Week 4 - Descriptive Stats and Observation Studies

Causal Inference without Randomization

Umberto Mignozzetti

March 12

Week 4 - Descriptive Stats and Observation Studies

Causal Inference without Randomization

Umberto Mignozzetti

March 12

Today's Agenda

- Quantiles
- Standard deviation
- Observational studies vs RCTs (QSS 2.5-2.6)
- Differences in Differences
- Leader assassination DiD
- Internal and external validity
- Replication crisis

Revision

- What is the fundamental problem of causal inference?
- How can we (try to) solve that problem?
- Randomization!

Randomization

- Randomization ensures the treatment is not correlated with any other variable
- Randomization generates balance between treatment and control groups
- Treatment and control groups identical on average
- Allows us to estimate the average treatment effect simply as a difference in means: $\frac{1}{n}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$

Social Pressure Example

- August 2006 Primary State-wide Election in Michigan
- Send postcards with different, randomly-assigned messages
    - no message (control group)
    - civic duty message
    - "you are being studied" message (Hawthorne effect)
    - neighbourhood social pressure message

Social Pressure Example

```
tapply(social$yearofbirth, social$messages, mean) #
```

```
## Civic Duty    Control  Hawthorne  Neighbors
##   1956.341   1956.186   1956.295   1956.147
```

```
tapply(social$hhsize, social$messages, mean) # House
```

```
## Civic Duty    Control  Hawthorne  Neighbors
##   2.189126   2.183667   2.180138   2.187770
```

```
tapply(social$primary2006, social$messages, mean) #
```

```
## Civic Duty    Control  Hawthorne  Neighbors
##  0.3145377  0.2966383  0.3223746  0.3779482
```

```
tapply(social$primary2006, social$messages, mean)[-2
    mean(social$primary2006[social$messages == "Cont
```

```
## Civic Duty  Hawthorne  Neighbors
## 0.01789934 0.02573631 0.08130991
```

Quantiles

- Sometimes it is useful to look at the distribution of a given variable
- We can split a variable in many ways:
    - Quartiles
    - Quantiles
    - Percentiles
- Which quantile is the median?

Quantiles

- What is the median of {2, 5, 6, 10}?
- What is the median of {1, 2, 3, 4, 20}?
- Interquartile range (IQR): the difference between the 75th and the 25th percentile

Standard Deviation

- Average distance of each data point to the mean
- $SD = (\sqrt{\frac{1}{n} \sum_{i=1}^{N} (x_i - \bar{x})^2})$
$SD = (\sqrt{\frac{1}{n} \sum_{i=1}^{N} (x_i - \bar{x})^2})$
- where $\bar{x}$$\bar{x}$ indicates the sample mean, that is, $\bar{x} = \frac{1}{n} \sum_{i=1}^{N} x_i$ $\bar{x} = \frac{1}{n} \sum_{i=1}^{N} x_i$, and $n$$n$ is the sample size
- Almost all data entries are located within 2 or 3 standard deviations

R Examples

```
median(leaders$age)
```

```
## [1] 52.5
```

```
IQR(leaders$age)
```

```
## [1] 16.75
```

```
quantile(leaders$age)
```

```
##     0%   25%   50%   75%  100%
## 18.00 45.00 52.50 61.75 81.00
```

```
quantile(leaders$age, probs = seq(0, 1, by = 0.1)) #
```

```
##    0%  10%  20%  30%  40%  50%  60%  70%  80%  90% 100%
## 18.0 39.0 43.0 47.0 50.0 52.5 57.0 60.0 64.0 70.0 81.0
```

```
quantile(leaders$age, probs = c(.34, .55, .93)) # 34
```

```
## 34% 55% 93%
##  48  55  71
```

R Examples

```r
mean(leaders$age)
```

```
## [1] 53.524
```

```r
sd(leaders$age)
```

```
## [1] 12.03982
```

```r
summary(leaders$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   18.00   45.00   52.50   53.52   61.75   81.00
```

Observational Studies

- Often, it is unethical or infeasible to randomize the treatment.
  - Smoking and cancer
  - Gender and salary
  - Death penalty and crime
- Observational data: the treatment is naturally assigned

Observational Studies

- Why is it harder to make causal inference with observational data?
- Confounders/confounding variables: variables that are associated both with the treatment and the outcome
- If the treatment is not independent from pre-treatment variables, we cannot be sure about what causes what

Examples of

- Are incumbents more likely to win elections?
  - Maybe, but they also receive more campaign donations
- Are democratic countries more peaceful than authoritarian ones?
  - Maybe, but they also tend to be richer
- Selection bias: groups are not directly comparable

Observation Studies

- Pre-treatment variables have to be controlled for, that is, *held constant* in the analysis

- Statistical control: we use statistical methods to create balance between treatment and control and emulate an experiment as best as we can. E.g:

    - Instrumental variables (IV)
    - Regression discontinuity designs (RDD)
    - Differences-in-differences (DD/DiD/Diff-in-diff)

Differences-in-

- Compare trends before and after the treatment across the same units
- Controls for the initial conditions
- Requires data for two time periods (before and after treatment) and at least two units
- Parallel time trends assumption

Differences-in-

The **difference-in-differences** (DiD) design uses the following estimate of the sample average treatment effect for the treated (SATT):

$$\text{DiD estimate} = \underbrace{\left(\overline{Y}^{\text{after}}_{\text{treated}} - \overline{Y}^{\text{before}}_{\text{treated}}\right)}_{\text{difference for the treatment group}} - \underbrace{\left(\overline{Y}^{\text{after}}_{\text{control}} - \overline{Y}^{\text{before}}_{\text{control}}\right)}_{\text{difference for the control group}} .$$

The assumption is that the counterfactual outcome for the treatment group has a time trend parallel to that of the control group.

Minimum Wage and

- How does the increase in minimum wage affect unemployment?
- Economists believe the effect is positive: higher wages lead to higher unemployment
- Difficult to randomize minimum wage legislation
- In 1992, NJ minimum wage increased from \$4.25 to \$5.05…
- … but neighbouring PA stays at \$4.25
- NJ and (eastern) PA are similar, and fast food chains are similar too: wages, prices, products, etc

Minimum Wage and

Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania

By DAVID CARD AND ALAN B. KRUEGER*

On April 1, 1992, New Jersey's minimum wage rose from $4.25 to $5.05 per hour. To evaluate the impact of the law we surveyed 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. Comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant) provide simple estimates of the effect of the higher minimum wage. We also compare employment changes at stores in New Jersey that were initially paying high wages (above $5) to the changes at lower-wage stores. We find no indication that the rise in the minimum wage reduced employment. (*JEL* J30, J23)

Minimum Wage and

```r
minwage <- read.csv("https://raw.githubusercontent.c
str(minwage)
```

```
## 'data.frame':    358 obs. of  8 variables:
##  $ chain     : Factor w/ 4 levels "burgerking","kfc",..
##  $ location  : Factor w/ 5 levels "centralNJ","northNJ
##  $ wageBefore: num  5 5.5 5 5 5.25 5 5 5 5 5.5 ...
##  $ wageAfter : num  5.25 4.75 4.75 5 5 5 4.75 5 4.5 4.
##  $ fullBefore: num  20 6 50 10 2 2 2.5 40 8 10.5 ...
##  $ fullAfter : num  0 28 15 26 3 2 1 9 7 18 ...
##  $ partBefore: num  20 26 35 17 8 10 20 30 27 30 ...
##  $ partAfter : num  36 3 18 9 12 9 25 32 39 10 ...
```

Minimum Wage and

```r
# Subset the data into NJ and PA
minwageNJ ← subset(minwage, subset = (location ≠ "
minwagePA ← subset(minwage, subset = (location == "
```

```r
# Compute the proportion of restaurants whose wage i
mean(minwageNJ$wageBefore < 5.05) # NJ before
```

```
## [1] 0.9106529
```

```r
mean(minwageNJ$wageAfter < 5.05) # NJ after
```

```
## [1] 0.003436426
```

```r
mean(minwagePA$wageBefore < 5.05) # PA before
```

```
## [1] 0.9402985
```

```r
mean(minwagePA$wageAfter < 5.05) # PA after
```

```
## [1] 0.9552239
```

Minimum Wage and

```r
# Are the NJ and PA Restaurants Comparable?
mean(minwageNJ$wageBefore)
```

```
## [1] 4.609966
```

```r
mean(minwagePA$wageBefore)
```

```
## [1] 4.651343
```

```r
# Compute the proportion of full-time employees afte
minwageNJ$fullPropBefore <- minwageNJ$fullBefore / (
minwagePA$fullPropBefore <- minwagePA$fullBefore / (
mean(minwageNJ$fullPropBefore) # Proportion full-tim
```

```
## [1] 0.2965262
```

```r
mean(minwagePA$fullPropBefore) # Proportion full-tim
```

```
## [1] 0.3099657
```

Minimum Wage and

```
# Compute the proportion of full-time employees afte
minwageNJ$fullPropAfter ← minwageNJ$fullAfter / (mi
minwagePA$fullPropAfter ← minwagePA$fullAfter / (mi
mean(minwageNJ$fullPropAfter) # Proportion full-time
```

```
## [1] 0.320401
```

```
mean(minwagePA$fullPropAfter) # Proportion full-time
```

```
## [1] 0.2722821
```

Minimum Wage and

```
# Compare NJ before and after the change
NJdiff ← mean(minwageNJ$fullPropAfter) - mean(minwa
NJdiff
```

```
## [1] 0.02387474
```

However, the region as a whole could have changed too. So we use PA as a control case

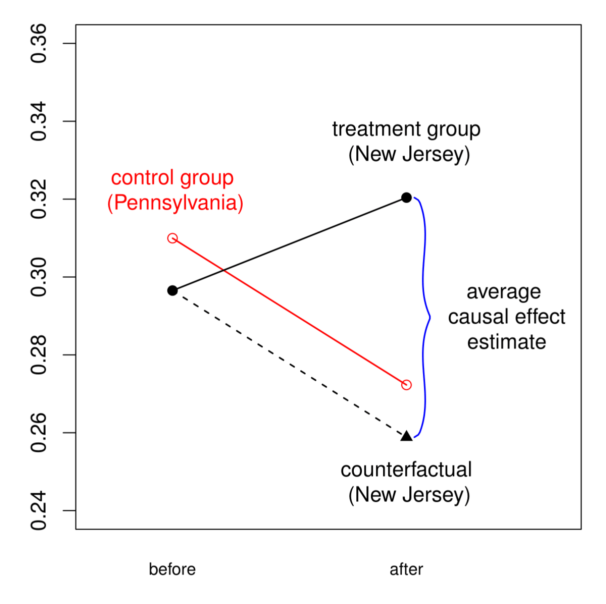
```
# PA before/after difference to control for possible
PAdiff ← mean(minwagePA$fullPropAfter) - mean(minwa
PAdiff
```

```
## [1] -0.03768357
```

```
# Difference in difference
NJdiff - PAdiff
```

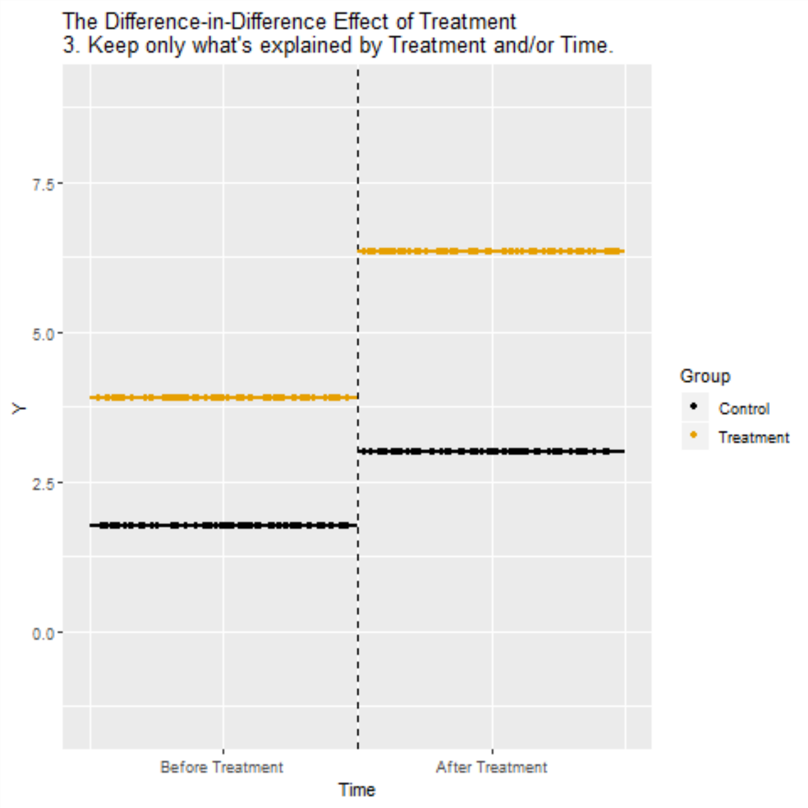
```
## [1] 0.06155831
```

Minimum Wage and



- Parallel trends assumption: NJ would have followed the same trend as PA had it not received the treatment. However, *we can't test that.*

Animation of DiD



Questions?

Leader Assassination DD

- **Question 5**
- Does successful leader assassination cause democratization?
- Does successful leader assassination lead countries to war?
- Answer these questions by analyzing the data. Be sure to state your assumptions and provide a brief interpretation of the results.

Leader Assassination DD

```r
str(leaders)
```

```
## 'data.frame':    250 obs. of  11 variables:
##  $ year          : int  1929 1933 1934 1924 1931 1968
##  $ country       : Factor w/ 88 levels "Afghanistan",.
##  $ leadername    : Factor w/ 196 levels "Abdullah Al-H
##  $ age           : int  39 53 50 29 36 41 73 48 76 77
##  $ politybefore  : num  -6 -6 -6 0 -9 -9 -2 1 2 2  ...
##  $ polityafter   : num  -6 -7.33 -8 -9 -9  ...
##  $ interwarbefore: int  0 0 0 0 0 0 0 0 0 0  ...
##  $ interwarafter : int  0 0 0 0 0 0 0 0 0 0  ...
##  $ civilwarbefore: int  1 0 0 0 0 0 0 0 0 0  ...
##  $ civilwarafter : int  0 0 0 0 0 0 1 0 0 0  ...
##  $ result        : Factor w/ 10 levels "dies between a
```

```r
# create success variable
leaders$success <- ifelse(leaders$result == "dies be
                            leaders$result == "dies
                            leaders$result == "dies
                            leaders$result == "dies,
```

Leader Assassination DD

```
# polity score before and after assassination attemp
diff.pol.succ <- mean(leaders$polityafter[leaders$su
                        mean(leaders$politybefore[leaders

diff.pol.succ
```

```
## [1] -0.05864198
```

```
diff.pol.unsucc <- mean(leaders$polityafter[leaders$
                          mean(leaders$politybefore[leade

diff.pol.unsucc
```

```
## [1] -0.1513605
```

```
## difference in differences
diff.pol.succ - diff.pol.unsucc
```

```
## [1] 0.09271857
```

Leader Assassination DD

```
# create variable for warbefore and warafter
leaders$warbefore <- ifelse(leaders$interwarbefore =
                              leaders$civilwarbefore
leaders$warafter <- ifelse(leaders$interwarafter ==
                             leaders$civilwarafter =

## compare war before to war after among successful
diff.war.succ <- mean(leaders$warafter[leaders$succe
                   mean(leaders$warbefore[leaders$su

diff.war.unsucc <- mean(leaders$warafter[leaders$suc
                     mean(leaders$warbefore[leaders$

diff.war.succ - diff.war.unsucc # difference in diff
```

```
## [1] -0.07161754
```

Using the difference-in-difference approach, we find very little difference in the contries' polity score and in the proportion of countries engaged in war. Leader assassination does not seem to cause countries to democratise or engage

Internal and External

- Because of randomization, we know that RCTs have strong *internal validity*

- Internal validity: the degree to which we can attribute the results to the treatment and not to other factors

- However, observational studies have greater *external validity*

- External validity: the extent to which the results can be

Replication Crisis

Website: https://theconversation.com/you-cant-characterize-human-nature-if-studies-overlook-85-percent-of-people-on-earth-106670

Replication Crisis

SCIENCE

Psychology's Replication Crisis Is Running Out of Excuses

Another big project has found that only half of studies can be repeated. And this time, the usual explanations fall flat.

ED YONG NOV 19, 2018

The Thinker, by Auguste Rodin (JASON LEE / REUTERS)

MORE STORIES

Science Is Getting Less Bang for Its Buck
PATRICK COLLISON AND MICHAEL NIELSEN

Why Rich Kids Are So Good at the Marshmallow Test
JESSICA MCCRORY CALARCO

Psychology's Replication Crisis Can't Be Wished Away
ED YONG

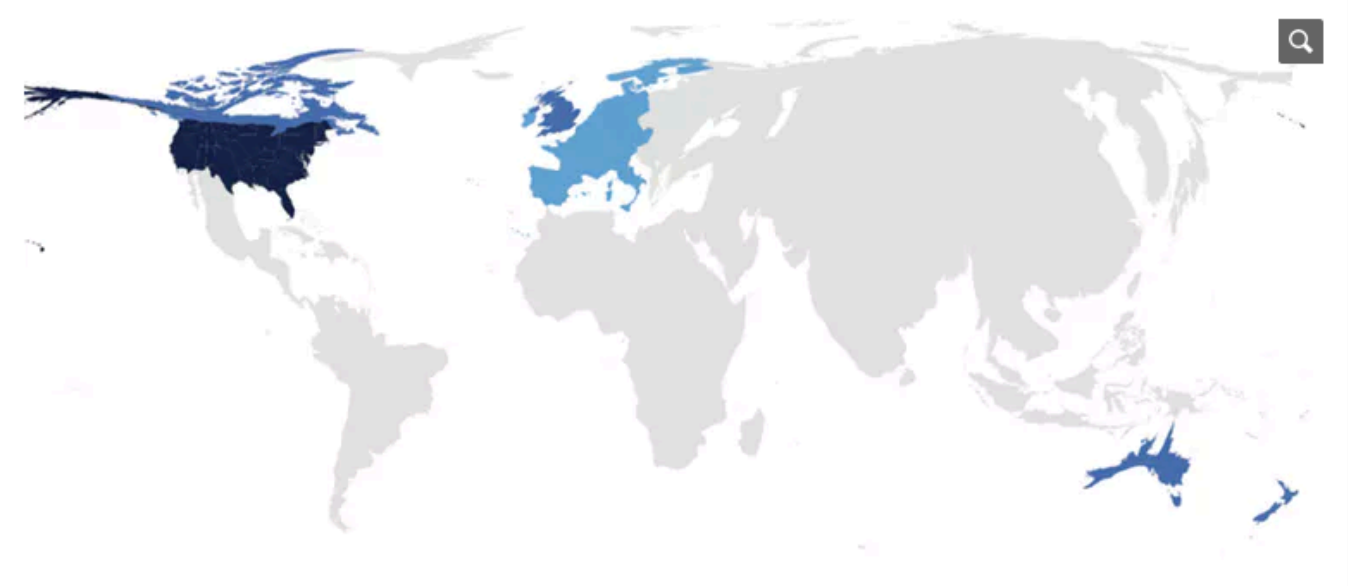
In Science, There Should Be a Prize for Second Place
ED YONG

Website:

https://www.theatlantic.com/science/archive/2018/11/psycho

replication-crisis-real/576223

Replication Crisis



Blue countries represent the locations of 93 percent of studies published in Psychological Science in 2017. Dark blue is U.S., blue is Anglophone colonies with a European descent majority, light blue is western Europe. Regions sized by population.

Website: https://theconversation.com/you-cant-characterize-human-nature-if-studies-overlook-85-percent-of-people-on-earth-106670

What Should We Do?

- Replications: see whether the same results hold under different conditions
- Field experiments: conduct experiments in realistic settings
- Larger sample sizes: large samples tend to be more representative of the underlying population
- Open methods and open data: share your code and datasets so other can verify them
- Pre-registration: state your

We are Getting Better

All that said, there are some promising signs that social science is getting better. More and more scientists are **preregistering their study designs**. This prevents them from cherry-picking results and analyses that are more favorable to their favored conclusions. Journals are getting better at demanding larger subject pools in experiments and are increasingly **insisting that scientists share all the underlying data** of their experiments for others to assess.

"The lesson out of this project," Nosek says, "is a very positive message of reformation. Science is going to get better."

Website: https://www.vox.com/science-and-health/2018/8/27/17761466/psychology-replication-crisis-nature-social-science

Wrap-up

- Not all experiments are true, many don't replicate
- If possible, pre-register your hypotheses and make your data and code available (RMarkdown can help!)
- Replicate your and other people's work
- Is science wrong? *No*, but there are many wrong findings maskerading as science
- Keep those things in mind while

To do during the

- Watch this video, it is very good:
  - https://youtu.be/dSlCBJSh96w
- John Oliver on the replication crisis: https://youtu.be/0Rnq1NpHdmw
- BBC podcast on the same problem: https://pca.st/n5b3

Survey

See you next week!