Appendix: Sample syllabus

PPOL 560
Accelerated Statistics for Public Policy
MS in Data Science for Public Policy
Georgetown University

Fall 2018
Professor Michael Bailey
Michael.Bailey@georgetown.edu

Course Description

Statistical analysis is a core element in the data science toolkit. Effective statistical analysis goes beyond merely calculating mathematical quantities and focuses on understanding the conditions under which observed statistical relationships actually reflect true relations between variables rather than incidental or spurious correlation.

This course provides the foundation for effective statistical analysis of policy data by introducing students to multivariate linear regression, with a particular emphasis on distinguishing causal effects from spurious relationships.

By the end of the semester students will be able to:

1) Understand research methods and be able to formulate statistically testable hypotheses;
2) Interpret the results of multivariate linear regressions and think critically about the potential problems that arise when trying to draw conclusions from such results, including omitted variable bias, multicollinearity, interaction effects and transforming variables;
3) Use R to conduct statistical analyses.

Texts and Materials

Required text:

Bailey, Michael. 2015. *Real Stats: Using Econometrics for Political Science and Public Policy*. Oxford University Press.

Other materials:

I will provide slides for all lectures via Canvas. Please make sure you can access this and let me know if you have any problems. I recommend that you print out the notes before class as this will free you from copying equations, output or figures. (I recommend printing as "handouts" with 4 to a page to save trees!)

Course Requirements and Grading

Course grades will be based on the following:

Problem sets	15%
Quizzes	10%
Participation/labs	10%
Final Project	15%
Midterm exam	25%
Final exam	25%

Problem sets will be available in the assignments folder in Canvas. They are due in class on the day indicated. You are strongly encouraged to work on these problem sets in study groups; however, each student must prepare and submit his or her own write-up for each problem set (including R code and (concise!) output when appropriate). If you work with other people, we ask that you indicate who you worked with on the problem set you hand in.

We will provide solutions for the problem sets after they are turned in. No credit will be given for problem sets submitted after the stated due dates.

The **quizzes** are based on the readings. I will let you know when they are ahead of time and the questions will be based on the summary questions at the end of each chapter in the textbook. These are meant to encourage reading ahead and to offer me a gauge how well you are understanding the readings.

Participation grades are based on behaviors that reflect the seriousness of purpose required to excel in policy analysis at Georgetown. Signs of positive engagement include attending class, asking questions, focusing on substance rather than grades, doing the reading and fully participating in the discussion sections. For each chapter, we will have a lab session in which we analyze a data set; students will provide a lab worksheet for each of these sessions.

Attendance at all class lectures and discussion sections is required. If something arises that will disrupt your ability to attend multiple classes, I encourage you to talk to me.

The **final project** will allow you to delve deeper into a topic of interest. Option 1: a research design in which you describe a research question and work through various issues you would need to confront in answering the question statistically. Option 2: find an existing data set of interest and replicate it (I will point you to replication archives), describing what the author did and exploring alternative specifications or diagnostics.

How to succeed in this class

1) Come prepared to class. Do the readings. Think about the readings on their own terms, but also in terms of how the concepts apply to things you are interested in. This is a great time to be greedy: ask "how will this help me?"

2) Ask questions.
    a) Formulating a question helps you engage with the material much more deeply. We all have experienced the thought "this doesn't make sense." Taking that thought to the next level and formulating what doesn't make sense pushes us to deeper understanding.
    b) If you have a question, it's almost certain that others do too; asking a question will not only help yourself, but you will help others.
    c) Asking questions helps keep the class on track. If there are lots of questions, we'll slow down and get things figured out. If there are few questions, we'll charge ahead.
    d) Asking questions helps you develop a voice in statistics. You'll start to have opinions about the right way to do things and will ask questions resisting other ways. That's great! That's how we learn.
    e) Questions make things interesting. We have to take a concept that may seem straightforward and apply it on the fly to a new situation or probe what exactly is going on. That's fun.

3) Use the discussion questions in the readings. We'll do some as exercises in class. Work through the others on your own or with your study groups. Statistics needs to be learned by experience.

You won't really know if you understand something until you are able to work with it, apply it to new context or do something active. Simply reading and highlighting is not enough.

4) Work in groups, but do so wisely. Working with classmates is a great way to learn from each other. Often, classmates will be able to explain things in a way that clicks for you. And, more often, the act of explaining something to someone else will help you understand the material better. Group work only works, though, if you prepare by yourself first. If you show up and wait for classmates to do the work, you can probably muddle through the exercises, but you'll have trouble participating in classes and may fall behind as the material we cover cumulates and needs to be understood at each step.

5) Start assignments early. Sometimes the data doesn't cooperate; you don't want to find this out at 11pm the night before the homework is due. Also, the more you are doing exercises, the more you will be able to follow the lectures.

Technology in the classroom
Laptops, iPads, and iPhones are not allowed during lectures or exams except when we are doing lab or group work. Assume that computers/phones are not allowed until I tell you otherwise.

Academic honesty
Plagiarism or other acts of academic dishonesty will not be tolerated. Cases of suspected academic dishonesty will be handled according to the university's honor code. Please review Georgetown's guide to understanding and avoiding plagiarism.

COURSE OUTLINE
Note: the dates for the midterm and final exams are set and cannot be changed. Plan accordingly. The dates for the other items on the schedule may change modestly, which I will announce during class.

Aug. 29 **Topic 1: Introduction to course**
Reading: *Real Stats*, Chapter 1
Achen. 2002. Advice for Students Taking a First Political Science Graduate Course in Statistical Methods. *The Political Methodologist* Vol. 10, No. 2. Spring 2002.
Wolfers and Stephens.2013. How to Separate Lies from Statistics. BloombergView
Topics: Randomness, causality, endogeneity, exogeneity, experiments, randomization

Sept. 3 **No class (Labor Day)**

Sept. 5, 10 **Topic 2: Working with data in R**
Reading: *Real Stats*, Chapter 2
Topics: Types of research (observational data, experiments, natural experiments), internal validity, external validity, good data practices

Sept. 12, 17, 19 **Topic 3: Bivariate regression**
Reading: *Real Stats*, Chapter 3

	Topics*:*	Estimation and interpretation of OLS coefficients; distribution of estimates, sources of randomness, exogeneity and bias, precision of estimates, probability limits and consistency, heteroscedasticity, goodness of fit, R^2, outliers
Sept. 24, 26	**Topic 4: Hypothesis testing**	
Oct. 1, 3	Reading:	*Real Stats*, Chapter 4
		What a nerdy debate about p-values shows about science, *Vox*
	Topics*:*	Hypothesis tests, *t* tests, *p* values, statistical power, critiques of hypothesis testing, confidence intervals
Oct. 8	**No class (Columbus Day)**	
Oct. 10- 22	**Topic 5: Multivariate OLS**	
	Reading:	*Real Stats*, Chapter 5
	Topics*:*	Motivation and estimation of multivariate OLS, controlling for variables, use in experiments vs. observational studies, omitted variable bias; precision, multicollinearity, inclusion of irrelevant variables, model specification
Oct. 24	**Midterm exam**	
Oct. 29, 31	**Topic 6: Dummy independent variables and interactions**	
Nov. 5	Reading:	*Real Stats*, Chapter 6
	Topics*:*	Difference of means (bivariate and multivariate), nominal variables, dummy variable interactions
Nov. 7-19	**Topic 7: Transforming variables and *F* tests**	
	Reading:	*Real Stats*, Chapter 7
	Topics*:*	Quadratic and polynomial models, logged models, standardized coefficients, *F* tests about multiple coefficients
Nov. 21	**No class (Thanksgiving)**	
Nov. 26, 28	**Topic 8: MLE/Probit**	
Dec. 3, 5	Reading:	*Real Stats*, Chapter 12
	Topics*:*	Linear Probability Model (LPM), logit and probit, maximum likelihood estimation, interpretation of coefficients, hypothesis testing
Dec. 10	**Course wrap-up**	
	Reading:	*Real Stats*, Chapter 16

Final Exam
Thursday, December 13 from 12:30 to 2:30