POLS 207

Problem Set 1

Villaseñor-Derbez, J.C.

Problem 1: Thinking about Potential Outcomes

a) Consider binary treatment $D_i \in \{0,1\}$, observed outcome variable Y_i. What is the meaning of Y_{1i} and Y_{0i}? Describe both in words, and choose an example to illustrate.

Y_{1i} represents the *potential* value that we would observe for a variable of interest Y if unit i receives the treatment (*i.e.* $D_i = 1$). Converseley, Y_{0i} represents the *potential* value that we would observe for a variable of interest Y if unit i **does not** receive the treatment (*i.e.* $D_i = 0$).

These potential outcomes can be exemplified by a variable of interest such as Y_i which represents the total fishing hours by vessel i. The binary treatment D_i represents a policy intervention (implementation of Individual Transferable Quotas: ITQ) randomly applied to the vessels. Y_{0i} represents the potential fishing hours of vessel i if it is not within the randomly selected group of vessels that operate under ITQ. However, if that same vessel were randomly selected to participate in the ITQ program, its fishing hours woulod be characterized by Y_{1i}. The implementation of ITQs provides fishers with a sense of ownership, which reduces "the race to fish". Therefore we would expect $Y_{1i} < Y_{0i}$.

b) What is the difference between Y_{1i} and "Y_i for a unit that actually received the treatment"? Explain the difference using the example you began in section a.

For a unit that actually received the treatment, $Y_i = Y_{1i}$. The observed outcome can be written as:

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$$

The statement "a unit that actually received the treatment" implies $D = 1$. Therefore:

$$\begin{aligned}
Y_i &= D_i Y_{1i} + (1 - D_i) Y_{0i} \\
&= (1) Y_{1i} + (1 - (1)) Y_{0i} \\
&= (1) Y_{1i} + (0) Y_{0i} \\
&= (1) Y_{1i} + 0 \\
Y_i &= Y_{1i}
\end{aligned}$$

c) Define the average treatment effect (ATE) and average treatment effect among the treated (ATT) using potential outcomes notation. Describe in words what each quantity means.

ATE

The average treatment effect in the potential outcomes notation is given by:

$$ATE = \mathbb{E}[Y_{1i} - Y_{0i}]$$

For an individual unit i, the treatment effect τ_i is just the difference between its potential outcomes Y_{1i} and Y_{0i}. For the entire population of interest, the average treatment effect is given by the expectation of this τ_i value across all units. In other words,the ATE is given by the difference of potential outcomes for all i's, independent on the treatment status D_i.

ATT

The average treatment effect on the treated is given by:

$$ATT = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1]$$

This estimand focuses *only* on treated units (*i.e.* units for which $D_i = 1$). It represents the expected difference in potantial outcomes for *treated* units only.

d) When will the ATT and the ATE be the same? Prove it.

As stated above, ATE represents the expected differences in potential outcomes for all units, while ATT represents this same measure but for treated units only. If D_i is assigned at random, one can expect $ATE = ATE$, the assertion is that:

$$\mathbb{E}[Y_{1i} - Y_{0i}] = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1]$$

The observed difference in means is given by:

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0]$$

Which can be decomposed into:

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] + (\mathbb{E}[Y_{0i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0])$$

Since Y_{0i} and D_i are independent, the last term on the right-hand side ($\mathbb{E}[Y_{0i}|D_i = 0]$) can be expressed as $\mathbb{E}[Y_{0i}|D_i = 1]$, such that the equation becomes:

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] + (\mathbb{E}[Y_{0i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 1])$$
$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] + (0)$$

The right-hand side of the above equation is now just the ATT. Given the independence, the left-hand side can also be rewritten such that:

$$\mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1]$$
$$\mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 1] = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1]$$
$$\mathbb{E}[Y_{1i} - Y_{0i}] = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1]$$

This now states that the $ATE = ATT$.

Problem 2: Potential Outcomes with Interference

a) You decide to conduct a get-out-the-vote experiment. For your treatment, you encourage a randomly assigned set of voters to put up yard signs supporting a particular candidate. Your hypothesis is that people who are asked to publicly declare their support for a candidate are more likely to remember to vote. Is this experiment likely to violate the SUTVA assumption? Why or why not?

Yes, this is likely to violate SUTVA if the untreated units can also see the signs. When treated individuals put up yard signs and declare their support for a candidate, this is likely to also incentivize neighbours to go vote. An estimation of the ATE would result in a downard biased estimate of the true effect of putting up the signs.

b) Now suppose you decide to conduct a different get-out-the-vote experiment. With this experiment, you send a randomly assigned set of voters a mailer. These mailers tell each treated individual whether their neighbors voted or not in the last election. Is this experiment likely to violate the SUTVA assumption? Why or why not?

This experiment is less likely to violate SUTVA. Since the mailer is received privately and individuals are chosen at random, it is unlikely that untreated units will be affected by the mailer that a neighbour might have received. Of course, communication between neighbours may still lead to a violation of SUTVA.

c) Imagine you have the following study population. For this population, calculate the ATE, ATT, and ATC. Then calculate the ATE for the subgroup of odd-numbered cases.

```r
# create a tibble with the data
data <- tibble(i = 1:5,
               Di = c(0, 1, 1, 0, 1),
               Y1i = c(6, -1, 2, 5, 2),
               Y0i = c(5, 3, 2, 2, -3)) %>%
  mutate(tau = Y1i - Y0i) # Add tau

# Print the table
data %>%
  knitr::kable(col.names = c("i", "$D_i$", "$Y_{1i}$", "$Y_{0i}$", "$\\tau$"),
               escape = F,
               booktabs = T,
               caption = "Data for potential outcomes of 5 observational units.")
```

Table 1: Data for potential outcomes of 5 observational units.

i	D_i	Y_{1i}	Y_{0i}	τ
1	0	6	5	1
2	1	-1	3	-4
3	1	2	2	0
4	0	5	2	3
5	1	2	-3	5

ATE

$$
\begin{aligned}
ATE &= \mathbb{E}[Y_{1i} - Y_{0i}] \\
&= \mathbb{E}[\tau_i] \\
&= \frac{1 - 4 + 0 + 3 + 5}{5} \\
&= \frac{5}{5} \\
&= 1
\end{aligned}
$$

```r
# The mean of tau
mean(data$tau)
```

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

ATT

$$
\begin{aligned}
ATT &= \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] \\
&= \mathbb{E}[\tau_i|D_i = 1] \\
&= \frac{-4 + 0 + 5}{3} \\
&= \frac{1}{3} \\
&= 0.333
\end{aligned}
$$

```r
# Calculate the mean of tau for treated units only
data %>%
  filter(Di == 1) %$%
  mean(tau)
```

```
## [1] 0.3333333
```

ATC

$$
\begin{aligned}
ATC &= \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 0] \\
&= \mathbb{E}[\tau_i|D_i = 0] \\
&= \frac{1 + 3}{2} \\
&= \frac{4}{2} \\
&= 2
\end{aligned}
$$

```
# Calculate the mean of tau for untreated units only
data %>%
  filter(Di == 0) %$%
  mean(tau)
```

```
## [1] 2
```

d) Assume some arbitrary treatment. Now suppose that a small boy's treatment status depends on the treatment status of his sister, his mother and his father. Write out the full set of potential outcomes for this boy. How many such potential outcomes are there? How many τ_i (*e.g.* unit-level treatment effects) can be defined for this situation?

Let $D_{i,s}, D_{i,m}, D_{i,f}$ denote the treatment status of the sister, mother, and father, respectively. The small boy's treatment status (D_i) is therefore given by $D_i(D_{i,s}, D_{i,m}, D_{i,f})$. Therefore, $Y_i(D_i, D_{i,s}, D_{i,m}, D_{i,f})$.

Is D_i a function of D_ij? Is treatment status

```
s <- c(0, 1)
m <- c(0, 1)
f <- c(0, 1)

expand.grid(s, m, f)
```

```
##   Var1 Var2 Var3
## 1    0    0    0
## 2    1    0    0
## 3    0    1    0
## 4    1    1    0
## 5    0    0    1
## 6    1    0    1
## 7    0    1    1
## 8    1    1    1
```

References