

## Problem Set 3 – Causal Inference 1

*Estimated time: 60–75 minutes. Type your responses and submit a single file \*ON CANVAS\*. Most answers here are a sentence or two – there is no length requirement, and padding will not help you. Where a question asks you to sign a bias or compute an estimate, show the two pieces you used to get there.*

### Part 1: Reading Coefficients

1. A researcher has data on every 8th grader in a state. She regresses math score (a 0–100 scale) on an indicator equal to 1 if the student’s teacher holds a master’s degree, and gets:

$$\widehat{Score} = 62.0 + 2.4 \cdot Masters$$

- (a) Interpret the 2.4 in one sentence.
  - (b) Because *Masters* is either 0 or 1, the 2.4 is the difference between two averages. What are they?
  - (c) Name two things this coefficient does *not* establish.
2. She now adds controls for family income (measured in \$10,000s) and an indicator for the school being high-poverty:

$$\widehat{Score} = 51.2 + 0.9 \cdot Masters + 1.6 \cdot Income - 3.4 \cdot HighPoverty$$

- (a) Interpret the 0.9. Hint: Your sentence should differ from your answer to 1(a)
- (b) Interpret the 1.6.
- (c) The coefficient on *Masters* fell from 2.4 to 0.9 when these controls were added. What does that tell you about where teachers with master’s degrees tend to teach?
- (d) Does the fact that the coefficient moved so much mean 0.9 is the causal effect of a master’s degree? Explain.

### Part 2: Omitted Variable Bias

3. For each of the following, state the likely **sign** of the omitted variable bias. You need two pieces of information each time – the relationship between the omitted variable and the outcome, and the relationship between the omitted variable and the treatment. Give both, then give the sign.

- (a) *Outcome*: college completion. *Treatment*: grant aid. *Omitted*: family income.
  - (b) *Outcome*: test scores. *Treatment*: district per-pupil spending. *Omitted*: family income. Districts in this state are funded mostly through local property taxes.
  - (c) *Outcome*: test scores. *Treatment*: class size. *Omitted*: whether the student is placed in a remedial track. The district places remedial-track students in smaller classes.
  - (d) *Outcome*: student achievement. *Treatment*: years of teaching experience. *Omitted*: the poverty rate of the school where the teacher works.
  - (e) *Outcome*: test scores. *Treatment*: hours of tutoring received. *Omitted*: the student's achievement before tutoring began.
  - (f) *Outcome*: test scores. *Treatment*: tutoring. *Omitted*: student motivation. The district *randomly assigned* students to tutoring.
4. One of the cases above produces no bias. Which one, and why? State the general condition in one sentence.

### Part 3: Experiments

5. A community college offers an optional summer bridge program. Students who attend go on to earn higher first-year GPAs than students who don't.
- (a) Give one reason this comparison might **overstate** the effect of the program.
  - (b) Give one reason it might **understate** it.
  - (c) What would you need to observe to know the true effect? Can you observe it?
6. A district randomly assigns half of its elementary schools to serve free breakfast in the classroom at the start of the school day. A year later, the treated schools have higher attendance and higher test scores than the control schools.
- (a) Can we say test scores went up *because of* the increase in attendance? Why or why not?
7. You run an experiment on a college advising program and find no effect on graduation for the sample as a whole. The funder wants you to show some effect and encourages you to look for a group it worked for. You split the

sample into 20 groups and find significant effects for students from Houston and for students who played a varsity sport.

- (a) What is the problem with these results?
- (b) Do they suggest the college should keep the program, but only in Houston? Explain.
- (c) Name one thing you could have done *before* running the study that would have protected you here.

#### **Part 4: Instrumental Variables**

8. An oversubscribed magnet school admits students by lottery. Of the students offered a seat, 85% enroll. Of the students not offered a seat, 10% enroll anyway, through the waitlist or a later transfer. Students who are offered a seat score 4.2 points higher on the state exam than students who are not.

- (a) What is the first stage?
- (b) What is the reduced form, and what does it mean in words?
- (c) What is the estimated effect of *attending* the magnet school?
- (d) Why is your answer to (c) larger than your answer to (b)?

9. Staying with the magnet school lottery:

- (a) What fraction of applicants are always takers, never takers, and compliers?
- (b) Your estimate is the effect for which of those groups? Who does it tell you nothing about?
- (c) A board member reads your report and says it shows the magnet school raises scores by that amount for any student who enrolls. What is wrong with that statement?