EE19MTECH01008

Monday, 12 April 2021 12:27 AM

Assignment 2

1) a) SUTVA or Stable Unit Treatment Assumption →

There are two assumptions :

i) No interference - The potential outcomes for any unit do not vary with treatments assigned to other units. A subject's potential outcome is not affected by other subject's exposure to treatment.

ii) No hidden variations in treatment - For each unit there are no different versions of each treatment level, which lead to different potential outcomes. Unambiguous definition of exposure. Eg: 15 min of exercise or 30 mins of exercise. There is no clear definition.

b) Large Sample Size - It can help with two things

i) Sampling variability - This helps in better and robust selection of samples from a given population.

ii) non deterministic outcome - In cases where we have very random outcomes when applying the same treatment on the same individual, we often need a very large sample size in order to see the full distributions of these outcomes. This way we can be more sure of the true parameters in the population distributions of these outcomes. This way we can be more sure of the true parameters in the population are.

As sample size increases the size of the confidence interval decreases which mean we can be more confident about a specific outcome

c) <u>No Measurement Error</u> – A fundamental assumption in all statistical analysis is that all the observations are correctly measured. In many situation this assumption is violated. There can be several reasons such as, variables are not measurable, variable are not clearly defined etc.
In all these cases true value of the variable cannot be recorded. Instead it is observed with some error. The difference b/w the observed & true values of the variable is called as measurement error.

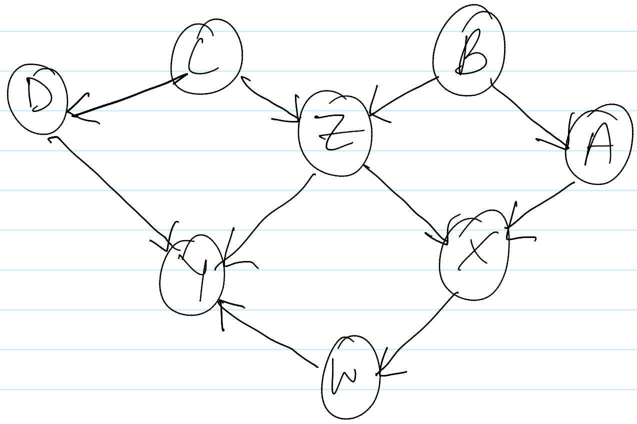
d) Double blindedness – It eliminates the placebo effect i.e. the patients don't necessarily know that they are being treated with the original ingrediant or not. The doctors also don't know that whom they are treating

$$\begin{tikzcd} A \arrow[r, "\text{causal}"] \arrow[r, bend left=60, "\text{placebo}"] \arrow[r, bend right=60, "\text{scientist prefrence.}"'] & Y \end{tikzcd}$$

All experiments should be double-blind to get the best causal estimate of a treatment

All experiment should be double-blind to get the best causal estimate of a treatment.

2)



a) ($\cancel{Z}$) (A, B, Z) (Z, C, D) (A, B, Z, C, D). (Z, A) (Z, B)
(Z, C) (B, C, Z)

b) (~~Z, A~~) (Z, B) (Z, C) (Z, D)

c) (Z, C) (Z, B) (Z, A) (A, X) (C, X) (B, X)
(Z, W)

d) (Z, B) (Z, A) (C, X)

3) a) $P_Y = \mathcal{N}(0, 17)$

b) $P_{Y|X=k} = \mathcal{N}(4k, 1)$

c) $P_{Y|do(X=k)} = \mathcal{N}(4k, 1)$

d) $P_{X|Y=k} = \mathcal{N}(k/4, 1)$

e) $P_{X|do(Y=k)} = \mathcal{N}(0, 1)$ (since X & Y are independent now)

4) a)

X	Y	Z	P(x, y, z)
0	0	0	$(1-p_1)(1-q_1)(1-r)$

0	0	0	$(1-p_1)(1-q_1)(1-r)$
0	0	1	$(1-p_3)(1-q_2)\,r$
0	1	0	$(1-r)\,p_1(1-q_1)$
0	1	1	$r(1-q_2)\,p_3$
1	0	0	$(1-r)\,q_1(1-p_2)$
1	0	1	$r\,q_2(1-p_4)$
1	1	0	$(1-r)\,q_1\,p_2$
1	1	1	$r\,q_2\,p_4$

b) $P(y=1 \mid x=1) - P(y=1 \mid x=0)$

$$= \frac{(1-r)q_1p_2 + rp_4q_2}{(1-r)q_1p_2 + rp_4q_2 + rq_2(1-p_4) + (1-r)q_1(1-p_2)} - \frac{r(1-q_2)p_3 + (1-r)(1-q_1)p_1}{r(1-q_2)p_3 + (1-r)(1-q_1)p_1 + (1-p_3)(1-q_2)r + (1-p_1)(1-q_1)(1-r)}$$

$$= \frac{(1-r)q_1p_2 + rp_4q_2}{(1-r)q_1 + rq_2} - \frac{r(1-q_2)p_3 + (1-r)(1-q_1)p_1}{r(1-q_2) + (1-r)(1-q_1)}$$

c) $p_4 - p_3$

d) $p_2 - p_1$

f) $P(y \mid do(x))\ \forall\, x \,\&\, y.$

$P(y=1 \mid do(x=1)) = (1-r)p_2 + rp_4$

$P(y=1 \mid do(x=0)) = (1-r)p_1 + rp_3$

$P(y=0 \mid do(x=1)) = (1-r)(1-p_2) + r(1-p_4)$

$P(y=0 \mid do(x=0)) = (1-r)(1-p_1) + r(1-p_3)$

g) $P(y=1 \mid do(x=1)) - P(y=1 \mid do(x=0))$

g) $P(y=1 \mid do(x=1)) - P(y=1 \mid do(x=0))$

$= \big((1-r)p_2 + rp_4\big) - \big((1-r)p_1 + rp_3\big)$

$= (p_2 - rp_2 + rp_4 - p_1 + rp_1 - rp_3$

$= p_2 - p_1 - r[p_2 - p_1 + p_3 - p_4]$

$= (1-r).(p_2 - p_1) - r(p_4 - p_3)$

This quantity does not contain the effect of Z
The quantity calculated in (b) is more effective since it captures the effect of the syndrome in the estimation process

e) $r = 0.5$, $p_1 = 0.6$, $p_4 = 0.1$, $p_3 = 0.2$
$q_1 = 0.8$, $q_2 = 0.1$, $p_2 = 0.5$

(b) equals 0.18
(c) equals -0.1
(d) equals -0.1