POS 6737: Political Data Analysis

University of Florida

Syllabus: Autumn 2020

Instructor:	Professor Drew Rosenberg	Class location:	The World Wide Web
Office:	210 Anderson Hall	Class time:	Th, 08:30–11:30
Email:	andrewrosenberg@ufl.edu	Office hours:	W, 13.00–16.00, by appointment.
TA:	Helen Jang	Recitation loc:	The World Wide Web
Office:	330 Anderson Hall	Recitation time:	M, 13.00–14.00
Email:	hrjang52@ufl.edu	Office hours:	M, 14.00–15.00, by appointment

Schedule a meeting with me: https://calendly.com/asrosenberg.
Schedule a meeting with Helen: https://calendly.com/hrjang52.

Course Description

Political Science 6737 is an introduction to probability and statistics targeted toward Political Science PhD students. A primary purpose of the course is to build a strong foundation for regression and generalized linear models, which will be studied in great depth in Political Science 6747 and beyond. To accomplish this goal, we will study the basics of probability theory, properties of random variables, asymptotic approximations, methods for developing and evaluating statistical estimators, and hypothesis testing. In addition, the course will provide a hands-on introduction to statistical computing.

The course will be taught as a combination of lectures by the instructor and practical exercises at the computer.

Course Goals:

1. Students will gain experience using and understanding the basic foundations of all quantitative social science. This experience will allow them to conduct their own projects, progress to more advanced courses on regression and causal inference, and evaluate published work.

   Expected Learning Outcomes:

- Learn basic probability axioms and rules and the moments of discrete and continuous random variables as well as be familiar with common named discrete and continous random variables.
- Use summary statistics and graphs to conduct exploratory data analysis.
- Calculate probabilities, and derive the marginal and conditional distributions of bivariate random variables.
- Conduct hypothesis tests.
- Understand the fundamental problem of causal inference.

2. Students will develop modern statistical computing skills that will allow them to conduct data analysis at the PhD level.

   Expected Learning Outcomes:

   - Read data into `R` and do basic data cleaning.
   - Learn good coding etiquette.
   - Download packages, write functions, and debug code.
   - Run a basic simulation study to solve probability exercises.
   - Run and interpret simple regressions.
   - Use R Markdown to present research findings in a beautiful format.

Course Materials:

Textbook

We will use three textbooks in this course. They are all very good and should be much more affordable than the average textbook (one is free if you use it online!).

> Sean Gailmard. 2014. *Statistical Modeling and Inference for Social Science*. New York, NY: Cambridge University Press.
>
> Kosuke Imai. 2018. *Quantitative Social Science: An Introduction*. Princeton, NJ: Princeton University Press.
>
> Hadley Wickham and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol, CA: O'Reilly Media, Inc. (https://r4ds.had.co.nz/).

There are countless introductory textbooks on probability and statistics. You may wish to consult any of these for this class, but they are by no means required:

> George Casella and Roger L. Berger. 2002. *Statistical Inference*. Pacific Grove, CA: Duxbury Press.

Morris H. DeGroot and Mark J. Schervish. 2012. *Probability and Statistics*. New York, NY: Addison-Wesley.

Paul M. Kellstedt and Guy D. Whitten. 2018. *The Fundamentals of Political Science Research*. New York, NY: Cambridge University Press.

Richard J. Larsen and Morris L. Marx. 2011. *An Introduction to Mathematical Statistics and its Applications*. Englewood Cliffs, NJ: Prentice-Hall.

Statistical Software

NB: This is the section that will seem overwhelming at first. Please bear with me; I promise that the computer skills you learn now will pay massive dividends later.

Proficiency in political data analysis requires one to analyze political data! With this end in mind, we will be playing with data from the very first week of this course. We will use the open source and free statistical software **R** in our course: http://www.r-project.org/.

What is R and why use it?

- Widely-used in academia and industries
- Open-source and free
- Power and flexibility
- Graphical capabilities
- Learning R = learning basic programming
- When you accomplish things, it will feel awesome

The *New York Times* described 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."[1]

You might want to consider Microsoft R Open https://mran.revolutionanalytics.com/open/. This is a version of R that automatically leverages multiple cores.

I recommend that you also install the free RStudio interface (`http://www.rstudio.com/`), which makes working with **R** a little easier. The first problem set will walk you through the process of installing **R** / R Open and RStudio on your own computer/laptop.

[1] Vance, Ashlee. 2009. "Data Analysts Captivated by R 's Power." New York Times, January 6.

However, if the installation stuff overwhelms you, first take a deep breath. R is free, but quite annoying to install. To make things easier at first, you can use RStudio cloud (https://rstudio.cloud/), which lets you run a full version of R in an internet browser. This means you don't have to install anything to get going with this course.

This is a useful guide to using **R** that will come in handy throughout the semester:

- SimpleR – Using R for Introductory Statistics.

There are plenty of other free resources for **R** to be found on the internet. Google will get you very far in many instances. I want to recommend in particular the new R package swirl. It This is an add-on to R that contains a number of self-guided lessons that show you how to do basic data and regression analysis in R. This package is a great complement to our own exercises, and it is what is used in the Imai (2018) book.

R Markdown

Markdown is a simple mark-up language that allows you to use simple syntax to make beautiful documents. R Markdown is a special version of Markdown that allows you to embed data analysis into text documents. This might not seem cool, but it is: it allows you to create reproducible analyses that anyone can replicate. When you use R Markdown in RStudio, you can output PDF, HTML, and Microsoft World documents. It's pretty cool. Here is an introduction: https://rmarkdown.rstudio.com/lesson-1.html. We will also spend the first problem set getting you up and running!

Assignments:

- Weekly Problem Sets (60%): There will be problem sets almost every week. The problem sets typically consist of a set of theoretical and conceptual questions and a hands-on data analysis portion. The purpose of the problem sets is to give you practice doing data analysis and engaging with each week's concepts. Students are encouraged to work on the problem sets in small groups, i.e. you should discuss possible answers with your fellow students. It is good practice to first try to develop answers on your own and then meet in a group setting to discuss potential difficulties. While group discussion and work is explicitly encouraged, you are required to write and hand in your own computer code and final write-up of the answers. DO NOT simply copy computer code or answers from your classmates. Write-ups have to be provided in a well-formatted, electronic format (e.g. R Markdown). I will not accept any late homework assignments. The write-up and code have to be submitted on Thursdays before class the following week. To accommodate your busy schedules, **I will drop your two lowest homework scores when calculating your final grades**. These will be pretty hard, so don't worry. See below for more reassurance.
- Problem Set Grading Procedure
    - ✓+ (40 points; 100%) Problem set is 100% complete. Every question was attempted and answered, and all are correct. Document is clean and easy to follow.

Code is well-written. Work is exceptional. *These are rare.*

- ✓ (37 points; 93%) Problem set is 75—99% complete and most answers are correct. *This is the expected level of performance.*
- ✓− (25 points; ∼ 63%) Problem set is less than 75% complete and/or most answers are incorrect. This indicates that you need to improve next *and make an appointment to come talk to me.* This is not an indictment of your ability to do well in this course!

- Take-Home Midterm Exam (15%): The midterm will be a big problem set, but *you are not allowed to collaborate.*
- Take-home final (due December 10) (25%): The take-home final will be similar to the take-home midterm. I will distribute the exam on Canvas on December 3rd and you will have one week to complete it. The exam will be very similar to prior problem sets, but it will be cumulative. Again, you must complete the exam without the help of other students. Write-ups and computer code are due on December 10th at 11:59pm.
- Attendance Policy: We will meet once a week during the semester. You can expect me to be prepared, give the lecture and answer questions. When you come to class, please also be prepared. I will not require attendance, but class is a resource to *you*. The classroom is a great place to exchange ideas, meet your classmates, and ask questions. Regular attendance is also encouraged because lectures and practical sessions are tightly linked to weekly assignments, the midterm and final. If you do not attend regularly, it will be difficult to pass the class. Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies that can be found at: https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx. *I understand that virtual lectures/classes are not fun. I will do my best to make this an enjoyable and rewarding semester.*
- Summary of most important dates:
  - Oct 22: Midterm
  - Dec 3: Take-home final made available
  - Dec 10: Take-home final due
- Grading Schedule:

		A	93–100%	A−	90–92%
B+	87–89%	B	84–86%	B−	80–83%
C+	77–79%	C	74–76%	C−	70–73%
D+	67–69%	D	64–66%	D−	60–63%
E	$< 60\%$				

Policies and procedures

Communication and logistics: Microsoft Teams

Microsoft Teams is an online collaboration platform (like Slack) that we will use to communicate throughout the semester. I will invite you to a "team" and there will be various "channels" we will use to share readings, have discussion, and (perhaps) host online meetings. I am still figuring out the best way to use this platform, but it is far superior to constantly spamming each other with email.

Communication and logistics: Email

You can also email me if you want.

Office Hours

I hold three hours of office hours per week, but you may arrange a meeting outside of those hours if you are unavailable during this time. Please make use of office hours, as that is the time I allocate to be 100% available to you. If you have any questions or are having difficulty completing course requirements, please come see me as soon as possible. *Use the Calendly link at the top of this syllabus and on my website to book a meeting.*

Academic misconduct

UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code." On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code (http://www.dso.ufl.edu/sccr/process/student-conduct-honorcode/) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Disability services

Students with disabilities requesting accommodations should first register with the Disability Resource Center (352-392-8565, www.dso.ufl.edu/drc/) by providing appropriate documentation. Once registered, students will receive an accommodation letter which must be presented to the instructor when requesting accommodation. Students with disabilities should follow this procedure as early as possible in the semester.

Health and Wellness Resources

- U Matter, We Care: If you or a friend is in distress, please contact umatter@ufl.edu or 352-392- 1575 so that a team member can reach out.
- Counseling and Wellness Center: https://counseling.ufl.edu/, 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.
- Sexual Assault Recovery Services (SARS) Student Health Care Center, 392-1161. University Police Department, 392-1111 (or 9-1-1 for emergencies). http://www.police.ufl.edu

Online Course Evaluations

Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations at https://evaluations.ufl.edu. Evaluations are typically open during the last two or three weeks of the semester, but students will be given specific times when they are open. Summary results of these assessments are available to students at: https://evaluations.ufl.edu/results/.

Course Overview and Schedule:

Week 1 (Sept 3): Introduction and Descriptive Statistics I

Introduction to the class, general requirements, and logistics.

- Core readings:
  - M.A. Schwartz. 2011. "The Importance of Stupidity in Scientific Research." *Seismological Research Letters* 82 (1): 3–4. doi:10.1785/gssrl.82.1.3. https://doi.org/10.1785/gssrl.82.1.3
  - Gailmard, Ch. 1–2
  - Wickham and Grolemund, 1–3
  - Imai, 10–28
- Assignments:
  - Download PS-1, due on Thursday Aug 10 at 830a.

Week 2 (Sept 10): Descriptive Statistics II

- Core readings:
  - Imai, 32–46, 63–69, Ch. 3
  - Wickham and Grolemund, Ch.4–6
- Assignments:
  - Download PS-2, due on Thursday Sep 17.

Week 3 (Sept 17): DGPs and Probability I

- Core readings:
  - Gailmard, Ch. 3–4.2
  - Wickham and Grolemund, Ch.7–8
- Assignments:
  - Download PS-3, due on Thursday Sep 24.

Week 4 (Sep 24): DGPs and Probability II

- Core readings:
  - Gailmard, Ch. 4.2–4.7
  - Imai, Ch. 4

  - Wickham and Grolemund, Ch.10–12
- Assignments:
  - Download PS-4, due on Thursday Oct 1.

Week 5 (Oct 1): Expectation and Moments

- Core readings:
  - Gailmard, Ch. 5
- Assignments:
  - Download PS-5, due on Thursday Oct 8.

Week 6 (Oct 8): Probability and Models

- Core readings:
  - Gailmard, Ch. 6
- Assignments:
  - Download PS-6, due on Thursday Oct 15.

Week 7 (Oct 15): Sampling Distributions

- Core readings:
  - Gailmard, Ch. 7
- Assignments:
  - Download PS-7, due on Thursday Oct 22.

Week 8 (Oct 22): MIDTERM

This speaks for itself! It's a closed-book exam and everything will be ok!

Week 9 (Oct 29): Hypothesis Testing I

- Core readings:
  - Gailmard, Ch. 8
  - Imai, Ch. 6
- Assignments:
  - Download PS-8, due on Thursday Nov 5.

Week 10 (Nov 5): Hypothesis Testing II

- Core readings:
  - Gailmard, Ch. 8
  - Check this out.
  - *Economist* video on problems with false positives and negatives in scientific research
- Assignments:
  - Download PS-9 from Carmen, due on Thursday Nov 12.

Week 11 (Nov 12): Estimation I

- Core readings:
  - Gailmard, Ch. 9
- Assignments:
  - Download PS-10 from Carmen, due on Thursday Nov 19.

Week 12 (Nov 19): Estimation II

- Core readings:
  - Gailmard, Ch. 9
  - Imai, Ch. 7
- Assignments:
  - Download PS-11, due on Thursday Dec 3.

Week 13 (Nov 26): THANKSGIVING BREAK

Enjoy!

Week 13 (Dec 3): Final Wrap-up

- Core readings:
  - -
- Assignments:
  - Final take-home exam will be posted on Dec 3, submissions are due on Dec 10.