Studying Social Inequality with Data Science

## Interventions to Promote Equality: Causal inference

Learning goals for today

By the end of class, you will be able to

► define causal effects using potential outcomes

► understand the fundamental problem of causal inference

Class module: Interventions to promote equality

- don't just study the world
- learn how we might change it

Education: A tool to promote opportunity
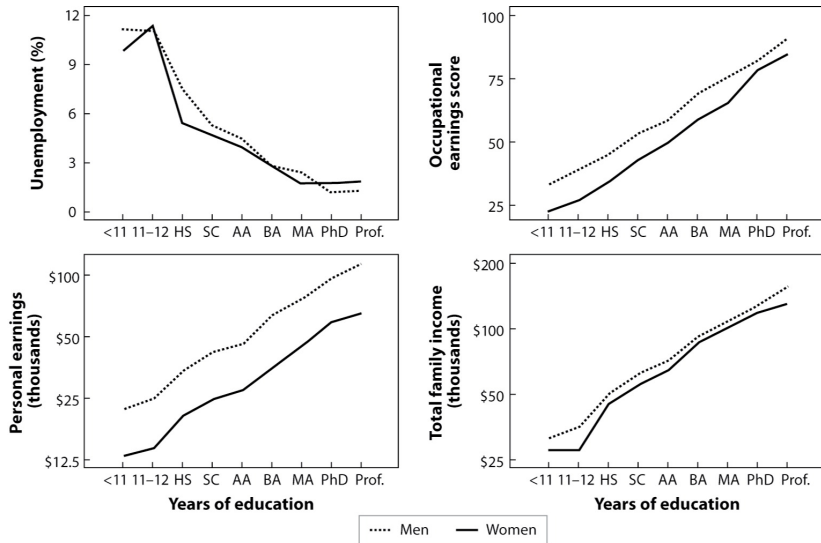
Two claims

People with
college degrees
earn more

A college degree
causes
higher earnings

Two claims

People with
college degrees
earn more

A college degree
causes
higher earnings

What policies would each claim support?

Two claims

People with
college degrees
earn more
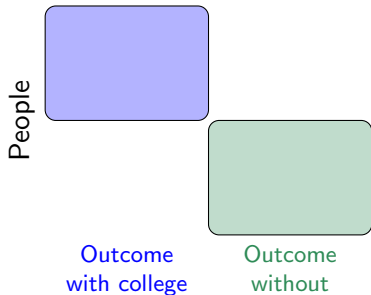
A college degree
causes
higher earnings

Two sets of people
Two treatments

Two claims

People with
college degrees
earn more

A college degree
causes
higher earnings

Two sets of people
Two treatments

People

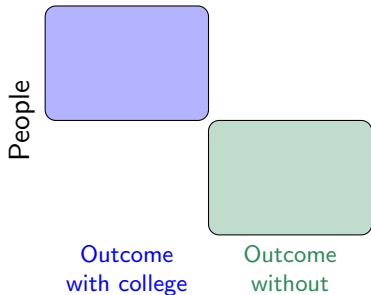Outcome
with college

Outcome
without

Two claims

People with
college degrees
earn more

A college degree
causes
higher earnings

Two sets of people
Two treatments

Same people
Two treatments

People

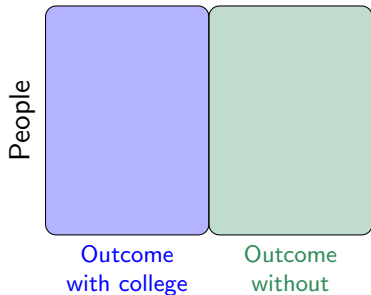Outcome
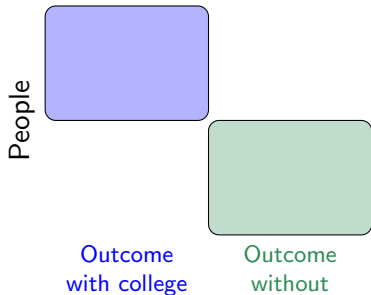with college

Outcome
without

Two claims

People with
college degrees
earn more

A college degree
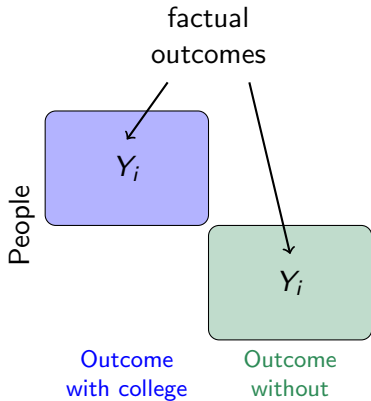causes
higher earnings

Two sets of people
Two treatments

Same people
Two treatments

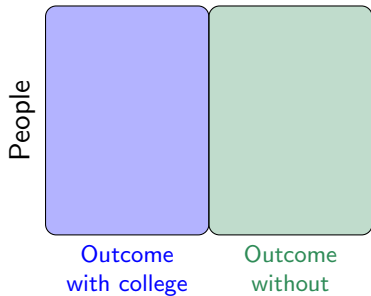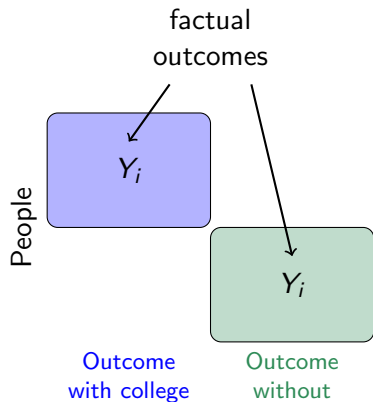# Two claims

People with
college degrees
earn more

A college degree
causes
higher earnings

Two claims

People with
college degrees
earn more

A college degree
causes
higher earnings

factual
outcomes

Y_i

Y_i

People

Outcome
with college

Outcome
without

potential
outcomes

Y_i^{College}

$Y_i^{\text{Non-college}}$

People

Outcome
with college

Outcome
without

Two claims

People with
college degrees
earn more

A college degree
causes
higher earnings

factual
outcomes

potential
outcomes

Imbens &
Rubin
2015

People

Y_i

Y_i

Outcome
with college

Outcome
without

People

Y_i^{College}

$Y_i^{\text{Non-college}}$

Outcome
with college

Outcome
without

Translate causal statements to potential outcomes

Translate causal statements to potential outcomes

1) Sarah would have been hired
   if she had submitted her
   application on time

Translate causal statements to potential outcomes

1) Sarah would have been hired
   if she had submitted her
   application on time

$$Y_{Sarah}^{On\ time} = 1$$
$$Y_{Sarah}^{Late} = 0$$

Translate causal statements to potential outcomes

1) Sarah would have been hired
   if she had submitted her
   application on time

   $Y_{Sarah}^{On\ time} = 1$
   $Y_{Sarah}^{Late} = 0$

2) David would have been
   hired even if he hadn't
   had a college degree

3) The hiring manager discriminatorily
   chose Emily over Lakisha
   because of their names

4) Juan never even applied because
   he has a newborn infant

Translate causal statements to potential outcomes

1) Sarah would have been hired if she had submitted her application on time

$$Y_{\text{Sarah}}^{\text{On time}} = 1$$
$$Y_{\text{Sarah}}^{\text{Late}} = 0$$

2) David would have been hired even if he hadn't had a college degree

$$Y_{\text{David}}^{\text{College}} = 1$$
$$Y_{\text{David}}^{\text{Non-college}} = 1$$

3) The hiring manager discriminatorily chose Emily over Lakisha because of their names

4) Juan never even applied because he has a newborn infant

Translate causal statements to potential outcomes

1) Sarah would have been hired if she had submitted her application on time

$$Y_{\text{Sarah}}^{\text{On time}} = 1$$
$$Y_{\text{Sarah}}^{\text{Late}} = 0$$

2) David would have been hired even if he hadn't had a college degree

$$Y_{\text{David}}^{\text{College}} = 1$$
$$Y_{\text{David}}^{\text{Non-college}} = 1$$

3) The hiring manager discriminatorily chose Emily over Lakisha because of their names

$$Y_{\text{Lakisha}}^{\text{Named Emily}} = 1$$
$$Y_{\text{Lakisha}}^{\text{Named Lakisha}} = 0$$

4) Juan never even applied because he has a newborn infant
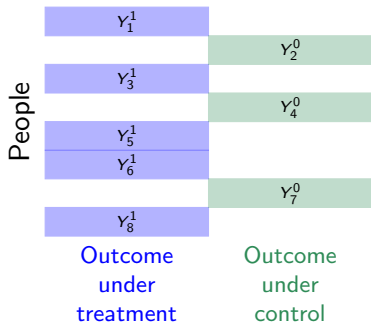
Translate causal statements to potential outcomes

1) Sarah would have been hired if she had submitted her application on time

$$Y_{\text{Sarah}}^{\text{On time}} = 1$$
$$Y_{\text{Sarah}}^{\text{Late}} = 0$$

2) David would have been hired even if he hadn't had a college degree

$$Y_{\text{David}}^{\text{College}} = 1$$
$$Y_{\text{David}}^{\text{Non-college}} = 1$$

3) The hiring manager discriminatorily chose Emily over Lakisha because of their names

$$Y_{\text{Lakisha}}^{\text{Named Emily}} = 1$$
$$Y_{\text{Lakisha}}^{\text{Named Lakisha}} = 0$$

4) Juan never even applied because he has a newborn infant

$$Y_{\text{Juan}}^{\text{Infant}} = 0$$
$$Y_{\text{Juan}}^{\text{No infant}} = 1$$

Translate causal statements to potential outcomes
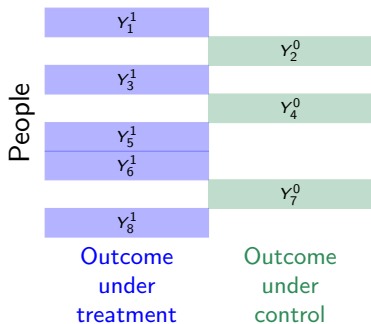
Is it possible to know if each statement is true?

1) Sarah would have been hired if she had submitted her application on time

$$Y_{\text{Sarah}}^{\text{On time}} = 1$$
$$Y_{\text{Sarah}}^{\text{Late}} = 0$$

2) David would have been hired even if he hadn't had a college degree

$$Y_{\text{David}}^{\text{College}} = 1$$
$$Y_{\text{David}}^{\text{Non-college}} = 1$$

3) The hiring manager discriminatorily chose Emily over Lakisha because of their names

$$Y_{\text{Lakisha}}^{\text{Named Emily}} = 1$$
$$Y_{\text{Lakisha}}^{\text{Named Lakisha}} = 0$$

4) Juan never even applied because he has a newborn infant

$$Y_{\text{Juan}}^{\text{Infant}} = 0$$
$$Y_{\text{Juan}}^{\text{No infant}} = 1$$

The data

People

| Outcome under treatment | Outcome under control |

Could we ever know the effect for person 1?

The data

People

Y_1^1	
	Y_2^0
Y_3^1	
	Y_4^0
Y_5^1	
Y_6^1	
	Y_7^0
Y_8^1	

Outcome under treatment

Outcome under control
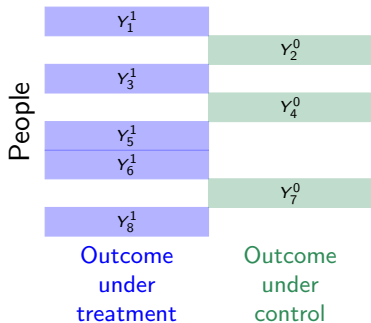
Holland 1986

Could we ever know the effect for person 1?
For anyone?

The data

Outcome under treatment

Outcome under control
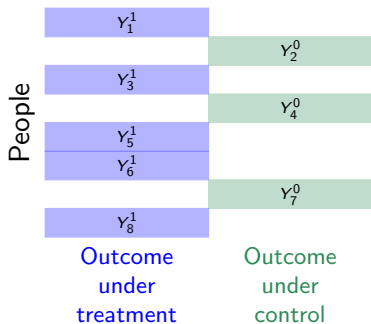
Could we ever know the effect for person 1?
For anyone?
Can anything be known?

The data

People

Y_1^1

Y_2^0

Y_3^1

Y_4^0

Y_5^1

Y_6^1

Y_7^0

Y_8^1

Outcome
under
treatment

Outcome
under
control

Holland 1986

Average causal effects

The data

People

Y_1^1
Y_2^0
Y_3^1
Y_4^0
Y_5^1
Y_6^1
Y_7^0
Y_8^1

Outcome under treatment · Outcome under control

The claim

People

Average $\bar{Y}^1$ · Average $\bar{Y}^0$

Outcome under treatment · Outcome under control
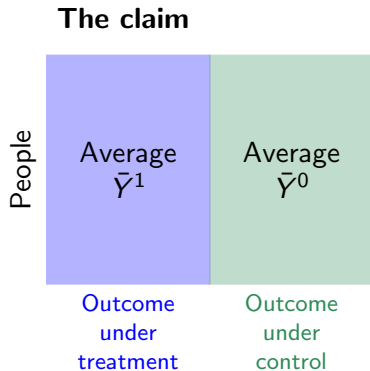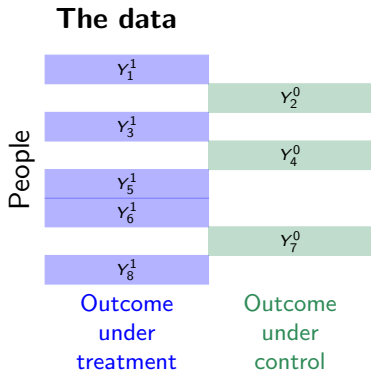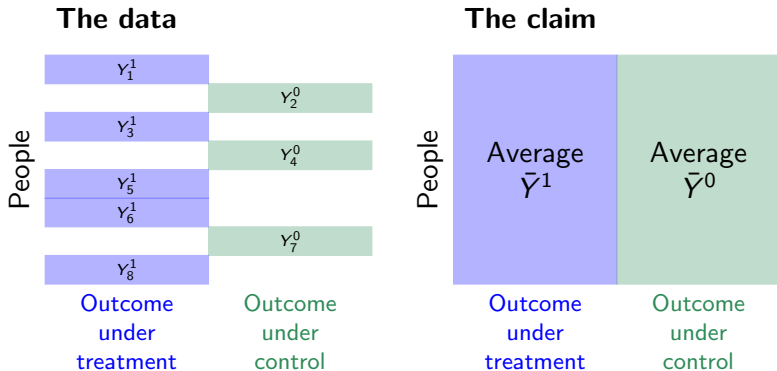
Average causal effects

If treatment is randomized,
then the observed cases
in each treatment condition
are a simple random sample

Learning goals for today

By the end of class, you will be able to
- ▶ define causal effects using potential outcomes
- ▶ understand the fundamental problem of causal inference

Bonus slides

For each statement, draw a table of potential outcomes.

► Who are the people being considered?

► Under what treatment conditions?

► What potential outcome is obseverved?

Statements

1. On average, those who attend college have higher annual earnings than what they would have earned without college

2. On average, those who did not attend college would earn more if they attended college

3. On average, the causal effect of college on earnings is higher for those whose parents did not attend college

Challenge:

Visitors to the Cornell Dairy Bar would be more impressed if they ordered Cascadilla Cookies & Cream than if they ordered French Vanilla, but they would be most impressed if they ordered Big Red Bear Tracks.