

Introduction to Causal Inference

Academia Sinica ISSM — Day 1, Worksheet 1: Potential Outcomes

Name: _____

1. What is the causal question?

Which best captures the causal question in the potential-outcomes framework?

- (a) Compare average outcomes between treated and untreated groups
- (b) Compare outcomes for the same unit under treatment vs. control
- (c) Compare treated units to similar control units
- (d) Compare before-and-after outcomes for treated units

2. Observability

For a given unit i , classify each as *always observable* (AO), *sometimes observable* (SO), or *never observable* (NO):

- $Y_i(1)$ and $Y_i(0)$ (both at once)
- D_i
- Y_i
- $\tau_i = Y_i(1) - Y_i(0)$

3. Difference in means

Consider $\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$. Which is correct?

- (a) Always equals the ATE
- (b) Always equals the ATT
- (c) Equals the ATT plus selection bias
- (d) Equals the ATC

4. Direction of bias

Suppose $Y_i(1) = Y_i(0)$ for all i (no treatment effect for anyone), but treatment is more likely for units with higher $Y_i(1)$. What do you expect for

$$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] ?$$

- (a) Zero
- (b) Positive
- (c) Negative
- (d) Cannot tell

5. Science table

i	D_i	$Y_i(1)$	$Y_i(0)$
1	1	10	4
2	1	1	2
3	0	3	3
4	0	5	2

Compute the two quantities below.

- ATE:
- ATT:
- Under $Y_i(1)$ and $Y_i(0)$, circle the numbers that are quantities you would actually observe in a real experiment. Then compute the difference-in-means estimator $\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$.

6. Bring it together: The big identification result.

Suppose treatment D_i is randomly assigned, so that $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp D_i$. Starting from the ATE, fill in the missing expressions.

$$\begin{aligned}
 \tau_{\text{ATE}} &= \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\
 &= \mathbb{E}[Y_i(_) \mid D_i = _] - \mathbb{E}[Y_i(_) \mid D_i = _] \quad (\text{ignorability}) \\
 &= \mathbb{E}[_ \mid D_i = 1] - \mathbb{E}[_ \mid D_i = 0] \quad (\text{consistency}) \\
 &\approx \bar{Y}_{D=1} - \bar{Y}_{D=0} \quad (\text{difference-in-means estimator})
 \end{aligned}$$

In one line: which single step relied on the treatment being randomized?