Quantiative Political Methods II
L32 582

CLASS MEETING
Monday/Wednesday
10:00am - 11:20am
Zoom ID: 999 4834 5919
Zoom link: https://wustl.zoom.us/j/99948345919

OFFICE HOURS: Montgomery
Tuesday 2:30pm-4:00pm
Friday: 2:30pm-4:00pm and by appointment
https://wustl.zoom.us/my/jacobmontgomery

Instructor Information

Jacob M. Montgomery, Ph.D.
Associate Professor, Department of Political Science
E-mail: jacob.montgomery@wustl.edu

Teaching Mentee

Ivan Liao
Division of Computational and Data Sciences
E-mail: iliao@wustl.edu
OFFICE HOURS: Liao
Tuesday 12:00pm-2:00pm
https://wustl.zoom.us/j/97378948623

1 Course Description

This is an advanced quantitative methods course for graduate students in the social sciences. The goal of this class is to give you the vocabulary and skills to (1) make you an independent consumer of contemporary methods in the social science, and (2) to prepare you for more advanced coursework in quantitative methods, statistics, and machine learning. The core idea behind this class is that it is not possible to teach you every statistical model or approach you need to know. But if you can "speak methods" to a certain degree, you will be able of either self-teach or take more advanced courses to tackle any problem.

Conceptually, therefore, the goal of this course is not to make you fluent in any area of methods. Instead, the course is designed to introduce you to the grammar of statistics and model building. How do we learn from data? What assumptions go into a model and what can we learn from it? What makes a model "good" and how would we measure that?

1.1 Disciplinary boundaries

There is a lot of variation in how different fields teach and think about data analysis. This class is taught from my own vantage of working in a political science department but having significant interests and statistics and machine learning. We will spend a lot less time on sandwhich estimators than you might get in an Econometrics. I won't even mention ANOVA, which is a workhorse model for psychology. And I will spend more time on interpretation and less time on proofs than you would get in a math department.

Most students in this course are taking it as part of the required methods sequence for the political science Ph.D. program. However, a sizable number will also be taking this as a first-semester course via the Division of Computational Data Sciences. So there is a lot of heterogeneity in backgrounds and interest. Try to make that a virtue. This is a good opportunity to build your network and maybe even find some collaborators. Try and reach out beyond your own program and ask lot's of questions if I reference something you don't understand.

1.2 Prerequisites

Formally, this class is open to all graduate students. However, the course was designed assuming students have had the following two courses:

- PS 5052: Mathematical Modeling in Political Science. This course covers single-variable calculus and portions of multi-variate calculus, linear algebra, and probability theory.
- PS 581: Quantitative Political Methods I. This is equivalent to a rigorous graduate level course in linear models. This is not as rigorous as something you would get in a math department, but teaches linear models assuming a background in probability, matrix algebra, and calculus.

This is an R based class. It's good for you. Students should be able to write their own functions, organize datasets, execute iterative loops and more with limited guidance. I have added two free books to the list of required texts and you should rely on these for doing problem sets and other tasks. I am also developing a set of instructional videos that go over needed materials, but those may not be complete on time. It is your responsibility to look at these texts and start to work through them at your own pace throughout the semester if you are not already familiar with this software.

2 Detailed Description

Unit 1: Eat your vegetables

We will introduce some of the basic concepts and tools that you need to even engage in advanced quantitative materials. For example, if you don't understand the expected value operator, your understanding of causal inference is going to be pretty superficial. And if you don't understand the relationship between bias and variance for estimators, your work in machine learning is going to be pretty bad.

This is material that would normally be considered as part of a probability and/or a mathematical-statistics course. We obviously cannot cover all of that material in depth, but we will try and touch on the core concepts that you will need later in the course and in future training/reading. The topics covered will be:

- Probability definition, axioms of probability, conditional probability
- What is a probability distribution? PDFs & PMFs & CDFS, Joint/marginal/conditional distributions, Bayes rule

- Expected value, variance, covariance, and moments.
- Convergence of a sequence, the law of large numbers, the central limit theorem
- Bias, efficience and MSE
- Likelihood function, information, and information lower bound

Unit 2: What is it that you think you are even doing?

The second part of the course tries to step back and use these tools to ask a very simple question: what is statistical inference? In your previous exposure to this idea it was probably presented as if making inferences about the world from data was simple and easy. We tend to introduce statistics by saying, "just follow these three steps and you're done!" But it's not that simple, and it's important for you to understand the basic approaches to tackling this difficult problem as well as their strengths and limitations.

In this section we will talk about four "schools" of thought for making inferences about the world from observed data.

- Frequentist statistics: Estimation, confidence intervals, and hypothesis testing
- Maximum likelihood: Estimation, confidence intervals, and hypothesis testing
- Bayesian statistics: Estimation, credible intervals, and Bayes factors
- Non-parametric statistics: Distribution estimation and confidence intervals.

Unit 3: Just another linear model

Despite the proliferation of models in the literature, the vast majority of methods you see every day are variations on the linear model. In this unit, we will develop a conceptual toolbox for this broader class of linear models. This part of the class is more in line with a standard "textbook" approach, and we will lean more heavily on Faraway's *Extending the Linear Model with R*. We will look at the following models as time permits focusing on implementation. Building on Unit 2, this will also give us a chance to work with models in each of the "families" of inference.

- Binary/count data models
- Generalized linear models
- Hierarchical linear models
- Kernel regression
- Tree models

Unit 4: Problems in model building

Regardless of the model you use, there are a set of common problems that you will encounter. What features/covariates should you include? What should you do about missing data? What if the data is "grouped" in some specific way? What if there are non-linearities? How do you interpret your model? The final unit of class will attempt to tackle these issues. Topics covered may very from year to year, but may include:

- Variable selection/regularization

- Model evaluation
- Missing data
- Clustering and time
- Interpretation

3 Course structure (2020)

This course will be entirely online and the structure outlined below is highly experimental. We may need to make some adjustments as the course goes on. But here is an idealized version of how I hope to the class will procede. Please help me help you by giving feedback. I can't always change everything on the fly, but I will at least think about it.

General outline

The basic idea is that for each class period you will do the following:

- Do some related reading from a text
- Watch pre-recorded content
- Complete "in class" exercises during the synchronous sessions.
- Work on problem sets in groups and with assistance from us.

Technology

Canvas: Canvas will serve as a home base for all materials and assignments. This should be your first stop if you are looking for some particular piece of material.

Zoom: All synchronous content will be done via Zoom and recorded/posted to Canvas. However, you are strongly encouraged to come during these times to the extent possible. If you have already done the problems, that's great. Come and help your fellow students understand so you can *master* the material. I will also hold office hours via Zoom at the time posted above. PLEASE NOTE: If you want to have a private meeting (where other students can't just appear in the discussion) we need to set up a separate meeting. Please email me or DM me via Slack.

Slack: All course communications will take place via Slack. This will also be a place you and your fellow students can post questions and work together. You should be able to join the course slack channel using the following link: `https://join.slack.com/t/qpm2-2020/shared_invite/zt-hc4vtdwk-xOk~cm3hpDM1AFj7wYF6vA`

Jamboard/Whiteboard: During class sessions you will be doing a lot of work in groups on some sort of whiteboard. I am still working this out, but assume for the first class we will be using Google Jamboard. You should register for an account before the first class. (`https://jamboard.google.com/`).

Hardware: In additional to the standard hardware for an online course, you will need the ability to "draw" on digital whiteboards. This will help you be able to do math in groups, in office hours, etc. I realize you may not have this for the first class sessions, but do take care of it. You can get a tablet, but you can also buy a much cheaper attachment (like a mouse). If you are taking more than one mathematically oriented class, I recommend you invest in something good. It will save you a lot of frustration.

Software: You need to install R, Rstudio, and Stan on your machines. More information on that will be posted on Canvas later in the semester.

Rmarkdown: All problem sets will need to be turned in using R Markdown. I expect to have one clean document. But please don't send me 90 page documents. Take the time to clean it up and get to the point and not produce every output you think might be helpful.

Textbooks

There will be a number of readings posted to Canvass for each section with chapters from various texts. You should check canvas for each class session well in advance to give yourself time to review this content. When appropriate, I may also give you additional readings and resources. If you are finding a particular area difficult or confusing, just let me know and I'll try and find more material out there to help you out.

Required text:

Faraway, Julian J. *Extending the Linear Model with R* (second edition) CRC Press

R Programming for Data Science. Free online. Roger D. Peng. `https://bookdown.org/rdpeng/rprogdatascience/`

R for Data Science. Free online. Garrett Grolemund and Hadley Wickham. `https://r4ds.had.co.nz/`

Additional texts:

Aronow, Peter M. and Benjamin T. Miller. 2019. Foundations of Agnostic Statistics. *Cambridge University Press*.

Casella, George and Roger L. Berger. 2002. *Statistical Inference* (2nd edition). Duxbury/Thomson Learning

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis* (Third Edition). CRC Press.

McCullagh, P. and John A. Nelder. 1990. *Generalized Linear Models* (Second Edition). Chapman & Hall/CRC.

Evaluation

Evaluation will come in four parts. First, you will be have problem sets approximately ever other week (30% of your final grade). Second, you will be graded based on your participation during class time (10% of your final grade). Third, you will have two exams (40% of your final grade – 20% for each). One exam will take place at the completion of Unit 2 and one will be at the completion of Unit 3. These will both be open note, open book, open Internet takehome exams.

Finally, you will work with a group to complete a final research project (20%). The goal for this will be for you to use the skills from this class to conduct your own data analysis using an advanced method. The paper will be something similar to a "methods/results" section of a social science journal article. If you are not familiar with that format, please go find a recent issue of *The American Political Science Review* and look through

Specific instructions and details about all of these will provided later in the class. But this semester is going to be a like a firehose of work, so please feel free to ask about these in advance if you want or need more information to schedule your life.

4 Class policies

Academic Honesty

Cheating and plagiarism will not be tolerated. All students are expected to adhere to high standards of academic integrity. In this class especially, that means that all work presented as original must, in fact, be original, and the ideas and contributions of others must always be appropriately acknowledged. Quotations must, of course, be acknowledged, but so must summaries, paraphrases, and the ideas of others. If you have any doubts or questions about documentation requirements, **please ask me**. Don't guess.

Religious observances

Some students may wish to take part in religious observances that occur during this academic term. If you have a religious observance that conflicts with your participation in the course, please meet with me before the end of the second week of the term to discuss appropriate accommodations.

Students with disabilities

Students with disabilities enrolled in this course who may need disability-related classroom accommodations are encouraged to make an appointment to see me before the end of the second week of the term.

Grading scale

The course is graded on the 10 point scale below. There will be no exceptions. Don't ask.

Score	Grade	Score	Grade	Score	Grade	Score	Grade
$\geq$94	A	$\geq$83	B	$\geq$ 73	C	$\geq$63	D
$\geq$90	A-	$\geq$80	B-	$\geq$ 70	C-	$\geq$60	D-
$\geq$87	B+	$\geq$77	C+	$\geq$ 67	D+	<60	Fail

5 Tentative Schedule

September 14:	Course structure and introduction	*Class survey due*
September 16:	Sets, sample spaces, axioms, and theorems	
September 21:	Conditional probability and Bayes law	
September 23:	Introduction to random variables	*Problem set 1 due*
September 28:	Moments	
September 30:	Introduction to estimation	
October 5:	Properties of estimators	
October 7:	Likelihood and information	*Problem set 2 due*
	End of unit 1	
October 12:	Frequentist statistics	
October 14:	Testing	
October 19:	Maximum likelihood estimation	
October 21:	Bayesian statistics	*Problem set 3 due*
October 26:	Approximating posteriors	
October 28:	Non-parametric statistics	
	End of unit 2	
November 2:	The basic linear model	
November 4:	Binary outcomes	*Problem set 4 due*
November 9:	Count data	*Take home exam 1 due*
November 11:	Hierarchical models	
November 16:	Estimating HLMs	
November 18:	Kernel regression *Problem set 5 due*	
November 23:	Tree models	
	End of unit 3	
November 25:	Variable selection and regularization	
November 30:	Interactions & interpretation	
December 2:	Missing data	*Problem set 6 due*
December 7:	Model evaluation	*Take home exam 2 due*
December 9:	Clustering	
December 14:	Time	
December 16:	Non-linearities	*Problem set 7 due*
January (TBD)	Final project due	