

Recitation, Week 4

Ye Wang

New York University

POL-850

Spring 2018

Outline

Recitation, Week 4

shortname

Causality

Comparable Cases

Random Assignment

Internal/External
Validity

Stata

1. Causality
2. Comparable Cases
3. Random Assignment
4. Internal/External Validity

Causality

Recitation, Week 4

shortname

Causality

Comparable Cases

Random Assignment

Internal/External
Validity

Stata



unit

Causality

Recitation, Week 4

shortname

Causality

Comparable Cases

Random Assignment

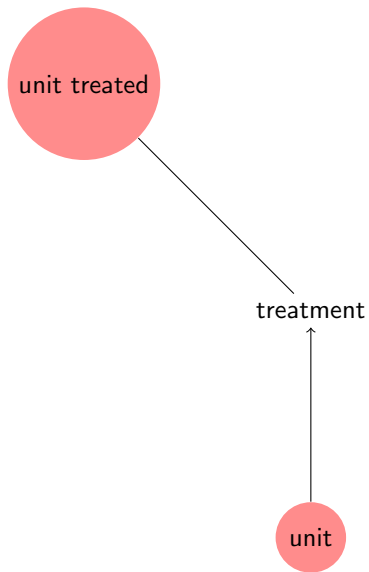
Internal/External
Validity

Stata

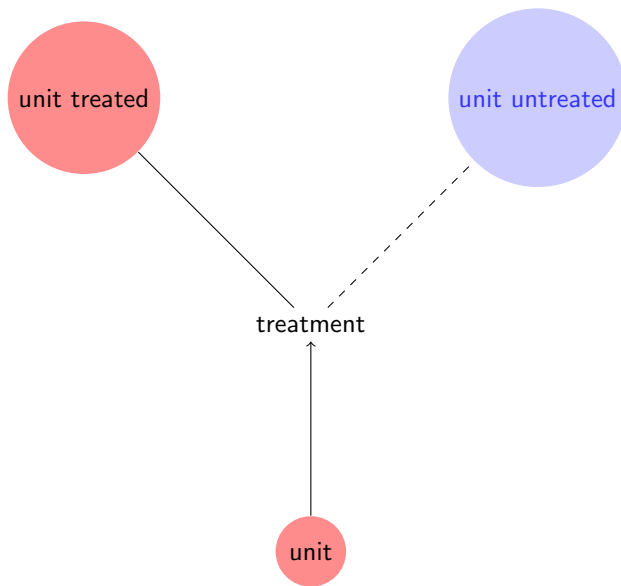
treatment



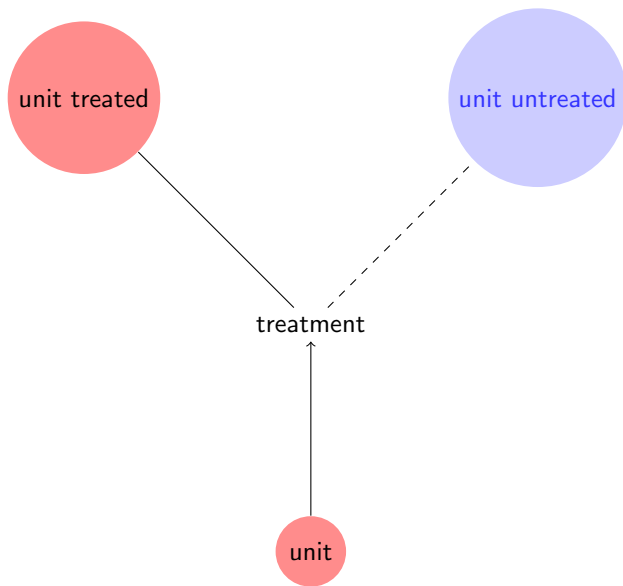
Causality



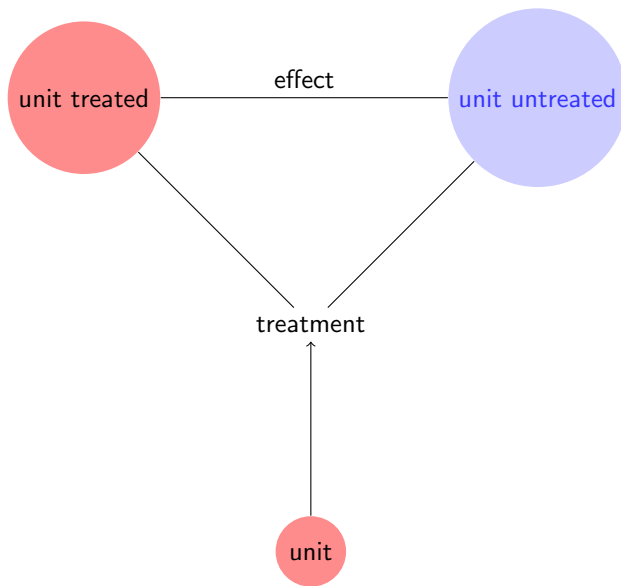
Causality



Causality



Causality



The causal effect of an independent variable X is the difference between what actually happened and what would have happened if we had changed X (**counterfactual**).

The fundamental problem of causal inference is that we never observe both of these events for a single person, at a single time.

The Rubin Model

Suppose we are interested in the effect of charter school on a student's performance.

- ▶ Let Y denote our outcome (dependent variable, performance)

The Rubin Model

Suppose we are interested in the effect of charter school on a student's performance.

- ▶ Let Y denote our outcome (dependent variable, performance)
- ▶ Let (1) mean **received a treatment (charter school)**

The Rubin Model

Suppose we are interested in the effect of charter school on a student's performance.

- ▶ Let Y denote our outcome (dependent variable, performance)
- ▶ Let (1) mean received a treatment (charter school)
- ▶ Let (0) mean **received no treatment (other schools)**

The Rubin Model

Suppose we are interested in the effect of charter school on a student's performance.

- ▶ Let Y denote our outcome (dependent variable, performance)
- ▶ Let (1) mean received a treatment (charter school)
- ▶ Let (0) mean received no treatment (other schools)
- ▶ Also $Y(1)$ be our outcome of interest under **treatment** (performance in a charter school)

The Rubin Model

Suppose we are interested in the effect of charter school on a student's performance.

- ▶ Let Y denote our outcome (dependent variable, performance)
- ▶ Let (1) mean received a treatment (charter school)
- ▶ Let (0) mean received no treatment (other schools)
- ▶ Also $Y(1)$ be our outcome of interest under treatment (performance in a charter school)
- ▶ Then $Y(0)$ be our outcome of interest under control (performance in a non-charter school)

Suppose we are interested in the effect of charter school on a student's performance.

- ▶ Let Y denote our outcome (dependent variable, performance)
- ▶ Let (1) mean received a treatment (charter school)
- ▶ Let (0) mean received no treatment (other schools)
- ▶ Also $Y(1)$ be our outcome of interest under treatment (performance in a charter school)
- ▶ Then $Y(0)$ be our outcome of interest under control (performance in a non-charter school)
- ▶ What is the **causal effect**? $\rightarrow Y(1) - Y(0)$

The Rubin Model

- ▶ This is called the "Rubin Model"



Let's call the subject of interest Mike. In an ideal world (the world of parallel universe), we would see this:

Let's call the subject of interest Mike. In an ideal world (the world of parallel universe), we would see this:

Unit	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	90/100	84/100	?

Let's call the subject of interest Mike. In an ideal world (the world of parallel universe), we would see this:

Unit	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	90/100	84/100	6

But in the real world, we only see this:

But in the real world, we only see this:

<i>Unit</i>	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	90/100	?	

But in the real world, we only see this:

<i>Unit</i>	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	90/100	?	Impossible to calculate

But in the real world, we only see this:

<i>Unit</i>	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	90/100	?	Impossible to calculate

Donald Rubin: "Causal inference is a missing data problem."

Causality

Comparable Cases

Random Assignment

Internal/External
Validity

Stata

Note:

Note:

- ▶ If $Y(1)$ is observed, $Y(0)$ is the **counterfactual**

Note:

- ▶ If $Y(1)$ is observed, $Y(0)$ is the counterfactual
- ▶ If $Y(0)$ is observed, $Y(1)$ is the counterfactual

Group exercise 1

Discuss (1) Treatment, (2) Outcome of interest, (3) Counterfactual, (4) Actual, and (5) Causal effect for two following examples.

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment:
- ▶ Outcome of interest:
- ▶ Counterfactual:
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment: attending section
- ▶ Outcome of interest:
- ▶ Counterfactual:
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment: attending section
- ▶ Outcome of interest:
- ▶ Counterfactual:
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment: attending section
- ▶ Outcome of interest: midterm score
- ▶ Counterfactual:
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment: attending section
- ▶ Outcome of interest: midterm score
- ▶ Counterfactual:
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment: attending section
- ▶ Outcome of interest: midterm score
- ▶ Counterfactual: Her midterm score *if she had not attended section*
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment: attending section
- ▶ Outcome of interest: midterm score
- ▶ Counterfactual: Her midterm score *if* she had not attended section
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment: attending section
- ▶ Outcome of interest: midterm score
- ▶ Counterfactual: Her midterm score *if* she had not attended section
- ▶ Actual: Her midterm score after she *did* attend section; 95
- ▶ Causal effect:

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment: attending section
- ▶ Outcome of interest: midterm score
- ▶ Counterfactual: Her midterm score *if* she had not attended section
- ▶ Actual: Her midterm score after she *did* attend section; 95
- ▶ Causal effect:

Group exercise 1

1. A POL-850 student wants to know whether recitation helps her get a good grade for the mid-term exam. Being an enterprising student, she attended section. And she received 95/100.

- ▶ Treatment: attending section
- ▶ Outcome of interest: midterm score
- ▶ Counterfactual: Her midterm score *if* she had not attended section
- ▶ Actual: Her midterm score after she *did* attend section; 95
- ▶ Causal effect: $Y(1) - Y(0) = 95 - ?$

Group exercise 1

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment?
- ▶ Outcome of interest?
- ▶ Counterfactual?
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment? **higher minimum wage**
- ▶ Outcome of interest?
- ▶ Counterfactual?
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment? higher minimum wage
- ▶ Outcome of interest?
- ▶ Counterfactual?
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment? higher minimum wage
- ▶ Outcome of interest? the change of unemployment rate
- ▶ Counterfactual?
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment? higher minimum wage
- ▶ Outcome of interest? the change of unemployment rate
- ▶ Counterfactual?
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment? higher minimum wage
- ▶ Outcome of interest? the change of unemployment rate
- ▶ Counterfactual? the change of unemployment rate if the policy remained the same
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment? higher minimum wage
- ▶ Outcome of interest? the change of unemployment rate
- ▶ Counterfactual? the change of unemployment rate if the policy remained the same
- ▶ Actual:
- ▶ Causal effect:

Group exercise 1

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment? higher minimum wage
- ▶ Outcome of interest? the change of unemployment rate
- ▶ Counterfactual? the change of unemployment rate if the policy remained the same
- ▶ Actual: unemployment rate rises by 2%
- ▶ Causal effect:

Group exercise 1

2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment? higher minimum wage
- ▶ Outcome of interest? the change of unemployment rate
- ▶ Counterfactual? the change of unemployment rate if the policy remained the same
- ▶ Actual: unemployment rate rises by 2%
- ▶ Causal effect:

Group exercise 1

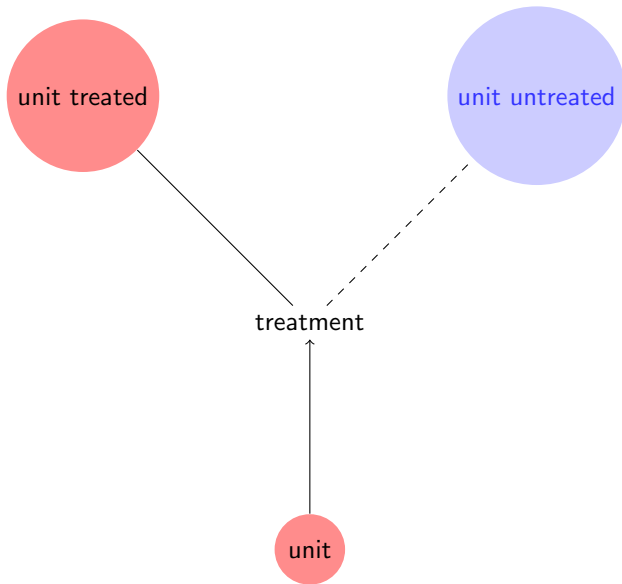
2. Suppose that the governor of New Jersey announces to increase the state-level minimum wage from \$8.6 per hour to \$10 per hour. After the implementation of the policy, unemployment rate rises by 2%.

- ▶ Treatment? higher minimum wage
- ▶ Outcome of interest? the change of unemployment rate
- ▶ Counterfactual? the change of unemployment rate if the policy remained the same
- ▶ Actual: unemployment rate rises by 2%
- ▶ Causal effect: $Y(1) - Y(0) = 2\% - ?$

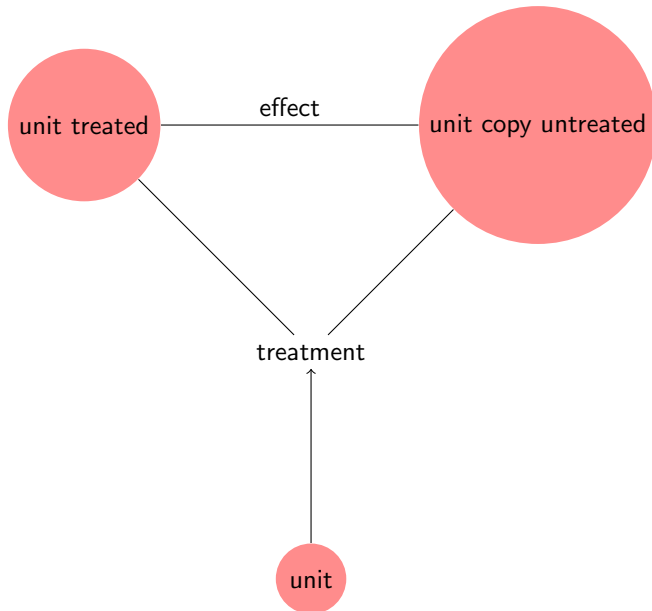
Comparable Cases

Fundamental problem of causal inference means
we have to find **good comparison cases**

Comparable cases



Comparable cases



Can we make inferences when we have Mike's identical twin Bob and give him no treatment?:

- ▶ They were raised up in the same environment and attended the same primary school
- ▶ Their IQs and grades in the primary school are same
- ▶ They share a lot of common hobbies

Can we make inferences when we have Mike's identical twin Bob and give him no treatment?:

- ▶ They were raised up in the same environment and attended the same primary school
- ▶ Their IQs and grades in the primary school are same
- ▶ They share a lot of common hobbies

<i>Unit</i>	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	90/100	?	?
Bob	?	86/100	?

What about $Y_{Mike}(1) - Y_{Bob}(0)$?

$$Y_{Mike}(1) - Y_{Bob}(0) =$$

Can we make inferences when we have Mike's identical twin Bob and give him no treatment?:

- ▶ They were raised up in the same environment and attended the same primary school
- ▶ Their IQs and grades in the primary school are same
- ▶ They share a lot of common hobbies

<i>Unit</i>	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	90/100	?	?
Bob	?	86/100	?

What about $Y_{Mike}(1) - Y_{Bob}(0)$?

$$Y_{Mike}(1) - Y_{Bob}(0) =$$

Can we make inferences when we have Mike's identical twin Bob and give him no treatment?:

- ▶ They were raised up in the same environment and attended the same primary school
- ▶ Their IQs and grades in the primary school are same
- ▶ They share a lot of common hobbies

<i>Unit</i>	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	90/100	?	?
Bob	?	86/100	?

What about $Y_{Mike}(1) - Y_{Bob}(0)$?

$$Y_{Mike}(1) - Y_{Bob}(0) = 90/100 - 86/100$$

Can we make inferences when we have Mike's identical twin Bob and give him no treatment?:

- ▶ They were raised up in the same environment and attended the same primary school
- ▶ Their IQs and grades in the primary school are same
- ▶ They share a lot of common hobbies

<i>Unit</i>	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	90/100	?	?
Bob	?	86/100	?

What about $Y_{Mike}(1) - Y_{Bob}(0)$?

$$\begin{aligned}
 Y_{Mike}(1) - Y_{Bob}(0) &= 90/100 - 86/100 \\
 &= 4/100
 \end{aligned}$$

Group exercise 2

- ▶ What is a good comparison in the New Jersey minimum wage senario? Do you think your answer can help us overcome the fundamental problem of causal inference?
- ▶ An economist finds that in the same period, minimum wage in Philadelphia didn't change, and the unemployment rate rises by 1%. He thus concludes that the impact of higher minimum wage must be negative. Do you agree with his argument?
- ▶ What is the counterfactual of being female?

Group exercise 2

- ▶ What is a good comparison in the New Jersey minimum wage senario? Do you think your answer can help us overcome the fundamental problem of causal inference?

Towns near the border of New Jersey where the minimum wage was not changed.

- ▶ An economist finds that in the same period, minimum wage in Philadelphia didn't change, and the unemployment rate rises by 1%. He thus concludes that the impact of higher minimum wage must be negative. Do you agree with his argument?
- ▶ What is the counterfactual of being female?

Group exercise 2

- ▶ What is a good comparison in the New Jersey minimum wage senario? Do you think your answer can help us overcome the fundamental problem of causal inference?

Towns near the border of New Jersey where the minimum wage was not changed.

- ▶ An economist finds that in the same period, minimum wage in Philadelphia didn't change, and the unemployment rate rises by 1%. He thus concludes that the impact of higher minimum wage must be negative. Do you agree with his argument?
- ▶ What is the counterfactual of being female?

Group exercise 2

- ▶ What is a good comparison in the New Jersey minimum wage senario? Do you think your answer can help us overcome the fundamental problem of causal inference?

Towns near the border of New Jersey where the minimum wage was not changed.

- ▶ An economist finds that in the same period, minimum wage in Philadelphia didn't change, and the unemployment rate rises by 1%. He thus concludes that the impact of higher minimum wage must be negative. Do you agree with his argument?

Philadelphia is not a comparable case, since it is a large city.

- ▶ What is the counterfactual of being female?

Group exercise 2

- ▶ What is a good comparison in the New Jersey minimum wage senario? Do you think your answer can help us overcome the fundamental problem of causal inference?

Towns near the border of New Jersey where the minimum wage was not changed.

- ▶ An economist finds that in the same period, minimum wage in Philadelphia didn't change, and the unemployment rate rises by 1%. He thus concludes that the impact of higher minimum wage must be negative. Do you agree with his argument?

Philadelphia is not a comparable case, since it is a large city.

- ▶ What is the counterfactual of being female?

It is hard to find one.

Twin Example Revisited

$$\begin{aligned}Y_{Mike}(1) - Y_{Bob}(0) &= 90/100 - 86/100 \\ &= 4/100\end{aligned}$$

Can we convincingly conclude that **charter schools have positive effect on exam scores**?

Twin Example Revisited

$$\begin{aligned}Y_{Mike}(1) - Y_{Bob}(0) &= 90/100 - 86/100 \\ &= 4/100\end{aligned}$$

Can we convincingly conclude that **charter schools have positive effect on exam scores**?

No because of the **unobserved confounding variables**. Thus the relationship between X and Y could have been spurious.

Twin Example Revisited

- ▶ Let's see how an **unobserved confounding variable** Z could produce a **spurious relationship**
- ▶ For example, say Mike finds the course very interesting and invests much more time than Bob (not observed)
- ▶ Suppose charter schools have in fact no effect on the outcome

<i>Unit</i>	<i>Z</i>	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Mike	20 hours	90/100	?	?
Bob	10 hours	?	86/100	?

$$\begin{aligned}
 Y_{Mike}(1) - Y_{Bob}(0) &= 90/100 - 86/100 \\
 &= 4/100
 \end{aligned}$$

- ▶ This leads to a wrong conclusion: **charter schools have positive impact on Y**

Random assignment

- ▶ Need **treatment** and **control** groups to be as similar as possible
- ▶ **Randomization in treatment assignment** is key
- ▶ e.g. Toss a coin to decide the assignment of treatment
- ▶ **Two groups “on average” will be the same on all other factors**
- ▶ **This applies to even unobserved confounders Z**
- ▶ That way we can make sure that variation in Y is due **only to X**
- ▶ So any difference in Y is due to X (alone) given the fact that two groups are identical

Group exercise 3

Discuss whether assignment of treatment is as good as random in the following examples. If not, what could be a confounder (Z)? (X: a new medicine and Y: health condition)

Give the new medicine

1. on first-come first-serve basis
2. to patients whose room number is odd
3. to patients who volunteer for the experiment
4. to patients who win a lottery within the hospital

Random assignment

- ▶ A major problem of causal inference in observational studies (not experiments) is that units are allowed to **select into treatment** (self-selection)
- ▶ Is this a problem in the minimum wage case?

Random assignment

- ▶ A major problem of causal inference in observational studies (not experiments) is that units are allowed to **select into treatment** (self-selection)
- ▶ Is this a problem in the minimum wage case?
People may move to towns with higher wage

Internal/External Validity

- ▶ _____ is when a study isolates the effect of treatment from other plausible explanations.
- ▶ _____ is when a study's findings can be generalized.

Internal/External Validity

- ▶ **internal validity** is when a study isolates the effect of treatment from other plausible explanations.
- ▶ _____ is when a study's findings can be generalized.

Internal/External Validity

- ▶ **internal validity** is when a study isolates the effect of treatment from other plausible explanations.
- ▶ **external validity** is when a study's findings can be generalized.

Internal/External Validity

1. Suppose we conducted a randomized study of the effect of chocolate on attentiveness in an astrophysics lab at MIT. Would this have high internal or external validity?

Internal/External Validity

1. Suppose we conducted a randomized study of the effect of chocolate on attentiveness in an astrophysics lab at MIT. Would this have high internal or external validity?

Internal/External Validity

2. Suppose we conducted a large survey of the world and asked about chocolate consumption and productivity. Would this have high internal or external validity?

Internal/External Validity

2. Suppose we conducted a large survey of the world and asked about chocolate consumption and productivity. Would this have high internal or external validity?

Internal/External Validity

Recitation, Week 4

shortname

Causality

Comparable Cases

Random Assignment

Internal/External
Validity

Stata

Tradeoff. Why?

Key takeaways

You should be able to answer these questions:

1. Why do we care about the **counterfactual**?
2. What's the **fundamental problem of causal inference**?
3. Why is the **random assignment** is so important?
4. What do we randomly assign?
5. Why is it typically hard to make a causal inference with observational data?

- ▶ Go to <http://guides.nyu.edu/stata>
- ▶ Find out the average pizza expenditure of female respondents
- ▶ Find out the average income of people whose highest degree is college