Week 4, Class 8

Modern Causal Inference

Sean Westwood

Modern Causal Inference

How do we know if X really causes Y?
Major Topics:

- The fundamental problem of causal inference
- Average Treatment Effect (ATE) calculation
- Difference-in-differences designs
- More on Internal vs. external validity

The Fundamental Problem of Causal Inference

A Voter and a Mailer

We want to know if X causes Y. Do campaign mailers cause people to vote?
The Scenario: A voter (Sarah) receives a campaign mailer three days before election day
The Question: Did the mailer cause her to vote?
The Challenge: We can only observe one reality - the voter either received the mailer OR didn't receive it, but not both

> **The fundamental problem**: We can never observe what would have happened to the same person under different conditions

Counterfactuals: A Missing Reality

Formal notation for potential outcomes:

- $Y_i(1)$ = Outcome for person i **if treated** (Sarah gets mailer)
- $Y_i(0)$ = Outcome for person i **if not treated** (Sarah gets no mailer)

Individual causal effect: $\tau_i = Y_i(1) - Y_i(0)$

The fundamental problem: We observe either $Y_i(1)$ OR $Y_i(0)$, but never both

- **For Sarah specifically**:
  - **What we observe**: $Y_{Sarah}(1) = 1$ (Sarah got mailer, voted)
  - **What we need but can't see**: $Y_{Sarah}(0) = ?$ (Would Sarah have voted without mailer?)
  - **Sarah's causal effect**: $\tau_{Sarah} = Y_{Sarah}(1) - Y_{Sarah}(0) = 1 - ?$ ::: {.warning-box} **Causal inference is fundamentally about estimating unobservable counterfactuals** :::

The Solution

Since can't observe both $Y_{Sarah}(1)$ and $Y_{Sarah}(0)$ for the same person, we need an alternative

Compare groups of similar people instead!

- If we can't see what Sarah would do without the mailer…
- Maybe we can find people **just like Sarah** who didn't get the mailer
- Then compare Sarah (who got mailer) to these similar people (who didn't get mailer)
- The difference tells us the likely effect of the mailer

If the groups are truly similar in every way except treatment, then:

- Group differences ≈ Individual treatment effects
- $\bar{Y}_{treated} - \bar{Y}_{control} \approx$ Average of $Y_i(1) - Y_i(0)$

::: {.highlight-box} **The challenge**: How do we ensure the groups are truly similar? :::

Randomized Controlled Trials: The Gold Standard

The Logic of Randomization

Random assignment is how we ensure the groups are truly similar!

- If we randomly assign who gets the mailer, then treated and control groups should be similar on average.

What we want to estimate: The Average Treatment Effect (ATE)

The ATE answers the question: "On average, how much does the treatment change the outcome?"

- **ATE = Average difference in outcomes if everyone got treated vs. if no one got treated**
- **In our example**: How much more likely is the average person to vote if they get a mailer vs. if they don't?

Green & Gerber Experiment

The Question: Do get-out-the-vote (GOTV) efforts actually increase turnout?
The Design:

- Randomly selected households to receive GOTV contact
- Control group received no contact
- Measured actual voting behavior from public records

Analyzing Real Experimental Data

```r
# Load GOTV experiment data
gotv_data <- read_csv("../../data/gotv_experiment.csv")

gotv_data %>%
  group_by(treatment) %>%
  summarise(
    count = n(),
    baseline_prob = mean(baseline_turnout_prob, na.rm = TRUE),
    .groups = "drop"
  )
```

```
# A tibble: 4 × 3
  treatment      count baseline_prob
  <chr>          <int>         <dbl>
1 Control         1241         0.566
2 Personal Visit  1267         0.569
3 Phone Call      1254         0.559
4 Postcard        1238         0.560
```

Calculating the Average Treatment Effect (ATE)

What we estimate with random assignment:

$$\widehat{\text{ATE}} = \bar{Y}_{treated} - \bar{Y}_{control}$$

Where:

- $\bar{Y}_{treated}$ = average outcome for treated group
- $\bar{Y}_{control}$ = average outcome for control group

Calculating the Average Treatment Effect (ATE)

```r
turnout_by_group <- gotv_data %>%
  group_by(treatment) %>%
  summarise(
    count = n(),
    turnout_rate = mean(voted_2022, na.rm = TRUE),
    .groups = "drop"
  )

print(turnout_by_group)
```

```
# A tibble: 4 × 3
  treatment      count turnout_rate
  <chr>          <int>        <dbl>
1 Control         1241        0.564
2 Personal Visit  1267        0.687
3 Phone Call      1254        0.632
4 Postcard        1238        0.624
```

```r
treated_rate <- turnout_by_group %>%
  filter(treatment == "Phone Call") %>%
  pull(turnout_rate)

control_rate <- turnout_by_group %>%
  filter(treatment == "Control") %>%
  pull(turnout_rate)

ate <- treated_rate - control_rate

print(paste("Average Treatment Effect:", round(ate, 2)))
```

```
[1] "Average Treatment Effect: 0.07"
```

Interpreting the ATE

Linking notation to results:

$$\widehat{\text{ATE}} = \bar{Y}_{treated} - \bar{Y}_{control} = 0.63 - 0.56 = 0.07$$

What this means:

- On average, receiving GOTV contact increased the probability of voting by 7 percentage points
- This is the causal effect of the treatment
- We can make this causal claim because of random assignment

Visualizing the GOTV Treatment Effect

ATE Example 2: Resume Audit Study

Research Question: Do employers discriminate against Black job applicants?
Experimental Design (Bertrand & Mullainathan, 2004):

- **Treatment**: Resume with Black-sounding name (e.g., “Lakisha”, “Jamal”)
- **Control**: Resume with White-sounding name (e.g., “Emily”, “Greg”)
- **Outcome**: Whether employer called back for interview (1 = yes, 0 = no)

ATE Example 2: Resume Audit Study Results

```r
# Load real resume audit experiment data
resume_data <- read_csv("../../data/resume.csv")

callback_by_race <- resume_data %>%
  group_by(race) %>%
  summarise(
    count = n(),
    callback_rate = mean(call),
    .groups = "drop"
  )
print(callback_by_race)
```

```
# A tibble: 2 × 3
  race  count callback_rate
  <chr> <int>         <dbl>
1 black  2435        0.0645
2 white  2435        0.0965
```

```r
ate_discrimination <- callback_by_race %>%
  filter(race == "white") %>%
  pull(callback_rate) - callback_by_race %>%
  filter(race == "black") %>%
  pull(callback_rate)
```

White-sounding names received 3.2 percentage points more callbacks than Black-sounding names

Visualizing Employment Discrimination

ATE Example 3: STAR Class Size Experiment

Research Question: Does reducing class size improve high school graduation rates?
Experimental Design (Tennessee STAR Experiment):

- **Treatment**: Small classes (13-17 students) in grades K-3
- **Control**: Regular classes (22-25 students) in grades K-3
- **Outcome**: High school graduation (1 = graduated, 0 = did not graduate)

ATE Example 3: STAR Class Size Experiment Results

```r
star_data <- read_csv("../../data/STAR.csv")

class_comparison <- star_data %>%
  filter(classtype %in% c(1, 2), !is.na(hsgrad)) %>%
  mutate(class_size = ifelse(classtype == 1, "Small Class", "Regular Class")) %>%
  group_by(class_size) %>%
  summarise(
    count = n(),
    graduation_rate = mean(hsgrad)
  )
print(class_comparison)
```

```
# A tibble: 2 × 3
  class_size    count graduation_rate
  <chr>         <int>           <dbl>
1 Regular Class  1081           0.825
2 Small Class     902           0.836
```

```r
# Calculate ATE
ate_class_size <- class_comparison %>%
  filter(class_size == "Small Class") %>%
  pull(graduation_rate) -  class_comparison %>%
  filter(class_size == "Regular Class") %>%
  pull(graduation_rate)
```

Small classes increased high school graduation rates by 1.1 percentage points

STAR Results Across All Outcomes

Visualizing treatment effects on graduation, math, and reading:

- **Graduation**: Small classes increased graduation by 1.1 percentage points
- **Math**: Small classes improved 4th grade math scores by -0.3 points
- **Reading**: Small classes improved 4th grade reading scores by 3.5 points

When Experiments Aren’t Possible

Difference-in-Differences Designs

Sometimes we can't randomly assign treatments
Example: Studying the effect of minimum wage increases on employment

Find a "natural experiment" where treatment assignment is as good as random

Card & Krueger's Minimum Wage Study

In 1992 New Jersey raised the minimum wage, Pennsylvania didn't We want to leverage this to know if there were costs to the policy.

Compare fast-food employment in NJ vs PA before and after the policy change

Why this works:

- NJ and PA are similar in many ways
- The policy change was plausibly exogenous
- Any differences should be due to the minimum wage increase

DiD Logic: Removing Confounders

Simple comparisons don't work:

- **Just compare NJ before vs after?** → Could be due to economy-wide trends, not minimum wage
- **Just compare NJ vs PA after policy?** → States are different in many ways

The DiD Solution - Think of it as a "Double Subtraction":

- Calculate how NJ changed over time
  - NJ After - NJ Before = "NJ Change"
  - *But this includes both policy effect AND general trends*
- Calculate how PA changed over time
  - PA After - PA Before = "PA Change"
  - *This captures only general trends (no policy change)*
- Subtract PA's change from NJ's change
  - DiD Effect = (NJ Change) - (PA Change)
  - *This removes general trends, leaving only the policy effect*

DiD Logic: Removing Confounders

Imagine these employment levels:

- NJ Before: 20 workers per restaurant
- NJ After: 22 workers per restaurant
- PA Before: 19 workers per restaurant
- PA After: 20 workers per restaurant

Step 1: NJ Change = 22 - 20 = +2 workers

Step 2: PA Change = 20 - 19 = +1 worker

Step 3: DiD Effect = (+2) - (+1) = +1 worker

DiD Logic: Removing Confounders

Interpretation: After removing general trends, NJ's minimum wage increase caused employment to rise by 1 worker per restaurant

Key Assumption: NJ and PA would have followed the same trends if there had been no policy change (called "parallel trends")

Understanding Trends: Global vs Parallel

What is a "Global Trend"?

A change that affects *everyone* similarly over time:

- Economic recession → employment falls everywhere
- Technological change → productivity rises everywhere
- Seasonal effects → tourism drops in winter everywhere

What are "Parallel Trends"?

When two groups change at the *same rate* over time (but can start at different levels):

Understanding Trends: Global vs Parallel

Example: If the economy is booming, employment might rise in both NJ and PA

Why Global Trends Matter for DiD:

- Global trends affect both treatment (NJ) and control (PA) groups
- By comparing changes between the groups, we “cancel out” these global trends
- What’s left is the policy effect

Understanding Trends: Global vs Parallel Examples

Understanding Trends: Global vs Parallel

Why Parallel Trends Matter:

- If NJ and PA have different underlying trends, we can't tell whether differences are due to:
  - The policy (what we want to measure), OR
  - Different natural growth rates (confounding)
- Parallel trends ensure PA is a good "counterfactual" for NJ

> **DiD only works if trends would have been parallel without the policy!**

Summary: Take the difference over time, then difference across states

Understanding Trends: Global vs Parallel with Real Data

```r
minwage_data <- read_csv("../../data/minwage.csv")

did_table <- minwage_data %>%
  mutate(
    state  = if_else(str_detect(location, "NJ"), "NJ", "PA"),
    Before = fullBefore + partBefore,
    After  = fullAfter  + partAfter
  ) %>%
  group_by(state) %>%
  summarise(
    change = mean(After,  na.rm = TRUE) -
             mean(Before, na.rm = TRUE),
    .groups = "drop"
  )

did_effect <- did_table %>%
  summarise(DiD = change[state == "NJ"] - change[state == "PA"]) %>%
  pull(DiD)

did_table
```

```
# A tibble: 2 × 2
  state change
  <chr>  <dbl>
1 NJ     0.189
2 PA    -1.78
```

```r
print(paste("DiD effect:", round(did_effect, 2)))
```

```
[1] "DiD effect: 1.97"
```

Interpreting the DiD Result

Interpretation: After accounting for general trends (PA), the minimum wage increase in NJ increased employment by 1.97 jobs per restaurant

DiD removes confounders that affect both groups equally over time

Another DiD Example: Voting Access Reforms

The Research Question

Do voting access expansions increase Democratic vote share?
The Challenge: We can’t randomly assign voting reforms to states
The Natural Experiment: Some states expanded voting access between 2016-2020, others maintained status quo

The DiD Design

Treatment Group: Western states (expanded mail-in voting, early voting between 2016-2020)
Control Group: States in other regions with no major voting access changes
Pre-treatment Period: 2016 presidential election
Post-treatment Period: 2020 presidential election

The DiD Logic Applied

Interpretation: After accounting for general trends, voting access reforms increased Democratic vote share by 1.7 percentage points
What this controls for: National trends affecting all states (Trump presidency effects, COVID-19 impacts, national mood shifts, etc.)

Key DiD Assumptions

Parallel Trends: Western and other states would have had similar vote share trends without voting access reforms
No Other Differences: The regional groups didn't experience other major policy changes at the same time
Why This Example Works:

- **Clear treatment**: Voting access reforms (mail-in voting, early voting expansion) have specific implementation periods
- **Comparable groups**: Non-Western states serve as controls for national trends
- **Measurable outcome**: Vote shares are precisely recorded
- **Controls for trends**: Accounts for national political changes (candidate effects, national issues)

> Real research would need to test these assumptions carefully and consider alternative explanations!

Formal DiD Notation

The Mathematical Framework

The DiD estimator in formal notation:

$$\hat{\delta}_{DiD} = (\bar{Y}_{1,1} - \bar{Y}_{1,0}) - (\bar{Y}_{0,1} - \bar{Y}_{0,0})$$

Where:

- $\bar{Y}_{i,t}$ = Average outcome for group i in time period t
- $i = 1$: Treatment group (e.g., NJ, Voter ID states)
- $i = 0$: Control group (e.g., PA, Non-voter ID states)
- $t = 1$: After treatment period
- $t = 0$: Before treatment period

Breaking Down the Formula

Step 1 - Treatment Group Change: $(\bar{Y}_{1,1} - \bar{Y}_{1,0})$

- How much the treatment group changed from before to after
- *Includes both treatment effect AND time trends*

Step 2 - Control Group Change: $(\bar{Y}_{0,1} - \bar{Y}_{0,0})$

- How much the control group changed from before to after
- *Captures only time trends (no treatment)*

Step 3 - The Difference: $\hat{\delta}_{DiD}$

- Subtract control change from treatment change
- *Removes time trends, leaving only treatment effect*

Alternative Notation

You might also see DiD written as:

$$\hat{\delta}_{DiD} = \Delta\bar{Y}_{Treatment} - \Delta\bar{Y}_{Control}$$

Where Δ means "change from before to after"

Practical Guidelines

When to Trust Causal Claims

Strong evidence:

- Randomized controlled trials with good compliance
- Natural experiments with plausible exogeneity
- Multiple studies with consistent findings
- Transparent pre-registered analyses

Weak evidence:

- Simple correlational studies
- Cherry-picked comparisons
- Post-hoc explanations for findings
- Studies with obvious confounders

Building Your Causal Intuition

Always ask:

1. Could something else explain this relationship?
2. How were the groups formed?
3. What would the counterfactual look like?
4. Is the timing right for causation?
5. How robust are the findings?

Healthy skepticism is essential for good causal inference

Conclusions

What We’ve Learned Today

The fundamental problem: We can’t observe counterfactuals
The gold standard: Randomized controlled trials eliminate selection bias
The alternative: Natural experiments and difference-in-differences
The goal: Estimate average treatment effects
The challenge: Balancing internal and external validity

Next Class Preview

Linear regression as a tool for causal analysis

- How OLS fits lines through data
- Interpreting coefficients causally
- When regression can and cannot identify causal effects
- Building intuition for multivariate analysis

Coming up: We'll connect causal thinking to statistical modeling

Speaker notes