protocol for naming your files, `xxxProblemSetX.Rmd` and `xxxProblemSetX.pdf`.

The problem sets are "open book," which means that you can use you book, other books, things on the Internet etc. However, the problem set collaboration policy (described below) is in effect.

9.3 Quizzes (20 %)

You will have two in-class, closed-book quizzes. These quizzes will be used to test your understanding of concepts and ideas introduced during lecture.

9.4 Final (40 %)

You will be given a take-home final in January. The final will resemble a problem set. The final will be distributed Wednesday, January 17 at 9am, and it will be due Monday, January 22 at 5pm.

The final exam is "open book," which means that you can use you book, other books, things on the Internet etc. However, no collaboration at all is allowed on the final exam.

9.5 Late Policy

Any QSS assignment handed in late, without permission of the preceptor, will not receive credit. Any problem set or final handed in late, and without permission from an instructor, will be penalized 30% a day. Any quiz that is missed, without permission of an instructor, will not receive credit.

10 Instructor and Preceptors

Instructors:

Marc Ratkovic; 035 Corwin Hall; Office Hours: 1:30-3:30 M, or see WASS; ratkovic@princeton.edu

Matthew Salganik; 145 Wallace Hall; Office hours: 4:30-5:30 M. mjs3@princeton.edu

Head Preceptor (Office hours and location are on piazza):

Xander Slaski aslaski@princeton.edu

Preceptors (Office hours and location are on piazza):

Naoki Egami negami@princeton.edu

Chris Felton cfelton@princeton.edu

Galileu Kim galileuk@princeton.edu
John Maldonado jam14@princeton.edu
Jose Maria Rodriquez Valadez jvaladez@princeton.edu
Daniela Urbina Julio djulio@princeton.edu
Elsa Voytas evoytas@princeton.edu

11 Resources

Aside from our regularly scheduled office hours, we offer several additional sets of resources to help you through this course.

1. **McGraw Tutors** In previous semesters, McGraw Study Halls and Tutors have provided crucial assistance to many struggling students.
2. **Review Sessions** We will offer a mid-semester review session, in order to help students who are falling behind. We will also offer a final review session. Occasional problem set sessions will be offered as time and resources allow.
3. **R COMPASS Workshops** We are pleased to use a wonderful resource designed for this course, COMPASS workshops. These workshops provide valuable education, practice, and extensions of the **R** language. The workshops are not required but are highly recommended. They are fantastic. The sessions will be from 7:30–9:00 PM in Room 307 of McGraw in Frist. The workshops are open to the broader Princeton community, so please help spread the word!
4. **R Programming Drop-In Hours:** The drop-in sessions are designed to give you a chance to ask specific questions about issues you are having with the **R** programming language. They will be staffed by students who have taken and done well in multiple courses using **R** . Unless otherwise announced, **R** programming drop-in hours will be available on Sundays and Mondays from 7:30pm to 10:30pm at the McGraw Center in the Frist Campus Center.
5. **Piazza:** We will host a Piazza forum for this course. Substantive questions will be addressed here, with both the instructors and preceptors managing the discussion.

12 Problem Set Collaboration

This language is adapted from the policies for COS 126 (link).

Programming is an individual creative process much like composition. You must reach your own understanding of the problem and discover a path to its solution. During this time, discussions with other people 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. If you have a question about how to use some feature of **R** , you can certainly ask your friends or the teaching assistants, but specific questions about code you have written must be treated more carefully. For each assignment, you must specifically describe in your **R** file, whatever help (if any) that you received from others and tell us the names of any individuals with whom you collaborated. This includes help from friends, classmates, lab TAs, and course staff members.

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. This includes adapting solutions or partial solutions to assignments from any offering of this course or any other course. 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 or make your source files available 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. Not only does plagiarized code quickly identify itself as part of the grading process, but also we can turn to software packages for automatic help.

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. Violators will be referred to the Committee on Discipline for review; if found guilty, you will receive an F as a course grade plus whatever disciplinary action the Committee imposes.

In regards to this class, we give the following guidelines. First, we understand precept materials and assignments are done collaboratively and in groups. So, for precept assignments, you should not cut and paste someone else's code. Second, problem sets are a more substantial part of your grade, so you will not be able to look at others' code during the problem sets. You can discuss the problem set questions, though. Third, during the final, there is no looking at each others' work or discussion at all–the final needs to wholly represent your individual work. Please come to us with any specific questions.

13 Regrade policy

If you believe there was a mistake with a grade you received for an assignment, you must submit, in writing (email), a detailed and clearly stated argument for what you believe is incorrect and why. You should also attach a scanned copy of the entire graded assignment (if you do not have access to a scanner, let us know). This must be submitted to Professor Ratkovic and Professor Salganik no later than the beginning of class one week after the assignment was returned. For example, if the assignment were returned to the class on Wednesday, your re-grade request would have to be submitted before the start of class on the next Wednesday. Requests for a re-grade after this time will not be accepted.

If you request a re-grade within the appropriate timeframe, a written response will be provided within one week of your request. This re-grade and written response is final. Please note that a request for a re-grade of a specific problem may result in a re-grade of the entire assignment. Therefore, a re-grade request may result in an increase or decrease of your overall score for the assignment. This policy is modeled after the policy of Rocco Servidio.

14 Schedule

14.1 Weekly Schedule

Below is a normal week in POL 345/SOC 305:

- 1:30–3:30 M: Office Hours (Ratkovic, 035 Corwin)
- 4:30–5:30 M: Office Hours (Salganik, 145 Wallace)

- 3:30–4:20 MW: Lecture
- 9 am F: Precept exercise for *next* week posted
- 9 am F: Answers for *current* precept exercise posted

14.2 Course Schedule

Below are the dates of problem sets and quizzes in POL 345/SOC 305:

- Problem Set 1: posted F 9/29 due W 10/11
- Quiz 1: W 10/18
- Problem Set 2: posted W 11/8; due W 11/15
- Quiz 2: M 11/20
- Problem Set 3: posted W 12/6; due W 12/13
- Final: Distributed Wednesday, January 17 at 9am, Due Monday, January 22 at 5pm.

15 Week-by-Week Schedule

Dates	Lecture	Readings	Precept	Other
Week 2: 9/18 – 9/22	Experiments Observational Studies	QSS 2.1-4 (FPP 1) QSS 2.5-6 (FPP 2)	Installing R	
Week 3: 9/25 – 9/29	Causality: Potential Outcomes Density Plots	The Colbert Bump QSS 3.1–3.5 (FPP 3)	Summarizing Univariate Data	**PSet 1** **Out: F 9/29**
Week 4: 10/2 – 10/6	Central Tendency and Variability The Normal Approximation	QSS 3.6 – 3.8 (FPP 4 , 6) QSS 3.6 – 3.8 (FPP 5)	Summarizing Bivariate Data; FPP 7	**PSet 1** **In: W 10/11**
Week 5: 10/9 – 10/13	Correlation and Causation Regression 1	QSS 4.1– 4.2 (FPP 8 – 9) QSS 4.2 (FPP 10 – 11)	Summary Statistics	
Week 6: 10/16 – 10/20	Regression 2 — **Quiz 1** —	QSS 4.2. – 4.3 (FPP 12)	Conditional Statements	**Review Session**
Week 7: 10/23 – 10/27	Expectation and Variance Text and Network Data	QSS 6.3 (FPP 16 – 17) QSS 5.1–5.2	**None – Midterm Week**	
10/30 – 11/3		— **Fall Break** —		
Week 8: 11/6 – 11/10	The Central Limit Theorem Confidence Intervals	QSS 6.4 (FPP 18) QSS 7.1 (FPP 21, 23)	Loops 1	**Pset 2** **Out: W 11/8**
Week 9: 11/13 – 11/17	Hypothesis Testing Statistical Significance	QSS 7.2 (FPP 24 – 25) QSS 7.2 (FPP 26)	Loops 2	**Pset 2** **Due: W 11/15**
Week 10: 11/20 – 11/24	— **Quiz 2** — — **No Class** —			
Week 11: 11/27 – 12/1	Inference with Regression Issues with Regression	QSS 7.1 (FPP 27) QSS 7.3 (FPP 29)	Regression 1	
Week 12: 12/4 – 12/8	More with Multivariate Regression Regression with Causal Inferenc	Instructor Provided Instructor Provided	Regression 2	**Pset 3** **Out: W 12/6**
Week 13: 12/11 – 12/15	Regression and Causal Inference II Student-Driven	Instructor Provided		**Pset 3** **Due: W 12/13**

Note: **QSS** is Imai's *Quantitative Social Science* and is required reading. **FPP** denotes supplemental readings from Freedman, Pisani, and Purves.