

Defining causal effects

STSCI / INFO / ILRST 3900: Causal Inference

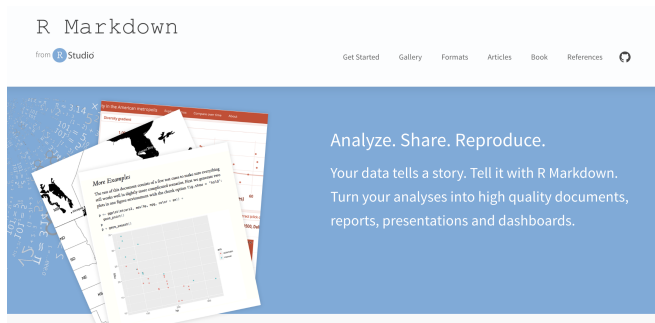
28 Aug 2025

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

Typesetting



As soon as possible, you should

- ▶ **Install R** (statistical software)
- ▶ **Install RStudio** (user interface)
- ▶ **Bookmark the RMarkdown cheat sheet** (documentation)

Causal claims hinge on arguments, not just data



Left photo: By Fernando Frazão/Agência Brasil - http://agenciabrasil.ebc.com.br/sites/_agenciabrasil2013/files/fotos/1035034-_mg_0802_04.08.16.jpg, CC BY 3.0, <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 just data

1. Statistical evidence

- ▶ Simone Biles swung on the uneven bars. She won a gold medal.

Causal claims hinge on arguments, not just 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 just 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 just 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.

What do we mean when we say “cause”?

Causal claims hinge on arguments, not just data

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

Causal claims hinge on arguments, not just data

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

Causal claims hinge on arguments, not just data

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

Causal claims hinge on arguments, not just data

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

Causal claims hinge on arguments, not just data

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

Causal claims hinge on arguments, not just data

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

Causal claims hinge on arguments, not just data

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

Causal claims hinge on arguments, not just data

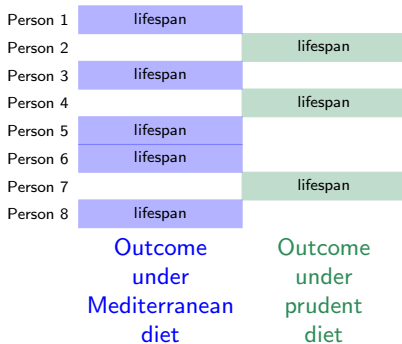
		Do you win gold if you:		Causal effect
		Swing	Do not swing	of swinging
Simone Biles	Yes (1)	No (0)		+1
Sam	No (0)	No (0)		0



Fundamental problem of causal inference

Holland 1986

Descriptive evidence



Fundamental problem of causal inference

Holland 1986

Descriptive evidence



Person 1	lifespan	
Person 2		lifespan
Person 3	lifespan	
Person 4		lifespan
Person 5	lifespan	
Person 6	lifespan	
Person 7		lifespan
Person 8	lifespan	

Outcome
under
Mediterranean
diet

Outcome
under
prudent
diet

Causal claim



lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan

Outcome
under
Mediterranean
diet

Outcome
under
prudent
diet

Fundamental problem of causal inference

Holland 1986

Descriptive evidence



Causal claim



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
prudent
diet

lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan

Outcome
under
Mediterranean
diet

Outcome
under
prudent
diet

Fundamental problem of causal inference

Holland 1986

Descriptive evidence



Causal claim



Causal inference is a **missing data** problem

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
prudent
diet

lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan

Outcome
under
Mediterranean
diet

Outcome
under
prudent
diet

Mathematical notation: Potential outcomes¹

¹Capital letters and lowercase letters mean different things!

Mathematical notation: Potential outcomes¹

Y_i Outcome

Whether person i survived

¹Capital letters and lowercase letters mean different things!

Mathematical notation: Potential outcomes¹

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

¹Capital letters and lowercase letters mean different things!

Mathematical notation: Potential outcomes¹

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

¹Capital letters and lowercase letters mean different things!

Mathematical notation: Potential outcomes¹

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:

If assigned prudent diet	If assigned mediterranean diet
Died	Survived

¹Capital letters and lowercase letters mean different things!

Mathematical notation: Potential outcomes¹

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:

If assigned prudent diet	If assigned mediterranean diet
Died	Survived

$Y_{\text{Sam}} = \text{survived}$ We observe that Sam survived

¹Capital letters and lowercase letters mean different things!

Mathematical notation: Potential outcomes¹

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:

If assigned prudent diet	If assigned mediterranean diet
Died	Survived

$Y_{\text{Sam}} = \text{survived}$ We observe that Sam survived

$A_{\text{Sam}} = \text{MedDiet}$ We observed that Sam ate a Mediterranean diet

¹Capital letters and lowercase letters mean different things!

Mathematical notation: Potential outcomes¹

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:

If assigned prudent diet	If assigned mediterranean diet
Died	Survived

$Y_{\text{Sam}} = \text{survived}$ We observe that Sam survived

$A_{\text{Sam}} = \text{MedDiet}$ We observed that Sam ate a Mediterranean diet

$Y_{\text{Sam}}^{\text{MedDiet}} = \text{survived}$ If Sam had been assigned a Mediterranean diet he would have survived

¹Capital letters and lowercase letters mean different things!

Mathematical notation: Potential outcomes¹

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:

If assigned prudent diet	If assigned mediterranean diet
Died	Survived

$Y_{\text{Sam}} = \text{survived}$ We observe that Sam survived

$A_{\text{Sam}} = \text{MedDiet}$ We observed that Sam ate a Mediterranean diet

$Y_{\text{Sam}}^{\text{MedDiet}} = \text{survived}$ If Sam had been assigned a Mediterranean diet
he would have survived

$Y_{\text{Sam}}^{\text{PruDiet}} = \text{died}$ If Sam had been assigned a prudent diet
he would have died

¹Capital letters and lowercase letters mean different things!

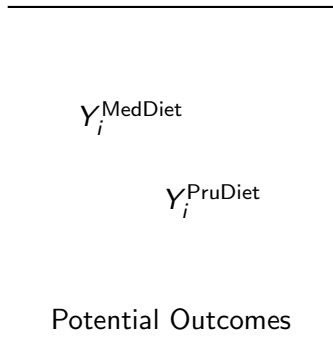
Practice

Using the slip of paper you received and the diet you follow, what is

- ▶ Y_i
- ▶ A_i
- ▶ Y_i^{MedDiet}
- ▶ Y_i^{PruDiet}

The consistency assumption

The consistency assumption



The consistency assumption



A diagram consisting of a horizontal line at the top and a vertical line on the right, forming an L-shape. Inside the corner, the text Y_i^{MedDiet} is positioned higher and to the left, and Y_i^{PruDiet} is positioned lower and to the right.

Y_i^{MedDiet}

Y_i^{PruDiet}

Potential Outcomes



A diagram consisting of a horizontal line at the top and a vertical line on the left, forming an L-shape. Inside the corner, the text Y_i is centered.

Y_i

Factual Outcomes

The consistency assumption

Consistency Assumption

$$Y_i^{A_i} = Y_i$$

Y_i^{MedDiet}

Y_i^{PruDiet}

Potential Outcomes

Y_i

Factual Outcomes

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

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

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

► Average earnings are higher among those with college degrees

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

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

- ▶ Average earnings are higher among those with college degrees

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

- ▶ The average earning if everyone was assigned to get a degree is higher than the average earnings if everyone was assigned to not get a degree
- ▶ On average, getting a degree causes higher earnings

Practice:

1. On average, individuals who eat a Mediterranean diet survive more/less than those who eat a prudent diet:
2. On average, eating a Mediterranean diet causes people to survive more/less

Practice:

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:

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

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

2. On average, doing the homework causes more learning

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

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

You can now

- ▶ Read Chapter 1 of [Hernán and Robins 2020](#)
- ▶ Start Problem Set 1 (due Sep 9)