Defining causal effects

Cornell STSCI / INFO / ILRST 3900: Causal Inference
Fall 2023

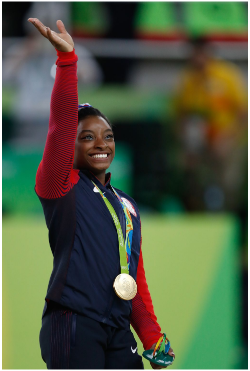
24 Aug 2022

Learning goals for today

By the end of class, you will be able to

- explain the fundamental problem of causal inference and the need for causal arguments
- define potential outcomes
- recall mathematical concepts from probability
  - random variables
  - expectation
  - conditional expectation

Causal claims hinge on arguments, not on data



Left photo: By Fernando Frazão/Agência Brasil -
http://agenciabrasil.ebc.com.br/sites/_agenciabrasil2013/files/fotos/1035034-_mg_0802_04.08.16.jpg,CCBY3.0br,https://commons.wikimedia.org/w/index.php?curid=50548410
Right photo: By Agencia Brasil Fotografias - EUA levam ouro na ginástica artística feminina; Brasil fica em 8 lugar, CC BY 2.0, https://commons.wikimedia.org/w/index.php?curid=50584648

Causal claims hinge on arguments, not on data

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.

Causal claims hinge on arguments, not on data

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.

Causal claims hinge on arguments, not on data

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

Causal claims hinge on arguments, not on data

1. Statistical evidence
    - Simone Biles swung on the uneven bars. She won a gold medal.
    - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
    - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you:		Causal effect
	Swing	Do not swing	of swinging
Simone Biles	Yes (1)	?	?
Ian	?	No (0)	?

Causal claims hinge on arguments, not on data

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you:		Causal effect
	Swing	Do not swing	of swinging
Simone Biles	Yes (1)	No (0)	?
Ian	?	No (0)	?

Causal claims hinge on arguments, not on data

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you: Swing	Do not swing	Causal effect of swinging
Simone Biles	Yes (1)	No (0)	$+1$
Ian	?	No (0)	?

Causal claims hinge on arguments, not on data

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you: Swing	Do not swing	Causal effect of swinging
Simone Biles	Yes (1)	No (0)	$+1$
Ian	No (0)	No (0)	?

Causal claims hinge on arguments, not on data

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you: Swing	Do not swing	Causal effect of swinging
Simone Biles	Yes (1)	No (0)	$+1$
Ian	No (0)	No (0)	0


EXTRAVERGINE
olio
olio
olio
OLIVAOLIVA

Fundamental problem of causal inference

Holland 1986

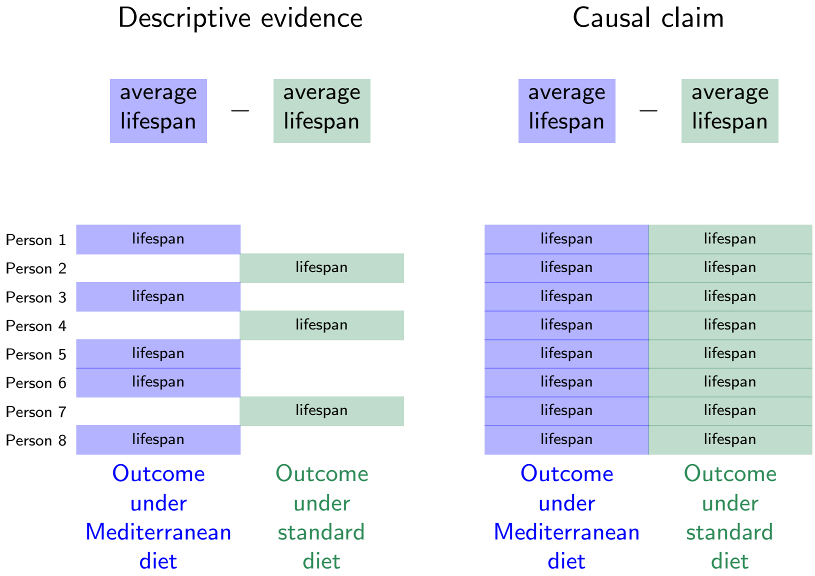
Descriptive evidence

average lifespan − average lifespan

	Outcome under Mediterranean diet	Outcome under standard diet
Person 1	lifespan	
Person 2		lifespan
Person 3	lifespan	
Person 4		lifespan
Person 5	lifespan	
Person 6	lifespan	
Person 7		lifespan
Person 8	lifespan	

Fundamental problem of causal inference

Holland 1986



Fundamental problem of causal inference

Holland 1986

Descriptive evidence

average lifespan − average lifespan

	Outcome under Mediterranean diet	Outcome under standard diet
Person 1	lifespan	missing
Person 2	missing	lifespan
Person 3	lifespan	missing
Person 4	missing	lifespan
Person 5	lifespan	missing
Person 6	lifespan	missing
Person 7	missing	lifespan
Person 8	lifespan	missing

Causal claim

average lifespan − average lifespan

Outcome under Mediterranean diet	Outcome under standard diet
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan

Fundamental problem of causal inference

Holland 1986

Descriptive evidence

average lifespan − average lifespan

Causal claim

average lifespan − average lifespan

Causal inference is a **missing data** problem

	Outcome under Mediterranean diet	Outcome under standard diet
Person 1	lifespan	missing
Person 2	missing	lifespan
Person 3	lifespan	missing
Person 4	missing	lifespan
Person 5	lifespan	missing
Person 6	lifespan	missing
Person 7	missing	lifespan
Person 8	lifespan	missing

Outcome under Mediterranean diet	Outcome under standard diet
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan

Mathematical notation: Potential outcomes

Mathematical notation: Potential outcomes

Y_i	Outcome	Whether person i survived

Mathematical notation: Potential outcomes

Y_i	Outcome	Whether person i survived
A_i	Treatment	Whether person i ate a Mediterranean diet

Mathematical notation: Potential outcomes

Y_i	Outcome	Whether person i survived
A_i	Treatment	Whether person i ate a Mediterranean diet
Y_i^a	Potential Outcome	Outcome person i would realize if assigned to treatment value a

Mathematical notation: Potential outcomes

Y_i	Outcome	Whether person i survived
A_i	Treatment	Whether person i ate a Mediterranean diet
Y_i^a	Potential Outcome	Outcome person i would realize if assigned to treatment value a

Examples:

$Y_{\text{Ian}} = \texttt{survived}$	Ian survived

Mathematical notation: Potential outcomes

Y_i	Outcome	Whether person i survived
A_i	Treatment	Whether person i ate a Mediterranean diet
Y_i^a	Potential Outcome	Outcome person i would realize if assigned to treatment value a

Examples:

$Y_{\text{Ian}} = \texttt{survived}$	Ian survived
$A_{\text{Ian}} = \texttt{MediterraneanDiet}$	Ian ate a Mediterranean diet

Mathematical notation: Potential outcomes

Y_i	Outcome	Whether person i survived
A_i	Treatment	Whether person i ate a Mediterranean diet
Y_i^a	Potential Outcome	Outcome person i would realize if assigned to treatment value a

Examples:

$Y_{\text{Ian}} = \texttt{survived}$	Ian survived
$A_{\text{Ian}} = \texttt{MediterraneanDiet}$	Ian ate a Mediterranean diet
$Y_{\text{Ian}}^{\text{MediterraneanDiet}} = \texttt{survived}$	Ian would survive on a Mediterranean diet

Mathematical notation: Potential outcomes

Y_i	Outcome	Whether person i survived
A_i	Treatment	Whether person i ate a Mediterranean diet
Y_i^a	Potential Outcome	Outcome person i would realize if assigned to treatment value a

Examples:

$Y_{\text{Ian}} = \texttt{survived}$	Ian survived
$A_{\text{Ian}} = \texttt{MediterraneanDiet}$	Ian ate a Mediterranean diet
$Y_{\text{Ian}}^{\text{MediterraneanDiet}} = \texttt{survived}$	Ian would survive on a Mediterranean diet
$Y_{\text{Ian}}^{\text{StandardDiet}} = \texttt{died}$	Ian would die on a standard diet

Mathematical notation: Potential outcomes

Y_i	Outcome	Whether person i survived
A_i	Treatment	Whether person i ate a Mediterranean diet
Y_i^a	Potential Outcome	Outcome person i would realize if assigned to treatment value a

Examples:

$Y_{\text{Ian}} = \texttt{survived}$	Ian survived
$A_{\text{Ian}} = \texttt{MediterraneanDiet}$	Ian ate a Mediterranean diet
$Y_{\text{Ian}}^{\text{MediterraneanDiet}} = \texttt{survived}$	Ian would survive on a Mediterranean diet
$Y_{\text{Ian}}^{\text{StandardDiet}} = \texttt{died}$	Ian would die on a standard diet

Discuss.

Which potential outcome is observed?

Which is counterfactual?

The consistency assumption

The consistency assumption

$Y_i^{\text{MediterraneanDiet}}$

$Y_i^{\text{StandardDiet}}$

Potential Outcomes

The consistency assumption

$Y_i^{\text{MediterraneanDiet}}$

$Y_i^{\text{StandardDiet}}$

Potential Outcomes

Y_i

Factual Outcomes

The consistency assumption

Consistency Assumption

$$Y_i^{A_i} = Y_i$$

$Y_i^{\text{MediterraneanDiet}}$

$Y_i^{\text{StandardDiet}}$

Potential Outcomes

Y_i

Factual Outcomes

Mathematical notation: Potential outcomes are fixed

A person's potential outcome is a **fixed quantity**

Mathematical notation: Potential outcomes are fixed

A person's potential outcome is a **fixed quantity**

$$Y_{\text{Ian}}^{\text{MediterraneanDiet}} = \texttt{survived}$$

Mathematical notation: Potential outcomes are fixed

A person's potential outcome is a **fixed quantity**

$$Y_{\text{Ian}}^{\text{MediterraneanDiet}} = \texttt{survived}$$

The outcome for a random person is a **random variable**

Mathematical notation: Potential outcomes are fixed

A person's potential outcome is a **fixed quantity**

$$Y_{\text{Ian}}^{\text{MediterraneanDiet}} = \texttt{survived}$$

The outcome for a random person is a **random variable**

- Draw a random person from the population

Mathematical notation: Potential outcomes are fixed

A person's potential outcome is a **fixed quantity**

$$Y_{\text{Ian}}^{\text{MediterraneanDiet}} = \texttt{survived}$$

The outcome for a random person is a **random variable**

- Draw a random person from the population
- Assign them a Mediterranean diet

Mathematical notation: Potential outcomes are fixed

A person's potential outcome is a **fixed quantity**

$$Y_{\text{Ian}}^{\text{MediterraneanDiet}} = \texttt{survived}$$

The outcome for a random person is a **random variable**

- Draw a random person from the population
- Assign them a Mediterranean diet
- The outcome $Y^{\text{MediterraneanDiet}}$ is a random variable:
  - takes the value `survived` if we randomly sample some people
  - takes the value `died` if we randomly sample others

Mathematical notation: Potential outcomes are fixed

A person's potential outcome is a **fixed quantity**

$$Y_{\text{Ian}}^{\text{MediterraneanDiet}} = \texttt{survived}$$

The outcome for a random person is a **random variable**

- Draw a random person from the population
- Assign them a Mediterranean diet
- The outcome $Y^{\text{MediterraneanDiet}}$ is a random variable:
  - takes the value `survived` if we randomly sample some people
  - takes the value `died` if we randomly sample others

Check for understanding:

Does it make sense to write $\text{V}(Y_i^a)$? How about $\text{V}(Y^a)$

Notation: Expectation operator

The **expectation operator** E() denotes the population mean

$$\mathrm{E}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} Y_i^a$$

The quantity Y^a inside the expectation must be a random variable

Notation: Expectation operator

The **expectation operator** E() denotes the population mean

$$\text{E}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} Y_i^a$$

The quantity Y^a inside the expectation must be a random variable

A **conditional expectation** is denoted with a vertical bar

$$\text{E}(Y \mid A = a) = \frac{1}{n_a}\sum_{i:A_i=a} Y_i$$

Practice: How would you say this in English?

We might wonder how a person's earnings relate to whether they hold a college degree

1. $\text{E}(\text{Earnings} \mid \text{Degree} = \texttt{TRUE}) > \text{E}(\text{Earnings} \mid \text{Degree} = \texttt{FALSE})$

2. $\text{E}(\text{Earnings}^{\text{Degree}=\texttt{TRUE}}) > \text{E}(\text{Earnings}^{\text{Degree}=\texttt{FALSE}})$

Practice: How would you say this in English?

We might wonder how a person's earnings relate to whether they hold a college degree

1. $\text{E}(\text{Earnings} \mid \text{Degree} = \texttt{TRUE}) > \text{E}(\text{Earnings} \mid \text{Degree} = \texttt{FALSE})$
   - Average earnings are higher among those with college degrees

2. $\text{E}(\text{Earnings}^{\text{Degree}=\texttt{TRUE}}) > \text{E}(\text{Earnings}^{\text{Degree}=\texttt{FALSE}})$

Practice: How would you say this in English?

We might wonder how a person's earnings relate to whether they hold a college degree

1. $\text{E}(\text{Earnings} \mid \text{Degree} = \texttt{TRUE}) > \text{E}(\text{Earnings} \mid \text{Degree} = \texttt{FALSE})$
    - Average earnings are higher among those with college degrees
2. $\text{E}(\text{Earnings}^{\text{Degree}=\texttt{TRUE}}) > \text{E}(\text{Earnings}^{\text{Degree}=\texttt{FALSE}})$
    - On average, a degree causes higher earnings

Practice: How would you write this in math?

1. On average, students who do the homework learn more than those who don't.

2. On average, doing the homework causes more learning.

Practice: How would you write this in math?

1. On average, students who do the homework learn more than those who don't.

$$E(\text{Learning} \mid \text{HW} = \texttt{TRUE}) > E(\text{Learning} \mid \text{HW} = \texttt{FALSE})$$

2. On average, doing the homework causes more learning.

Practice: How would you write this in math?

1. On average, students who do the homework learn more than those who don't.

$$\text{E}(\text{Learning} \mid \text{HW} = \texttt{TRUE}) > \text{E}(\text{Learning} \mid \text{HW} = \texttt{FALSE})$$

2. On average, doing the homework causes more learning.

$$\text{E}(\text{Learning}^{\text{HW}=\texttt{TRUE}}) > \text{E}(\text{Learning}^{\text{HW}=\texttt{FALSE}})$$

Learning goals for today

By the end of class, you will be able to

- explain the fundamental problem of causal inference and the need for causal arguments
- define potential outcomes
- recall mathematical concepts from probability
  - random variables
  - expectation
  - conditional expectation