PLSC 512 Spring 2017 Midterm Exam

Alexander Coppock

This exam will be administered in two parts. You will have 30 minutes to complete Part I and the remainder of the class period (approx 45 minutes) to complete Part II. If you wish to start on Part II early, that's fine, you just can't go back to Part I after beginning Part II.

Part I of the exam is closed-book except for 1 page of notes (your cheat sheet). You must hand-write your answers in a blue book, which will be collected at the end of 30 minutes.

Part II of the exam is open-book. You may use any materials you like, including notes and passive internet queries. That is, you may not interact with other students online, nor may you ask other humans on the internet any questions you might want answered. Turn the second part as a pdf to Canvas.

Part I

1. Consider an experiment with binary random assignment Z, outcome Y, and binary covariate X that is measured **posttreatment**. Show that the difference-in-means estimate of the average effect of Z on Y conditional on $X = 1$ is prone to bias away from the average treatment effect among the subgroup of subjects for whom $X_i(1) = X_i(0) = 1$.
2. What are the assumptions you need in order to believe that the CACE estimator will consistently estimate the CACE? I count five of them. You can say them in words, math, or both (just not neither).
3. Consider the following hypothetical experiment. Imagine you've partnered with an advocacy group that organizes protest marches. You got them to randomize their outreach activities – good on you! You want to see if coming to people's houses and giving them protest sign-making materials (posterboard, good sharpies, etc) will cause them to turn out to march more. You did a good job randomizing whether subjects were assigned to be contacted or not. However, some of those assigned to be contacted were not contacted. Additionally, some of the people who were assigned to NOT be contacted nevertheless were because there were enthusiastic interns who figured that some of the supplies were going to waste, so they decided to go and treat their friends in the darn control group! The organizers "took attendance" at the march. Here are the resulting data. Please estimate the effect of contacting subjects on their probability of marching, at least among compliers.

	Assigned Contact	Not Assigned Contact
number of contacted subjects who marched	170	50
number of uncontacted subject who marched	20	100
number of contacted subjects who didn't march	30	30
number of uncontacted subjects who didn't march	80	120

Part II

Please read the following [lightly-edited] abstract from Duflo, Esther, Rema Hanna and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." American Economic Review, 102(4): 1241-78. Please turn in a PDF with your answers to canvas.

> In the rural areas of developing countries, teacher absence is a widespread problem. This paper tests whether a simple incentive program based on teacher presence can reduce teacher absence, and whether it has the potential to lead to more teaching activities and better learning. In 60 informal one-teacher schools in rural India, randomly chosen out of 120 (the treatment schools), a financial incentive program was initiated to reduce absenteeism. Teachers were given a camera with a tamper-proof date and time function, along with instructions to have one of the children photograph the teacher and other students at the beginning and end of the school day. The time and date stamps on the photographs were used to track teacher attendance. A teacher's salary was a direct function of his attendance. The remaining 60 schools served as comparison schools. The introduction of the program resulted in an immediate decline in teacher absence. The absence rate (measured using unannounced visits both in treatment and comparison schools) changed from an average of 43 percent in the comparison schools to 24 percent in the treatment schools. The program positively affected child achievement levels: a year after the start of the program, test scores in program schools were 0.17 standard deviations higher than in the comparison schools and children were 40 percent more likely to be admitted into regular schools.

1. Please download the dataset from canvas and load it into R.

- How many rows are there? (rows are the unannounced visits)
- How many schools are there? (schools are the units of assignment)
- How many blocks are there? (this experiment was blocked)
- Is the probability of assignment constant across blocks?

2. Conduct a school-level analysis.

- You'll need to collapse the dataset to the school level. I'd recommend using `group_by` and `summarize` from the `dplyr` package, but if you're old school, `tapply` might work…
- What is the estimated average effect of treatment within each block?
- What is the estimated average treatment effect? Please interpret this estimate.
- Bonus: what is the estimated standard error around this treatment effect?

3. Conduct a visit-level analysis.

- Consider the visits to be "clustered" within school.
- What is the estimated average treatment effect, where our target is the average effect of treatment on whether a *visit* finds an open or closed school. Please interpret this estimate.
- Hint: Be sure to account for differential probabilities of assignment due to the blocking procedure. You might consider using `randomizr::declare_ra` to declare a blocked-and-clustered trial, then use `randomizr::obtain_condition_probabilities` to get the right probabilities of assignment.
- Conduct randomization inference, using a one-tailed test under the sharp null.