A Primer on Causal Inference: Exploring the Effect of Living in an Urban Area on Wages

Mentees: Ridhika Madan, Ivena Yeung, Catherine McCulloch
Mentor: Xiaoya Wang

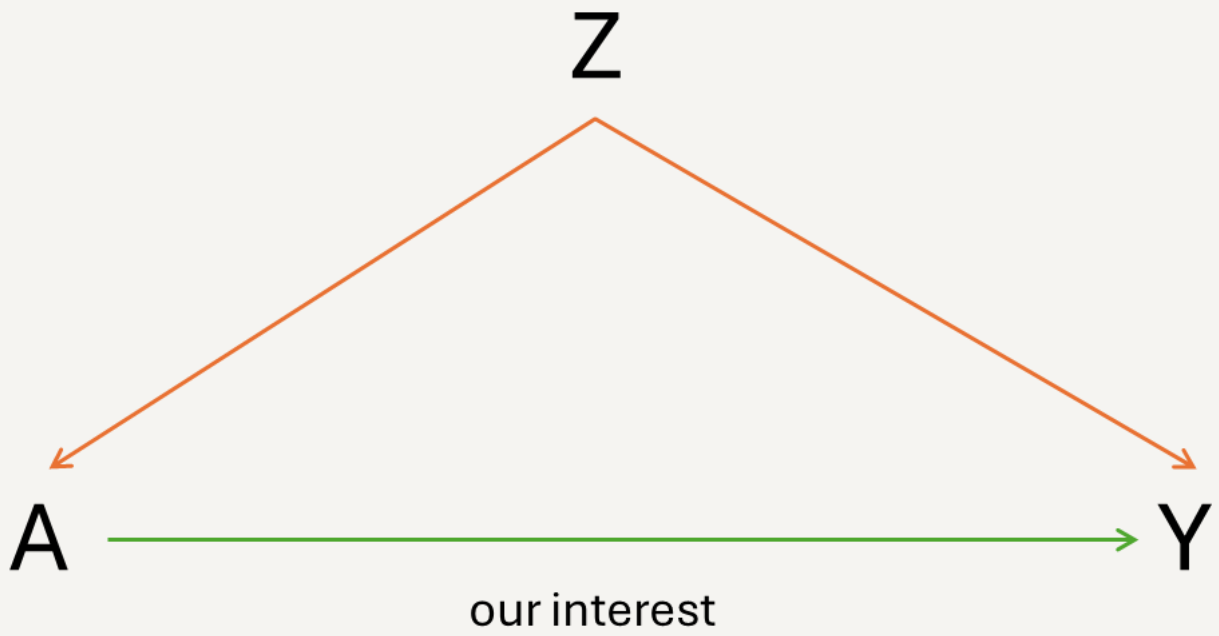
What is causal inference?

Goal: How does a particular phenomenon (a treatment) *actually* effect an outcome of interest?

Assumptions:

- Consistency
- Exchangeability
- Positivity

Causal Diagram



Where…

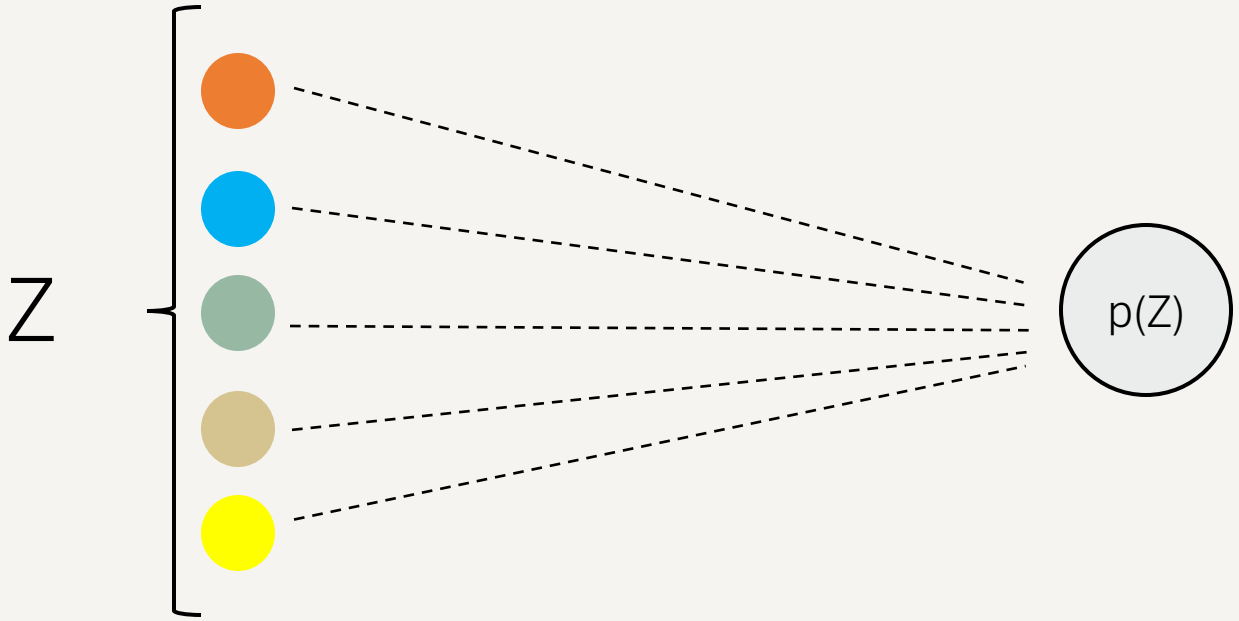
- A is our treatment,
- Y is our outcome, and
- Z is our potential confounder(s)

How to adjust for confounders?

Propensity score (Rosenbaum and Rubin, 1983) – denoted p(Z)

- Converts values for a set of confounders into 1 scalar value
- Probability of receiving treatment given a set of potential confounders, Z

$$p(Z) = P(A=1 \mid Z)$$



Research Question: What is the effect of living in an SMSA (Standard Metropolitan Statistical Area) on log wages?

Data: National Longitudinal Survey Young Men Cohort (1966 to 1981)

Treatment: SMSA (urban, 0 or 1)
Outcome: lwage (log wages)

Covariates (Confounders):
educ - Years of education
exper - Years of work experience
black - If the individual is black
south - In the southern United States
married - Is married
nearc4 - There is a four-year college in the county

Method: Propensity Score Regression Adjustment (Vansteelandt and Daniel, 2014)

Steps:

1. Estimate the propensity scores for each individual in the experiment
    - Logistic regression model between treatment variable and confounders
2. Fit a linear regression model to the outcome
    - Only need to consider the treatment values and propensity scores
    - $Y = A + p(Z)$

Method: Inverse Probability Weighting (Robins, 2000)

1. Estimate the propensity scores for each individual in the study
   - Use logistic regression model to predict the **propensity score** (ps), or the probability of being treated (living in an SMSA) given a set of covariates (education, experience, etc.)

2. Weight groups by the **inverse** of their probability of being treated
   - For treated groups: $\frac{1}{P(smsa=1)}$
   - For control groups: $\frac{1}{1-P(smsa=1)}$

3. Fit weighted regression model on outcome (lwage) using the IP weights
4. Estimate **Average Treatment Effect (ATE)** of being in an SMSA on log wages (lwage)

Method: Nearest Neighbour Matching (Rubins and Thomas, 1996)

1. Estimate the propensity scores for each individual in the experiment.

2. Pair each treated individual to one or more untreated individuals with the nearest propensity score(s).

3. Estimate ATE.

Results

$ATE_{regression}$ = 0.168; std. error = 0.0188

ATE_{IPW} = 0.161; std. error = 0.0155

$ATE_{matching}$ = 0.315; std. error = 0.0209

References

- Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In Statistical models in epidemiology, the environment, and clinical trials (pp. 95-133). New York, NY: Springer New York.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.
- Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. Biometrics, 249-264.
- Vansteelandt, S., & Daniel, R. M. (2014). On regression adjustment for the propensity score. Statistics in medicine, 33(23), 4053-4072.