EE19MTECH01008 - Assignment 3

Wednesday, 5 May 2021 10:07 AM

3) Exclusion Restriction → Z causal effect on Y is fully mediated by T. This assumption excludes Z from structural equation of Y and from any other structural equations that would make causal association flow from Z to Y without going through T.

Relevance → Z has causal effect on T. It corresponds to the existence of an active edge from Z to T in the causal graph.

Exogeneity → Given a simple regression model

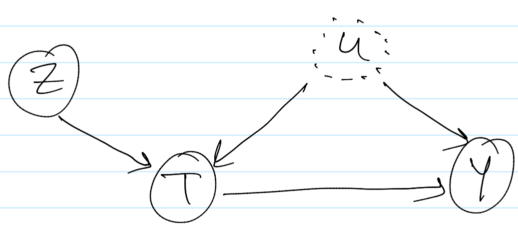
$$y = \beta_0 + \beta_1 x + u$$

Z is used as an instrumental variable for x

$Cov(z, u) = 0$ & $Cov(z, x) \neq 0$

x, u are uncorrelated then β_0 & β_1 are consistent

Instrument exogeneity means Z should have no partial effect on y (after x and omitted variables have been controlled for) & Z should be uncorrelated with omitted variables



Graph used for reference of the assumptions.

4) $P(X=1 \mid Z=z)$

$$P(X=1 \mid Z=1) = \frac{0.116 + 0.009}{0.116 + 0.009 + 0.334 + 0.051} = 0.245$$

$$0.116 + 0.009 + 0.334 + 0.051$$

$$P(X=1 \mid Z=0) = \frac{0.274 + 0.101}{0.274 + 0.101 + 0.079 + 0.036} = 0.765$$

We multiply the corresponding values by $\frac{1}{0.245}$ & $\frac{1}{0.765}$

$$P(X=0 \mid Z=1) = 0.755 \quad (\because 1 - 0.245)$$

$$P(X=0 \mid Z=0) = 0.235 \quad (\because 1 - 0.765)$$

$$\therefore \tau = E\left[Y(1) - Y(0)\right]$$

$$= E\left[\frac{\mathbb{I}(X=1)\,Y}{e(w)}\right] - E\left[\frac{\mathbb{I}(x=0)\,Y}{1-e(w)}\right]$$

$$= \left[\frac{0.116 \times 1}{0.245} + \frac{0.274 \times 1}{0.765} + \frac{0.009 \times 0}{0.245} + \frac{0.101 \times 0}{0.765}\right]$$

$$- \left[\frac{0.334 \times 1}{0.755} + \frac{0.079 \times 1}{0.235} + \frac{0.051 \times 0}{0.765} + \frac{0.031 \times 0}{0.235}\right]$$

$$= 0.053$$

5) The propensity score is given by

$$p(x) = P(T=1 \mid X=x)$$

For treated individuals, $w(x) = \frac{1}{\ldots}$

For treated individuals, $\omega(x) = \dfrac{1}{p(x)}$

For control individuals, $\omega(x) = \dfrac{1}{1-p(x)}$

In the presence of measured confounders, the causal effect is estimated by IPW $\rightarrow$

$$\tau = \frac{1}{n_T} \sum_{i:T_i=1} \frac{Y_i}{p(X_i)} - \frac{1}{n_C} \sum_{i:T_i=0} \frac{Y_i}{1-p(X_i)}$$

$$= \frac{1}{n} \sum_{i}^{n} \frac{T_i Y_i}{p(X_i)} - \frac{1}{n} \sum_{i}^{n} \frac{(1-T_i) Y_i}{1-p(X_i)}$$

$p(x)$ is logistic regression based estimator of propensity score

The original quantity $\tau = E[Y(1)] - E[Y(0)]$ is unbiased, we have to show

$$E\left[\frac{YT}{p(X)}\right] = E[Y(1)] \quad \& \quad E\left[\frac{Y(1-T)}{1-p(X)}\right] = E[Y(0)]$$

$$\therefore E\left[\frac{YT}{p(X)}\right] = E\left[E\left[\frac{YT}{p(X)} \,\middle|\, X\right]\right]$$

$$= E\left[E\left[\frac{Y(1)\,T}{p(X)} \,\middle|\, X\right]\right]$$

$$= E\left[\frac{E[Y(1)|X]\,E[T|X]}{p(X)}\right]$$

$$= E\left[E[Y(1)|X]\right] = E[Y(1)]$$

Similarly,

$$E\left[\frac{Y(1-T)}{1-p(X)}\right] = E[Y(0)]$$

$\left(1-p(x)\right)$

6) a) $E\left[Y_{x=1} \mid Z=1\right] = b\left(1 + \frac{P(U_1=0)P(U_2=1)}{P(U_1=0)P(U_2=1) + P(U_1=1)P(U_2=0)}\right)$

b)

U_1	U_2	X_{U_1,U_2}	Z_{U_1,U_2}	Y_{U_1,U_2}	$Z_{x=0,\,U_1,U_2}$	$Z_{x=1,\,U_1,U_2}$	$Y_{x=0,\,U_1,U_2}$	$Y_{x=1,\,U_1,U_2}$
0	0	0	0	0	0	a	0	ab
0	1	0	1	b	1	a+1	b	b(a+1)
1	0	1	a	ba	0	a	0	ab
1	1	1	a+1	b(a+1)	1	a+1	b	b(a+1)

7) τ be the slope of total effect of X on Y

$$\tau = E[Y \mid do(x+1)] - E[Y \mid do(x)]$$

for any evidence $E=e$,

$$E[Y_{X=x} \mid E=e] = E[Y \mid E=e] + \tau\left(x - E[X \mid E=e]\right)$$

The quantity can be calculated by best estimate of Y conditioned on evidence "e", $E[Y|e]$ and adding to the change Y when X is shifted from current best estimate $E[X \mid E=e)$ to hypothetical value x.

This is used in computing effect of treatment on treated

$$ETT = E[Y_1 - Y_0 \mid X=1]$$

evidence $e = \{X=1\}$

$$ETT = E[Y_1 \mid X=1] - E[Y_0 \mid X=1]$$
$$= E[Y \mid X=1] - E[Y \mid X=1] + \tau\left(1 - E[X \mid X=1]\right) - \tau\left(0 - E[X \mid X=1]\right)$$
$$= \tau$$

$\therefore$ The effect of treatement on treated is basically equal to the effect on treatement on entire population

∴ the effect of treatement on treated is basically equal to the effect of treatement on entire population.

2) a) For $W=0$ $X=0$

$$\frac{30+20}{200} = \frac{50}{200} = \frac{1}{4} = 25\%$$

$W=1$ $X=0$

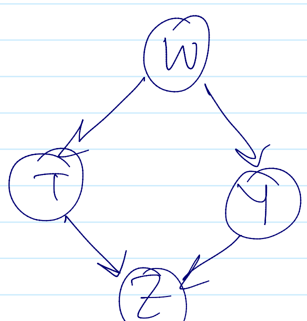
$$\frac{30+20}{200} = \frac{1}{4} = 25\%$$

So we can assume the treatement assignment is unconfounded.

b) $P(W=1 \mid X=1) = \frac{75}{100} = 0.75$

$P(W=1 \mid X=0) = \frac{30+20}{30+30+20+20} = 0.5$

1)



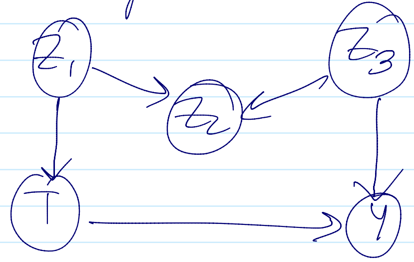
Condition on Z

If we condition on a descendant of T that isn't a mediator, it could unblock a path from T to Y that was blocked by a collider. In this figure conditioning on Z induces non causal association between T and Y, which biases the estimate of the causal effect. So conditioning on Z or any descendant of Z in a path like this, will induce collider bias.

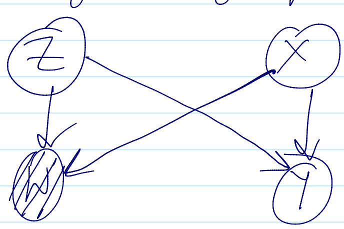
Selection bias → It is a distortion in measure of association due to sample selection that does not accurately reflect

Selection bias → It is a distortion in measure of association due to sample selection that does not accurately reflect the target population. It occurs when investigators use improper procedures for selecting a sample population

M-bias → If we condition on pre treatement covariates we can induce collider bias. In the following figure if we condition on Z_2 open a back door path for non causal association to flow. This is known as M-bias due to M-shape that this non causal association flows.



In the given graph, if we condition on W, there's Collider bias



$$P(Y=y \mid do(X=x), W=w)$$

$$= \sum_{z} P(Y=y \mid X=x, W=w, Z=z)$$

2C) $\tau = E[Y(1) - Y(0)]$

$$= E\left[\frac{\mathbb{I}(W=1)Y}{e(w)}\right] - E\left[\frac{\mathbb{I}(W=0)\,Y}{e(w)}\right]$$

$$= 7.80 - 6.07$$

$$= 1.7299$$