Homework 2

MA 590 Special Topics: Causal Inference

Aukkawut Ammartayakun

25 January, 2023

Problem 1

A

In a completely randomized experiment in which $n_1 = n_0 = n/2$, say $S_t^2 = S_c^2$ (the sets of potential outcomes have the same variance in the experimental sample). What is the smallest possible value for the true sampling variance of the difference in means estimator ($\mathbb{V}_{fp}(\hat{\tau}^{diff})$)? Justify your answer (with math)

(*hint: $S_{tc}^2 = S_t^2 + S_c^2 - 2\rho S_t S_c$, where $\rho = corr\{Y(1), Y(0)\}$, the sample correlation of potential outcomes, $S_t = \sqrt{S_t^2}$, and $S_c = \sqrt{S_c^2}$, and recall that $-1 \leq \rho \leq 1$*)

Solution

Given the expression here, we can say that $\mathbb{V}_{fp}(\hat{\tau}^{diff}) \leq \frac{2S_c^2+2S_t^2}{n}$ In another word, $\mathbb{V}_{fp}(\hat{\tau}^{diff}) \leq \frac{2S_{tc}^2+2\rho S_t S_c}{n}$. Since n, S_t, S_c and S_{tc} are stictly positive, the minimum of maximum value of $\mathbb{V}_{fp}(\hat{\tau}^{diff})$ is $\frac{2S_{tc}^2-2S_tS_c}{n}$ (i.e., negative correlation between two outcome). However, if there is no difference or constant treatment in two group. Namely, $y_i(0) = y_i(1), \forall i$, then $\mathbb{V}_{fp}(\hat{\tau}^{diff}) = 0$.

B

With $n = 6$ and binary Y , write out a table of potential outcomes for which $\mathbb{V}_{fp}(\hat{\tau}^{diff})$ attains its lowest value. What is the ATE?

Solution

```
##   y_1 y_0 Z_i
## 1   1   0   1
## 2   1   0   1
## 3   1   0   1
## 4   1   0   0
## 5   1   0   0
## 6   1   0   0
```

ATE in this case would be $\bar{\tau} = 1 - 1 = 0$.

Problem 2

In a Bernoulli randomized trial, each subject is randomized between $Z = 1$ and $Z = 0$ independently. Hence, n_0 and n_1 are random variables, and since $Pr(n_0 = 0) > 0$ and $Pr(n_1 = 0) > 0$, the difference-in-means estimator is not always defined. (The total sample size n remains fixed, not random.) If $\mathbb{E}[Z_i] = Pr(Z_i = 1) = p$ for every subject i, show that the following is an unbiased estimator for $\bar{\tau}$:

$$\hat{\tau}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Z_iY_i}{p} - \frac{(1-Z_i)Y_i}{1-p}\right\}$$

(IPW stands for "inverse probability weighted" and will become important when we talk about propensity scores.)

Solution

We can see that if $Y_i = Z_iy_i(1) + (1-Z_i)y_i(0)$. Then,

$$\hat{\tau}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Z_iy_i(1)}{p} - \frac{(1-Z_i)y_i(0)}{1-p}\right\}$$

Taking the expectation, we can see that

$$\mathbb{E}[\hat{\tau}_{IPW}] = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{py_i(1)}{p} - \frac{(1-p)y_i(0)}{1-p}\right\}$$

which turns out to be the ATE or $\bar{\tau}$. Thus, conclude the unbiasness of the estimator.

Problem 3

Had the Lady Tasing Tea experiment been a Bernoulli randomized trial, so that each cup would be randomized to tea-first or milk-first independently, what would be the p-value (using Fisher's method) had the lady gotten all 8 cups right?

Solution

Let assume that we don't care whether the trial we have will contains both tea-first and milk-first. Then, there are $2^8 = 256$ possible outcomes. Even so, there is only one way to orient the test to match with the given answer. The probability of getting all 8 cups to align with the guessing is $\frac{1}{256}$. Thus, the p-value is $\frac{1}{256}$.