

Introduction to Causal Inference

Academia Sinica ISSM — Day 2, Worksheet 2: Which X , and What Goes Wrong

Name: _____

1. Evaluate a conditional-ignorability (CI) claim

A researcher estimates the effect of college attendance on earnings by comparing college-goers to non-goers, “controlling for” high-school GPA, parental income, and region.

- (a) State what conditional ignorability requires *here* — what would have to be true about who goes to college, given those controls?

- (b) Name one plausible confounder *not* in their list, and say which way it would bias the estimate.

2. The post-treatment trap

A *randomized* job-training program (D) aims to raise earnings (Y). Noticing that some treated people dropped out of the training early, a researcher restricts the analysis to “completers” vs. controls.

- (a) Is “program completion” pre- or post-treatment?

- (b) Even though D was randomized, why can restricting to completers create bias? (Think about *who* completes vs. drops out.)

- (c) Draw a small DAG with D , completion, an unobserved “motivation,” and Y that shows the problem.

3. Which of these is a real argument for *identification*?

In the following statements, researchers try to defend a causal claim. For each statement, decide whether it is a good argument that the effect is *identified* (by CI) — and if not, name the trap.

1. “We adjusted for 40 covariates and the estimate barely moved — it’s robust.”
2. “After matching, treatment and control are balanced on every measured covariate, so unconfoundedness holds.”
3. “The 95% CI is $[0.18, 0.22]$ — very tight — so we can trust the effect.”
4. “We used a doubly-robust machine-learning estimator, so the result is credible.”
5. “The program was rolled out by lottery within each district, so treatment is as-good-as-random within district.”

4. **Balance: falsification, not a test**

A study claims CI given **age** and **location**. After subclassifying on $\text{age} \times \text{location}$: age and location come out perfectly balanced, but **parental education** is still imbalanced.

- (a) Why is the balance on age and location *unsurprising* — does it confirm anything?
- (b) What might the *residual* imbalance on parental education suggest?
- (c) Suppose instead everything you checked had balanced perfectly. Would that *prove* CI holds? Why or why not?