

Causality

Erick F Molina

Summer, 2024

ITAM

Potential outcomes

Potential outcomes

- A treatment (T) induces 2 *potential outcomes* for individual i .
 - The untreated outcome $Y_i(0)$
 - The treated outcome $Y_i(1)$
- The observed outcome:

$$\begin{aligned} Y_i &= \begin{cases} Y_i(1) & \text{if } T_i = 1 \\ Y_i(0) & \text{if } T_i = 0 \end{cases} \\ &= Y_i(0) + (Y_i(1) - Y_i(0)) T_i \end{aligned}$$

- The individual-level impact is $\delta_i = Y_i(1) - Y_i(0)$
- The fundamental problem of causal inference:
We can never observe both potential outcomes for the same individual.

Potential Outcomes

- We cannot simply compare treated individuals and untreated individuals.
- We observe $Y_i = Y_i(0) + \underbrace{(Y_i(1) - Y_i(0))}_{\delta_i = \text{impact}} T_i$
- If we compare both outcomes:

$$\begin{aligned} \underbrace{\mathbb{E}(Y_i | T_i = 1) - \mathbb{E}(Y_i | T_i = 0)}_{\text{Observed difference}} &= \mathbb{E}(Y_i(1) | T_i = 1) - \mathbb{E}(Y_i(0) | T_i = 1) + \\ &\quad \mathbb{E}(Y_i(0) | T_i = 1) - \mathbb{E}(Y_i(0) | T_i = 0) \\ &= \underbrace{\mathbb{E}(Y_i(1) | T_i = 1) - \mathbb{E}(Y_i(0) | T_i = 1)}_{\text{average treatment effect on the treated}} + \\ &\quad \underbrace{\mathbb{E}(Y_i(0) | T_i = 1) - \mathbb{E}(Y_i(0) | T_i = 0)}_{\text{selection bias}} \end{aligned}$$

Important definitions

- Average Treatment Effect (ATE)

$$\mathbb{E}[\delta_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$$

- Average Treatment Effect on the Treated (ATT) is the average treatment effect conditional on being in the treated group:

$$\begin{aligned}\mathbb{E}[\delta_i | T_i = 1] &= \mathbb{E}[Y_i(1) - Y_i(0) | T_i = 1] \\ &= \mathbb{E}[Y_i(1) | T_i = 1] - \mathbb{E}[Y_i(0) | T_i = 1]\end{aligned}$$

- Average Treatment Effect on the Untreated (ATU) is the average treatment effect conditional on being in the untreated group:

$$\begin{aligned}\mathbb{E}[\delta_i | T_i = 0] &= \mathbb{E}[Y_i(1) - Y_i(0) | T_i = 0] \\ &= \mathbb{E}[Y_i(1) | T_i = 0] - \mathbb{E}[Y_i(0) | T_i = 0]\end{aligned}$$

Correlation, causality, and false counterfactuals

Correlation, causality, and false counterfactuals

Correlation vs. causality:

- If I go to school (T), will I earn more money (Y)?
- Are more years of schooling associated with higher wages?
- $\frac{\text{Cov}(T, Y)}{\sigma_T \sigma_Y}$

We are analyzing false counterfactuals if:

- We compare participants vs. non-participants in a study.
- We do a simple pre/post comparison around treatment.

Correlation $\neq$ Causality: reverse causality

- If two variables A and B are correlated, which one causes which?
- Example: lice in the Middle Ages.
- It may also be that A causes B and B causes A. Example: depression and drug addiction.
- Spurious correlation.

The Experimental ideal

The experimental ideal

- We randomly assign T_i .
- T_i is independent of $Y_i(0)$ and $Y_i(1)$. Recall:

$$\begin{aligned}\underbrace{\mathbb{E}(Y_i|T_i=1) - \mathbb{E}(Y_i|T_i=0)}_{\text{Observed difference}} &= \mathbb{E}(Y_i(1)|T_i=1) - \mathbb{E}(Y_i(0)|T_i=1) + \\ &\quad \mathbb{E}(Y_i(0)|T_i=1) - \mathbb{E}(Y_i(0)|T_i=0) \\ &= \underbrace{\mathbb{E}(Y_i(1)|T_i=1) - \mathbb{E}(Y_i(0)|T_i=1)}_{\text{average treatment effect on the treated}} + \\ &\quad \underbrace{\mathbb{E}(Y_i(0)|T_i=1) - \mathbb{E}(Y_i(0)|T_i=0)}_{\text{selection bias}}\end{aligned}$$

- Given random assignment of treatment, the selection bias term disappears.¹

¹With random assignment, $\mathbb{E}(Y_i(0)|T_i=1) = \mathbb{E}(Y_i(0)|T_i=0)$.

The experimental ideal

- Since $\mathbb{E}(Y_i(1) - Y_i(0) | T_i = 1) = \mathbb{E}(Y_i(1) - Y_i(0))$, we can drop the “treated” conditioning:

$$\mathbb{E}(Y_i | T_i = 1) - \mathbb{E}(Y_i | T_i = 0) = \underbrace{\mathbb{E}(Y_i(1) - Y_i(0))}_{\text{ATE}}$$

- Despite randomization, we cannot identify δ_i , because the individual-level missing counterfactual remains.
- What randomization does is eliminate differences *on average*. In other words, the treatment and control groups are similar by the LLN.
- **Stable Unit Treatment Value Assumption (SUTVA):** The potential outcomes for each unit do not vary with the treatments assigned to other units.

Stable Unit Treatment Value Assumption: SUTVA

- It implies a homogeneous dose, i.e., the treatment is the same for everyone.
- Potential outcomes do not vary with who receives the treatment or how many receive it.
- Partial equilibrium.