POL 345/SOC 305/SPI 211:
Introduction to Quantitative Social Science

Fall 2021

Marc Ratkovic (Politics)

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 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 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. Upon completing POL 345, you may be interested in pursuing the University's Certificate in Statistics and Machine Learning.

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

All Majors

POL 345 is only offered this Fall.

Politics Majors

POL 345 satisfies the analytical requirement for the Politics Department. Other courses taught at the same level as POL 345 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, SPI 200, and SPI 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 SPI 200 are generally offered in the Spring, and both are comparable but less intensive than POL 345.

Policy School Majors

POL 345 satisfies the statistics requirement for admission to the School of Policy and International Affairs. Other courses that satisfy this requirement are SPI 200 and POL 346. SPI 200 is generally offered in the Spring, and is comparable but less intensive than POL 345.

Sociology Majors

Juniors majoring in Sociology are required to take POL 345.

Economics Majors

POL 345 along with POL 346 satisfies the statistics requirement for Economics majors. Please note that POL 346 is only offered in the Spring and that POL 345 and POL 346 should not be taken concurrently.

Distribution Requirements

POL 345 satisfies the quantitative reasoning requirement. ORF245/EGR 245, PSY 251, SPI 200 can be used to replace this course.

Certificates

POL 345 can be used towards the **Statistics and Machine Learning** certificate. POL 345 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.

POL 345 may also be used towards the **Program in Political Economy** certificate. Please see the Program's website for details.

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. Utilize McGraw.
4. *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.
5. Start the problem sets the day we send them out.
6. Speak up in precept with any questions you have.

6 Hopefully Not Necessary: Adjustments for COVID

Hopefully we are back to normal and genuinely post-shutdown. If things arise, please:

- Be patient and flexible; we will do the same. If something is not working, tell us, and we will return the favor.
- Due dates will be extended as necessary, such that testing positive for COVID will NOT impact your grade.
- If either an illness (of you or a family member) or civic engagement requires some flexibility on an assignment, be in touch as early as possible.

7 Software

We will be relying on three different software programs:

- **CampusWire**: We will use the website CampusWire to post lecture slides, announcements, and communicate with you. You will receive an invitation to CampusWire after the first lecture.
- R and **RStudio**: You will find a set of instructions for installing software for this course at https://github.com/ratkovic/pol345.student. You will also find instructions on how to download and access our weekly assignments.
- **DataCamp**: You will be enrolled in DataCamp and 5% of your course grade will rely on completing two DataCamp courses, details to follow. DataCamp can be accessed here: https://www.datacamp.co

8 Textbook

Optional

- Imai, Kosuke. (2018). *Quantitative Social Science: An Introduction*. Princeton University Press.
- 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 Handouts (5%)

There will be eight 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. This component will be graded. It normally takes some time for everyone to get the proper routines and rhythm in for this section, so we will be flexible the first two weeks.

Problem Sets ($2 \times 20\% = 40\%$)

You will have three problem sets due through the semester. You will have roughly three days to finish each assignment. 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 Canvas.

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

Quizzes ($2 \times 15 = 30\%$)

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.

9.2 DataCamp (5%)

You will be enrolled in DataCamp, an online resource for learning coding. You will be expected to complete two courses from DataCamp during this semester, each taking about five hours. This will be practice designed to augment the course material. You will have the option for one point extra credit on your course grade for completing an additional course. These will be due the Fridays before Fall Break and the end of the Semester. Details will follow in the coming week.

Final (20%)

You will be given an in-class Final Examination during the time assigned by the Registrar. The final exam will resemble a quiz, but will be cumulative.

Late Policy

Any weekly 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

Instructor

Marc Ratkovic

Office: 306 Fisher Hall
Email: ratkovic@princeton.edu
Office Hours: TBA

Head Preceptor

Tolgahan Dilgin

Office: TBD
Email: dilgin@princeton.edu
Office Hours: TBD

Preceptors

- Gregory Amusugamusu@princeton.edu
- Anna Houk ahouk@princeton.edu
- Di Jia Su dsu@princeton.edu
- William Wen wwen@princeton.edu

Preceptors office hours will be posted on Campuswire.

11 Problem Set Collaboration

This language is adapted from the policies for COS 126.

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.

12 Regrade Policy

Regrade requests fall into two categories. The first involves a simple mis-tabulation of your grade, where we did not properly add up your points. Just confirm with your preceptor, cc the Professor in an e-mail, and we will make the change.

For more substantive concerns, where you believe there was a mistake with a grade you received for an assignment, you must fill out and submit a regrade request through our Google form. 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 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.

All grades prior to Dean's Date will be frozen after Dean's Date. Please check your grades prior to Dean's Date.

13 Schedule

Weekly Schedule

Below is a normal week in POL 345:

- Precepts run on a Monday-Friday cycle
- 11:00–11:50 TR: Lecture

Course Schedule

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

- Problem Set 1: posted M 9/27 due R 9/30
- Quiz 1: R 10/6
- Problem Set 2: posted M 11/29; due R 12/2
- Quiz 2: R 11/17
- Final: As given by Registrar

14 Week-by-Week Schedule

Dates	Lecture	Readings	Precept	Other
Week 1: 9/5 – 9/9	Introduction Experiments	QSS 2.1-4 (FPP 1)		
Week 2: 9/12 – 9/16	Observational Studies Empirical Studies of Race	QSS 2.5-6 (FPP 2)	Installing R	
Week 3: 9/19 – 9/23	Causality: Potential Outcomes Density Plots	The Colbert Bump QSS 3.1–3.5 (FPP 3)	Summarizing Univariate Data	
Week 3: 9/26 – 9/30	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	**Pset 1 Out: T 9/27** **In: F 9/30**
Week 4: 10/3 – 10/7	Correlation and Causation **— Quiz 1 —**	QSS 4.1– 4.2 (FPP 8 – 9)	Summary Statistics	
Week 5: 10/10 – 10/14	Regression 1 Regression 2	QSS 4.2 (FPP 10 – 11) QSS 4.2. – 4.3 (FPP 12)	**None – Midterm Week**	**Datacamp Due Fri**
Week 6: 10/17 – 10/21	**— Fall Break —**			
Week 7: 10/24 – 10/28	Text and Network Data Expectation and Variance	QSS 5.1–5.2 QSS 6.3 (FPP 16 – 17)	Conditional Statements	
Week 8: 10/31 – 11/4	The Central Limit Theorem Confidence Intervals	QSS 6.4 (FPP 18) QSS 7.1 (FPP 21, 23)	Loops 1	
Week 9: 11/7 – 11/11	Hypothesis Testing Statistical Significance	QSS 7.2 (FPP 24, 25) QSS 7.2 (FPP 6)	Loops 2	
Week 10: 11/14 – 11/18	Difference-in-Means **— Quiz 2 —**	QSS 7.1 (FPP 27)	Regression 2	
Week 11: 11/21 – 11/25	**— Thanksgiving —**			
Week 12: 11/28 – 12/2	More on Significance Fixed-Effects Regression	QSS 7.3 (FPP 29) Instructor Provided	Regression 1	**Pset 3 Out: T 11/29** **In: F 12/2**
Week 13: 12/5 – 12/9	Multivariate Regression Summary and Conclusion	Instructor Provided	Regression 2	**Datacamp Due Fri**

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