How Economists Learn from Data

EC 201: Principles of Microeconomics

Kyle Raze

Winter 2020

Prologue

Last Time

1. Why bother learning from data?
    - Figure out whether policies work or don't work.
    - Test theories.
2. Why does correlation $\neq$ causation?
    - Selection bias!
3. When can correlation $\implies$ causation?
    - Randomized contrial trials (experiments).

Today

1. Regression analysis.
   - The workhorse of data science.
2. Natural experiments.
   - Sometimes we get lucky.

Regression

Correlation

Correlation coefficient

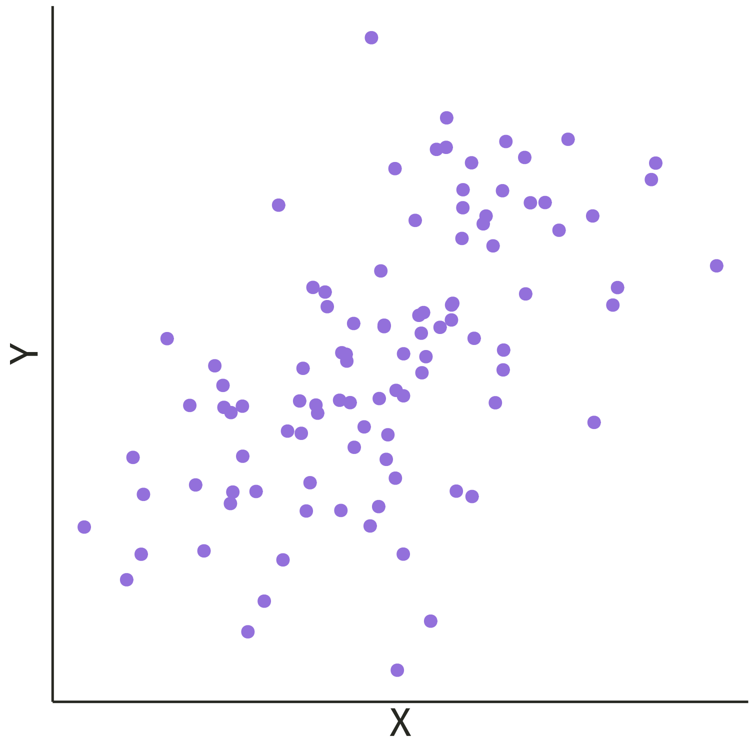
> A measure of the strength of a relationship between two variables, denoted by ρ.

$-1 \leqslant \rho < 0 \implies$ negative correlation.

$\rho = 0 \implies$ no correlation (unrelated).

$0 < \rho \leqslant 1 \implies$ positive correlation.

Correlation coefficient = 0.58



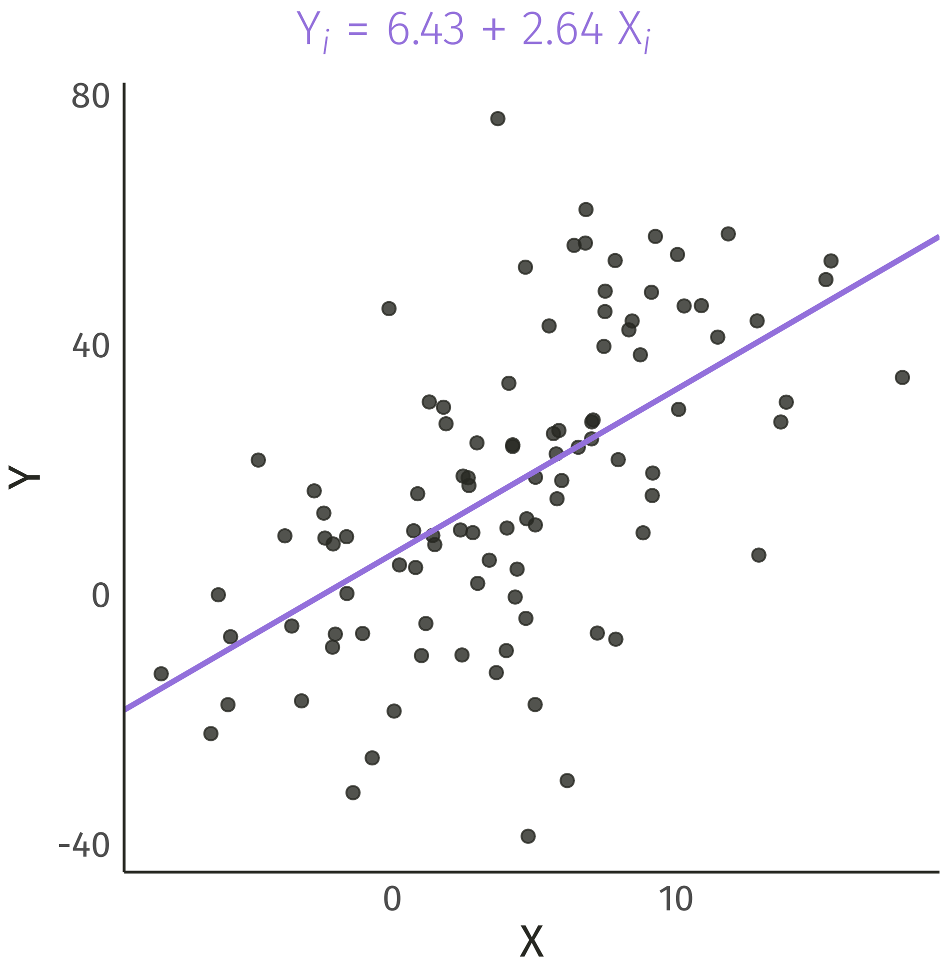
Regression

Goal: Identify the effect of a treatment variable X on an outcome variable Y while **controlling** for potential confounders.

Economists often rely on regression analysis for statistical comparisons.

- Regression analysis facilitates *other things equal* comparisons by explicitly controlling for certain variables.
- Failure to control for confounding variables leads to **omitted-variable bias**, a close cousin of selection bias.

Simple Linear Regression



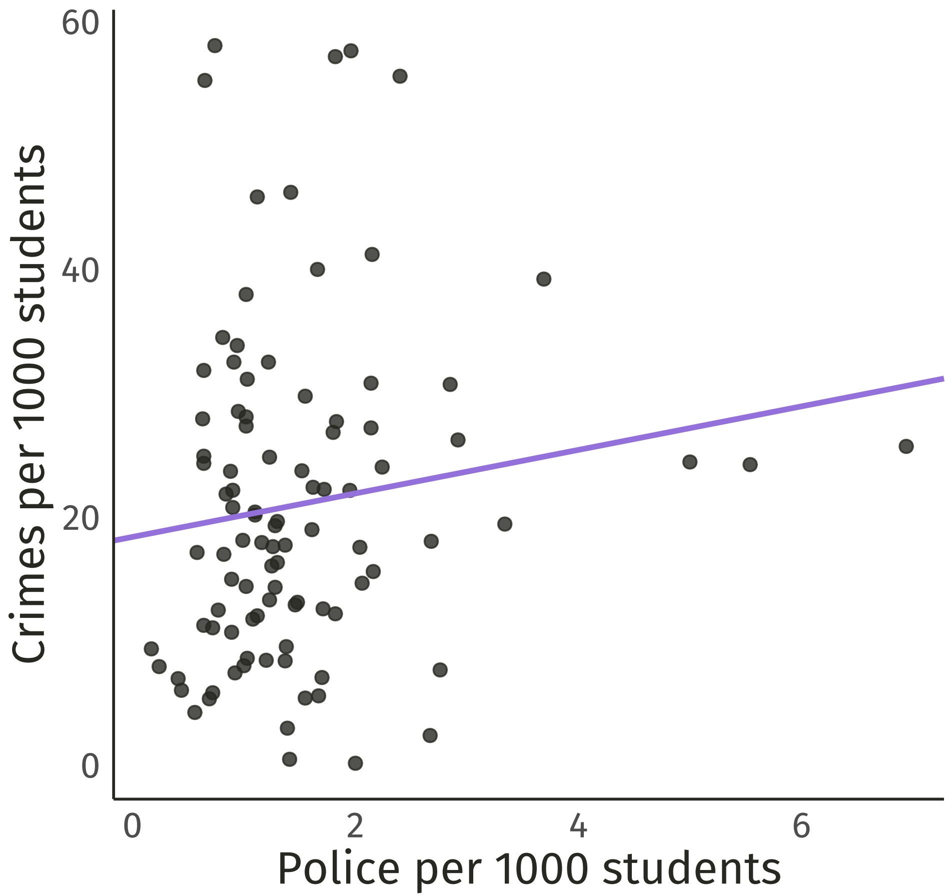
Model

$$Y_i = \beta_1 + \beta_2 X_i + e_i$$

- β_1 = intercept
- β_2 = slope
- e_i = error term

Simple Linear Regression

$\text{Crime}_i = 18.41 + 1.76\ \text{Police}_i$



Q: Do 👮 *cause* crime!?

A: Probably not

$\longrightarrow$ Colleges experiencing high crime rates probably respond by hiring more police.

Example: Returns to Education

The optimal investment in education by students, parents, and legislators depends in part on the monetary *return to education*.

Thought experiment:

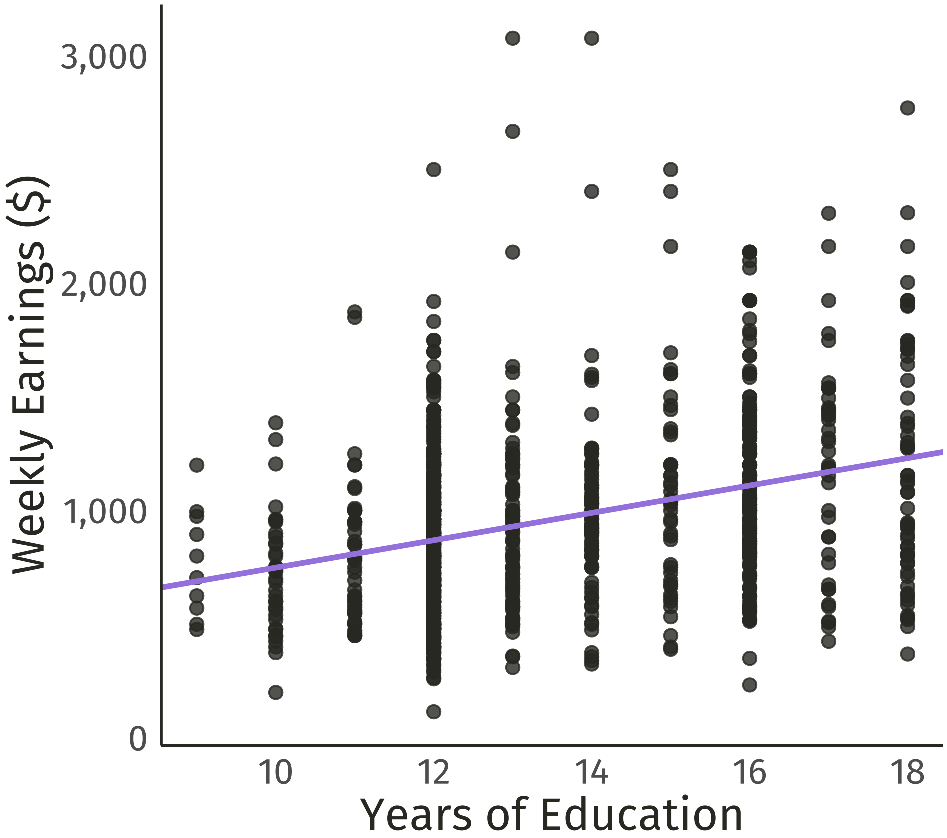
- Randomly select an individual.
- Give her an additional year of education.
- How much do her earnings increase?

The change in her earnings describes the **causal effect** of education on earnings.

Causality

Example: Returns to Education

$\text{Earnings}_i = 146.95 + 60.21 \text{ Schooling}_i$



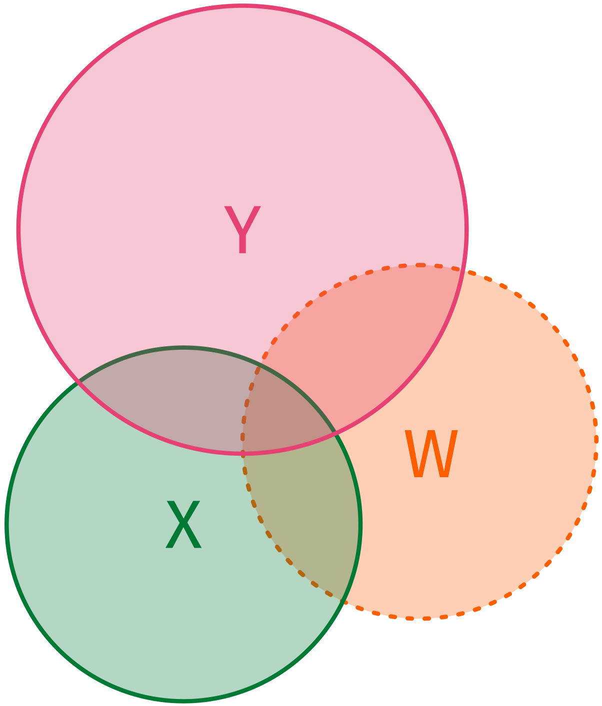
Q: Does the slope isolate the causal effect of an additional year of education on weekly earnings?

A: Probably not

$\longrightarrow$ There could be other variables that influence earnings and schooling.

Omitted Variables

Bias



Y = Outcome

X = Treatment

W = Omitted variable

If **W** is correlated with both **X** and **Y** $\longrightarrow$ omitted variable bias –$\rightarrow$ regression fails to isolate the causal effect of **X** on **Y**.

Controlling for Confounders

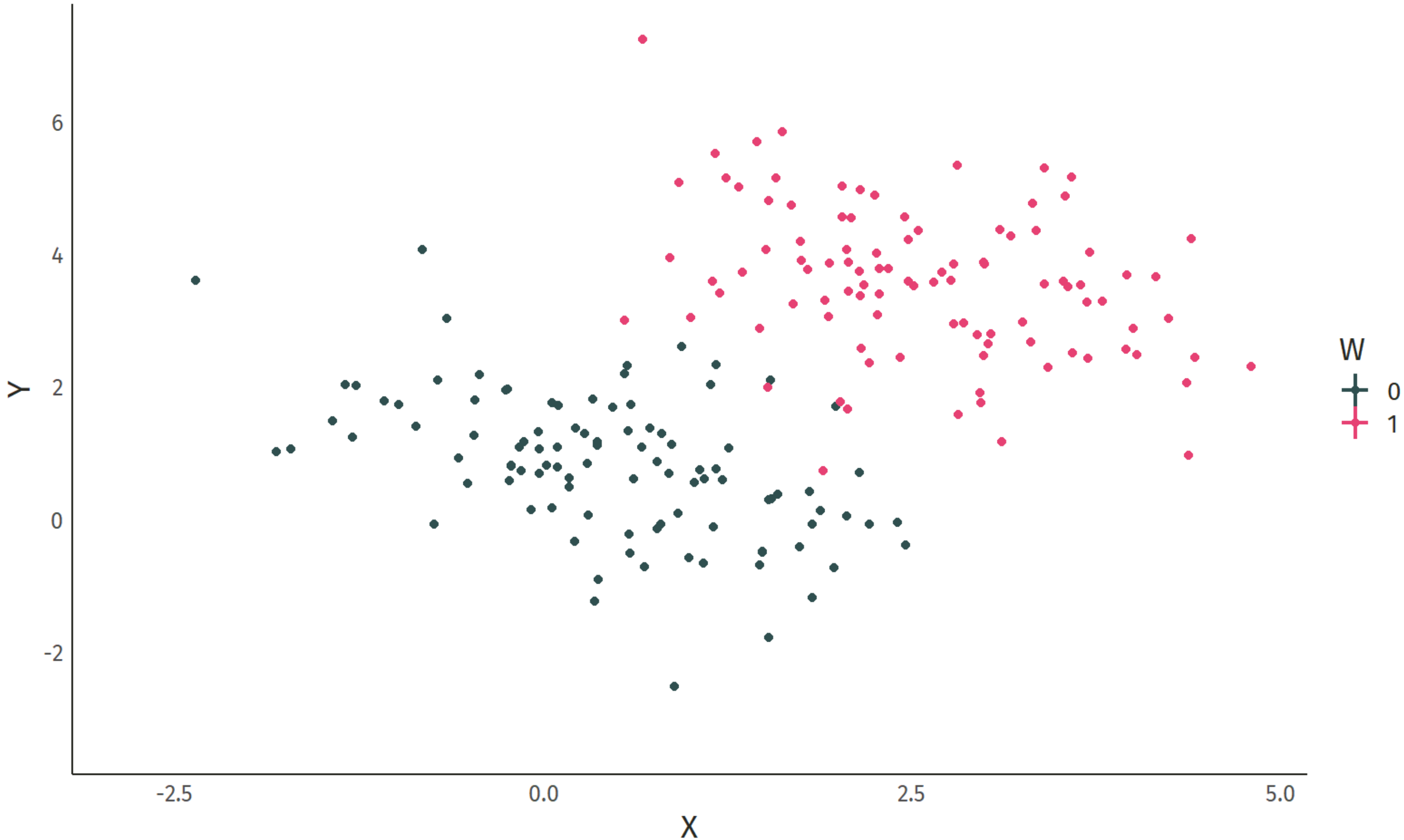
Economists can control for a confounder W by including it in the regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + e_i$$

- W_i is a **control variable**.
- By including W_i, adjusts the data to account for confounding effects of W.
- **Note:** The model doesn't care whether a right-hand side variable is a treatment or control variable, but we do.

Controlling for Confounders

The relationship between Y and X, controlling for a binary variable W
1. Start with raw data. Correlation between X and Y: 0.395



Controlling for Confounders

Example: Returns to Education

Two regressions of earnings on schooling. The second regression controls for IQ score, a proxy for ability.

Outcome: Weekly Earnings

Parameter	**1**	**2**
Intercept	146.95	**-128.89**
	(77.72)	**(92.18)**
Schooling (Years)	60.21	**42.06**
	(5.70)	**(6.55)**
IQ Score (Points)		**5.14**
		(0.96)

Standard errors in parentheses.

Bias from omitting IQ score

$$\begin{aligned} &= \text{"short"} - \text{"long"} \\ &= 60.21 - 42.06 \\ &= 18.15 \end{aligned}$$

The first regression mistakenly attributes some of the influence of intelligence to education.

Natural Experiments

Causality

Q: Given that selection bias and omitted variables are ubiquitous, how can economists estimate the returns to education and other causal effects of other interventions?

Option 1: Run an **experiment**.

- Randomly assign education (might be difficult/unethical).
- Randomly encourage education (might work).
- Randomly assign programs that affect education (*e.g.*, mentoring).

Option 2: Look for a **natural experiment** (*e.g.*, a policy or accident in society that arbitrarily increased education for one subset of people).

- Admissions cutoffs.
- Lottery enrollment and/or capacity constraints.

Oregon Medicaid Experiment

Background

As of 2016, 27 million Americans do not have health insurance.

- Down from 46.5 million in 2010.
- US is the only developed country without universal coverage.

Healthcare spending accounts for a growing share of the economy.

- Almost 18% of GDP or $10,000 per person per year!
- US spends more on healthcare than any other developed country.

Oregon Medicaid Experiment

Background

Medicaid: A social assistance program that provides health insurance to families on welfare, the disabled, other children from low-income families, and low-income pregnant women.

- Federal program run by states.

Policy Question: Should we expand Medicaid to cover more of the uninsured?

Research Questions

- Would Medicaid expansion reduce costly emergency room visits?
- Would Medicaid expansion improve health?

Oregon Medicaid Experiment

Natural Experiment

In 2008, Oregon decided to expand its version of Medicaid, called Oregon Health Plan (OHP).

- **Problem:** 75,000 applicants, but only 30,000 spots!
- **Solution:** Ration spots by lottery.

Lottery = random assignment!

- **Treatment group:** 30,000 lottery winners.
- **Control group:** 45,000 people who did not win medicaid lottery.

Effect of OHP on Coverage and Healthcare Use

Outcome	Control Mean	Treatment Effect	Standard Error	N
Ever on Medicaid?	0.141	0.256	0.004	74922
Any hospital admissions?	0.067	0.005	0.002	74922
Any emergency room visits?	0.345	0.017	0.006	24646
Emergency room visits	1.020	0.101	0.029	24646
Outpatient visits	1.910	0.314	0.054	23741
Any prescriptions?	0.637	0.025	0.008	23741

Informal Rule: Estimate of treatment effect more than twice its standard error $\Longrightarrow$ effect is statistically distinguishable from zero.

Effect of OHP on Health and Personal Finances

Outcome	Control Mean	Treatment Effect	Standard Error	N
Good Health?	0.548	0.039	0.008	23741
Physical health index	45.500	0.290	0.210	12229
Mental health index	44.400	0.470	0.240	12229
Cholesterol	204.000	0.530	0.690	12229
Systolic blood pressure	119.000	-0.130	0.300	12229
Big medical expenditures?	0.055	-0.011	0.005	12229
Any medical debt?	0.568	-0.032	0.010	12229

Informal Rule: Estimate of treatment effect more than twice its standard error $\Longrightarrow$ effect is statistically distinguishable from zero.

Differences-in-Differences

Minimum Wage

Research Question: Do binding minimum wage laws cause unemployment?

- Theory predicts that binding minimum wage laws reduce employment levels.
- **Q:** How could we test this prediction?

Idea 1: Compare employment levels in states with binding minimum wage laws to those without.

- **Q:** Is this a good idea? Would it isolate the causal effect?
- **A:** Probably not. States with binding minimum wages laws are different than those without $\longrightarrow$ selection bias!

Minimum Wage

Research Question: Do binding minimum wage laws cause unemployment?

- Theory predicts that binding minimum wage laws reduce employment levels.
- **Q:** How could we test this prediction?

Idea 2: Compare employment levels in a state before and after it increases the minimum wage?

- **Q:** Is this a good idea? Would it isolate the causal effect?
- **A:** Probably not. Other things might coincide with the policy change (*e.g.*, a recession) $\longrightarrow$ omitted variable bias!

Minimum Wage

Research Question: Do binding minimum wage laws cause unemployment?

- Theory predicts that binding minimum wage laws reduce employment levels.
- **Q:** How could we test this prediction?

Idea 3: Two wrongs make a right?

- Compare employment levels in a state that raises its minimum wage with a state that doesn't, before and after the policy change.
- A difference-in-differences comparison.

Card and Krueger (1994)

Influential study of the impact of minimum wage laws on fast-food workers.

Natural Experiment: New Jersey increased its minimum wage in 1992, but neighboring Pennsylvania did not.

- **Control group:** Fast food restaurants in Pennsylvania.
- **Treatment group:** Fast food restaurants in New Jersey.

Differences-in-Differences

Card and Krueger (1994)

Effect of Minimum Wage on Employment

Outcome: Number Full-Time Workers

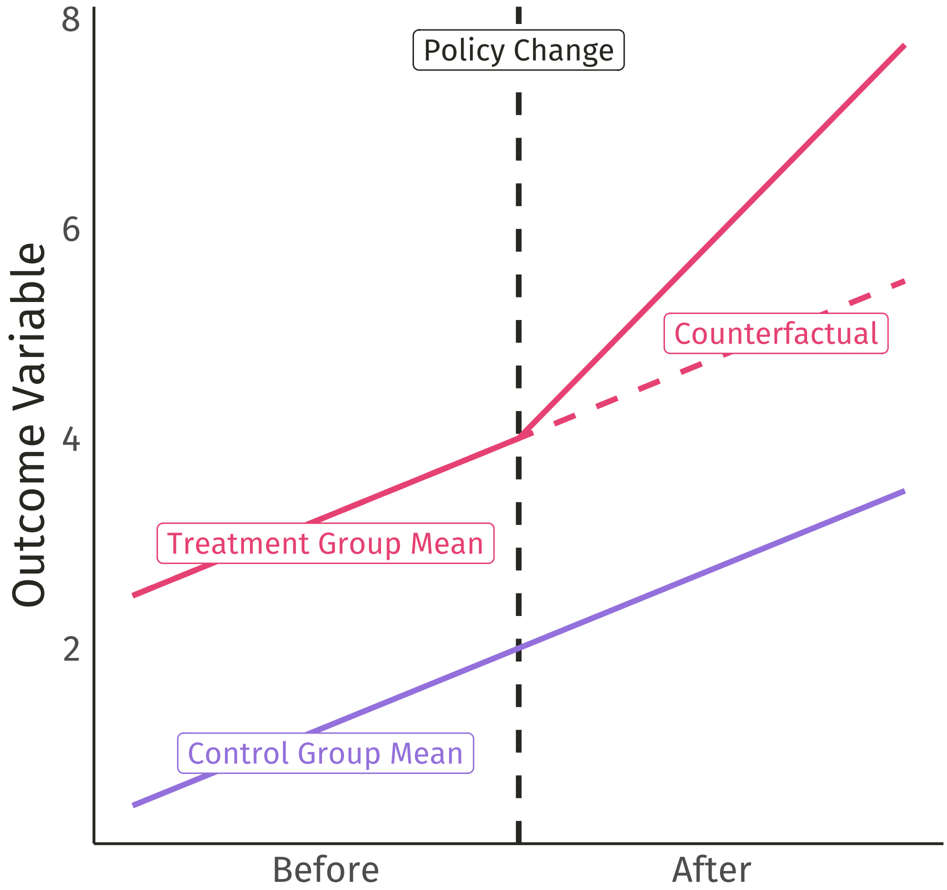
Group	After	Before	Difference
Treatment (NJ)	21.03	20.44	0.59
Control (PA)	21.17	23.33	-2.16

Difference-in-differences = 0.59 − -2.16

= 2.75.

Result: Increasing the minimum wage did not reduce employment!

Differences-in-Differences



Internal Validity

Q: When should we trust a difference-in-differences comparison?

A: When we believe that the comparison groups exhibit **parallel trends** in the absence of the policy change.

Podcast

Podcast Question: According to Raj Chetty,

> **A.** No social assistance program pays for itself in the long run, on average.
>
> **B.** All social assistance programs pay for themselves in the long run, on average.
>
> **C.** Social assistance programs that target adults tend to pay for themselves in the long run, but those targeted toward children do not, on average.
>
> **D.** Social assistance programs that target children tend to pay for themselves in the long run, but those targeted toward adults do not, on average.