Advanced Methods of Political Analysis

Basics

What? POLI706, University of South Carolina, Spring 2017.

Who? Tobias Heinrich, `heinrict@mailbox.sc.edu`, GAMB333. No official office hour for a graduate class; just chat me up when I'm around or email me to schedule a meeting.

Who else? Wendell Mayes, .

When and where? Wednesdays, Gambrell 302, 6–8.45p.

Outline

Unless you are a political philosopher or 100% game theorist, skills of statistical analysis are the *sine qua non* for a modern social scientist. This course is one of the early and many formal and informal stepping stones on your way to learning these skills. Perhaps it is the most important of these as linear regression, the main topic of this course, is at the intersection of almost all statistics (as practiced by social scientists).

The demands on statistical sophistication in order to be taken seriously and to get published are rising rapidly. Whereas in 1970s, analyzing correlations was fine, by the 1980s simple regressions were the norm. Within the 1990s, likelihood approaches become the ubiquitous. Since the 2000s, structural estimation, Bayesian statistical inference, and causal inference became more prominent albeit not necessarily required. In recent years, machine-learning approaches crept up. Today, articles even in lower ranked subfield journals will feature technique unknown to very respectable social scientists in the 1990s.

This course helps you along to learning the tools to carry out empirical social science research. The focus will be on the modern, practical aspects of this. To this end, the course features four main building blocks, namely the rigorous introduction of the basics of the linear regression model; classical and modern social scientific inference; and carrying out data analyses.

If you participate actively and take this course *very* seriously, you will learn the following:

- understand the basics of the linear regression model;
- read and comprehend social science research that uses linear regressions as a means for inference;
- learn how to perform inference in a traditional as well as modern way.

Required material

The only material that requires a purchase is Jeffrey M. Wooldridge's *Introductory Econometrics: A Modern Approach*, 5th (or later) edition. All other readings and code examples (in R) will be provided.

It makes sense to have a few statistics books around to read alongside what we are doing. There are many different ways covering and approaching the same material, and maybe someone else's exposition will help things click more efficiently. The following books are likely to come in handy and are great resources in any case.

- Andrew Gelman & Jennifer Hill, *Data Analysis Using Regression and Multilevel/ Hierarchical Models*.
- Hadley Wickham, *ggplot2: Elegant Graphics for Data Analysis*.
- Fumio Hayashi, *Econometrics*.
- Joshua D. Angrist & Jörn-Steffen Pischke, *Mostly Harmless Econometrics*.
- William H. Greene, *Econometric Analysis*.
- Peter Kennedy, *A Guide to Econometrics*.
- Damodar Gujarati & Dawn Porter, *Basic Econometrics*.

The entire course will revolve around R and LaTeX as the pieces of software to carry out analyses and write up output, respectively. RStudio is the recommended GUI for R.

Expectations

Lest this is not perfectly clear: unless you participate actively and take every task *very* seriously, you will fail to learn to the material, and thus by the end will not have acquired the skills to conduct the inferential part of modern social science research. This negatively affect your performance in political science at USC and (if you're getting that far) at the dissertation and job market stage.

The problems stemming from such a failure ought to be obvious. It is the expectation that, if you take this course, that you will immerse yourself in it. "Immerse yourself"? It means that you carefully read the assigned chapters, immediately raise points that are not clear, devote numerous hours to homework assignments, midterm, and final, and consult with instructor and TA. It ought to go without saying that you should be in class on time and every time.

There are also specific expectations about everything that will be handed in the course. It is this instructor's firm belief that anything you produce and that you deem to be worthy to be shown to someone else should be packaged in a way that it concisely and clearly conveys the content. This means, most often, the use of graphs, formatted tables, properly labeled output, etc.[1]

[1] For example, variable names that you use within R (*lngdppc*, e.g.) are not formatted for easy reading; don't report the mean of variable capturing, say, human height as 1.743232535cm, but round it to something sensible (1.75cm).

ASSISTANCE

The instructor and the TA are here to help you learn the material–make extensive use of us. That said, for many, many, many, many, many problems you will encounter (in particular with R and LaTeX), you will trivially find the answer through a simple search on the internet. Do this, please, before you consult us. After all, learning efficient self-help on these matters will save many headaches in the future; this course is just a semester, but good self-help will make your life easier for many, many years.

When you contact us with questions, it has to be obvious to us that you have grappled with the question. "I don't know where to begin. How does X work?" is not an acceptable email inquiry. You need to say at which point understanding broke down, which things are unclear, and so on. If you have coding troubles, send us code that we can run; just copy-pasting a chunk of code without your data, knowing which packages were loaded, and which code preceded the code chunk make helping a thorough mess.

See here for guidance:

- `http://stackoverflow.com/questions/5963269/how-to-make-a-great-r-reproducible-`
- `http://stackoverflow.com/help/mcve`
- `http://adv-r.had.co.nz/Reproducibility.html`
- `http://sscce.org`

And: If in doubt about anything, talk to the TA or me at any point. And do so early. We're there to help you understand this essential material; we are *not* there to remedy laziness.

Grading, Rules, etc.

There are three sets of works that will determine your grade: (roughly) weekly exercises, midterm, and a final exam. Their weighting is 2, 2, and 5, respectively.

Roughly every week, you have the chance to work through exercises that connect to the lecture and previous lectures. Each contributes equally. You are highly encouraged to collaborate on this and consult material from outside of the course. What you turn in, however, has to be your own work using your own chosen data. What matters is that you understand the material; it does not matter how exactly you arrived at the comprehension. Consequently, for any exercise, you have the option(s) to fix your errors *within two weeks of the due date*. That is, you may redo the parts of the homework that you got wrong and show it to the TA. Your grade then gets adjusted if that happens within two weeks of the due date. The final date for getting a grade adjusted is a week after the last day of class.

The midterm (roughly after Week 6) will feature data analysis with interpretation, graphs, etc. It will focus on basics and classically-done inference. The same re-grading rule applies. Grading will be done by the instructor.

The final exam focuses on publication-quality and modern inference, statistics, and presentation. It will be given about two weeks before class ends. More details will be provided as the semester progresses. All issues related to the final exam will be handled by the instructor. The final exam is due a week after the last day of class.

In line with the goal to help you learn modern statistical inference, the tools used to carry out analyses should reflect the goal. You are strongly urged to adopt LaTeX and R as the tools to carry out and summarize your work throughout the course. If you opt to use for some other software, the instructor and TA will be unlikely to be able to offer help.

I will use the following grading scale to map between percentages and letter grades. A 92-100, B+ 87-91, B 80-86, C+ 77-79, C 70-76, D+ 67-69, D 60-66, and F 0-59. The standard rules of rounding are applied. The University of South Carolina Honor Code applies.

Reasonable accommodations are available for students with a documented disability. If you have a disability and may need accommodations to fully participate in this class, contact the Office of Student Disability Services: 777-6142, TDD 777-6744, email `sasds@mailbox.sc.edu`, or stop by LeConte College Room 112A. All accommodations must be approved through the Office of Student Disability Services.

Course schedule

This is a rough schedule of the sequence of what will be covered in class. The instructor will most probably amend things quite a bit. So, always check what will be next.

Week 1

Why are we doing this? Why statistical analysis? What do we want?

- Wooldridge, Chapter 1.
- Jonah Weiner, "Jerry Seinfeld Intends to Die Standing Up", *New York Times*, December 20, 2012. URL: `http://goo.gl/2MlunX`.
- Jenny Woodward, "Jerry Seinfeld: How to Write a Joke", *New York Times*, December 20, 2012. URL: `http://goo.gl/F2XZ8n`.
- Serious Jibber-Jabber with Conan O'Brien, "Rocker Jack White". January 10, 2013. URL: `http://goo.gl/mYNkw7`.
- Optional: Netflix, "Grant Achatz", *Chef's Table*, Season 2, Episode 1.
- Optional: Peter Thiel, *Zero to one*.
- Optional: Leander Kahney, 2013. *Jony Ive: The Genius Behind Apple's Greatest Products*.

Weeks 2 and 3

The fundamentals of linear regression.

- Wooldridge, Chapters 2 and 3.

WEEK 4

Including qualitative explanatory variables.

- Wooldridge, Chapter 7.

WEEK 5

Working with interactions in your regression.

- Thomas Brambor, William Clark, and Matt Golder, 2006. "Understanding Interaction Models: Improving Empirical Analyses." Political Analysis 14(1).
  URL: `http://pan.oxfordjournals.org/content/14/1/63.short`

WEEK 6

Statistical inference, done classically. Mister p.

- Wooldridge, Chapter 4.

WEEKS 7–9

Statistical inference, done in a modern way. Nonparametric and parametric bootstrapping.

- Gary King, Michael Tomz, and Jason Wittenberg, 2000. "Making the Most of Statistical Analyses." American Journal of Political Science 44(2).
  URL: `http://www.jstor.org/discover/10.2307/2669316`

WEEKS 10 AND 11

Model fit, checking models. All models are wrong, some are useful.

WEEKS 12 AND BEYOND

From minimizing sums-of-squares to maximizing the likelihood.