PSY 503: Foundations of Psychological Methods
Lecture 17: Causality, Experimental Design, and Regression

Robin Gomila

Princeton

November 4, 2020

Predictive inference, causal inference, and regression

Predictive inference, causal inference, and regression

- Regression is often used for **predictive inference**

Predictive inference, causal inference, and regression

- Regression is often used for **predictive inference**
  - Once we estimate the CEF or BLP, we can use regression to predict outcomes based on predictors
  - Focus of inference is **between units.**

Predictive inference, causal inference, and regression

- Regression is often used for **predictive inference**
  - Once we estimate the CEF or BLP, we can use regression to predict outcomes based on predictors
  - Focus of inference is **between units.** i.e., What values of Y do we expect for different values of X?

Predictive inference, causal Inference, and regression

Predictive inference, causal Inference, and regression

- What questions could prediction researchers ask about protests?

Predictive inference, causal Inference, and regression

- What questions could prediction researchers ask about protests?
- Are people who attend protests. . .

Predictive inference, causal Inference, and regression

- What questions could prediction researchers ask about protests?
- Are people who attend protests...
  - ... more identified with the cause?

Predictive inference, causal Inference, and regression

- What questions could prediction researchers ask about protests?
- Are people who attend protests...
  - ... more identified with the cause?
  - ... more concerned with social justice?

Predictive inference, causal Inference, and regression

- What questions could prediction researchers ask about protests?
- Are people who attend protests. . .
  - . . . more identified with the cause?
  - . . . more concerned with social justice?
  - . . . more likely to publicly display a sign for that cause outside of their home?

Predictive inference, causal Inference, and regression

- What questions could prediction researchers ask about protests?
- Are people who attend protests. . .
  - . . . more identified with the cause?
  - . . . more concerned with social justice?
  - . . . more likely to publicly display a sign for that cause outside of their home?
  - . . . more likely to go out to vote?

Predictive inference, causal Inference, and regression

- What questions could prediction researchers ask about protests?
- Are people who attend protests. . .
  - . . . more identified with the cause?
  - . . . more concerned with social justice?
  - . . . more likely to publicly display a sign for that cause outside of their home?
  - . . . more likely to go out to vote?
  - . . . more likely to accept to donate to charity?

Predictive inference, causal Inference, and regression

- What questions could prediction researchers ask about protests?
- Are people who attend protests. . .
  - . . . more identified with the cause?
  - . . . more concerned with social justice?
  - . . . more likely to publicly display a sign for that cause outside of their home?
  - . . . more likely to go out to vote?
  - . . . more likely to accept to donate to charity?
  - . . . more likely to accept to volunteer for an organization?

Predictive inference, causal Inference, and regression

- Prediction research is consequential!
  - e.g., spend money on promoting a cause on those who are most likely to react

Predictive inference, causal Inference, and regression

- Prediction research is consequential!
  - e.g., spend money on promoting a cause on those who are most likely to react
- We could also imagine how prediction researchers would want to estimate the CEF of Y given a battery of predictors (e.g., demographic characteristics)

Predictive inference, causal Inference, and regression

- Again, predictive inference is about predicting how an outcome varies **between units**
  - i.e., prediction compares **different units**
  - e.g., What values of Y do we expect for different values of X?

Predictive inference, causal Inference, and regression

- Again, predictive inference is about predicting how an outcome varies **between units**
  - i.e., prediction compares **different units**
  - e.g., What values of Y do we expect for different values of X?
- Causal inference compares different treatments if applied to the **same unit**
  - Question is: *What would have happened under a different treatment option?*

Predictive inference, causal Inference, and regression

- What questions could causality researchers ask about protests?

Predictive inference, causal Inference, and regression

- What questions could causality researchers ask about protests?
- Does attending a protest **make** people...

Predictive inference, causal Inference, and regression

- What questions could causality researchers ask about protests?
- Does attending a protest **make** people...
  - ... more identified with the cause?

Predictive inference, causal Inference, and regression

- What questions could causality researchers ask about protests?
- Does attending a protest **make** people...
  - ... more identified with the cause?
  - ... more concerned with social justice?

Predictive inference, causal Inference, and regression

- What questions could causality researchers ask about protests?
- Does attending a protest **make** people...
  - ... more identified with the cause?
  - ... more concerned with social justice?
  - ... more likely to publicly display a sign for that cause outside of their home?

Predictive inference, causal Inference, and regression

- What questions could causality researchers ask about protests?
- Does attending a protest **make** people...
  - ... more identified with the cause?
  - ... more concerned with social justice?
  - ... more likely to publicly display a sign for that cause outside of their home?
  - ... more likely to go out to vote?

Predictive inference, causal Inference, and regression

- What questions could causality researchers ask about protests?
- Does attending a protest **make** people. . .
  - . . . more identified with the cause?
  - . . . more concerned with social justice?
  - . . . more likely to publicly display a sign for that cause outside of their home?
  - . . . more likely to go out to vote?
  - . . . more likely to accept to donate to charity?

Predictive inference, causal Inference, and regression

- What questions could causality researchers ask about protests?
- Does attending a protest **make** people...
  - ... more identified with the cause?
  - ... more concerned with social justice?
  - ... more likely to publicly display a sign for that cause outside of their home?
  - ... more likely to go out to vote?
  - ... more likely to accept to donate to charity?
  - ... more likely to accept to volunteer for an organization?

Predictive inference, causal Inference, and regression

- Causal effects: Comparison between different potential outcomes of **what might have occurred under different scenarios**

Predictive inference, causal Inference, and regression

- Causal effects: Comparison between different potential outcomes of **what might have occurred under different scenarios**
  - Comparison between counterfactuals
  - Comparison between what did happen and what could have happened

Causality:

Review of potential outcomes framework

Running example: Protesting and attitudes

Question:

What is the impact protesting (vs. not protesting) for a cause on people's attitudes towards that cause?

Running example: Protesting and attitudes

- Suppose we had a list of individuals who:

Running example: Protesting and attitudes

- Suppose we had a list of individuals who:
    1. Live about an hour away from San Francisco

Running example: Protesting and attitudes

- Suppose we had a list of individuals who:
  1. Live about an hour away from San Francisco
  2. Want to go to a protest

Running example: Protesting and attitudes

- Suppose we had a list of individuals who:
  1. Live about an hour away from San Francisco
  2. Want to go to a protest
  3. Unfortunately do not have the means and ability to go to that protest
     - e.g., no way to get to the protest (e.g., no car, no public transportation); getting an unpaid leave of absence on the protest day is too costly

Potential outcomes

- Let Y_i be individual i's attitude towards the cause after the protest, measured on a 7-point scale ($1 =$ not at all important, $7 =$ very important)

Potential outcomes

- Let Y_i be individual i's attitude towards the cause after the protest, measured on a 7-point scale ($1 =$ not at all important, $7 =$ very important)
- Let D_i be a bernoulli random variable indicating whether individual i actually went to the protest, such that $D_i = 1$ if individual i went to the protest and $D_i = 0$ if individual i did not go to the protest

Potential outcomes

- Let Y_i be individual i's attitude towards the cause after the protest, measured on a 7-point scale (1 = not at all important, 7 = very important)
- Let D_i be a bernoulli random variable indicating whether individual i actually went to the protest, such that $D_i = 1$ if individual i went to the protest and $D_i = 0$ if individual i did not go to the protest
- Let $Y_i(d_i)$ be the attitude of individual i towards the cause after the protest for $D_i = d_i$
  - i.e., we consider two potential outcomes for each individual: $Y_i(0)$ and $Y_i(1)$

Potential outcomes

- Each individual i has a schedule of **potential outcomes** for Y_i, **conditional** on D_i

Potential outcomes

- Each individual i has a schedule of **potential outcomes** for Y_i, **conditional** on D_i
  - If going to the protest **does not have a causal effect** on individual i's attitude towards the cause:

$$Y_i(0) = Y_i(1)$$

  - If going to the protest **does have a causal effect** on individual i's attitude towards the cause:

$$Y_i(0) \neq Y_i(1)$$

Hypothetical schedule of potential outcomes

individual i	$Y_i(0)$	$Y_i(1)$	τ_i
1	5	6	+1
2	5	7	+2
3	4	4	0
4	6	7	+1
5	7	6	-1
6	7	7	0
7	4	7	+3
8	4	6	+2

Average treatment effect

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

- For these 8 individuals, we have $ATE = 1$

Fundamental problem of causal inference

We do not have access to the full schedule of potential outcomes for an individual i, therefore we cannot calculate τ_i

Possible solution to fundamental problem of causal inference

- Experimental studies: randomly assign individuals to treatment vs. control
  - Let Z_i be an indicator of random assignment for individual i
  - For now, we assume full compliance: $z_i = d_i$

Possible solution to fundamental problem of causal inference

- Experimental studies: randomly assign individuals to treatment vs. control
  - Let Z_i be an indicator of random assignment for individual i
  - For now, we assume full compliance: $z_i = d_i$
- Observe $\hat{Y}_i(D_i = z_i)$ for each individual i

Possible solution to fundamental problem of causal inference

- Experimental studies: randomly assign individuals to treatment vs. control
  - Let Z_i be an indicator of random assignment for individual i
  - For now, we assume full compliance: $z_i = d_i$
- Observe $\hat{Y}_i(D_i = z_i)$ for each individual i
- Focus on estimating the ATE

Hypothetical experimental dataset

individual i	Z_i	$\widehat{Y}_i$
1	0	5
2	0	5
3	1	4
4	0	6
5	1	6
6	1	7
	...	
200	0	4

Hypothetical experimental dataset with potential outcomes

individual i	Z_i	$Y_i(0)$	$Y_i(1)$	τ_i
1	0	5	?	?
2	0	5	?	?
3	1	?	4	?
4	0	6	?	?
5	1	?	6	?
6	1	?	7	?
		...		
200	0	4	?	?

Random assignment and unbiased inference

- Causal inference is a "missing data" problem
- Randomization addresses the missing data problem by creating two groups of observations that are, in expectation, identical prior to application of the treatment
- **In expectation: treatment and control groups have the same potential outcomes**

Random assignment and unbiased inference

- Under random assignment, we have

$$\mathbb{E}[Y_i(1)|D_i = 0] = \mathbb{E}[Y_i(1)|D_i = 1] = \mathbb{E}[Y_i(1)]$$

Random assignment and unbiased inference

- Under random assignment, we have

$$\mathbb{E}[Y_i(1)|D_i = 0] = \mathbb{E}[Y_i(1)|D_i = 1] = \mathbb{E}[Y_i(1)]$$

$$\mathbb{E}[Y_i(0)|D_i = 0] = \mathbb{E}[Y_i(0)|D_i = 1] = \mathbb{E}[Y_i(0)]$$

Random assignment and unbiased inference

- Under random assignment, we have

$$\mathbb{E}[Y_i(1)|D_i = 0] = \mathbb{E}[Y_i(1)|D_i = 1] = \mathbb{E}[Y_i(1)]$$

$$\mathbb{E}[Y_i(0)|D_i = 0] = \mathbb{E}[Y_i(0)|D_i = 1] = \mathbb{E}[Y_i(0)]$$

- This implies that the randomly assigned values of D_i do not convey any information about the potential values of $Y_i(0)$ and $Y_i(1)$
  - The randomly assigned values of D_i determine which value of Y_i we actually *observe*,

Random assignment and unbiased inference

- Under random assignment, we have

$$\mathbb{E}[Y_i(1)|D_i = 0] = \mathbb{E}[Y_i(1)|D_i = 1] = \mathbb{E}[Y_i(1)]$$

$$\mathbb{E}[Y_i(0)|D_i = 0] = \mathbb{E}[Y_i(0)|D_i = 1] = \mathbb{E}[Y_i(0)]$$

- This implies that the randomly assigned values of D_i do not convey any information about the potential values of $Y_i(0)$ and $Y_i(1)$
  - The randomly assigned values of D_i determine which value of Y_i we actually *observe*, but they are independent of the *potential outcomes* $Y_i(0)$ and $Y_i(1)$

Random assignment and unbiased inference

- Implication: We can estimate ATE using the difference between the observed average outcome among the treated and the observed average outcome among the control
- We have

$$\begin{aligned} \mathbb{E}[\mu_{Y(1)} - \mu_{Y(0)}] &= \mathbb{E}[\mu_{Y(1)}] - \mathbb{E}[\mu_{Y(0)}] \\ &= \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] \\ &= \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\ &= \mathbb{E}[\tau_i] \\ &= ATE \end{aligned}$$

- This demonstrates that when units are randomly assigned, a comparison between average outcomes in treatment and control groups is an **unbiased estimator** of the ATE

Keep in mind two core assumptions

- Excludability
- Non-interference

Threat of selection bias when no random assignment

- Without random assignment, this identification strategy unravels!
  - Treatment and control groups are not anymore a random subset of all units in the sample

Threat of selection bias when no random assignment

- Without random assignment, this identification strategy unravels!
  - Treatment and control groups are not anymore a random subset of all units in the sample
- We confront a **selection problem**
  - Receiving the treatment may be systematically related to potential outcomes

Threat of selection bias when no random assignment

- Without random assignment, this identification strategy unravels!
  - Treatment and control groups are not anymore a random subset of all units in the sample
- We confront a **selection problem**
  - Receiving the treatment may be systematically related to potential outcomes
- Under non-random assignment, what does our identification strategy from the previous slide actually yield?

Threat of selection bias when no random assignment

- To understand the issue, we can subtract and add $\mathbb{E}[Y_i(0)|D_i = 1]$ from the expected difference between treated and untreated outcomes $\mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$. We have

$$\mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$$
$$= \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] + \mathbb{E}[Y_i(0)|D_i = 1]$$
$$= \text{ATE among the treated} - \text{selection bias}$$

Regression analysis for experimental design

Since experimental treatments are bernoulli random variables, the CEF $\mathbb{E}[Y|D]$ is inherently linear

Regression analysis for experimental design

Since experimental treatments are bernoulli random variables, the CEF $\mathbb{E}[Y|D]$ is inherently linear and under **random assignment**, **excludability**, and **non-interference**,

Regression analysis for experimental design

Since experimental treatments are bernoulli random variables, the CEF $\mathbb{E}[Y|D]$ is inherently linear and under **random assignment**, **excludability**, and **non-interference**, we can use simple linear regression with robust SEs to estimate the ATE,

Regression analysis for experimental design

Since experimental treatments are bernoulli random variables, the CEF $\mathbb{E}[Y|D]$ is inherently linear and under **random assignment**, **excludability**, and **non-interference**, we can use simple linear regression with robust SEs to estimate the ATE, and test the alternative hypothesis that it is different from 0.