POS 6737: Political Data Analysis

University of Florida

Syllabus: Autumn 2022

Instructor:	Dr. Drew Rosenberg	Class location:	Matherly Hall 0012
Office:	210 Anderson Hall	Class time:	Tu, 08.30–11.30
Email:	andrewrosenberg@ufl.edu	Office hours:	W, 12.00–15.00.
TA:	Muharrem Bagriyanik	Recitation loc:	Anderson Hall 0134
Office:	330 Anderson Hall	Recitation time:	F, 16.00–18.00
Email:	mbagriyanik@ufl.edu	Office hours:	M, 12.00 - 15.00

Office Hours Links
Rosenberg: https://calendly.com/asrosenberg.
Bayriyanik: https://calendly.com/mbagriyanik/pos6737-ta-mu-muharrem-office-hours.

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

I create very detailed lecture slides and make them available each week. You should think about these slides and the lectures as the weekly course readings. I will hand out selections from different textbooks to compliment the weekly concepts. This approach might seems weird to some of you, but there are thousands of possible books for this course, and you probably won't like most of them. If you are someone who learns best from reading textbooks and are having difficulty with the slides, then please come talk to me and I will find you a reading that will suit the topic.

To be sure, it is often helpful to have a general reference book. Here is my current book of choice:

> Sean Gailmard. *Statistical Modeling and Inference for Social Science*. New York: Cambridge University Press, 2014.

We may also use selections from another more advanced textbook, but I will provide them to you when necessary. This book tends to present some ideas better than the Gailmard book.

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

Third, this is a useful guide to using **R** that we will use. I will assign modules from this guide:

fasteR: Fast Lane to Learning R.

Finally, this is one of the best books available on **R** programming, and I would use it if I taught a semester-long data science course.

Jared P. Lander. *R for Everyone: Advanced Analytics and Graphics*. 2nd ed. Addison-Wesley, 2017.

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

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

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

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

Hadley Wickham and Garrett Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol: O'Reilly Media, Inc., 2016 (FREE at `https://r4ds.had.co.nz/`).

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

- And more!

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.

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.

There are plenty of 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.

R Markdown

Markdown is a basic 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!

LaTeX

Although I encourage you to use R Markdown, you can also familiarize yourself with the LaTeX typesetting environment. LaTeXdocuments look a bit like scary computer code, but they allow you to make beautiful documents with perfect citations. However, R Markdown actually uses LaTeX"under the hood," so they are equivalent. In addition, https://www.overleaf.com/ is an online LaTeXeditor that makes typesetting and collaboration quite easy. Please come talk to me more if you need help.

Overleaf's introduction is very good and there are many other resources online:

https://www.overleaf.com/learn/latex/Free_online_introduction_to_LaTeX_(part_1)

Assignments:

- Weekly Problem Sets (50%): 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!
- Weekly Check-ins (5%): After class each week you will be responsible for letting me know what was confusing from lecture and/or what you wanted to hear more about. You will submit this feedback on my website: https://www.asrosenberg.com/political-data-analysis.
- Midterm Exam (15%): There will be an in-class, closed-note midterm. Please note that this exam represents a small portion of your grade.
- Take-home final (due December 13) (30%): I will distribute the exam on Canvas on December 2rd 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 13th 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 18: Midterm
  - Dec 6: Take-home final made available
  - Dec 13: 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	$< 60\%$				

How to Succeed in this Class

Statistics classes are really hard. Most people find this course challenging and we cover a lot of stuff. But you can do it; you have nothing to fear! All successful people struggle with this class. It's not just you.

Your responsibility is to work hard, do your best, and communicate with me. You cannot learn this stuff if you don't put in the time. I can't help you if you don't turn in assignments. I can't help you if I don't know there is a problem. Here are some more specific resources.

1. Lectures. I will lecture during most of our time together. I will speak over the slides and often provide off-the-cuff examples and explanations. The goal is to help you understand the material, so please let me know in class how I can help. Participate! Ask me stuff!
2. Weekly Check-ins. See above. I can't help you unless you let me know what you need help with.
3. Lecture slides. The slides will be available on Canvas after class. You have to read them and ask questions about them if there are things you don't understand.
4. Example Code. I will provide example code and in-class exercises to help you learn R. Please come to class prepared to participate.
5. Recitation. Our TA will host an optional recitation to help reinforce the weekly concepts. You should go to them.
6. Office hours. Don't be afraid to come talk to me about broader conceptual issues and specific things you don't understand.
7. The Internet. There are infinite free resources online that will help with the conceptual and computing aspects of the course.

8. Your classmates. You will learn more from each other than from me. Form a study group!

If you have any questions or can think of anything else that would be useful for you, then please come talk to me. To reiterate: if you work hard and put in the time, then I can provide help that meets your needs.

Policies and procedures

Communication and logistics: Email

Please email me with any pressing questions or concerns. However, do not expect immediate replies. I often do not check my email on the weekends or in the evening.

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.*

Computers

I will provide example computer code and in-class exercises to help you learn R. I encourage you to bring your computer to class each week in order to participate. Please reach out to me as soon as possible if you have trouble downloading R on your computer.

Collaboration Policy

I encourage students to work together on the problem sets, but you must write your own solutions (this includes code). However, I *strongly* suggest that you try all the problems before consulting others. The midterm will be very difficult if you have no experience working on your own.

Assignment dispensation policy

If a student is unable to complete an assignment, they will be allowed to turn it in late only if the absence is due to a *documented* medical, family, or similar serious emergency, observance of religious holy days (which requires written notification to the instructor at least 14 days prior to the due date), or properly documented University-sponsored planned activities. *Incomplete assignments or exams in all other cases will result in a score of zero.* If you become aware that you will not be able to complete an assignment or final project ahead of time, please contact the instructor and seek permission for an extension as soon as possible.

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/.

In-Class Recording

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or

(3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A "class lecture" is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To "publish" means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

Course Overview and Schedule:

Week 1 (Aug 30): Introduction

Introduction to the class, general requirements, and logistics.

- R readings:
  - Lesson 1 of fastR.
- Assignments:
  - PS-1

Week 2 (Sept 6): Descriptive Statistics I

- R readings:
  - Lessons 2–3 of fastR.
- Assignments:
  - PS-2

Week 3 (Sept 13): Descriptive Statistics II

- R readings:
  - Lessons 4–7 of fastR.
- Assignments:
  - PS-3

Week 4 (Sep 20): DGPs and Probability I

- R readings:
  - Lessons 8–10 of fastR.
- Assignments:
  - PS-4

Week 5 (Sep 27): DGPs and Probability II

- R readings:
  - lessons 11–13 of fastR.
- Assignments:
  - PS-5

Week 6 (Oct 4): DGPs and Probability III

- R readings:
  - lessons 14–15 of fastR.
- Assignments:
  - PS-6

Week 7 (Oct 11): Moments I: Expectation

- R readings:
  - Lesson 16 of fastR.
- Assignments:
  - PS-7

Week 8 (Oct 18): MIDTERM

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

Week 9 (Oct 25): Moments II: Variance

- R readings:
  - Lesson 17 of fastR.
- Assignments:
  - PS-8

Week 10 (Nov 1): Probability and Models

- R readings:
  - Lesson 18 of fastR.
- Assignments:
  - PS-9

Week 11 (Nov 8): Sampling Distributions

- R readings:
  - Lesson 20 of fastR.
- Assignments:
  - PS-10

Week 12 (Nov 15): Hypothesis Testing I (Analytic)

- R readings:
  - Lesson 21 of fastR.
- Assignments:
  - PS-11

Week 13 (Nov 22): Hypothesis Testing II (Computational)

- R readings:
  - Lesson 24 of fastR.
- Assignments:
  - PS-12

Week 14 (Nov 29): Estimation I (Intervals)

- R readings:
  - Lesson 25 of fastR.
- Assignments:
  - PS-13

Week 15 (Dec 6): Estimation II (Points)

- R readings:
  - Lesson 26–27 of fastR. (you should finish the rest of the lessons in your free time!)
- Assignments:
  - Final take-home exam will be posted on Dec. 6 after class and submissions are due on Dec. 13.