Causal Inference

Shirin Mojarad, Ph.D., Data Scientist

Outline

- Introduction
- Randomized control trials
- Causal inference basic concepts (counterfactuals, potential outcomes, etc)
- Causal Inference Methods
- Hands-on Session

Outline

Introduction

- Why?
  - Why do we care about finding causality?

- What?
  - What is causal inferencing?
  - What are the current methods and assumptions?

Why finding causality?

Many of the critical questions are causal in nature:

- Does having health insurance make you healthier?
- Does the new curriculum help students learn better?
- Does the new marketing campaign attract more sales?
- Does the new diet help you lose weight?

Causal Inference

- Does a relation from cause to effect exist?
- The most challenging empirical questions in educational research involve cause-effect relationships
    - Does the new curriculum improve student engagement?
    - Does the new ITS improve learning outcomes?
    - What drives student success?

McGraw Hill Education

Prediction vs. Causality

- Predictions can be made without finding the causality
- In treatment assignment, we want to investigate how the treatment has affected the outcome

Identifying causal impact

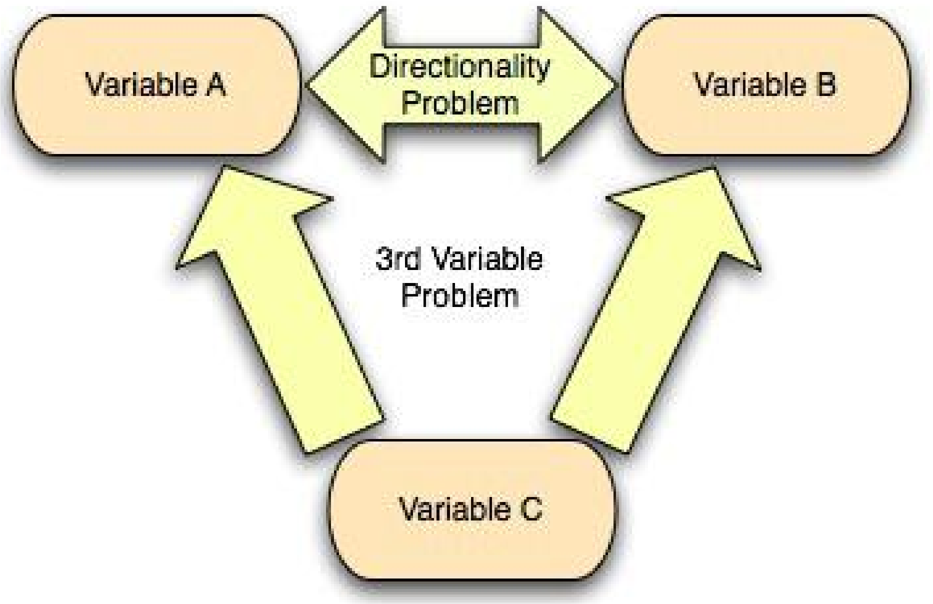
- Evaluate the impact/effect of a program or an intervention on some outcomes of interest
- By how much did treatment/intervention change the outcome?

Why not correlation?

- Basis of many scientific hypothesis
- Causation imply correlation
- Complex correlational designs allow for some limited causal inference

Correlation or Causation?

- Directionality problem
- Third-variable problem



Assumptions in Causality

- An effect could have multiple causes
- Each cause might have different weights
- There might be hidden or unknown causes

How to Establish Causality?

What is the effect of an intervention/treatment T on outcome Y ?

Example: What is the effect of a new curriculum (T) on student learning outcomes (Y)?

Impact of T =

Student outcome (Y) for a students using the new curriculum

–

Student outcome (Y) for the same students not using the new curriculum (at the same point in time)

Fundamental Problem of Causality

- We observe learning outcome (Y) for a student using the new curriculum
- But we do not observe learning outcome (Y) for the same student in the absence of the new curriculum

Fundamental problem: We never observe the same individual with and without the treatment at the same point in time

Solution

- Estimate a proxy for what would have happened to outcome Y in the absence of treatment T
- For example, compare the students who ‘look’ exactly like those who were exposed to the new curriculum at the same point of time

In other words, we must find a valid **Counterfactual** or **Control group**

Finding a Control Group

Understand the process by which treatment is determined:

- How are students assigned?
- The counterfactual must be similar in terms of the likelihood of treatment/program participation
- The treated observation and the counterfactual should have **identical characteristics**, except for benefiting from the intervention
  only reason for different outcomes between treatment and counterfactual is the treatment/intervention (T)

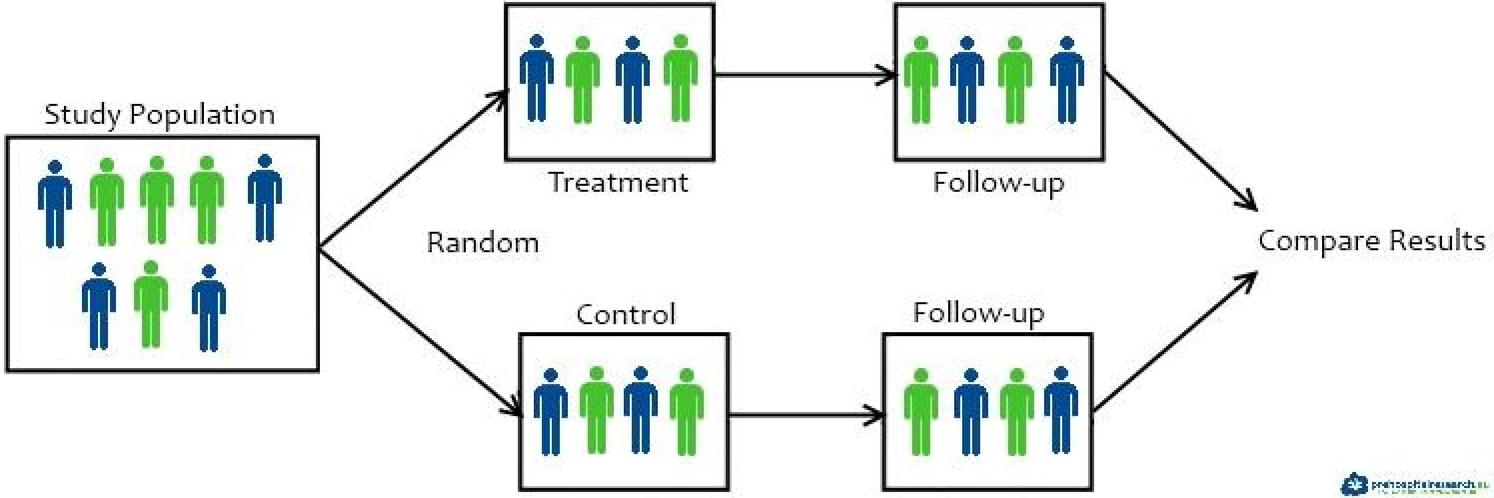
Finding a Control Group

- Guarantee comparability of treatment and control groups
- ONLY remaining difference is intervention

How?

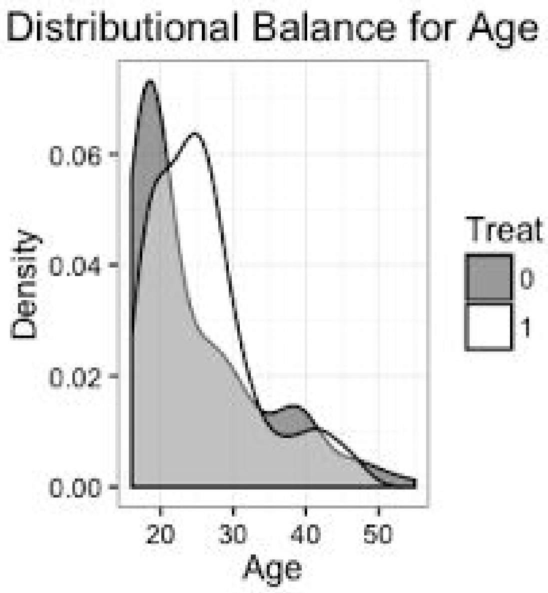
- Experimental design (randomized controlled trial)
- Non-experimental/ Quasi-experimental design (observational studies)

Randomized Controlled Trial



Why Randomization?

Randomization balances out all other causes (covariates), besides treatment, between treatment and control groups.



Elements of Causality

- Covariate: any causes, besides treatment, that can affect outcome
- Covariate balance: treated and control groups with similar covariate distributions

Challenges with RCT

- Expensive
- Bias
- Covariate balance
- Ethics

Quasi-experimental Design

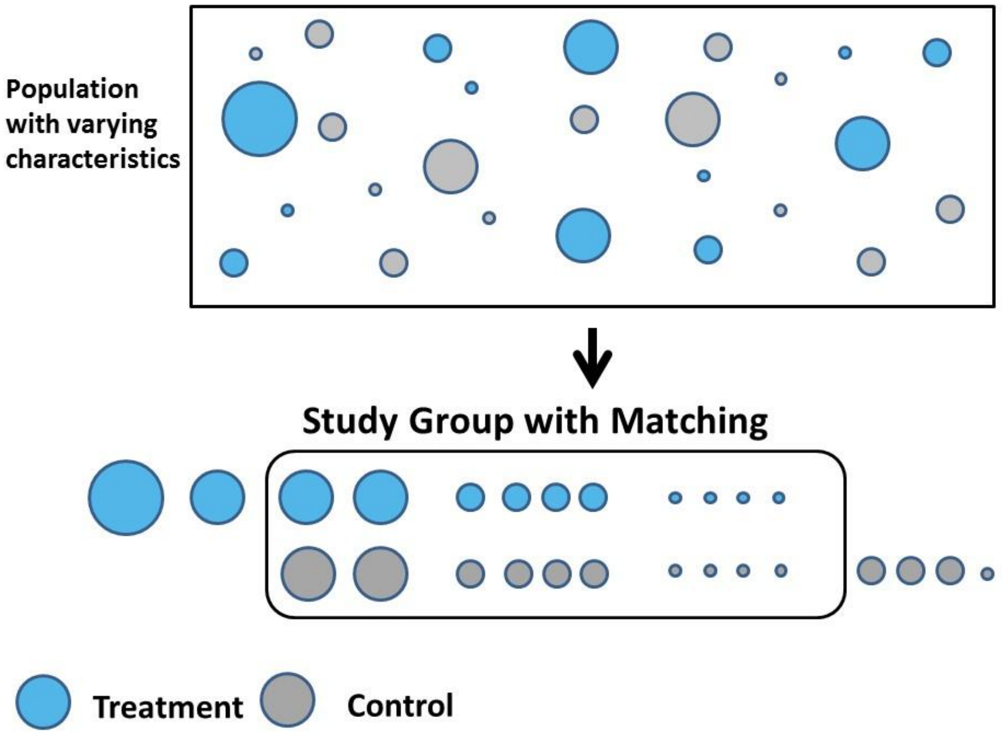
- Find a control group, that is as similar as possible to the group who received the treatment, the only difference is the treatment
- Guarantee comparability of treatment and control groups
- Similar to the RCT, but lacks “random assignment”
- Easily and more frequently implemented

McGraw Hill Education

Quasi-experimental Design: Matching

Matching methods attempt to replicate, as closely as possible, the ideal of randomized experiments when using observational data.

Matching


Population with varying characteristics
Study Group with Matching
Treatment
Control

Your Toolkit For Finding Causality

T-test

- A form of hypothesis testing, testing the difference between two populations averages.
- The t-value measures the size of the difference relative to the variation in your sample data.
- H0 (null hypothesis) claims “no difference”
- Ha (alternative hypothesis) contradicts the null

Statistical Significance (P-value)

A significance indicates whether or not the difference between two groups' averages most likely reflects a "real" difference in the population from which the groups were sampled

Smaller-and-smaller P-values → stronger-and-stronger evidence against H0

$P > .10$ evidence against H0 not significant
$.05 < P \leq .10$ evidence marginally significant
$.01 < P \leq .05$ evidence against H0 significant
$P \leq .01$ evidence against H0 very significant

Effect Size

- Effect size is a quantitative measure of the strength of a relationship.
- Effect size or Cohen's d is the difference in the two groups' means divided by the average of their standard deviations:

$$d = \frac{M_{group1} - M_{group2}}{SD_{pooled}}$$

where pooled standard deviation is:

$$SD_{pooled} = \sqrt{(SD^2_{group1} + SD^2_{group2})/2}$$

0.2 is considered a 'small', 0.5 represents a 'medium' , and 0.8 is a 'large' effect size.

Matching Metrics

Matching methods examine how to best choose treated and control subjects for comparison.

Matching Metrics

- Propensity score matching: matching treated and control subjects using propensity score:

$$P(X) = Pr\ (Tr=1|X)$$

- Mahalanobis distance matching: matching treated and control subjects using Mahalanobis distance. Mahalanobis distance between two vectors, x and y, where S is the covariance matrix is:

$$d_M(x,y) = \sqrt{(x-y)^T S^{-1} (x-y)}$$

Caliper

A match for person *i* is selected only if their propensity score difference is below a specified level (ε):

$$|Pi-Pj| < \varepsilon$$

Larger differences will not result in matches, and all units whose differences lie within the caliper's radius will be chosen.

Recommended caliper size: 0.25 σp (pooled standard deviation of the logit of the propensity score)

Sensitivity Test

- A technique to measure the impact that any external factor not taken into account in a study could have on the results
- Rbounds is an R package to run sensitivity tests
- Rbounds reference paper: http://www.personal.psu.edu/ljk20/rbounds%20vignette.pdf