JPM031: Research Methods

Jakub Stauber

Estimating Causal Effects with Observational Data and the Problem of Confounders

Plan for Today

- Review: Causation and Randomized Experiments
- Observational Studies
- Confounding Variables or Confounders
  - Why Are Confounders a Problem?
  - Why Don't We Worry About Confounders in Randomized Experiments?
- How Can We Estimate Causal Effects with Observational Data?
  - Interpretation of $\widehat{\beta}$ When X Is the Treatment Variable and Y Is the Outcome Variable

Review: Causation

- To measure causal effects, we need to compare the factual outcome with the counterfactual outcome
  - Fundamental problem: We can never observe the counterfactual outcome
- To estimate causal effects, we must find or create a situation in which the treatment and control groups are **comparable** with respect to all the variables that might affect the outcome other than the treatment variable itself
- Only when that assumption is satisfied can we use the factual outcome of one group as a good proxy for the counterfactual outcome of the other, and vice versa, thus, bypassing the fundamental problem of causal inference

Review: Randomized Experiments

- In randomized experiments, we can rely on the **random assignment of treatment** to make treatment and control groups, on average, identical to each other in terms of all observed and unobserved pre-treatment characteristics
- Thus, we can estimate the average treatment effect with the **difference-in-means estimator**

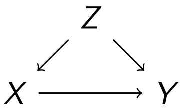
$$\overline{Y}_{\text{treatment group}} - \overline{Y}_{\text{control group}}$$

Observational Data

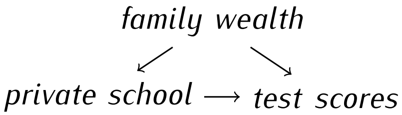
- But, what happens when we cannot conduct a randomized experiment and have to analyze observational data?
  - Observational data: data collected about naturally occurring events (i.e., researchers do not get to assign the treatment)
- We can no longer assume that treatment and control groups are comparable
- We need to identify and measure any relevant differences between treatment and control groups (known as confounding variables or confounders)
- Then, we will need to statistically control for them so that we can make the two groups comparable *after statistical controls are applied*

Confounders or Confounding Variables

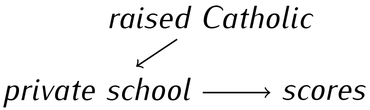
- A **confounding variable** is a variable that affects both
  - (i) the likelihood to receive the treatment X and
  - (ii) the outcome Y
- In mathematical notation, we represent a confounding variable as Z

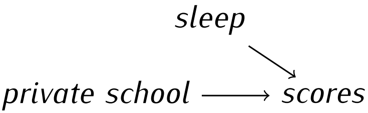


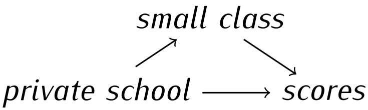
- Let's look at a simple example. Suppose we are interested in the average causal effect of attending a private school, as opposed to a public one, on students test performance
  - What is the treatment variable X?
  - What is the outcome variable Y?
  - Can you think of a confounder Z?



- Not confounders:

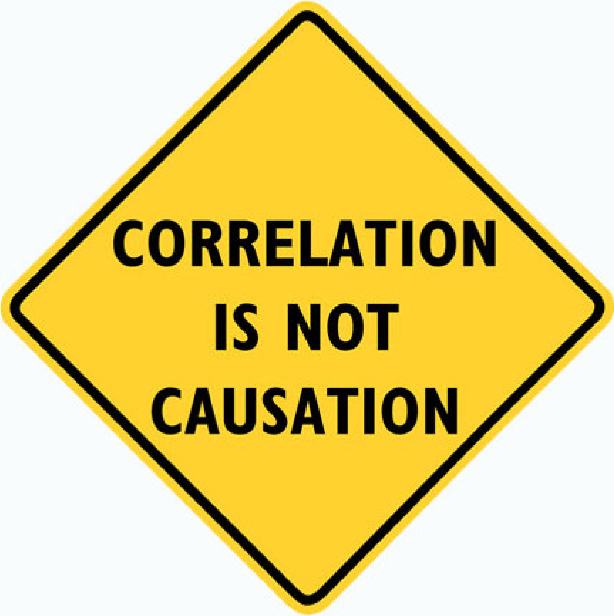




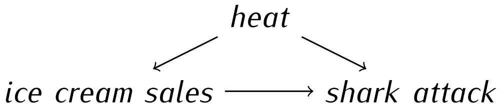


Why Are Confounders a Problem?

- They obscure the causal relationship between X and Y
- In the example above, if we observed that, on average, private school students perform better than public school students, we would not know whether it is
  - because they attended a private school or
  - because they came from wealthier families that could afford to provide them with after-school help
- We would not know what portion of the observed differences in test score performance (the difference-in-means estimator), if any, could be attributed to the causal effect of the treatment (attending a private school) and what portion could be attributed to the confounding variable (coming from a wealthy family)

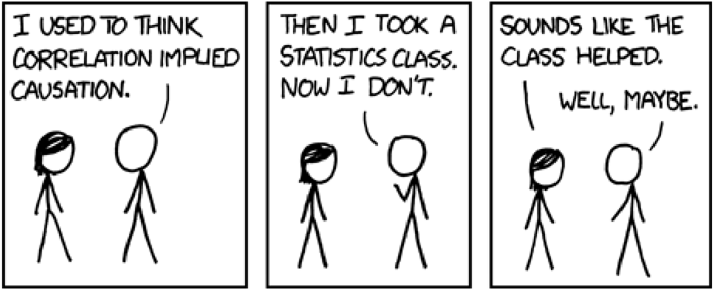

CORRELATION
IS NOT
CAUSATION

- In the presence of confounders, correlation does not necessarily imply causation
- Just because we observe two variables highly correlated with each other—when we observe one increase, we usually observe the other increase or decrease—it does not automatically mean that one causes the other
  - There could be a third variable that causes both
- For example, ice cream sales and shark attacks are highly correlated with each other. Does this mean that eating ice cream increases the probability that a shark attacks you?



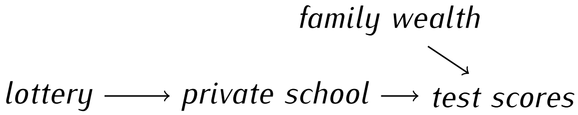
IN THE PRESENCE OF CONFOUNDERS

- correlation does NOT necessarily imply causation
- the treatment and control groups are NOT comparable
- the difference-in-means estimator does NOT provide a valid estimate of the average treatment effect



Why Don't We Worry About Confounders in Randomized Experiments?

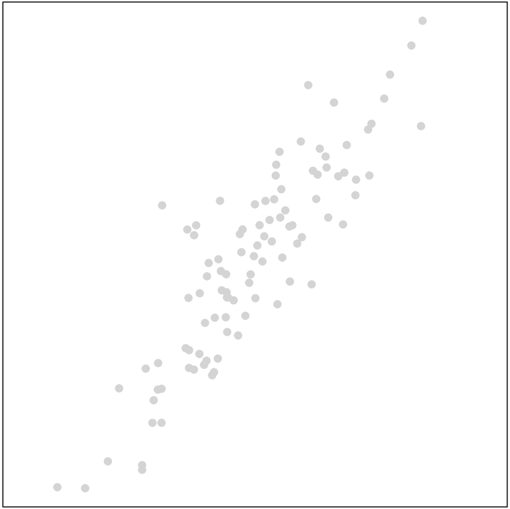
- Randomization of treatment assignment eliminates all potential confounders
- It ensures that treatment and control groups are comparable by breaking the link between any potential confounder and the treatment
- If we assign who attends a private school at random, we ensure that nothing related to the outcome is also related to the likelihood of receiving the treatment



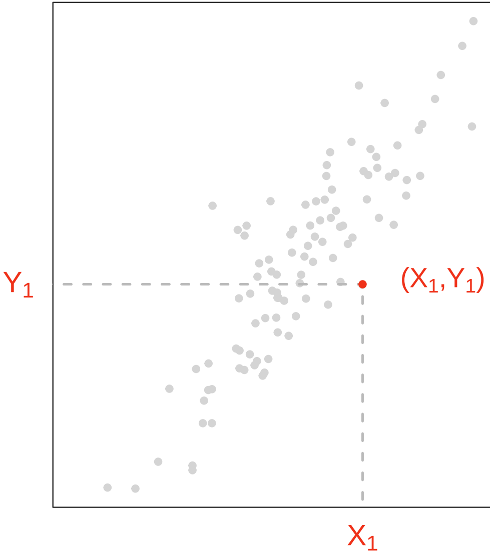
How Can We Estimate Causal Effects with Observational Data?

- We cannot rely on random treatment assignment to eliminate potential confounders
- We need to identify and measure all confounding variables and statistically control for them
- Before we learn how to do that, we should learn how to fit a simple linear regression model to produce an estimated coefficient equivalent to the difference-in-means estimator
- Let's quickly review how we fit a line and interpret the estimated coefficients

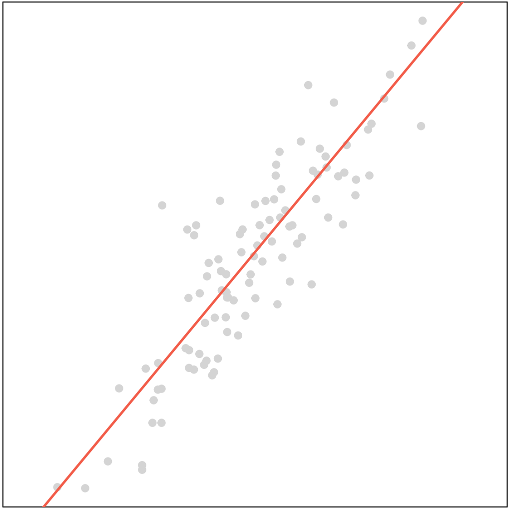
scatter plot where every dot is an observation



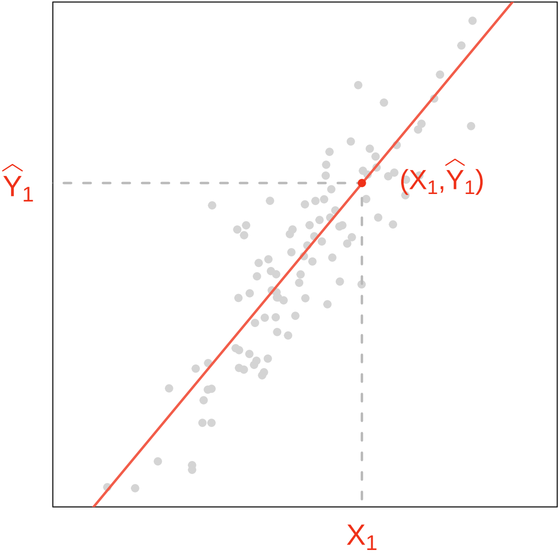
first observation: (X_1, Y_1)



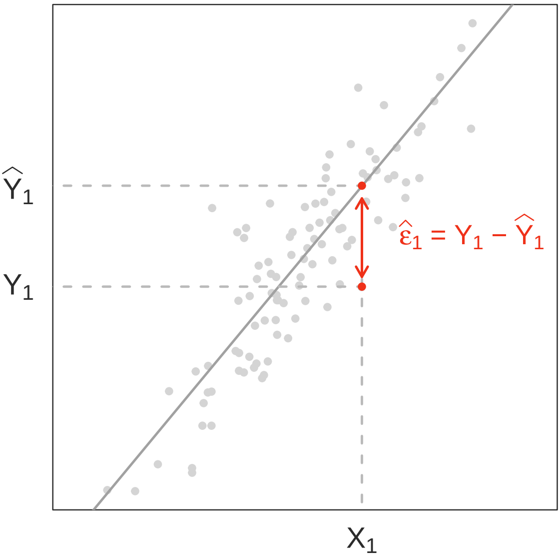
if we summarize the relationship between X and Y with a line



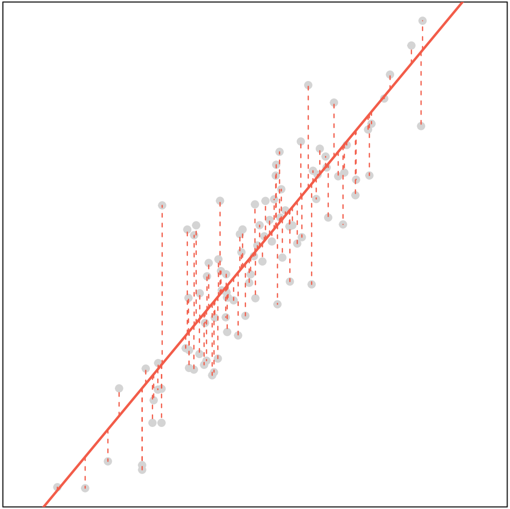
we can use the fitted line, to compute $\widehat{Y}$ for every value of X



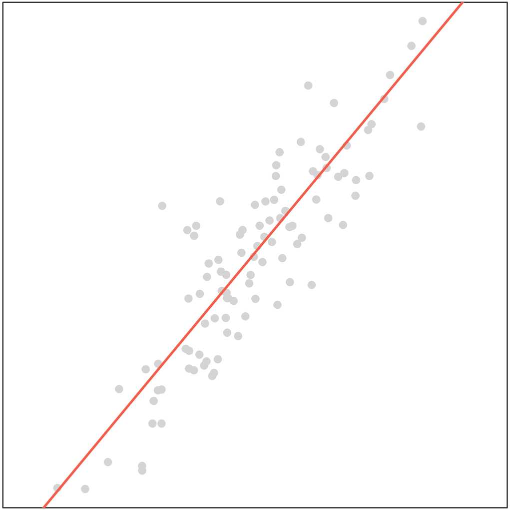
prediction errors = vertical distance between dots and line



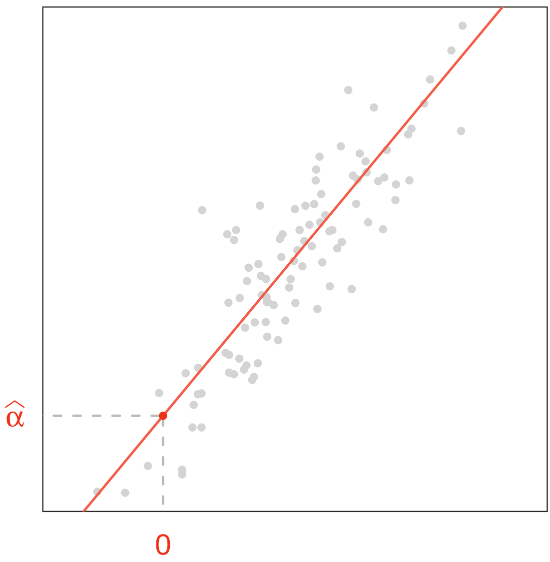
we choose the line with the smallest possible errors



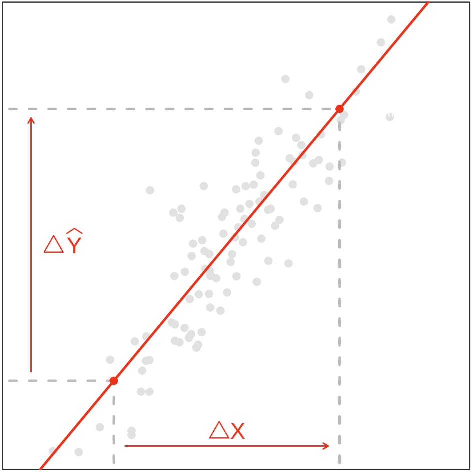
the fitted line: $\widehat{Y} = \widehat{\alpha} + \widehat{\beta}X$



estimated intercept ($\widehat{\alpha}$): $\widehat{Y}$ when X=0



estimated slope ($\widehat{\beta}$): $\triangle \widehat{Y}$ associated with $\triangle X$=1



Using the Simple Linear Model to Compute the Difference-in-Means Estimator

When X is the treatment variable and Y is the outcome variable of interest, the estimated slope coefficient ($\widehat{\beta}$) is equivalent to the difference-in-means estimator.

- Let's examine this ...

- Mathematical definition of $\widehat{\beta}$: $\triangle \widehat{Y}$ associated with $\triangle X$=1

$$
\begin{aligned}
\widehat{\beta} &= \triangle \widehat{Y} && \text{(if } \triangle X=1) \\
&= \widehat{Y}_{\text{final}} - \widehat{Y}_{\text{initial}} && \text{(if } \triangle X=1)
\end{aligned}
$$

- If X is the treatment variable:
  - $\triangle X$=1 is equivalent to changing from the control group (X=0) to the treatment group (X=1)
  - the control group is the initial state, and the treatment group is the final state

$$\widehat{\beta} = \widehat{Y}_{\text{treatment group}} - \widehat{Y}_{\text{control group}}$$

$$\widehat{\beta} = \widehat{Y}_{\text{treatment group}} - \widehat{Y}_{\text{control group}}$$

- Recall: $\widehat{Y}$ are *average* predicted values. In this case: $\widehat{Y}_{\text{treatment group}} = \overline{Y}_{\text{treatment group}}$ and $\widehat{Y}_{\text{control group}} = \overline{Y}_{\text{control group}}$

$$\widehat{\beta} = \overline{Y}_{\text{treatment group}} - \overline{Y}_{\text{control group}}$$

- Conclusion: When X is the treatment variable and Y is the outcome variable of interest, the estimated slope coefficient ($\widehat{\beta}$) is equivalent to the difference-in-means estimator

- Let's return to the exercise from Lecture 6: Does Social Pressure Affect Turnout?
- We answer it by analyzing data from a randomized experiment where registered voters were randomly assigned to either (a) receive a message designed to induce social pressure, or (b) receive nothing

Does Social Pressure Affect Turnout?

(Based on Alan S. Gerber, Donald P. Green, and Christopher W. Larimer. 2008. “Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review*, 102 (1): 33-48.)

1. Load and look at the data

```
voting <- read.csv("voting.csv") # loads and stores data
```

```
head(voting) # shows first six observations
##   birth message voted
## 1  1981      no     0
## 2  1959      no     1
## 3  1956      no     1
## 4  1939     yes     1
## 5  1968      no     0
## 6  1967      no     0
```

2. Create treatment variable

```
voting$pressure <-
  ifelse (voting$message=="yes",
          1, 0) # creates treatment variable
```

- Make sure the new variable was created correctly by looking at the first few observations again:

```
head(voting) # shows first six observations
##   birth message voted pressure
## 1  1981      no     0        0
## 2  1959      no     1        0
## 3  1956      no     1        0
## 4  1939     yes     1        1
## 5  1968      no     0        0
## 6  1967      no     0        0
```

3. Compute difference-in-means estimator directly

```r
mean(voting$voted[voting$pressure==1]) -
    mean(voting$voted[voting$pressure==0])
## [1] 0.08130991
```

4. Alternatively, we can fit a linear model where X is the treatment variable and Y is the outcome variable

- Recall: the R function to fit a linear model is lm()
  - required argument: a formula of the type $Y \sim X$

```
lm(voting$voted ~ voting$pressure) # or
```

```
lm(voted ~ pressure, data=voting)
##
## Call:
## lm(formula = voted ~ pressure, data = voting)
##
## Coefficients:
## (Intercept)     pressure
##     0.29664      0.08131
```

- Fitted model: $\widehat{voted} = 0.30 + 0.08\ pressure$
- Note that $\widehat{\beta}$ has the same value as the difference-in-means estimator above (both equal 0.08)

Interpretation of $\widehat{\beta}$ When X Is the Treatment Variable and Y Is the Outcome Variable

- Start same as in predictive models
  - definition: $\widehat{\beta}$ is the $\triangle \widehat{Y}$ associated with $\triangle X$=1
  - here: $\widehat{\beta} = 0.08$ is the $\triangle \widehat{voted}$ associated with $\triangle pressure$=1
  - in words: receiving the message inducing social pressure (i.e., an increase in *pressure* of 1 by going from pressure=0 to pressure=1) is associated with a predicted increase in the probability of voting of 8 percentage points, on average
- unit of measurement of $\widehat{\beta}$? same as $\triangle \overline{Y}$; here, Y is binary so $\triangle \overline{Y}$ is measured in p.p and so is $\widehat{\beta}$ (after x 100)

- Now, since here X is the treatment variable and Y is the outcome variable of interest, $\widehat{\beta}$ is equivalent to the difference-in-means estimator
- As a result, we can interpret $\widehat{\beta}$ using **causal langauge**
- Predictive language: We estimate that receiving the message inducing social pressure ***is associated with a predicted increase*** in the probability of voting of 8 percentage points, on average
- Causal language: We estimate that receiving the message inducing social pressure ***increases*** the probability of voting by 8 percentage points, on average

- This should be a valid estimate of the average treatment effect if there are no confounding variables present
  - if registered voters who received the message are comparable to the registered voters who did not
- Since the data come from a randomized experiment there should be no confounding variables
- And thus the difference-in-means estimator should produce a valid estimate of the average treatment effect

- Whether we compute the difference-in-means estimator directly or we fit a simple linear model where Y is the outcome variable and X is the treatment variable, we arrive to the same conclusion

- Conclusion: We estimate that receiving the message inducing social pressure increases the probability of voting by 8 percentage points, on average. This is a valid estimate of the average treatment effect if registered voters who received the message are comparable to the registered voters who did not (that is, if there a no confounding variables). Given that the data come from a randomized experiment, this is a reasonable assumption.

INTERPRETATION OF THE ESTIMATED SLOPE COEFFICIENT IN THE SIMPLE LINEAR MODEL:

- By default, we interpret $\widehat{\beta}$ using predictive language: It is the $\triangle\widehat{Y}$ *associated with* $\triangle X{=}1$.
- When X is the treatment variable, then $\widehat{\beta}$ is equivalent to the difference-in-means estimator and, thus, we interpret $\widehat{\beta}$ using causal language: It is the $\triangle\widehat{Y}$ *caused by* $\triangle X{=}1$ (the presence of the treatment). This causal interpretation is valid if there are no confounding variables present and, thus, the treatment and control groups are comparable.