# POS 6747: Topics in Political Methodology

## University of Florida

### Syllabus: Spring 2023


| | | | |
|---|---|---|---|
| Professor: | Dr. Drew Rosenberg | Class location: | 0013 Anderson Hall |
| Office: | 210 Anderson Hall | Class time: | M, 15:00–18:00 |
| Email: | andrewrosenberg@ufl.edu | Office hours: | W, 13:00–15:00 |

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


---

## Course Description

This course is a graduate-level introduction to statistical models, with a focus on ordinary least squares (OLS) regression and causal inference. The goal of this course is to give you the tools necessary to do high quality quantitative work. In other words, the purpose of this course is to transition you away from learning the brute facts of probability and statistics and toward answering your own questions. Conveniently, once you have the basic idea of statistical prediction, you can apply these principles beyond linear regression in many different ways and in many different contexts.

The course has three components. First, we will review key tools and concepts in mathematics, statistics, probability, and computing. This part sounds scary, but it will be straightforward if you took POS 6737 and are excited about data. Second, we will build linear models, use them to solve real problems, and evaluate their assumptions. Finally, we will introduce you to the basic principles of causal inference.

## Course Materials:

Class materials are available on Canvas.

### Readings

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. 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:

> Andrew Gelman, Jennifer Hill, and Aki Vehtari. *Regression and Other Stories*. New York: Cambridge University Press, 2020.

The Gelman, Hill, and Vehtari book has a specific ideological position on statistics: Bayesian inference *uber alles*. However, this book is brand new, and it is the best available that combines a modern approach to statistics with a rigorous treatment of computing. You can also use it in a MLE or causal inference course!

There are dozens of books on linear models and causal inference. Some are better than others. Here are several that you should consider consulting.

> Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press, 2008.
>
> Peter M. Aronow and Benjamin T. Miller. *Foundations of Agnostic Statistics*. New York: Cambridge University Press, 2019.
>
> Scott Cunningham. *Causal Inference: The Mixtape*. London: Yale University Press, 2021.
>
> John Fox. *Applied Regression Analysis and Generalized Linear Models*. New York: Sage Publications, 2015.
>
> Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press, 2006.
>
> Damodar N. Gujarati. *Basic Econometrics*. New York: McGraw-Hill, 2009.
>
> William H. Greene. *Econometric Analysis*. New York: Pearson, 2017.
>
> Gareth James et al. *An Introduction to Statistical Learning*. New York: Springer, 2013.
>
> Peter Kennedy. *A Guide to Econometrics*. New York: Wiley-Blackwell, 2008.
>
> Jan Kmenta. *Elements of Econometrics*. Ann Arbor: University of Michigan Press, 1997.
>
> Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach*. Boston: Cengage Learning, 2016.

There are also approximately 1 million books on calculus and linear algebra. Some of them are good, some of them are bad, and most of them go into more detail than you will need, even if you get super excited about political methodology. Here are two that you can draw from.

> Daniel Kleppner and Norman Ramsey. *Quick Calculus: A Self-Teaching Guide*. New York: John Wiley & Sons, 1985.
>
> William H. Moore and David A. Siegel. *A Mathematics Course for Political and Social Research*. Princeton: Princeton University Press, 2013.

Finally, this is a good guide for doing applied statistics in R.

> John Fox and Sanford Weisberg. *An R Companion to Applied Regression*. New York: Sage publications, 2018.

## Statistical Software

If you have taken POS 6737, you can disregard this section. 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

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]

I recommend that you also install the free RStudio interface (`http://www.rstudio.com/`), which makes working with **R** a little easier.

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.

## Making Pretty Documents: R Markdown & LaTeX

Appearance matters a lot in academia. You can think of academics as small business owners who sell their research in the marketplace of ideas. In the real marketplace, products that look nice tend to sell in spite of other failings. Apple products should come to mind. Accordingly, how your documents look matters a great deal in the academic marketplace. This

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

may seem facile, but I promise that any effort you put into making your papers look good will pay dividends.

To this end, I will encourage you to use one of two typesetting options in this class. The first option is the LaTeXtypesetting environment. LaTeXis nice because it produces pretty documents and it makes it much easier to produce nicely formatted homework assignments and research papers that include tables, graphs, and equations. I have used LaTeXto typeset everything I've written since my first year of graduate school. There are many resources online, and I would encourage you to google "latex + political science." Here is a good overview/introduction:

http://www.andyphilips.com/downloads/introduction%20to%20latex%20Philips.pdf

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.

Second, you may wish to use **rmarkdown** in RStudio. **rmarkdown** is based on markdown, a simple, plain text markup language. Many people find that it is easier than LaTeX, and you can embed **R** code and customize output just like you would in TeX. Rmarkdown is nice because it ensures that your work is *reproducible*, which will be a big topic in our course. You can even output to PDF, HTML, and Microsoft Word. I'm agnostic, I can support either, and I will provide a simple template for both. For an introduction, see http://rmarkdown.rstudio.com.

# Assignments:

I assess this course on the basis of two main components. The purpose of each component is to give you practice doing data analysis, to build up your practical skills, and to give you lots of low-stakes opportunities to figure out how you can improve. Rather than assign a large final paper, I divide the work evenly throughout the term and require you to do relatively smaller tasks more often.

- Problem Sets (50%): There will be six problem sets. Each problem set is meant to familiarize the student with essential concepts, how to *do* quantitative political science, and coding. Write-ups have to be provided in a well-formatted, electronic format (e.g. LaTeXor R Markdown). Computer code used for any data analysis has to be submitted as a supplement to the write-up.
- 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!

- Replication Project (45%): For this project, you will need to find a published political science paper that uses linear regression, replicate one table from it, and extend the analysis in some way. Beyond this technical requirement, I will ask you to break down the paper's research question, data sources, approach to causal identification, and research design to evaluate how the author went about doing the research. Don't be afraid! Here is the list of requirements:

  1. Identify a scholarly article to replicate. The article must use linear regression and you must be able to get the replication data. Upload a PDF of the article to Canvas along with a brief paragraph (200 words-ish) explaining your choice. (February 5)
  2. Turn in a short update that reports relevant descriptive statistics and replicates the main model of your analysis. (March 4)
  3. The final paper is due our last day of class (April 22).

- 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. I will circulate a Google form for this feedback.

- Summary of most important dates:

  - February 5: Finalize replication paper.
  - March 4: Replication project update.
  - April 22: Final replication project due.

# 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. Office hours. Don't be afraid to come talk to me about broader conceptual issues and specific things you don't understand.
5. The Internet. There are infinite free resources online that will help with the conceptual and computing aspects of the course.
6. 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 office hours on Wednesdays, 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.*

## 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/.

# Course Overview and Schedule:

**Week 01, 01/08:** Introduction to the class, general requirements, and logistics. Your first (?) regression. Where do data come from? Graphs.

**Week 02, 01/15:** MLK Day. No Class.

**Week 03, 01/22:** Review of probability and statistics concepts. Hypothesis testing. Statistical significance.

**Week 04, 01/29:** Regression basics. CEF. Best linear predictor. Interpretation.

**Week 05, 02/05:** Simple linear regression. Mechanics. Unbiasedness.

**Week 06, 02/12:** Inference. Hypothesis tests. Confidence intervals. Goodness-of-fit.

**Week 07, 02/19:** DAGs, Spuriousness, conditional interdependencies, "think before you regress."

**Week 08, 02/26:** Linear regression with two predictors. Multicollinearity. Categorical variables.

**Week 09, 03/04:** Post-treatment Bias, Collider Bias, Confounding, Backdoor Paths

**Week 10, 03/11:** Spring Break. No class.

**Week 11, 03/18:** Interactions. Omitted variables. Polynomials.

**Week 12, 03/25:** Multiple Regression. More hypothesis testing. F-test.

**Week 13, 04/01:** Diagnostics. Outliers. Influence. Leverage. Heteroskedasticity.

**Week 14, 04/08:** Potential Outcomes Model. Fundamental problem of causal inference. SUTVA.

**Week 15, 04/15:** Matching.

**Week 16, 04/22:** Repeated observations. Panel data. Fixed effects. Diff-in-diff.