POS 6747: Topics in Political Methodology

University of Florida

Syllabus: Spring 2020

Instructor:	Professor Drew Rosenberg	Class location:	0010 Matherly Hall
Office:	210 Anderson Hall	Class time:	Th, 15:00–18:00
Email:	andrewrosenberg@ufl.edu	Office hours:	W, 14.00–16.00
TA:	Elena Shatilova	Office hours:	TBD
Office:	230 Anderson Hall		
Email:	alenamore@ufl.edu		

Course Description

This course is a graduate-level introduction to statistical models, with a focus on ordinary least squares (OLS). We will spend a fair amount of time on this model, its assumptions, and violations thereof. In addition, a full presentation of the classic linear regression model requires some basic calculus and linear algebra, which may sound scary, but I promise you will be fine! We will also further develop our programming skills and use simulation to learn about what regression models actually do and when they break.

The goal of this course is to give you the tools necessary to do high quality quantitative work. To fully realize this goal, however, you should plan to continue your education in quantitative methods after this course. That is to say, this course will get you on your way, but it should also reveal the need for further training. This is completely normal: being a responsible practitioner of statistical methods is difficult and requires career-long learning and attention.

Course Materials:

Class materials are available on Canvas, as well as on Dropbox. You can access these files at the link below (you do not need Dropbox to access these files, but I would recommend it!).

Readings

We will use three textbooks in this course.

Andrew Gelman and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

Gareth James et al. 2013. *An Introduction to Statistical Learning*. New York, NY: Springer.

Jeffrey M. Wooldridge. 2016. *Introductory econometrics: A Modern Approach*. Boston, MA: Cengage Learning.

There are dozens of books on linear models. Some are better than others. Here are several that you should consider consulting. **Do not buy all of these**.

Damodar N. Gujarati. 2009. *Basic Econometrics*. New York, NY: McGraw-Hill.

Peter Kennedy. 2008. *A Guide to Econometrics*. New York, NY: Wiley-Blackwell.

William H. Greene. 2017. *Econometric Analysis*. New York, NY: Pearson.

Jan Kmenta. 1997. *Elements of Econometrics*. Ann Arbor, MI: University of Michigan Press.

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 we will need. Here are two that you can draw from.

Daniel Kleppner and Norman Ramsey. 1985. *Quick Calculus: A Self-Teaching Guide*. New York, NY: John Wiley & Sons.

William H. Moore and David A. Siegel. 2013. *A Mathematics Course for Political and Social Research*. Princeton, NJ: Princeton University Press.

Statistical Software

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]

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.

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

- SimpleR – Using R for Introductory Statistics.
- James E. Monogan. 2015. *Political Analysis Using R.* New York, NY: Springer.
- Hadley Wickham and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* Sebastopol, CA: O'Reilly Media, Inc.

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.

LaTeX

I recommend highly that you familiarize yourself with the LaTeX typesetting environment. It will make it much easier to produce nicely formatted homework assignments and research papers. Jamie Monogan's introduction is very good and there are many other resources online:

`http://spia.uga.edu/faculty_pages/monogan/latex.php`

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.

As a final alternative, 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. 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.

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

Assignments:

- Problem Sets (60%): There will be problem sets. The problem sets typically consist of a set of theoretical and conceptual questions and a hands-on data analysis portion. Each problem set is meant to familiarize the student with essential concepts, how to *do* quantitative political science, and coding. Students are encouraged to work on the problem sets in small groups, i.e. you should discuss possible answers and solution approaches 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. LaTeX). Computer code used for any data analysis has to be submitted as a supplement to the write-up. I will not accept any late homework assignments. The write-up and code have to be submitted on Thursdays before class in the following week.
- Replication Project (Apr 16) (40%): This project will require you to find a published political science paper that uses linear regression and replicate it. This will likely be something you have never done before, but it is important for learning how to conduct social science 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. (January 23)
  2. Turn in a short update that reports relevant descriptive statistics and replicates the main model of your analysis. This should only be a couple of pages. (February 27)
  3. The final paper is due our last day of class (April 16).
- Summary of most important dates:
  - Jan 23: Finalize replication article
  - February 27: Replication check-in due
  - April 16: Replication 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

Outside of class, the primary method of communication will be email. *Your University issued email will be used, so please be sure that you have access to that account and that you check it regularly.* I will try to answer any emails within 24 hours during the week, and 48 hours over the weekend. Thus, you should always prepare to write to me well in advance if you have questions about the course. I will make it a priority to respond as quickly as possible to emergencies and other extreme issues. In addition, I hold two 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.

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 (Jan 9): Introduction to Regression

Introduction to the class, general requirements, and logistics.

- Core readings:
  - Wooldridge, Ch. 1
  - Moore and Siegel, Ch. 1–4

Week 2 (Jan 16): Introduction to Linear Algebra

Scalars. Vectors. Matrices. Conformability. Inverses. Matrix operations. Systems of equations. Connecting matrices to data frames. Full rank.

- Core readings:
  - Moore and Siegel, Ch. 12–13
- Assignments:
  - Download PS-1, due on Thursday Jan 23.

Week 3 (Jan 23): Introduction to Calculus

Review the slope. Derivative = fancy slope. Optimization.

- Core readings:
  - Moore and Siegel, Ch. 5–6, 8
- Assignments:
  - Download PS-2, due on Thursday Jan 30.
  - Replication article and paragraph due.

Week 4 (Jan 30): Review of Probability and Statistics

Self-explanatory. Independence. Conditional Probability. Axioms.

- Core readings:
  - Gelman and Hill, Ch. 2
- Assignments:
  - Download PS-3, due on Thursday Feb 6.

Week 5 (Feb 6): Simple Linear Regression

Derive bivariate OLS. Basics of regression. Expected value and variance of OLS estimator.

- Core readings:
  - Wooldridge, Ch. 2
- Assignments:
  - Download PS-4, due on Thursday Feb 13.

Week 6 (Feb 13): Multiple Regression I

Derive multivariate OLS. More interpretation. Fitted values and residuals. Omitted variable bias.

- Core readings:
  - Wooldridge, Ch. 3
- Assignments:
  - Download PS-5, due on Thursday March 12.

Week 7 (Feb 20): Multiple Regression II

Inference. Sampling distribution of OLS estimator. Confidence intervals. F-test.

- Core readings:
  - Wooldridge, Ch. 4

Week 8 (Feb 27): Multiple Regression III

Asymptotics. Efficiency. Normality. Consistency. Outliers. Influential observations.

- Core readings:
  - Wooldridge, Ch. 5
- Assignments:

- Replication descriptives due.

Week 9 (Mar 12): Model Fit, Prediction, Bootstrapping

R^2. Mean squared error. Prediction. Parametric and non-parametric bootstrapping. Standard errors.

- Core readings:
  - Wooldridge, Ch. 6
  - James et al., Ch. 5
- Assignments:
  - Download PS-6, due on Thursday Apr 2.

Week 10 (Mar 19): Dummy Variables and Interactions

How to not make mistakes with dummy variables and interactions.

- Core readings:
  - Wooldridge, Ch. 7
  - Thomas Brambor, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political analysis* 14 (1): 63–82
  - **braumoeller04_int**

Week 11 (Mar 26): ISA, No Class

Week 12 (Apr 2): Heteroskedasticity

Variance. (Heteroskedasticity- and Cluster-) Robust standard errors. More diagnostics. Weighted least squares.

- Core readings:
  - Wooldridge, Ch. 8
- Assignments:
  - Download PS-7, due on Thursday Apr 16.

Week 13 (Apr 9): Simulation

Simulation of probability models and for model checking.

- Core readings:
  - Gelman and Hill, Ch. 7–8

Week 14 (Apr 16): Multilevel Models

Multilevel structures and the basics. Random intercepts and slopes. Five ways to write down the same model.

- Core readings:
  - Gelman and Hill, Ch. 11–12
- Assignments:
  - Replication papers due.