McCOMBS SCHOOL OF BUSINESS
Salem Center for Policy

Causality

David Puelz

September 2, 2021

Outline

Cause-and-effect

Potential outcomes and Counterfactuals

RCTs and Obs Studies

Causality and policy analysis?

These two ideas seem completely unrelated.

But, they are and steathly show up EVERYWHERE!

The Federal Unemployment Bonus Holds the Recovery Back

The Help Wanted signs came down in my town only when Arkansas opted out of the $300 supplement.

The Federal Unemployment Bonus Holds the Recovery Back

The Help Wanted signs came down in my town only when Arkansas opted out of the $300 supplement.

Covid-19 Rekindles Debate Over License Requirements for Many Jobs

Hair styling and medical fields are among the occupations where state rules can bar entry; Biden has pushed for changes

The Federal Unemployment Bonus Holds the Recovery Back

The Help Wanted signs came down in my town only when Arkansas opted out of the $300 supplement.

Covid-19 Rekindles Debate Over License Requirements for Many Jobs

Hair styling and medical fields are among the occupations where state rules can bar entry; Biden has pushed for changes

Beefed-Up Sanctions Could Limit the Damage in Afghanistan

The Taliban’s control of the government will significantly increase their wealth and influence.

The Federal Unemployment Bonus Holds the Recovery Back

The Help Wanted signs came down in my town only when Arkansas opted out of the $300 supplement.

Covid-19 Rekindles Debate Over License Requirements for Many Jobs

Hair styling and medical fields are among the occupations where state rules can bar entry; Biden has pushed for changes

Beefed-Up Sanctions Could Limit the Damage in Afghanistan

The Taliban's control of the government will significantly increase their wealth and influence.

Medicare Advantage Shows the Path Forward

The savings spill over into traditional Medicare and even into the nonelderly, commercially insured market.

The Federal Unemployment Bonus Holds the Recovery Back

The Help Wanted signs came down in my town only when Arkansas opted out of the $300 supplement.

Covid-19 Rekindles Debate Over License Requirements for Many Jobs

Hair styling and medical fields are among the occupations where state rules can bar entry; Biden has pushed for changes

Beefed-Up Sanctions Could Limit the Damage in Afghanistan

The Taliban's control of the government will significantly increase their wealth and influence.

Medicare Advantage Shows the Path Forward

The savings spill over in

Harvard's Katherine Baicker, Michael Chernew and Jacob Robbins have showed that as penetration of Medicare Advantage increases in different counties, hospital costs and length-of-stays decline not only for seniors enrolled in Medicare managed care plans, but also for beneficiaries still on the traditional program. For every 10% increase in the uptake of Medicare Advantage, inpatient spending among fee-for-service Medicare seniors falls by 5% to 10%. Similar findings are also observed among commercially-insured people under 65 in regions with rapid diffusion of Medicare Advantage.

The Federal Unemployment Bonus Holds the Recovery Back

The Help Wanted signs came down in my town only when Arkansas opted out of the $300 supplement.

Covid-19 Rekindles Debate Over License Requirements for Many Jobs

Hair styling and medical fields are among the occupations where state rules can bar entry; Biden has pushed for changes

Beefed-Up Sanctions Could Limit the Damage in Afghanistan

The Taliban's control of the government will significantly increase their wealth and influence.

Medicare Advantage Shows the Path Forward

The savings spill over in

Harvard's Katherine Baicker, Michael Chernew and Jacob Robbins have showed that as penetration of Medicare Advantage increases in different counties, hospital costs and length-of-stays decline not only for seniors enrolled in Medicare managed care plans, but also for beneficiaries still on the traditional program. For every 10% increase in the uptake of Medicare Advantage, inpatient spending among fee-for-service Medicare seniors falls by 5% to 10%. Similar findings are also observed among commercially-insured people under 65 in regions with rapid diffusion of Medicare Advantage.

How Often Should You Shower? Celebrities Ignite a Ferocious Debate

Hollywood types including Jake Gyllenhaal, Mila Kunis, Ashton Kutcher and Dax Shepard take a lax approach to hygiene, stoking a contentious uproar on how often one should bathe. It mirrors a similar discord in the medical community, and among everyday people.

Cause-and-effect $\iff$ policy impacts

These two paradigms in the title are one and the same! One is a general framework, and one is specific to the policy arena.

Cause is a statement of something being manipulated or changed
Effect is a measure of the change in an outcome of interest

Cause-and-effect $\iff$ policy impacts

These two paradigms in the title are one and the same! One is a general framework, and one is specific to the policy arena.

Cause is a statement of something being manipulated or changed
Effect is a measure of the change in an outcome of interest

- "Cause" is the same as a policy introduction or change
- "Effect" is the unique, independent measurement of how the cause modulated some other part of our system

Cause-and-effect: policy decisions as cause

We see causes all of the time

→ The federal government increases the minimum wage

Cause-and-effect: policy decisions as cause

We see causes all of the time

→ The federal government increases the minimum wage
→ The FAA mandates face coverings on planes

Cause-and-effect: policy decisions as cause

We see causes all of the time

→ The federal government increases the minimum wage
→ The FAA mandates face coverings on planes
→ The percentage of vaccinated individuals exceeds 50% of population

Cause-and-effect: policy decisions as cause

We see causes all of the time

$\rightarrow$ The federal government increases the minimum wage
$\rightarrow$ The FAA mandates face coverings on planes
$\rightarrow$ The percentage of vaccinated individuals exceeds 50% of population
$\rightarrow$ It's windy on an Austin afternoon

Cause-and-effect: policy decisions as cause

We see causes all of the time

$\rightarrow$ The federal government increases the minimum wage
$\rightarrow$ The FAA mandates face coverings on planes
$\rightarrow$ The percentage of vaccinated individuals exceeds 50% of population
$\rightarrow$ It's windy on an Austin afternoon
$\rightarrow$ Sweetgreen decreases its Kale caesar salad price by $2

Cause-and-effect: policy decisions as cause

We see causes all of the time

→ The federal government increases the minimum wage
→ The FAA mandates face coverings on planes
→ The percentage of vaccinated individuals exceeds 50% of population
→ It's windy on an Austin afternoon
→ Sweetgreen decreases its Kale caesar salad price by $2
→ States implement stay-at-home orders during the pandemic

Cause-and-effect: varying features (treatments) as cause

Causes might also just be variation in a population

$\rightarrow$ Are dog-lovers nicer people?

Cause-and-effect: varying features (treatments) as cause

Causes might also just be variation in a population

$\rightarrow$ Are dog-lovers nicer people?
$\rightarrow$ Does race affect hiring decisions?

Cause-and-effect: varying features (treatments) as cause

Causes might also just be variation in a population

→ Are dog-lovers nicer people?
→ Does race affect hiring decisions?
→ Does age affect COVID-19 mortality?

Cause-and-effect: varying features (treatments) as cause

Causes might also just be variation in a population

→ Are dog-lovers nicer people?
→ Does race affect hiring decisions?
→ Does age affect COVID-19 mortality?
→ Is there a "gender-gap" in salary?

Cause-and-effect $\iff$ policy impacts

The **difficult** question is, what about the effects?

Cause-and-effect $\iff$ policy impacts

The **difficult** question is, what about the effects?

- What outcomes do we look at?
- How do we measure them?
- Are there other variables that might affect the outcomes *and* the causes?

Increasing the minimum wage

The **difficult** question is, what about the effects?

- What outcomes do we look at? (lower class unemployment rate, income, ...)
- How do we measure them? (government data, surveys, ...)
- Are there other variables that might affect the outcomes *and* the causes? (current economic conditions, differences among states, ...)

Sweetgreen salad price increase

The **difficult** question is, what about the effects?

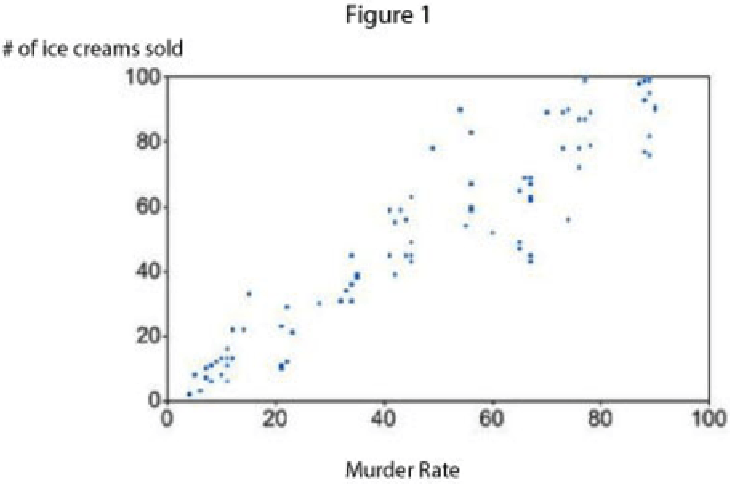
- What outcomes do we look at? (revenue, count of kale caesars sold, number of daily lunch visitors, ...)
- How do we measure them? (financial data, ...)
- Are there other variables that might affect the outcomes *and* the causes? (time of year (seasonality), temperature, weather, length of daily wait time, ...)

Racial discrimination in hiring?

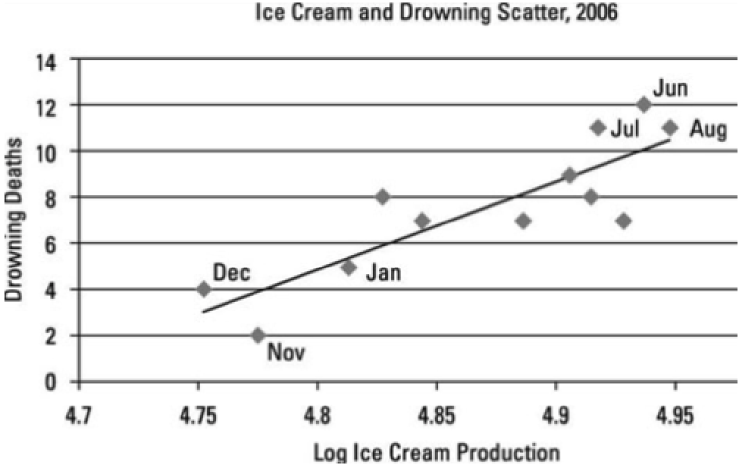
The **difficult** question is, what about the effects?

- What outcomes do we look at? (whether or not a job applicant receives a callback)
- How do we measure them? (follows directly from above ...)
- Are there other variables that might affect the outcomes *and* the causes? (other resume characteristics, average GPA, brand of university,...)

Ice cream and NYC murder rate



Ice cream and drownings



To sum up cause-and-effect

- Challenges are related to both the system of study and ability to gather the right data.
- With data in hand, you can start to formulate hypotheses and test them.
- There might be lurking variables driving an underlying relationship (ice cream). Only an expert (you!) can identify those and take them into account.

Let's formalize these ideas with some basic notation

Notation

- units of study are indexed by i (states, stores, customers, people)

Notation

- units of study are indexed by i (states, stores, customers, people)
- cause aka treatment aka policy for unit i is given by z_i

$$z_i = \begin{cases} 0 & \text{not treated (control)} \\ 1 & \text{treated} \end{cases}$$

Notation

- units of study are indexed by i (states, stores, customers, people)
- cause aka treatment aka policy for unit i is given by z_i

$$z_i = \begin{cases} 0 & \text{not treated (control)} \\ 1 & \text{treated} \end{cases}$$

- denote a outcome of interest for each i as Y_i, but let's write it as a function of the cause.

$$Y_i(z_i)$$

Notation

- units of study are indexed by i (states, stores, customers, people)
- cause aka treatment aka policy for unit i is given by z_i

$$z_i = \begin{cases} 0 & \text{not treated (control)} \\ 1 & \text{treated} \end{cases}$$

- denote a outcome of interest for each i as Y_i, but let's write it as a function of the cause.

$$Y_i(z_i) \leftarrow \textbf{ potential outcome } \text{for unit } i$$

Notation

- units of study are indexed by i (states, stores, customers, people)
- cause aka treatment aka policy for unit i is given by z_i

$$z_i = \begin{cases} 0 & \text{not treated (control)} \\ 1 & \text{treated} \end{cases}$$

- denote a outcome of interest for each i as Y_i, but let's write it as a function of the cause.

$$Y_i(z_i) \leftarrow \textbf{potential outcome} \text{ for unit } i$$

- with this notation, we have the building blocks to talk about policy effects!

Example: COVID-19 lockdowns

How did state lockdowns at the beginning of 2020 affect the spread of the virus?

Example: COVID-19 lockdowns

How did state lockdowns at the beginning of 2020 affect the spread of the virus?

What are the potential outcomes?

Example: COVID-19 lockdowns

Let's set up the structure of this problem.

Let i denote a state, so

$$i \in \{\text{New York, California, Florida, Texas, South Dakota, ... }\}$$

First, we define what z_i is:

$$z_i = \begin{cases} 0 & \text{no lockdown} \\ 1 & \text{lockdown} \end{cases}$$

Example: COVID-19 lockdowns

Let's set up the structure of this problem.

Let i denote a state, so

$$i \in \{\text{New York, California, Florida, Texas, South Dakota, ... }\}$$

First, we define what z_i is:

$$z_i = \begin{cases} 0 & \text{no lockdown} \\ 1 & \text{lockdown} \end{cases}$$

Second, we define our outcome: Y_i : let's choose the cases per capita (in state i) after lockdown or no lockdown.

Example: COVID-19 lockdowns

A brief aside:

Defining exactly what the treatment z_i is very hard! It could be a combination of many available data.

- masking
- bar and restaurant closures
- school closures
- curfews
- limits to exercise
- retail store closures

Example: COVID-19 lockdowns

A brief aside:

Defining exactly what the treatment z_i is very hard! It could be a combination of many available data.

- masking
- bar and restaurant closures
- school closures
- curfews
- limits to exercise
- retail store closures

Q: How would you define z_i?

Organizing our data: The Science Table

i **(state)**	z_i **(lockdown)**	$Y_i(0)$	$Y_i(1)$
New York			
Florida			
California			
Texas			
South Dakota			
Illinois			
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Organizing our data: The Science Table

i **(state)**	z_i **(lockdown)**	$Y_i(0)$	$Y_i(1)$
New York	1		
Florida	0		
California	1		
Texas	0		
South Dakota	0		
Illinois	1		
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Organizing our data: The Science Table (CDC, cases/100k)

i **(state)**	z_i **(lockdown)**	$Y_i(0)$	$Y_i(1)$
New York	1		.0034
Florida	0	.007	
California	1		.0014
Texas	0	.004	
South Dakota	0	.0028	
Illinois	1		.002
⋮	⋮	⋮	⋮

Organizing our data: The Science Table (CDC, cases/100k)

i **(state)**	z_i **(lockdown)**	$Y_i(0)$	$Y_i(1)$
New York	1	??	.0034
Florida	0	.007	??
California	1	??	.0014
Texas	0	.004	??
South Dakota	0	.0028	??
Illinois	1	??	.002
$\vdots$	$\vdots$	$\vdots$	$\vdots$

What is the ideal scenario?

We are able to know both of the potential outcomes for each state!

Y_{NY} () and Y_{NY} ()

Defining a causal effect for NY

We can define the causal effect of the “lockdown treatment” as the difference between the two potential outcomes.

$\tau_{\text{NY}} = Y_{\text{NY}}\big($  $\big) - Y_{\text{NY}}\big($  $\big)$

or written more generally:

$$\tau_i = Y_i(1) - Y_i(0)$$

The fundamental problem of causal inference

We only observe one of the two potential outcomes for New York and all other states. In general, we always only observe one of two potential outcomes for our units of study.

- economics of COVID policy: a state either locks down or doesn't
- drug trials: an individual either receives the medicine or the placebo
- gender wage gap: a person is either male or female

The unknown outcomes are called the **missing potential outcomes** or **counterfactuals**. This is what makes causality a nontrivial task ... it is a missing data problem.

Is all hope lost?

Is all hope lost?

Definitely not! The potential outcomes will **always** be used as a starting point. Depending on the data and question to be answered, there are several approaches:

- Randomization and the sample average treatment effect
- Observational data – before-and-after and DiD approaches
- Fancier (probabilistic) models to address confounding. Regression, etc. (the "Prediction" part of class).

The average causal effect across the sample

This is called the sample average treatment effect. In stats language, it is called an estimand. Let's suppose we have N units in our data.

$$\mathbf{SATE} = \frac{1}{N}\sum_{i=1}^{N}\tau_i$$
$$= \frac{1}{N}\sum_{i=1}^{N}\{Y_i(1) - Y_i(0)\}$$

We still don't know how to calculate this because of the fundamental problem of causal inference.

However, here's an idea ...

Estimator of the SATE

We have the **observed** outcome and treatment. Let's call them:

$$Y_{\text{obs}} = (Y_1, ..., Y_N)$$
$$Z_{\text{obs}} = (Z_1, ..., Z_N)$$

Let's define our estimator of the **SATE** as the simple difference-in-means between the treated and control units.

$$\widehat{\textbf{SATE}} = \frac{1}{\sum_i \mathbb{1}(Z_i = 1)} \sum_i \mathbb{1}(Z_i = 1) Y_i - \frac{1}{\sum_i \mathbb{1}(Z_i = 0)} \sum_i \mathbb{1}(Z_i = 0) Y_i$$

Q: When can this be reasonably interpreted as the average causal effect, when can it not?