Date: / /
Page No.:

Puneet-Mangla
CS17BTECH11029.

Q.1 a) SUTVA refers to stable unit-treatment value assumption which means:
i) there should be no interference → Treatment given to one unit should not affect outcome of other unit.
ii) There is only single version of each treatment level.

For eg: A and B lives in same home and B always cooks. A Blood pressure drug causes B to crave for salty foods, thus B cooks food with slightly higher level of salt. which in turn will increase A's blood pressure. This violates SUTVA assumption as treatment given to B affected A's outcome.

b) Large Sample size ensures that the Causal Estimate is more accurate and the confidence interval is small. It ensures that the population you are considering doesn't contain extreme value.

For eg: If you have sample size of 10 and coincidently the outcome variable (Y) is > 3 for all of them then E[Y] > 3.

But if you have large sample ~ 1000, It is possible that some have outcome variable (Y) < 3 which will give different value of E[Y] clearly. This value will be more accurate estimate.

c) No measurement error simply means that all the observations are correctly measured without any error. In many situations, this assumption could be violated.

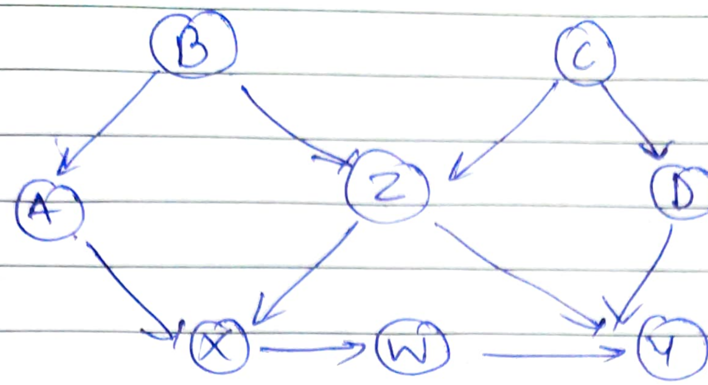
For example: The outcome you are measuring is abstract like taste, happiness, intelligence clearly one can't fully quantify these variables and we have to use dummy variables like IQ for intelligence.

If variables are not ~~observeded~~ observed without error, the causal estimate will also be erroneus. The error could get ~~est~~ amplified making it a unreliable estimate.

d) Double Blindedness means neither participants nor experimenters know who is receiving a particular treatment. It is to ensure that any bias due to demand or placebo effect is not present.

For eg: If a participant knows which treatment he/she will be given, it might be possible that he is not willing to take the treatment at end time.

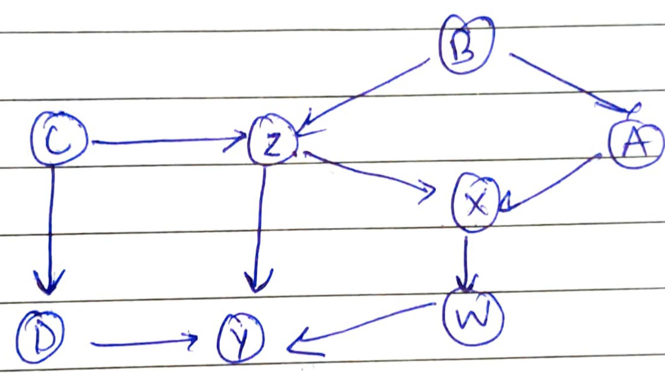
Q.2 ~~a~~)



a) {Z, A, D}, {Z, A}, {Z, D}, {Z, B}, {Z, C}
{Z, B, C}, {Z, A, B, C, D}, ~~In general~~ ~~{Z}∪P~~
{Z, A, B}, {Z, ABC}, {Z, ~~B~~, C, D} and so on.

b) {Z, A}, {Z, D}, {Z, B}, {Z, C}

c)



{C}, {Z, B, ~~A~~}, {Z, ~~B~~, A, ~~X~~}, {Z, X}, {Z, W}

d) {C, X}, {Z}

Q.3 (ii) $X := N_x$, $Y := 4X + N_y$ where $N_x, N_y \overset{i.i.d}{\sim} N(0,1)$

i) a) P_Y : $E[Y] = 4E[X] + E[N_y]$
$= 4E[N_x] + E[N_y] = 0$
$Var(Y) = 16 \cdot Var(X) + Var(N_y)$
$= 16 + 1 = 17$

b) $P_{Y|X=k}$ ~~$Y = 4k$~~
$E[Y|X=k] = E[4k] + E[N_y]$
$= 4k$
$Var[Y|X=k] = Var[4k] + Var[N_y]$
$= 0 + 1 = 1$

c) $P_{Y|do(X=k)}$ $E[Y|do(X=k)] = E[4k] + E[N_y] = 4k$
$Var[Y|do(X=k)] = Var[4k] + Var[N_y]$
$= 1$

d) $P_{X|Y=k}$ $k = 4X + N_y \Rightarrow X = \frac{k - N_y}{4}$
~~$X = \frac{k + N_y'}{4}$~~

~~where $N_y' \sim N(0,1)$~~ } ~~since gaussian is symmetric.~~

$E[X] = E[\frac{k}{4}] + E[\frac{N_y}{4}]$

$= k/4$

$Var[X] = 0 + \frac{1}{16} Var[N_y] = \frac{1}{16}$

e) $P_{X|do(Y=k)} = P_X$ $\quad E[X|do(y=k)] = 0$

$Var[X|do(Y=k)] = 1$

f) In jupyter notebook.

Q. 4

a)

	X	Y	Z	$p(x,y,z) = p(z) \cdot p(x \mid z) \cdot p(y \mid x,z)$
p_1'	0	0	0	$(1-\gamma) \cdot (1-q_1) \cdot (1-p_1)$
p_2'	0	0	1	$\gamma \cdot (1-q_2) \cdot (1-p_3)$
p_3'	0	1	0	$(1-\gamma) \cdot (1-q_1) \cdot p_1$
p_4'	0	1	1	$\gamma \cdot (1-q_2) \cdot p_3$
p_5'	1	0	0	$(1-\gamma) \cdot q_1 \cdot (1-p_2)$
p_6'	1	0	1	$\gamma \cdot q_2 \cdot (1-p_4)$
p_7'	1	1	0	$(1-\gamma) \cdot q_1 \cdot p_2$
p_8'	1	1	1	$\gamma \cdot q_2 \cdot p_4$

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

$$= \frac{p_7' + p_8'}{p_5' + p_6' + p_7' + p_8'} - \frac{p_3' + p_4'}{p_1' + p_2' + p_3' + p_4'}$$

$$= \frac{(1-\gamma) \cdot q_1 \cdot p_2 + \gamma \cdot q_2 \cdot p_4}{(1-\gamma) \cdot q_1 + \gamma \cdot q_2} - \frac{\gamma \cdot (1-q_2) \cdot p_3 + (1-\gamma) \cdot (1-q_1) \cdot p_1}{(1-\gamma) \cdot (1-q_1) + \gamma \cdot (1-q_2)}$$

c) $p_4 - p_3$

d) $p_2 - p_1$

e) $p_2 = 0.6,\ p_1 = 0.7,\ p_4 = 0.2,\ p_3 = 0.3$
$q_1 = 0.9,\ q_2 = 0.1,\ \gamma = 0.5$

Expression (b) evaluates to : 0.22
Expression (c) evaluates to : -0.1
" (d) " to : -0.1

f) $p(y \mid do(x)) = \sum_z p(y=y \mid x=x, z)$

$p(y=1 \mid do(x=1)) = \gamma \cdot p_4 + (1-\gamma) \cdot p_2$

$p(y=1 \mid do(x=0)) = \gamma \cdot p_3 + (1-\gamma) \cdot p_1$

$p(y=0 \mid do(x=1)) = \gamma \cdot (1-p_4) + (1-\gamma) \cdot (1-p_2)$

$p(y=0 \mid do(x=0)) = \gamma \cdot (1-p_3) + (1-\gamma) \cdot (1-p_1)$

g) $\gamma \cdot [p_4 - p_3] + [p_2 - p_1](1-\gamma)$

The expression is completly different from (b). Expression in (g) is more relevant in accessing the effectiveness as it takes into account the affect of compound Z and marginalize over it.