Prompts for Pre-Class Exercise

Class 3: Correlation vs Causation

Read this extract from "The Representation of Causality and Causation with Ontologies: A Systematic Literature Review." (Sawesi, Suhila, Mohamed Rashrash, and Olaf Dammann, 2022): [fill in link to abstract here]

Give a one-sentence working definition of causality or causation. Which of the seven aspects discussed in the excerpt appear in your answer and why? Your whole answer should be at most a paragraph or two long.

Optional: Here is the full paper for optional reading and more context: [link to full paper]

Class 4: Randomized Experiments

Read this paper by Fisher: R.A. Fisher and the Design of Experiments 1922 1926.pdf Download R.A. Fisher and the Design of Experiments 1922 1926.pdf

(Focus on the narrative and the use of randomness (Sections 1 & 2); you can ignore the stuff about "Latin squares" and "factorial block designs" for now).

Angus Deaton, 2015 Nobel laureate, wrote in "Instruments, randomization, and learning about development.":

"Randomized controlled trials cannot automatically trump other evidence, they do not occupy any special place in some hierarchy of evidence, nor does it make sense to refer to them as “hard” while other methods are “soft.” These rhetorical devices are just that; metaphor is not argument, nor does endless repetition make it so."

In a short paragraph, less than ten lines, what do you think Fisher's response to that statement would be? What part of the reading makes you think so? You may also indicate your own position, but the question is about what Fisher would think (and he is no longer alive, so we won't be able to check!)

Class 5: Problems with Experiments

Read this short version of the laptop experiment we discussed in class: laptops_in_the_classroom_consequences.pdf

Reflect on our discussion in class and think about how you would actually use this experiment for setting policy. Do you think that similar results would hold for other groups and outcomes? What do you do with the differential findings for men and women? Should we have a differential policy for different groups: men only, freshmen only? Was this experiment useful?

Class 6: Bandits (and other advanced experimental methods)

Read the introduction in the attached survey on Multi-Armed Bandits (you don't need to read the Preface or Bibliographic notes). Why do you think this material is relevant to our class? What sorts of causal questions might this help answer?

bandits_introduction.pdf

Class 7: UBI Intro: Permanent Income Hypothesis

Read this short introduction to universal basic income: what-is-universal-basic-income-basics.pdf Download what-is-universal-basic-income-basics.pdf.

What do you think about a universal basic income? Where do you come down on the six categories in the reading:

- Should it replace social protection?
- Should it be one time endowment or regular income?
- Should it be categorical or universal?
- Should it be conditional or unconditional?
- Should it be to households or individuals?
- Should it be a tax credit or a cash transfer?

Limit your answer to a short paragraph.

Class 8: Regression

Read the following summary of three studies on the Permanent Income Hypothesis: Studies on PIH.pdf.

Question: Based on the information you have, how might you explain the discrepancy between these three studies? Name one issue that might come up with each. Some things to think about include:

- What is being counted as "transitory income" in each of these? What is being counted as "consumption"? Are these in line with the meaning of these terms in the PIH?
- Are you worried about internal/external validity?
- Are you worried about construct validity?
- Are you worried about spillovers?

Class 9: Empirical Evidence on the Effect of Unearned Income

Visit this website. from Open Research which includes key findings and data visualizations from their Unconditional Income study conducted in two US states. After exploring the information and interacting with the visualizations, what surprised you most about the results and why? What do you take away from this about UBI as a policy?

Class 9.5 (new): Heterogeneous Treatment Effects and Random Forests

======================

Class 10: Returns to Education

We're going to begin the "Returns to Education" unit! The goal is to understand the effect of education (that is, an extra year of high school; or going to college; etc) on outcomes further down the road (for example, earnings at age 30).

Here are two first attempts to get at this effect, given that a randomized experiment is difficult/unethical.

Attempt 1: Look at data from the US census, and run a multivariate regression, where:

- The independent variable is X = years of education,
- The dependent variable is Y = log(earnings).

Imagine that the regression controls for a bunch of possible confounders: Parents' income, parents' education, race, gender, age, IQ tests taken as a child, …

Interpret the coefficient beta_hat on X as the “earnings returns to education.” That is, if someone gets one more year of education, you would expect that their log(earnings) would increase beta_hat, or alternatively that their earnings by would increase by a multiplicative factor of $e^{\text{beta\_hat}}$.

Attempt 2: Look at data from a few hundred pairs of identical twins. Each pair of twins have the same parents, race, gender, age, and so on. Run a regression where, for each pair of twins:

- The dependent variable is Y = log(earnings of twin 1) - log(earnings of twin 2)
- The independent variable is X = (years of education for twin 1) - (years of education for twin 2)

Interpret the coefficient on X as the “earnings returns to education” (in the same sense as in Attempt 1).

For both attempts, don’t worry for now about why we are measuring log(earnings) instead of earnings – we’ll talk about that in class!

Question: Which of these two attempts would you find more compelling and why? Also, what additional information about the two attempts would be most helpful in answering that question?

Class 11: Instrumental Variables (in Draft Lottery Setting)

Suppose we are doing a randomized control trial for a new drug, Stanfordizone. (It treats the compulsive urge to found a tech start-up; side effects may include wearing hoodies to formal events). We get a representative sample of 1000 people to participate, and split the sample randomly into a control group of 500 people and a treatment group of 500 people.

Unfortunately, after the experiment is over, follow-up surveys reveal that only about 400 people in the treatment group actually took the drug. (The other 100 people were presumably too busy with their start-ups and forgot).

The research team brainstorms and comes up with the following options:

1. The study was compromised, and we can't learn anything about the effect of Stanfordizone. We should throw out the data and re-do the study with better monitoring.
2. We should throw out the data from the 100 people who didn't take the drug, and analyze the rest of the data as normal. (That is, do "average among treatment group - average among control group" among the remaining 900 people). That should give us that ATE of Stanfordizone on start-up-starting, but the error bars might be larger since we have a slightly smaller sample.
3. We can't learn the ATE of Stanfordizone on start-up-starting on the original population we sampled the 1000 study participants from, but we may still be able to learn something from our data. We should think further about this.

What do you think about each of these options? For option 3, what do you think we can learn from this study, if anything? Can we get close to the ATE on the original study population? Or is there some other population for whom we can estimate the average treatment effect of Stanfordizone?

<u>Class 12: Extensions to IV</u>

Read the introduction of the attached paper "Does Compulsory School Attendance Affect Schooling and Earnings?" by Angrist and Krueger. (That is, read the first three pages or so – you can stop when Section I, "Season of Birth, Compulsory Schooling, and Years of Education", starts.): Pre-class reading Oct 31.pdf

What instrumental variable do the authors propose for measuring the returns to education? Do you find it compelling? Why or why not?

<u>Class 13: Regression Discontinuity Design</u>

Suppose that all students at a school take an end of the year exam in math and reading. If a student fails either exam (meaning, scoring less than 50 on each of the exams, on a scale of 0 to 100), they have to participate in a summer school. All students are also tested next year in math and reading.

We are interested in the effect of summer school on next year's test results. Suppose the data you see are, for each student: this year's math and reading scores; whether they participated in the summer school; and next year's math and reading scores. How would you use these data to estimate the effect of the summer school? Would your results be useful information for the school to make policy decisions, for example (a) whether to get rid of the summer school program or (b) to make it mandatory for everybody?

Class 14: Brexit and Panel Data Intro

Consider the following examples:

1. The UK left the EU in 2020, while 27 countries remained in the EU. What is the effect of belonging to the EU on GDP?
2. In the 1960's, there was an outbreak of terrorism in the Basque region of Spain. Other regions did not have terrorist attacks. What is the effect of terrorism on the per capita GDP of a region?
3. In 1989, California passed a proposition that heavily taxed tobacco products. Other states did not. What is the effect of anti-smoking legislation on the prevalence of smoking?
4. In 1980, the city of Miami FL experienced an influx of immigrants from Cuba. Other cities did not. What is the effect of immigration on wages for natives?

All of these examples have a common structure: They are a class of units who are treated for some time period(s), while other units are not treated or are treated in different time periods. For each of the four examples above, think about what are the units, what is the treatment, and which units are treated and when (you don't need to write anything down for this part, just think about it).

It might be tempting to get at the treatment effects in these examples by comparing treated units to non-treated units. However, in all of these examples, the treated units are different from non-treated units (they are treated at a different time, they come from a different population, etc). Think of at least one way you might try to correct for this. Describe what that correction would look like for at least one of the examples above.

Class 15: Difference-in-Differences (and some placebo analysis)

As one of the examples in Tuesday's lecture, we saw an example of the effect of the change in the minimum wage on employment in New Jersey. In that example, the authors (Card and Krueger) surveyed fast food restaurants in PA and NJ before and after a change in the minimum wage in NJ; there was no change in PA.

Suppose that we wanted to run a regression to get at this effect. Here are two possible regressions we could run:

Regression 1: $Y_2 - Y_1$ = a + tau* NJ + noise, where:

- Y_1 and Y_2 are the employment before and after the change, respectively
- NJ is an indicator variable that is 1 if the fast food restaurant is in NJ, and 0 otherwise

Regression 2: $Y_2 - Y_1$ = a + tau* NJ + b Y_1 + noise,

where the variables are the same as in Regression 1. That is, the only difference is that Regression 2 also includes Y_1 as a control variable.

Suppose we got tau_hat as our estimate for tau. How would you interpret tau_hat in each of these regressions? What are some reasons that the results might be very different? Which regression do you think would better capture the effect of raising the minimum wage in NJ and why?

Class 16: Sparsity and Low-Rankness

Suppose we have n units (for example, the 27 countries in the EU, **not** including the UK), observed over T time periods. For each unit and each time period, we observe:

- an outcome Y (for example, GDP)
- some covariates: X = population, W = chocolate consumption, Z = number of Nobel laureates per capita. (Pretend for the purposes of this response that these are the only variables that are important for GDP...)

Suppose we want to model the world (forget about causality for a moment) with a linear model, as in regression. Since we have t time periods, we can run t separate regressions:

Y_t = a_t + b_t * X + gamma_t * W + delta_t * Z+ noise,

where Y_t is the outcome (GDP) at time t. This gives us estimates a_hat, b_hat, gamma_hat and delta_hat for each time period .

Suppose that we also have data for the UK, even though we didn't include that in our regression – say the covariates are X_UK, W_UK, Z_UK. How would you interpret the quantity

a_t_hat + b_t_hat * X_UK + gamma_t_hat * W_UK + delta_t_hat * Z_UK?

Does your interpretation change if the time t is pre-Brexit or post-Brexit?

Class 17: Synthetic Controls

Difference-in-differences is based on the assumption that in the absence of the intervention, the treated and control units would follow parallel trends. Suppose we apply DiD to the "Anti-smoking campaign" example discussed last week in class. In this example, the state of CA funded an anti-smoking campaign starting January 1989, consisting of higher taxes on tobacco products; anti-smoking advertisements; and supporting local "clean indoor-air" initiatives. (It might be shocking now to realize that in the 1980s, it was allowed and normal to smoke in restaurants, planes, and all sorts of indoor spaces!) Seven other states passed some sort of legislation in the post-treatment period: MA, AZ, OR, FL passed their own state-wide programs between 1989 and 2000; and AL, HI, MD, MI, NJ, NY and WA significantly raised taxes on tobacco products. The remaining 38 states didn't pass any sort of smoking-related legislation. The outcome that we're interested in is state-wide smoking rates, as measured by cigarette sales per capita. Say we have data on this going back to 1970 and up to 2000, giving 19 years of pre-treatment data and 10 years of post-treatment data.

Do you think that the "parallel trends" assumption for DiD holds here, where the "control" is the average of the 38 states that did not pass any legislation, and the "treatment" is CA? What if we restricted to a subset of the other states, or to a time period going back less than 19 years or forward less than 10 years? If so, which other states/time periods would you pick and why? (Answer the question based on your intuition – no need to go get data or do any statistical analysis!)