POLI 30 D: Political Inquiry

Professor Umberto Mignozzetti
(Based on DSS Materials)

Lecture 05 | Does Social Pressure Increase Probability of Turning Out to Vote?

Before we start

Announcements:

- Quizzes and Participation: On Canvas.
- Github page: https://github.com/umbertomig/POLI30Dpublic
- Piazza forum: https://piazza.com/ucsd/winter2023/17221

Before we start

Recap:

- We learned the definitions of Theory, Scientific Theory, and Hypotheses.
- Data, datasets, variables, and how to compute means.
- Causal effect, treatments, outcomes, and randomization.

Great job!

- Do you have any questions about these contents?

Plan for Today

- Percentage Points
- Review: Unit of Measurement of Means
- Unit of Measurement of Diffs-in-Means

- In-Class Exercise: Does Social Pressure Increase the Probability of Turning Out To Vote?

What is a percentage point?

Unit of measurement for the arithmetic difference between two percentages:

% - % = p.p.

- Example: if a candidate's vote share increases from 50% to 60%, we would state that the vote share increased by

  $$\triangle \text{vshare} = \text{vshare}_{\text{final}} - \text{vshare}_{\text{initial}} = 60\% - 50\% = 10 \text{ p.p.}$$

- Why not 10%?
  - What is 10% of 50%? $0.10 \times 50 = 5$ p.p.

    $$\text{vshare}_{\text{final}} = \text{vshare}_{\text{initial}} + \triangle \text{vshare} = 50\% + 5 \text{ p.p.} = 55\%$$

REVIEW: Unit of Measurement of Means

interpretation of the mean of a variable

if variable is non-binary:	if variable is binary:
as an average, in the same unit of measurement as the variable	as a proportion, in % after multipying the result by 100

Unit of Measurement of the Diffs-in-Means Estimator

- Formula of the difference-in-means estimator?
  - *Average outcome* for the treatment group - *Average outcome* for the control group
- If the outcome variable is binary, in what unit of measurement will the average outcomes be?
  - percentages (after multiplying the decimal by 100)
- What will be the unit of measurement of estimator?
  - percentage points (% - % = p.p.)
- Do we need to multiply the result by 100?
  - yes!

Unit of Measurement of the Diffs-in-Means Estimator

unit of measurement of the difference-in-means estimator

if outcome variable is non-binary:	if outcome variable is binary:
in the same unit of measurement as the outcome variable	in percentage points (after multipying the result by 100)

Does Social Pressure Affect Turnout?

(Based on Alan S. Gerber, Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review*, 102 (1): 33-48.)

Does Social Pressure Affect Turnout?

- We will answer, by analyzing data from an experiment where registered voters were randomly assigned to either
  - (a) receive a message designed to induce social pressure to vote, or
  - (b) receive nothing
- The message told registered voters that after the election their neighbors would be informed about whether they voted in the election or not
  - What to you think the effect might be?

Does Social Pressure Affect Turnout?

- What do we need to calculate to estimate the average causal effect of receiving the message on the probability of turning out to vote?
  - the difference-in-means estimator
- Why does the difference-in-means estimator provide us with a *valid* estimate of the average treatment effect in this case?
  - because the data come from a randomized experiment (where treatment was randomly assigned)
  - as a result, treatment and control groups are comparable

Does Social Pressure Affect Turnout?

- In this case, the difference-in-means estimator is:
- Answer:

$$\overline{voted}_{\text{treatment group}} - \overline{voted}_{\text{control group}}$$

- $\overline{voted}_{\text{treatment group}}$: proportion of registered voters who voted among those who received the message
- $\overline{voted}_{\text{control group}}$: proportion of registered voters who voted among those who did not receive the message

In-Class Exercise

1. Open RStudio (RStudio will open R)
2. Open **inclass01.R** from within RStudio
   - RStudio: File >> Open File
3. Run code from steps 1-3
   - use the `setwd()` for your computer

```r
## STEP 1. Load the dataset
voting <- read.csv("https://raw.githubusercontent.com/umber
```

```r
## STEP 2. Look at the data
head(voting, 3) # shows the first three observations
```

```
##   birth message voted
## 1  1981      no     0
## 2  1959      no     1
## 3  1956      no     1
```

```r
## what does each observation represent?
## what is the outcome variable?
## what is the treatment variable?
```

STEP 4. Create binary treatment variable

- First, we need to learn how to use `==`and `ifelse()`
- The operator `==`
  - is used to create logical tests that evaluate whether the observations of a variable equal a particular value
    - values should be in quotes if text but without quotes if numbers
  - examples:
    - `data$variable==1`
    - `data$variable=="yes"`

- The function `ifelse()`
  - creates the contents of a new variable based on the values of an existing one
  - requires three arguments, separated by commas, in the following order:
    (1) logical test (using `==`),
    (2) return value if logical test is true,
    (3) return value if logical test is false
  - example: `ifelse(data$variable=="yes", 1, 0)`

```r
## STEP 4. Create binary treatment variable
## create variable pressure inside dataframe voting
voting$pressure <- # stores return values in new variable
  ifelse(voting$message=="yes", # logical test
         1, # return value if logical test is true
         0) # return value if logical test is false
```

- You need to run the code all at once (not line by line)
- Remember that R will ignore anything that follows the # sign, until the end of the line
- What would have happened had we not added `voting$`in front of `pressure`on the first line of code above?

- Whenever we create a new variable, we should make sure it was created correctly by looking at the first few observations of the dataframe again
  - What's the code to ask R to show the first observations of a dataframe?

```
head(voting, 3) # shows first observations
```

```
##   birth message voted pressure
## 1  1981      no     0        0
## 2  1959      no     1        0
## 3  1956      no     1        0
```

- Note that when *message* equals "yes", *pressure* equals 1; and when *message* equals "no", *pressure* equals 0

STEP 5. Compute the difference-in-means estimator

$$\overline{Y}_{\text{treatment group}} - \overline{Y}_{\text{control group}}$$

$\overline{Y}_{\text{treatment group}}$: average outcome for the treatment group
$\overline{Y}_{\text{control group}}$: average outcome for the control group

- In the voting experiment:

$$\overline{voted}_{\text{treatment group}} - \overline{voted}_{\text{control group}}$$

- $\overline{voted}_{\text{treatment group}}$: mean of *voted* for treatment group
- $\overline{voted}_{\text{control group}}$: mean of *voted* for control group

- Let's start by practicing computing and interpreting means

```
mean(voting$voted) # calculates the mean of voted
```

```
## [1] 0.3101759
```

- Interpretation?
  - 31% of *all* the registered voters who were part of the experiment voted
- Why in %?
  - Because *voted* is binary
  - Recall: The mean of a binary variable should be interpreted in % (after multiplying the output by 100)

- `mean(voting$voted)` computes the mean of *voted* for all the observations in the dataset (that is, for all the registered voters in the experiment)
- To compute the difference-in-means estimator, we need to calculate the mean of *voted* for subsets of observations
- Specifically, we need to compute:
  - the mean of *voted* for the treatment group (for which *pressure* equals 1)
  - the mean of *voted* for the control group (for which *pressure* equals 0)
- To do this, we need to learn how to use the `[]` operator

- Operator `[]`:
  - extracts a selection of observations from a variable
  - to its left, we specify the variable we want to subset
  - inside the square brackets, we specify the criteria of selection; for example, we can specify a logical test using the relational operator `==`; only the observations for which the test is true will be extracted
  - example: `data$var1[data$var2==1]`
    # extracts the observations of the variable *var1* for which the variable *var2* equals 1

- Compute the mean of *voted* for treatment and control groups

```
mean(voting$voted[voting$pressure == 1]) # treatment
```

```
## [1] 0.3779482
```

```
mean(voting$voted[voting$pressure == 0]) # control
```

```
## [1] 0.2966383
```

- Interpretation of the first mean?
  - 38% of the registered voters who received the message voted (38x100=38%)
- Interpretation of the second mean?
  - 30% of the registered voters who did *not* receive the message voted (30x100=30%)

- Now, we can compute the difference-in-means estimator:

```
mean(voting$voted[voting$pressure==1]) -
  mean(voting$voted[voting$pressure==0])
## [1] 0.08130991
```

- direction, size, and unit of measurement of the effect?
  - increase of 8 percentage points
- *increase* because we are measuring a change in Y and the number is positive
- *percentage points* because it is the result of subtracting two percentages
- *8* (and not 0.08) because we need to multiply the number by 100 to turn it into p.p. (because *voted* is binary)
- 38% - 30% = 8 p.p.

STEP 6. Write conclusion statement

- What's the assumption we are making when estimating the average causal effect?
  - registered voters who received the message are comparable to registered voters who did not
- Why is this a reasonable assumption?
  - data comes from a randomized experiment
- What's the treatment?
  - receiving the message inducing social pressure
- What's the outcome?
  - probability of voting
- What's the direction, size, and unit of measurement of the average causal effect?
  - an increase of 8 percentage points, on average

CONCLUSION STATEMENT

Assuming that [the treatment and control groups are comparable] (a reasonable assumption because ...),
we estimate that [the treatment] [increases/decreases] [the outcome] by [size and unit of measurement of the effect], on average.

- Assuming that registered voters who received the message are comparable to the registered voters who did not (a reasonable assumption because the data come from a randomized experiment), we estimate that receiving the message inducing social pressure increases the probability of voting by 8 percentage points, on average.

Today's Class

- Units of Measurement of Means and Diffs-in-Means
- In-Class Exercise: Does Social Pressure Increase the Probability of Turning Out To Vote?
- How to Write a Conclusion Statement
- R: `==`, `[]`, `ifelse()`

Next Class

- Survey Research
- Exploring One Variable at a Time
- **Bring your computers!**

Summary

- **Today's Class:**
  - Causal Effects
  - Treatment and Outcome Variables
  - Individual vs. Average Causal Effects
  - Randomized Experiments
  - Difference-in-Means Estimator
- Next class:
  - Hands on! We are going to analyze the voting experiment dataset!

Questions?

See you in the next class!