Inference for Average Treatment Effects

Kosuke Imai

Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2018

Social Pressure and Turnout (Gerber, et al. 2008. *Am. Political Sci. Rev.*)

- August 2006 Primary Election in Michigan
- Statewide elections: Governor, US Senator
- 180,000 households
- Send postcards with different messages

- Randomly assign each household to a group (or treatment)
  1. no message (control group)
  2. civic duty message
  3. "you are being studied" message (Hawthorne effect)
  4. household social pressure message
  5. neighborhood social pressure message

Neighborhood Social Pressure Message

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted i the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	______
9995 JENNIFER KAY SMITH		Voted	______
9997 RICHARD B JACKSON		Voted	______
9999 KATHY MARIE JACKSON		Voted	______

“You are being studied” Message

Dear Registered Voter:

YOU ARE BEING STUDIED!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

This year, we're trying to figure out why people do or do not vote. We'll be studying voter turnout in the August 8 primary election.

Our analysis will be based on public records, so you will not be contacted again or disturbed in any way. Anything we learn about your voting or not voting will remain confidential and will not be disclosed to anyone else.

DO YOUR CIVIC DUTY — VOTE!

Standard Empirical Analysis

Groups	Control	Civic duty	Hawthorne	Self	Neighbor
Turnout rate	29.7%	31.5%	32.2%	34.5%	37.5%
# of voters	191,243	38,218	38,204	38,218	38,201

- Neighborhood social pressure vs. Control

$$
\begin{aligned}
\hat{\tau} &= 37.5 - 29.7 = 7.8 \\
s.e. &= \sqrt{\frac{37.5 \times (100 - 37.5)}{38201} + \frac{29.7 \times (100 - 29.7)}{191243}} \approx 0.3 \\
95\%\mathrm{CI} &= [7.8 - 1.96 \times 0.3,\ 7.8 + 1.96 \times 0.3] = [7.2,\ 8.4]
\end{aligned}
$$

- This calculation ignores the fact that some households have multiple voters: we will discuss this issue later in the course
- How can we justify this standard difference-in-means analysis from the randomization perspective?

Theoretical Motivation

- Permutation inference has a difficult time handling heterogeneity
- Population inference is also difficult

- Can we address these limitations while keeping the design-based approach? $\rightsquigarrow$ inference about average treatment effects
- We will begin with randomization inference for sample survey: close connection to Neyman's randomization inference
- Causal inference as a missing data problem

Estimation of the Sample Average Treatment Effect

- Due to Neyman (1923) Neyman. 1990 (translated to English) *Stat. Sci.*
- Difference-in-means estimator:

$$\hat{\tau} \equiv \frac{1}{n_1}\sum_{i=1}^{n} T_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - T_i) Y_i$$

- Define $\mathcal{O} \equiv \{Y_i(0), Y_i(1)\}_{i=1}^{n}$
- Unbiasedness (over repeated treatment assignments):

$$\begin{aligned}
\mathbb{E}(\hat{\tau} \mid \mathcal{O}) &= \frac{1}{n_1}\sum_{i=1}^{n}\mathbb{E}(T_i \mid \mathcal{O}) Y_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\{1 - \mathbb{E}(T_i \mid \mathcal{O})\} Y_i(0) \\
&= \frac{1}{n}\sum_{i=1}^{n}(Y_i(1) - Y_i(0)) = \text{SATE}
\end{aligned}$$

Randomization Inference for SATE

- Variance of $\hat{\tau}$:

$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) \quad = \quad \frac{1}{n}\left(\frac{n_0}{n_1}S_1^2 + \frac{n_1}{n_0}S_0^2 + 2S_{01}\right),$$

where for $t = 0, 1$,

$$S_t^2 \quad = \quad \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(t) - \overline{Y(t)})^2 \quad \text{sample variance of } Y_i(t)$$

$$S_{01} \quad = \quad \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(0) - \overline{Y(0)})(Y_i(1) - \overline{Y(1)}) \quad \text{sample covariance}$$

- The variance is NOT *identifiable*

Details of Variance Derivation

1. Let $X_i = Y_i(1) + n_1 Y_i(0)/n_0$ and $D_i = nT_i/n_1 - 1$, and write

$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) = \frac{1}{n^2} \mathbb{E}\left\{ \left(\sum_{i=1}^{n} D_i X_i \right)^2 \middle| \mathcal{O} \right\}$$

2. Show

$$\mathbb{E}(D_i \mid \mathcal{O}) = 0, \quad \mathbb{E}(D_i^2 \mid \mathcal{O}) = \frac{n_0}{n_1},$$
$$\mathbb{E}(D_i D_j \mid \mathcal{O}) = -\frac{n_0}{n_1(n-1)}$$

3. Use ❶ and ❷ to show,

$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) = \frac{n_0}{n(n-1)n_1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

4. Substitute the potential outcome expressions for X_i

Conservative Variance Estimator

- The usual variance estimator is conservative on average:

$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) \;\; \leq \;\; \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} \; = \; \mathbb{E}\left(\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0} \mid \mathcal{O}\right)$$

  where

$$\hat{\sigma}_t \; = \; \frac{1}{n_t - 1} \sum_{i=1}^{n} \mathbf{1}\{T_i = t\}(Y_i - \overline{Y}_t)^2 \quad \text{for } t = 0, 1$$

- Under the constant additive unit causal effect assumption, i.e., $Y_i(1) - Y_i(0) = c$ for all i,

$$S_{01} \; = \; \frac{1}{2}(S_1^2 + S_0^2) \quad \text{and} \quad \mathbb{V}(\hat{\tau} \mid \mathcal{O}) \; = \; \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0}$$

- The optimal treatment assignment rule:

$$n_1^{opt} = \frac{n}{1 + S_0/S_1}, \quad n_0^{opt} = \frac{n}{1 + S_1/S_0}$$

Bounds on the Variance

- Use of the Cauchy-Schwartz inequality:
  1. Upper bound: sample correlation between $Y_i(1)$ and $Y_i(0)$ is 1
  2. Lower bound: sample correlation between $Y_i(1)$ and $Y_i(0)$ is -1

$$\frac{n_0 n_1}{n}\left(\frac{S_1}{n_1} - \frac{S_0}{n_0}\right)^2 \leq \mathbb{V}(\hat{\tau} \mid \mathcal{O}) \leq \frac{n_0 n_1}{n}\left(\frac{S_1}{n_1} + \frac{S_0}{n_0}\right)^2$$

- Constant additive unit causal effect $\rightsquigarrow$ sample correlation is 1

$$\frac{n_0 n_1}{n}\left(\frac{S_1}{n_1} + \frac{S_0}{n_0}\right)^2 = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0}$$

- Sharp bounds based on the entire marginal distributions $\rightsquigarrow$ application of Hoeffding's lemma (Aronow et al. 2015. *Ann. Stat.*)

Inference for Population Average Treatment Effect

- Assumption: simple random sampling from an infinite population
- Unbiasedness (over repeated sampling):

$$\mathbb{E}\{\mathbb{E}(\hat{\tau} \mid \mathcal{O})\} \;=\; \mathbb{E}(\text{SATE}) \;=\; \text{PATE}$$

- Variance:

$$\begin{aligned}\mathbb{V}(\hat{\tau}) &= \mathbb{V}(\mathbb{E}(\hat{\tau} \mid \mathcal{O})) + \mathbb{E}(\mathbb{V}(\hat{\tau} \mid \mathcal{O})) \\ &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}\end{aligned}$$

  where σ_t^2 is the population variance of $Y_i(t)$ for $t = 0, 1$
- Unbiased variance estimator:

$$\widehat{\mathbb{V}(\hat{\tau})} \;=\; \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0} \quad \text{where} \quad \mathbb{E}(\widehat{\mathbb{V}(\hat{\tau})}) = \mathbb{V}(\hat{\tau})$$

  for $t = 0, 1$

Asymptotic Inference for PATE

- Hold $k = n_1/n$ constant
- Rewrite the difference-in-means estimator as

$$\hat{\tau} \;=\; \frac{1}{n}\sum_{i=1}^{n}\underbrace{\left(\frac{T_i Y_i(1)}{k} - \frac{(1-T_i)Y_i(0)}{1-k}\right)}_{\text{i.i.d. with mean PATE \& variance } n\mathbb{V}(\hat{\tau})}$$

- Consistency via Law of large numbers:

$$\hat{\tau} \;\xrightarrow{p}\; \text{PATE}$$

- Asymptotic normality via the Central Limit Theorem:

$$\sqrt{n}(\hat{\tau} - \text{PATE}) \;\xrightarrow{d}\; \mathcal{N}\left(0, \frac{\sigma_1^2}{k} + \frac{\sigma_0^2}{1-k}\right)$$

- $(1-\alpha) \times 100\%$ Confidence intervals:

$$[\hat{\tau} - \text{s.e.} \times z_{\alpha/2},\ \hat{\tau} + \text{s.e.} \times z_{\alpha/2}]$$

Heated Exchange at the Royal Statistiacal Society

Neyman et al. (1935) *Suppl. of J. Royal Stat. Soc*

Neyman: *So long as the average yields of any treatments are identical, the question as to whether these treatments affect separate yields on single plots seems to be uninteresting*

Fisher: *It may be foolish, but that is what the* z *test was designed for, and the only purpose for which it has been used.*

Neyman: *I am considering problems which are important from the point of view of agriculture.*

Fisher: *It may be that the question which Dr. Neyman thinks should be answered is more important than the one I have proposed and attempted to answer. I suggest that before criticizing previous work it is always wise to give enough study to the subject to understand its purpose.*

Summary: Fisher vs. Neyman

- Like Fisher, Neyman proposed randomization-based inference
- Unlike Fisher,
  1. estimands are *average* treatment effects
  2. heterogenous treatment effects are allowed
  3. population as well as sample inference is possible
  4. asymptotic approximation is required for inference

- Suggested reading: IMBENS AND RUBIN, CHAPTER 6