PSY 503: Foundations of Psychological Methods

Lecture 3: Introduction to Causality, Potential Outcomes, and Experimental Design

Robin Gomila

Princeton

September 7, 2020

Causality

- Causal relationships are everywhere
- Psychologists are generally interested in **identifying** and **quantifying** causal relationships
  - What does this mean?

"Causes of Effects" or "Effects of Causes"?

“Causes of Effects” or “Effects of Causes”?

- **Effects** have **many causes**
  - brain chemistry
  - hormones
  - sensory cues
  - prenatal environment
  - early experiences
  - genes

    ...
  - social environment
  - ecological pressures

“Causes of Effects” or “Effects of Causes”?

- **Effects** have **many causes**
  - brain chemistry
  - hormones
  - sensory cues
  - prenatal environment
  - early experiences
  - genes

    ...
  - social environment
  - ecological pressures
- Psychology studies generally focus on the **“effects of causes”**
  - Primary contribution of statistics —concerned with measurement (Holland, 1986)

What do we learn from studies of the effects of causes?

- Causal relationships
  - "Manipulating X impacts Y"
- Direction of effects
  - "Manipulating X decreases / increases Y"
  - "Manipulating X decreases / increases the probability of Y"
- Magnitude of effects

What do we not necessarily learn from studies of the effects of causes?

- The *other causes* of the effect under study
- The *responsibility* held by the "cause" under study
- The *appropriate* course of action to impact Y in the real world

Illustration: Consequences of Racial Bias

- Resume studies (e.g., Bertrand & Mullainathan, 2004)
- Shooter bias (e.g., Correll et al., 2002)
- Race and perception of crime related objects (e.g., Eberhardt et al., 2004)

Illustration: Consequences of Racial Bias

- Resume studies (e.g., Bertrand & Mullainathan, 2004)
- Shooter bias (e.g., Correll et al., 2002)
- Race and perception of crime related objects (e.g., Eberhardt et al., 2004)
- Why are these studies important? What do we not learn from these studies? What should we not conclude?

Define causality

- Notion of causality is tied to an **action** applied to a **unit**

- In psychology:
  - action is called a **treatment**
  - unit is often an **individual**

Should California prison guards wear body cameras? Lawyers demand them in disability case

BY MATT KRISTOFFERSEN

JUNE 12, 2020 06:00 AM

Causality (a bit more) formally

- Let d_i be a treatment (e.g., wearing a body camera)
- Let Y_i be an outcome (e.g., use of force)

Definition

A treatment d_i has a causal effect on an outcome Y_i for individual i (e.g., a prison guard) if the action of d_i on individual i impacts Y_i (i.e., the extent to which prison guard i uses force)

The Rubin Causal Model

A Powerful Framework to Study Causal Effects and Threats to Causal Inference

Introduction

- Framework developed by Donald Rubin (Rubin, 1974, 1975)
- Mathematical definition of causal effects at the individual level
- Establishes the impossibility of measuring causal effects for an individual

Core concept: Potential Outcomes

- Each individual has different **potential outcomes** in alternative environments
- To measure the causal effect of a treatment d_i for individual i:
  - measure the outcome of interest Y_i for individual i in two environments E_0 and E_1 that differ on one aspect: d_i

Core concept: Potential Outcomes

- Each individual has different **potential outcomes** in alternative environments
- To measure the causal effect of a treatment d_i for individual i:
  - measure the outcome of interest Y_i for individual i in two environments E_0 and E_1 that differ on one aspect: d_i
- Why is this impossible?

Illustration

Definition

- E_0: $d_i = 0$, treatment was not applied to individual i
- E_1: $d_i = 1$, treatment was applied to individual i
- Imagine we can observe both $Y_i(0)$ and $Y_i(1)$ for the exact same individual i in E_0 and E_1, respectively.

For individual i, the causal effect τ_i of the treatment d_i is defined as the difference between two potential outcomes:

$$\tau_i = Y_i(1) - Y_i(0) \tag{1}$$

Implications

- If $\tau_i = 0$, wearing a body camera has no causal effect on Y_i

Implications

- If $\tau_i = 0$, wearing a body camera has no causal effect on Y_i
- If $\tau_i \neq 0$, wearing a body camera has a causal effect on Y_i

Implications

- If $\tau_i = 0$, wearing a body camera has no causal effect on Y_i
- If $\tau_i \neq 0$, wearing a body camera has a causal effect on Y_i
- The *magnitude* of the causal effect for individual i is τ_i, such that

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

Fundamental problem of causal inference (Holland, 1986)

- We cannot calculate τ_i

Fundamental problem of causal inference (Holland, 1986)

- We cannot calculate τ_i (Are we really interested in τ_i?)

Causal Effects in Populations

Population

- A *population* is a set of units defined a priori by the researcher
  - Think: *population of interest*

Definition of Population

The term population refers to all of the individuals from a *specified* group

Hypothetical scenario: set up

- Let our *population of interest* be a team of 8 prison guards ($N = 8$) from a NJ prison
- Let Y_i be the number of time guard i used force in one-on-one interactions with a prisoner in the past 30 days
- Let τ_i be the causal effect of wearing a body camera for individual i

Hypothetical schedule of potential outcomes

guard i	$Y_i(0)$	$Y_i(1)$	τ_i
1	12	10	-2
2	7	1	-6
3	1	1	0
4	25	27	2
5	5	0	-5
6	15	5	-10
7	0	1	1
8	13	9	-4

Average Treatment Effect (ATE) in the population

guard i	$Y_i(0)$	$Y_i(1)$	τ_i
1	12	10	-2
2	7	1	-6
3	1	1	0
4	25	27	2
5	5	0	-5
6	15	5	-10
7	0	1	1
8	13	9	-4

- ATE in this population?

Average Treatment Effect (ATE) in the population

guard i	$Y_i(0)$	$Y_i(1)$	τ_i
1	12	10	-2
2	7	1	-6
3	1	1	0
4	25	27	2
5	5	0	-5
6	15	5	-10
7	0	1	1
8	13	9	-4

- ATE in this population?
- Take the average of the last column

$$\text{ATE} = \frac{(-2) + (-6) + 0 + 2 + (-5) + (-10) + 1 + (-4)}{8} = -3$$

- **Conclusion:** on average in this population, wearing a body camera decreases use of force by 3 instances within the 30-day period

ATE: Formal defintion

$$\text{ATE} = \frac{1}{N}\sum_{i=1}^{N}\tau_i \tag{2}$$

ATE: Formal defintion

$$\text{ATE} = \frac{1}{N}\sum_{i=1}^{N}\tau_i \qquad (2)$$

- Population ATE is defined as the sum of τ_i divided by N, the number of individuals i the population
- Describes how the outcome of interest Y_i would change on average in the population if the treatment was applied to every single individual in the population.
- It is an extremely important concept in psychology
  - Identify and quantify average effects of treatments in populations

Problems?

guard i	$Y_i(0)$	$Y_i(1)$	τ_i
1	12	10	-2
2	7	1	-6
3	1	1	0
4	25	27	2
5	5	0	-5
6	15	5	-10
7	0	1	1
8	13	9	-4

Problems?

guard i	$Y_i(0)$	$Y_i(1)$	τ_i
1	12	10	-2
2	7	1	-6
3	1	1	0
4	25	27	2
5	5	0	-5
6	15	5	-10
7	0	1	1
8	13	9	-4

- τ_i is forever unknown
- We generally don't have access to the entire population of interest
- How do we **estimate** population average treatment effects?

Experimental Design

Why experiments?

- Identify the presence of causal effects
  - Does a causal effect exist at all?
  - Statistical significance
- Estimate the magnitude of causal effects
  - What is the direction of the causal effect?
  - Is the causal effect relevant?
  - Practical significance

Experimental set up

- Let our population of interest be all the prison guards of a specific U.S. prison ($N = 150$)
- Let's imagine that we can run an experiment on the entire population of interest
- **Random assignment:**
  - Control vs. Treatment condition
- Let z_i indicate assignment of guard i to an experimental condition
  - $z_i = 0$ if guard i was assigned to the control condition
  - $z_i = 1$ if guard i was assigned to the treatment condition
  - Assume *two sided compliance*: $d_i = z_i$

Hypothetical experimental dataset

Hypothetical experimental dataset

guard i	Z_i	Y_i
1	0	10
2	0	15
3	1	12
4	0	12
5	1	8
6	1	5
	...	
150	0	2

Hypothetical experimental dataset with potential outcomes

Hypothetical experimental dataset with potential outcomes

guard i	Z_i	$Y_i(0)$	$Y_i(1)$	τ_i
1	0	10	?	?
2	0	15	?	?
3	1	?	12	?
4	0	12	?	?
5	1	?	8	?
6	1	?	5	?
		...		
150	0	2	?	?

Potential and observed outcomes

- The previous slide makes it clear that $Y_i(1)$s are observed for individuals who are treated, and $Y_i(0)$s are observed for individuals who are not treated
- Causal inference is a missing data problem!

Potential and observed outcomes

- The previous slide makes it clear that $Y_i(1)$s are observed for individuals who are treated, and $Y_i(0)$s are observed for individuals who are not treated
- Causal inference is a missing data problem!
- We can express the connection between the observed outcome Y_i and the underlying potential outcomes through the “switching equation”:

$$Y_i = Y_i(1)z_i + Y_i(0)(1 - z_i) \tag{3}$$

Estimation of the ATE in experiments

Estimation of the ATE in experiments

First, let's work in Equation (2) to express the ATE with regards to $Y_i(0)$ and $Y_i(1)$:

$$
\begin{aligned}
\text{ATE} &= \frac{1}{N}\sum_{i=1}^{n}\tau_i \\
&= \frac{1}{N}\sum_{i=1}^{n}(Y_i(1) - Y_i(0)) \\
&= \frac{1}{N}\sum_{i=1}^{n}Y_i(1) - \frac{1}{N}\sum_{i=1}^{n}Y_i(0) \\
&= \mu_{Y(1)} - \mu_{Y(0)}
\end{aligned}
\tag{4}
$$

in which $\mu_{Y(1)}$ is the average value of $Y_i(1)$ for all individuals and $\mu_{Y(0)}$ is the average value of $Y(0)$ for all subjects.

Estimation of the ATE in experiments

In experimental studies, researchers estimate $\mu_{Y_i(1)}$ using the mean $\widehat{\mu}_{Y(1)}$ of all observed $Y_i(1)$ and $Y_i(0)$ using the mean $\widehat{\mu}_{Y(0)}$ of observed $Y_i(0)$. We have:

$$\widehat{\text{ATE}} = \widehat{\mu}_{Y(1)} - \widehat{\mu}_{Y(0)} \tag{5}$$

in which $\widehat{\text{ATE}}$ is the estimated ATE, $\widehat{\mu}_{Y(1)}$ is the estimated $\mu_{Y(1)}$, and $\widehat{\mu}_{Y(0)}$ is the estimated $\mu_{Y(0)}$.