PLSC 504: Advanced Quantitative Methods

Fall 2019
Preliminary syllabus subject to change

Course Personnel and Office Hours

- Instructor: Alex Coppock alex.coppock@yale.edu RKZ 135 Office Hours: Wednesday afternoons 1:30 to 3:30 or by appointment. Please email to schedule appointments, otherwise you might have to wait.
- Teaching Assistant: James Sundquist james.sundquist@yale.edu Office hours Thursdays, 2:00 - 4:00, RKZ 204.
- Teaching Assistant: Shikhar Singh shikhar.singh@yale.edu Office hours Fridays, 9:30 - 12:00, RKZ 204.

Course Meeting Times

- Lecture: Mondays and Wednesdays 9:00am to 10:15am in RKZ 05.
- Section: Time and location TBD. (Probably Thursday afternoons or Friday mornings)
- Attendence at all course meetings (lectures and sections) is required.

Objectives

PLSC 504 is the third course in the PhD level sequence in quantitative methods. The main goal all semester will be extracting causal inferences from data. We will focus mainly on observational settings with frequent comparison to experiments.

Along the way, we want to encourage:

1. A feel for research design
2. A love of code.
3. A habit of creating beautiful, reproducible documents.

Prerequisites

PLSC 503 is the only formal prerequisite. That course covered Aronow and Miller, chapters 1 - 5. The topics covered were probability, summarizing distributions, learning from random samples, regression, and parametric models. Everyone should read (or reread) these chapters.

Required Textbooks

1. Angrist, J. D., and Pischke, J. S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton.
2. Morgan, Stephen L. and Christopher Winship. 2010. *Counterfactuals and Causal Inference.* Cambridge.
3. Grolemund, Garret and Wickham, Hadley (2017) *R for Data Science.* Free online book here: http://r4ds.had.co.nz, physical copy here: http://amzn.to/2aHLAQ1
4. Additional weekly readings will be placed on canvas.yale.edu.

Recommended Textbooks

1. Aronow, P. M. and Miller, B. T. (2019) *Foundations of Agnostic Statistics.* Cambridge.
2. Gerber, Alan and Donald P. Green (2012) *Field Experiments: Design, Analysis, and Interpretation.* Norton.
3. Hernan and Robins (2019) *Causal Inference* Free online: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
4. Cunningham, Scott (2019) *Causal Inference: The Mixtape.* Free online: https://www.scunning.com/cunningham_mixtape.pdf

Software

We will be using the open-source statistical software `R`. In addition to `R`, please also download and install RStudio. Pro-tip: change your computer's defaults so that .R scripts open in RStudio, not R.

- Download `R` here: www.r-project.org
- Download RStudio here: www.rstudio.org

Problem Sets

This course will involve a relatively heavy workload, and students should be aware that this class will require sustained, serious effort all throughout the term. The weekly problem sets will be a mix of mathematical demonstrations, simulations, and data analysis. The documents that you turn in must satisfy the following criteria:

- It must be in .pdf format.
- Mathematical expressions must be nicely formatted.
- Tables must be nicely formatted.
- Figures must be well-labeled and placed in-text.
- The code you use and what it produces must be in the same place.
- It must have your name on it [important]

- All students must write up their problem sets individually. However, you may work in groups of up to three, though you are not required to work in groups at all. Please indicate the names of the other students you worked with that week. Do not "share" members across groups. Do not copy and paste the answers across group members.

This list of requirements is restrictive. They mean that you can't use Microsoft Word to do your homework. While this policy seems obnoxious now, I *promise* that you will be happier later in your graduate career when you've already practiced making professional documents. Practically speaking, you will use `rmarkdown`. `rmarkdown` is a way of weaving together text, code, and output in a single document. These documents are reproducible and transparent. If you want to use LaTeX+ `knitr`, that's OK too. Your TAs are good at `rmarkdown` and stand ready to help you. Please do not resist learning `R` and `rmarkdown`.

Exams

We will have two in-class exams, a midterm and a final They will be open-internet (no two-way communication) and you may bring a single page of handwritten notes. Because the class period is only 75 minutes long, the exams can't be that difficult!

Final Paper

For your final paper, you can either write a replication paper or do a piece of original quantitative analysis. They should have five sections: introduction, theory, design, results, conclusion. There's no need to be creative with form in this paper! All papers *must* be prepared in LaTeX. I'm adding this requirement because I think I know that it will make your lives better in the long run. If you don't yet know how to use LaTeX, please ask your TAs to help you get set up. Please do not resist learning LaTeXeither.

Most of the meat of the paper should be in the design and results sections. The theory section is important, but should be kept concise. There is *absolutely* no need to conduct a literature review or to do a "big frame" for these papers. I want you to distill the theory down to the parts that are the most important for defining the estimand and for supporting the statistical assumptions underlying the estimators. I don't really know how long these should be, but it's probably hard to do a good job in less than 15 pages. Please do not make me read 30 pages.

The paper is due on December 17th, the university-assigned final exam time. We'll have two mandatory check-ins along the way. On October 21 (right after we get back from Fall Break), you'll turn in a one-page proposal that must include some minimal proof of feasibility. On Monday, November 18th, you'll turn in a rough draft of the empirical section of the paper. This is the week before Thanksgiving, and we should find time to discuss your drafts in office hours before you head off for break so you can be making good progress.

Replication papers

In a replication paper, you download the data for a paper written by someone else and reanalyze it. Replication papers are usually extremely fun because you get to see all of the various analytic choices made by the authors up close. Good replication papers (a) successfully reproduce the main result (b) examine the plausibility of the underlying statistical assumptions and (c) offer modest extensions. I'm not particularly interested in "gotcha" replications, though occasionally replications turn up consequential errors. I would strongly suggest choosing a paper to replicate that follows a research design that you are considering implementing yourself.

Original analysis

You may choose instead to to an original piece of empirical quantitative analysis. You should choose a topic that is feasible and that you care about. You should be planning to incorporate the output of this term paper into a journal article you hope to write someday. This should be a new project in the sense that you will develop a new research design to study something that you care about. If you've written on the topic before or are currently writing something related for another course, that's fine, though please don't double count the same essay for two purposes. Coming up with good original research ideas is tough – let's please talk about your ideas early and often.

Grading Policy:

Problem Sets (35%), Midterm Exam (20%), Final Exam (15%), Final Paper (30%).

Academic Honesty:

To ensure that you do not accidentally violate Yale's academic honesty policies, please review these sites:

- Academic Honesty: http://bit.ly/2a6uTC5
- Understanding and Avoiding Plagiarism: http://bit.ly/29VnoN1

I would like to emphasize that it is a violation of the honesty policy to:

- Copy another student's problem set, just changing a few words here and there. Collaboration is encouraged, but at some point relying too much on your partner becomes a violation of academic integrity. Most cases are clear-cut; for cases that are ambiguous, ask.
- Copy and paste whole blocks of code from your partner that you didn't have a hand in writing.
- Copy whole sentences from the internet.

It is *not* a violation of the honesty policy to:

- Copy code from websites like stackoverflow or other online forums. This is not cheating, it's learning. Part of what makes it learning is that understanding code off the internet well enough to use it usually means that you at least sort of understand it. If you do copy such code, please include a link to the forum or site where you obtained the code in the comments. This is good practice anyway, as you will often forget where code came from!
- Discuss the problem sets with your partners and compare answers.

Preliminary Schedule

Week 1

- Wednesday, August 28. Introduction
- Friday, August 30: Make-up class for the Labor Day class we will miss.
- Readings:
  - R for Data Science, Chapters 1-3

Week 2

Potential outcomes model

- Wednesday, September 4: Class Meeting. Problem Set 0 due at 9:00am.
- Friday, September 6: Section.
- Readings:
  - Paul W. Holland (1986) Statistics and Causal Inference, Journal of the American Statistical Association, 81:396, 945-960
  - Mostly Harmless Chapters 1 - 2
  - Morgan and Winship Chapters 1 - 2
  - R for Data Science, Chapters 9 - 10 (skim 11, 12, 13)
  - Causal Inference Chapters 1 - 2 (suggested)
  - Gerber and Green Chapters 1 - 2 (suggested)=

Week 3

Causal Graphs

- Monday, September 9. Class Meeting. Problem Set 1 due at 9:00am.
- Wednesday, September 11. Class Meeting.
- Friday, September 13. Section.
- Readings:
  - Morgan and Winship Chapters 3 - 4
  - R for Data Science, Chapters 14 and 15
  - Hernan and Robbins Chapters 6-8 (suggested)

Week 4

Regression for causal inference 1

- Monday, September 16. Class Meeting. Problem Set 2 due at 9:00am.
- Wednesday, September 18. Class Meeting.
- Friday, September 22. Section.
- Readings:
  - Mostly Harmless Chapter 3
  - Aronow and Miller Chapter 4 and 7
  - Morgan and Winship Chapters 5 - 6
  - R for Data Science, Chapters 17 and 20

Week 5

Regression for causal inference 2

- Monday, September 23. Class Meeting. Problem Set 3 due at 9:00am.
- Wednesday, September 25. Class Meeting.
- Friday, September 27. Section.
- Readings:
  - Aronow, Peter and Cyrus Samii. 2016. Does Regression Produce Representative Estimates of Causal Effects?
  - Sekhon, Jasjeet: Opiates for the Matches: Matching Methods for Causal Inference

Week 6

IV for causal inference

- Monday, September 30. Class Meeting. Problem Set 4 due at 9:00am.
- Wednesday, October 2. Class Meeting.
- Friday, October 4. Section.
- Readings:
  - Mostly Harmless Chapter 4
  - Morgan and Winship Chapter 9
  - Sovey and Green (2010): Instrumental Variables Estimation in Political Science: A Readers Guide

Week 7

IV for causal inference

- Monday, October 7. Class Meeting. Problem Set 5 due at 9:00am.
- Wednesday, October 9. Class Meeting.
- Friday, October 11. Midterm Review
  - Gerber and Green Chapters 5 - 6 (suggested)
  - applications

Week 8

- Monday, October 14. In-Class midterm.
- Wednesday: Fall Break

Week 9

Regression Discontinuity

- Monday, October 21. Class Meeting. One page proposal for final paper due.
- Wednesday, October 23. Class Meeting.
- Friday, October 25. Section.
  - Mostly Harmless Chapter 6

Week 10

Regression Discontinuity

- Monday, October 28. Class Meeting. Problem Set 6 due at 9:00.
- Wednesday, October 30. Class Meeting.
- Friday, November 1. Section.
  - Applications

Week 11

Difference-in-difference / Panel Data

- Monday, November 4. Class Meeting. Problem Set 7 due at 9:00.
- Wednesday, November 6. Class Meeting.
- Friday, November 8. Section.
  - Mostly Harmless Chapter 5

Week 12

Difference-in-difference / Panel Data

- Monday, November 11. Class Meeting. Problem Set 8 due at 9:00.
- Wednesday, November 13. Class Meeting.
- Friday, November 15. Section
  - Applications

Week 13

Reflections

- Monday, November 18. Class Meeting. Problem Set 9 due at 9:00. Rough draft of the empirical section of the final paper due.
- Wednesday, November 20. Class Meeting.
- No Section!
  - Samii 2015: Causal Empiricism in Quantitative Research
  - Freedman: Statistical Models and Shoe Leather

Week 14: Thanksgiving

Week 15

- Monday, December 2. Class Meeting. In-class review session.
- Wednesday, December 4. In-class written final.

Exam Period

- Tuesday, December 17. Final Paper due by email to alex.coppock@yale.edu by 7pm (our scheduled final exam time). You can turn it in before then if you want.