Lab02: Causality Review

September 5, 2018

Today's Goals

1. Garner more intuition
    - We want *causal effect*
    - Can't observe it for any given unit
    - *Randomization* helps us get at effect *on average*

Today's Goals

2. Comfortably "speak the language" of randomized experiments
   - unit
   - treatment group/condition
   - control group/condition
   - outcome
   - potential outcomes
   - counterfactuals
   - causal effect
   - fundamental problem of causal inference
   - randomization
   - SATE
   - difference in means

Today's Goals

3. A little R

Running Example

- August 2006 Primary Statewide Election in Michigan
- Send postcards with randomly assigned message
  1. no message
  2. civic duty message
  3. "you are being studied" message
  4. **neighborhood social pressure message**
- Source: Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. "Social pressure and voter turnout: Evidence from a large-scale field experiment." American political Science review 102.1 (2008): 33-48.

Discuss

- What is a **unit** in this study?
- What is the **outcome**?
- What is the **control group**?
- What are the **treatment groups**?

Review

- What is a **unit** in this study?
  - a person
- What is the **outcome**?
  - whether or not the person voted, Y
- What is the **control group**?
  - subset of people who recieved no message
- What are the **treatment groups**?
  - subset of people who recieved
    - T_1: civic duty message
    - T_2: "you are being studied" message
    - T_3: social pressure message

$\rightarrow$ Let's focus on social pressure message as T of interest

Discuss

- What are **potential outcomes**?
- What are the potential outcomes as applied to this study?
- How do you find the true **casual effect**? Is this possible?
- What is the **fundamental problem of causal inference**?
- Explain the fundamental problem of causal inference as applied to this study.

Review

- What are **potential outcomes**?
  - $Y(1)$ outcome if you recieved treatment
  - $Y(0)$ outcome if you recieved control
- What are the potential outcomes as applied to this study?
- How do you find the true **casual effect**? Is this possible?
  - $Y(1) - Y(0)$
- What is the **fundamental problem of causal inference**?
  - can only observe one potential outcome
  - counterfactuals are not observed!
- Explain the fundamental problem of causal inference as applied to this study.

Discuss

- Explain this equation in words:
  $SATE = \frac{1}{n} \sum_{i=1}^{n} \{ Y_i(1) - Y_i(0) \}$ where i indexes each unit
- Can we get an answer to this equation? Why or why not?

Review

- Explain this equation in words, where i indexes each unit: $SATE = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$
- Can we get an answer to this equation? Why or why not?
  - no because counterfactuals aren't observed
  - i.e., the fundamental problem of causal inference!!
  - SATE is a theoretical concept so we must *estimate* it

Discuss

- How do we do research given the fundamental problem of causal inference?
- What did this study do?
- How do we *estimate* SATE?

Review

- How do we do research given the fundamental problem of causal inference?
  - Randomization
  - On average, treatment and control groups identical *but for which group they're assigned*
  - Compare *observed* outcomes across groups
- What did this study do?
  - Randomized, so groups were identical but for the message
  - Compare whether or not people voted across message groups
  - On average, was there a difference?
- How do we *estimate* SATE?
  - difference in means across groups

Discuss

- Remember we must *estimate* SATE.
- We do it using the difference in means.
- Explain how this equation works:
  - diff in means $= \frac{1}{n_1} \sum_{i=1}^{n} T_i Y_i - \frac{1}{n-n_1} \sum_{i=1}^{n} (1 - T_i) Y_i$
    - n_1 is the number of people in the treatment group
    - T_i is an indicator s.t. $T_i = 0$ if unit recieved control and $T_i = 1$ if treatment

Are you confortable “speaking the language” now?

- Hopefully this is getting repetative.
- If any questions remain. . . ask!

A little R

- What is each row?
- Which variable indicates treament?
- Which indicates outcome?
- What are the other variables for?

```
social <- read.csv("social.csv")
head(social)
```

```
##      sex yearofbirth primary2004   messages primary2006 hhsize
## 1   male        1941           0 Civic Duty           0      2
## 2 female        1947           0 Civic Duty           0      2
## 3   male        1951           0  Hawthorne           1      3
## 4 female        1950           0  Hawthorne           1      3
## 5 female        1982           0  Hawthorne           1      3
## 6   male        1981           0    Control           0      3
```

Estimate turnout rate (outcome) for each group

```
tapply(X = social$primary2006, INDEX = social$messages, FUN = mean)
```

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

```
t1 <- mean(social$primary2006[social$messages == "Civic Duty"])
control <- mean(social$primary2006[social$messages == "Control"])
t2 <- mean(social$primary2006[social$messages == "Hawthorne"])
t3 <- mean(social$primary2006[social$messages == "Neighbors"])

c(t1, control, t2, t3)
```

```
## [1] 0.3145377 0.2966383 0.3223746 0.3779482
```

Estimate the SATE for each group

What do these results tell us?

```
t1 - control
```

```
## [1] 0.01789934
```

```
t2 - control
```

```
## [1] 0.02573631
```

```
t3 - control
```

```
## [1] 0.08130991
```

Randomization makes groups "identical" but for treatment assignment

```
tapply(social$primary2004, social$messages, mean)
```

```
## Civic Duty    Control  Hawthorne  Neighbors
##  0.3994453  0.4003388  0.4032300  0.4066647
```

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

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

```
social$sex <- ifelse(social$sex == "female", 1, 0)
tapply(social$sex, social$messages, mean)
```

```
## Civic Duty    Control  Hawthorne  Neighbors
##  0.5001832  0.4989411  0.4990053  0.5000654
```