Section 1

Introduction and Potential Outcomes

Sooahn Shin

GOV 2003

Sept 9, 2021

Overview

• Logistics:

- Section: Thur 3:00 4:15 pm @ K262 @ K105
- TF Office Hours: Mon 1:30 2:30/Thur 4:30 5:30 pm @ TBD
- Pset 1 released! Due at 11:59 pm (ET) on Sept 15
- We encourage you to share your questions on Ed.
- By September 17: Find a collaborator for the project (check the open thread for finding partners on Ed).

Today's topics:

- 1. Identification and estimation
- 2. Example: Political canvassing

Identification and Estimation

• The fundamental problem of causal inference (Holland 1986)

We only observe one potential outcome per unit
 → How do we infer the missing potential outcomes (= counterfactual)?



Example: Political canvassing¹

- Study of *n* voters $\langle n_i \rangle$ canvassed • n_1 are canvassed • $n_0 = n - n_1$ are not canvassed • For each voter $i \in \{1, 2, ..., n\}$, observe: Treatment $D_2 = \int_0^{1} \frac{1}{16} \frac{1}{1$
- **Defined** \leftarrow Vote choice (observed outcome): $Y_i = 1$ if voter *i* cast ballot for candidate *A*, and 0 if the voter cast ballot for candidate *B*.

turn out • Turnout (observed selection): $S_i = 1$ if voter *i* turned out, and 0 otherwise.

- Canvassing (treatment): $D_i = 1$ if canvassed, and 0 otherwise.
- Causal question: does canvassing (D_i) affect vote choice (Y_i)?
- Selection on samples:
 - 1. canvassing may affect turnout (S_i) , and
 - 2. we only observe the vote choices of the voters who turned out \rightarrow post-treatment bias $E[\Upsilon_{i}[\mathfrak{d}_{i}=\langle \mathfrak{f}_{i}=\langle \mathfrak{f}_{i}=\rangle - E[\Upsilon_{i}[\mathfrak{d}_{i}=\circ,\mathfrak{f}_{i}=\rangle]$

¹Example adapted from 2021S STAT286/GOV2003 Review Question 1

Potential Outcomes and Principal Stratification

- 1. $D_i \rightarrow S_i$
 - S_i: Observed turnout





• $S_i(d)$ for $d \in \{0,1\}$: **Potential** turnout

Data:

- Recall the "consistency" assumption: $S_i = S_i(d)$ if $D_i = d$ (no hidden versions of treatment) counter example: Variation of amount / large
- If canvassed $[D_i = d]$, the potential turnout when the voter is canvassed $[S_i(d)]$ is the observed turnout $[S_i]$
- We have four principal strata defined by $(S_i(0), S_i(1))$
 - (1,1): turning out regardless of the canvassing
 - (0,1): turning out only when being canvassed
 - (1,0): turning out only when not being canvassed

•
$$(0,0)$$
: never turning out
 $first \rightarrow d = 0 < \begin{cases} S_{1}(0) = 0 \\ S_{1}(0) = 1 \end{cases}$ or 2
 $entry \rightarrow d = 1 < \begin{cases} S_{1}(0) = 0 \\ S_{1}(0) = 1 \end{cases}$ or 2
 $second \\ entry \rightarrow d = 1 < \begin{cases} S_{1}(1) = -0 \\ S_{1}(1) = 1 \end{cases}$ or 2
 $second \\ S_{1}(1) = 0 \\ S_{1}(1) = -0 \end{cases}$ or 2
 $second \\ entry \rightarrow d = 1 < \begin{cases} S_{1}(1) = -0 \\ S_{1}(1) = 1 \end{cases}$ or 2
 $second \\ S_{1}(1) = -0 \\ S_{1}(1) = -1 \end{cases}$ or 2
 $second \\ S_{1}(1) = -0 \\ S_{1}(1) = -1 \end{cases}$ or 2
 $second \\ S_{1}(1) = -0 \\ S_{1}(1) = -0 \\ S_{1}(1) = -1 \end{cases}$ or 2
 $second \\ S_{1}(1) = -0 \\ S_{1}(1) = -0$

Potential Outcomes and Principal Stratification

- 2. Vote choice does not exist if a voter i does not turn out
 - Y_i: **Observed** vote choice
 - $Y_i(d,s)$ for $d, s \in \{0,1\}$: **Potential** vote choice
 - $Y_i(1,0)$ and $Y_i(0,0)$ are not well defined
 - Y_i(1,0): Potential vote choice if the voter is canvassed and didn't turn out → does not exist
 - Y_i(0,0): Potential vote choice if the voter is not canvassed and didn't turn out → does not exist

" selection"

Estimands (Quantity of Interest) research guestion

- Suppose effect of interest is the effect among those who turn out regardless of the treatment.
- What is the individual causal effect of canvassing on voting for candidate A among always turnout? $(S_{i}(0), S_{i}(1))$ ind. causal effect $Y_{i}(1, 0)$ \rightarrow contradicts W/ $Y_{i}(1, 1) - Y_{i}(0, 1)$ $(S_{i}(0), 0)$ $(S_{i}(0), S_{i}(1))$ $(S_{$
 - What is the population average treatment effect of canvassing on voting for candidate A among always turnout?

$$\mathbb{E}[Y_i(1,1) - Y_i(0,1) \mid (S_i(0), S_i(1)) = (1,1)]$$

Estimands

- Vote share for candidate $A = \frac{\text{Number of votes for } A}{\text{Number of those who turn out}}$
- What is the group-level causal effect of canvassing on candidate A's vote share (among n voters in the study)? For all samples

$$Z(1) - Z(0) \text{ where } Z(t) = \frac{\sum_{i=1}^{n} Y_i(t) S_i(t)}{\sum_{i=1}^{n} S_i(t)} \text{ for } t \in \{0, 1\}$$

$$Q \text{ why not } \sum_{i=1}^{n} Y_i(t) \text{ for } new quantity everyone not conversed}$$

 $Lf S_{2}(t=0 \rightarrow Y_{2}(t))$ not well defined

$$\sum_{i=1}^{n} S_{i}(t) = # turnout - tenominator$$

$$\frac{1}{2}$$
 Y_i(t)S_i(t) = # votes for A among
those turned out

-D numerator 8

everyone