Homework 5

MA 590 Special Topics: Causal Inference

Aukkawut Ammartayakun

22 February, 2023

Problem 1

Imagine an experiment in which a group of smokers is randomized between two groups: one is encouraged to quit smoking, and the other is not. A year later, subjects in both treatment arms complete a survey asking whether they are still smoking, and they undergo a medical examination to determine the health of their lungs. In this case, receiving the treatment $W = 1$ if they quit smoking and $W = 0$ if they did not. The health of their lungs is the outcome of interest, Y. Describe possible violations of the three instrumental variable assumptions.

Solution

Assumption 1: $ITT_W > 0$

They don't have to quit smoking.

Assumption 2: $w(1) = w(0) = 0 \implies y(1) = y(0) = 0$

If they still smoke or not smoke from the beginning, they might develop lung cancer without the necessity of those treatments.

Assumption 3: $w(1) \geq w(0)$

They might quit smoking and then return to smoke again. Although in general, the observable defier in this case would not occur.

Problem 2

If A, B, and C are random variables, with A binary, show that

$$\frac{\operatorname{cov}(A,B)}{\operatorname{cov}(A,C)} = \frac{\mathbb{E}[B|A=1] - \mathbb{E}[B|A=0]}{\mathbb{E}[C|A=1] - \mathbb{E}[C|A=0]}$$

You may use the fact that for any two random variables x and y, $\operatorname{cov}(x,y) = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$

Solution

From the result from law of total probability, that is

$$\mathbb{E}[B] = p\mathbb{E}[B|A=1] + (1-p)\mathbb{E}[B|A=0]$$

given that $p = \mathbb{P}(A=1)$, and from,

$$\operatorname{cov}(x,y) = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$$

along with the definition of covariance, we have

$$\text{cov}(A, B) = p\mathbb{E}[(1-p)B|A=1] - (1-p)\mathbb{E}[pB|A=0]$$

Factorize this to get common factor of $p(1-p)$ which left with the LATE.

Problem 3

In a randomized trial within a piece of educational software, one random group of student working on a particular math problem are given the option of pushing a button that says "show explanation" that will show them a paragraph explaining how to do the problem, while another group are given the option of pushing on a button that says "show video" that includes a video of a teacher explaining the concepts behind the problem. The outcome of interest is students' performance on the next problem. In both conditions, students may solve the problem without requesting any help. Let $Z_i = 1$ if student i is assigned to an explanation and $Z_i = 0$ if i is assigned to the video condition. Let $W_i = 1$ if student i requested help (be it an explanation or a video, depending on Z_i) and 0 otherwise.

Part A

What are the principal strata (groups defined based on $w(0)$ and $w(1)$) in this problem?

Solution

There are four principal strata:

- $w(0) = 0, w(1) = 0$: Students who did not request help in either condition no matter what
- $w(0) = 1, w(1) = 0$: Students who requested help but not complies the assignment
- $w(0) = 0, w(1) = 1$: Students who requested help complies with the assignment
- $w(0) = 1, w(1) = 1$: Students who requested help in both conditions no matter what

Part B

What would be wrong with estimating average treatment effects using data from only students who requested help by pushing on either the explanation or video button? In what circumstances would this be justifiable?

Solution

The justification is that in this group, $y(1)$ and $y(0)$ are inside the strata and hopefully can determine the ATE. However, that does not tell anything because the test is violated.

Part C

Assume monotonicity, i.e. $w(1) \geq w(0)$. Give an estimate based on observed Z and W for the proportion of the dataset belonging to each of the four principal strata.

Solution

If monotonicity is assumed, and we know the proportion for each assignment along with the number of people who actually uses help, we can say that $\pi_1 = \pi_{W=0}$ and we can also say from monotonicity that $\pi_2 = 0$. In the case of π_3 and π_4 we know that $\pi_3 + \pi_4 = 1 - \pi_{W=0}$.