MATH 4824C: Causal Inference

♠ Solution of HW 1

Q1. Assignment 1.1 in textbook;

Answer
Problem statement: (Independence in two-by-two tables) Prove (1) and (2) in Proposition 1.1.

Proposition 1.1 (1) The following statements are all equivalent: $Z \perp Y$, $\mathrm{RD} = 0, \mathrm{RR} = 1$, and $\mathrm{OR} = 1$. (2) If p_{zy} 's are all positive, then $\mathrm{RD} > 0$ is equivalent to $\mathrm{RR} > 1$ and is also equivalent to $\mathrm{OR} > 1$.

Proof. To show (1), we show (1.a) $Z \perp Y \Rightarrow \mathrm{RD} = 0$, (1.b) $\mathrm{RD} = 0 \Rightarrow \mathrm{RR} = 1$, (1.c) $\mathrm{RR} = 1 \Rightarrow \mathrm{OR} = 1$ and (1.d) $\mathrm{OR} = 1 \Rightarrow Z \perp Y$.
(1.a) When $Z \perp Y$, we have

$$\begin{aligned}\mathrm{RD} &= \mathrm{pr}(Y = 1 \mid Z = 1) - \mathrm{pr}(Y = 1 \mid Z = 0)\\ &= \mathrm{pr}(Y = 1) - \mathrm{pr}(Y = 1) = 0.\end{aligned}$$

(1.b) If $\mathrm{RD} = 0$, we have

$$\mathrm{RR} - 1 = \frac{\mathrm{pr}(Y = 1 \mid Z = 1) - \mathrm{pr}(Y = 1 \mid Z = 0)}{\mathrm{pr}(Y = 1 \mid Z = 0)} = 0.$$

Thus $\mathrm{RR} = 0$.
(1.c) If $\mathrm{RR} = 1$, we have

$$\begin{aligned}\mathrm{OR} &= \frac{\mathrm{pr}(Y = 1 \mid Z = 1)/\,\mathrm{pr}(Y = 0 \mid Z = 1)}{\mathrm{pr}(Y = 1 \mid Z = 0)/\,\mathrm{pr}(Y = 0 \mid Z = 0)}\\ &= \frac{\mathrm{pr}(Y = 1|Z = 1)(1 - \mathrm{pr}(Y = 1|Z = 0))}{\mathrm{pr}(Y = 1|Z = 0)(1 - \mathrm{pr}(Y = 1|Z = 1))}\\ &= \frac{\mathrm{pr}(Y = 1|Z = 1)(1 - \mathrm{pr}(Y = 1|Z = 1))}{\mathrm{pr}(Y = 1|Z = 1)(1 - \mathrm{pr}(Y = 1|Z = 1))} = 1.\end{aligned}$$

(1.d) If $\mathrm{OR} = 1$, then

$$\begin{aligned}\mathrm{OR} - 1 &= \frac{\mathrm{pr}(Y = 1|Z = 1)(1 - \mathrm{pr}(Y = 1|Z = 0))}{\mathrm{pr}(Y = 1|Z = 0)(1 - \mathrm{pr}(Y = 1|Z = 1))} - 1\\ &= \frac{\mathrm{pr}(Y = 1|Z = 1)(1 - \mathrm{pr}(Y = 1|Z = 0)) - \mathrm{pr}(Y = 1|Z = 0)(1 - \mathrm{pr}(Y = 1|Z = 1))}{\mathrm{pr}(Y = 1|Z = 0)(1 - \mathrm{pr}(Y = 1|Z = 1))}\\ &= \frac{\mathrm{pr}(Y = 1|Z = 1) - \mathrm{pr}(Y = 1|Z = 0)}{\mathrm{pr}(Y = 1|Z = 0)(1 - \mathrm{pr}(Y = 1|Z = 1))} = 0\end{aligned}$$

Thus, $\mathrm{pr}(Y = 1|Z = 1) = \mathrm{pr}(Y = 1|Z = 0)$. Since $Z \in \{0, 1\}$, $\mathrm{pr}(Z = 1) + \mathrm{pr}(Z = 0) = 1$. Thus

$$\begin{aligned}\mathrm{pr}(Y = 1) &= \sum_{z=0,1} \mathrm{pr}(Y = 1|Z = z)\\ &= \mathrm{pr}(Y = 1|Z = 1)(\mathrm{pr}(Z = 1) + \mathrm{pr}(Z = 0))\\ &= \mathrm{pr}(Y = 1|Z = 1) = \mathrm{pr}(Y = 1|Z = 0).\end{aligned}$$

Similarly one also have

$$\mathrm{pr}(Y = 0) = \mathrm{pr}(Y = 1|Z = 1) = \mathrm{pr}(Y = 1|Z = 0).$$

Therefore, for any $y = 0, 1$ and $z = 0, 1$, we have

$$\mathrm{pr}(Y = y, Z = z) = \mathrm{pr}(Y = y)\,\mathrm{pr}(Z = z).$$

That is, $Y \perp Z$.

To show (2), we show when $p'_{zy}s$ are all positive, (2.a) RD $> 0 \Rightarrow$ RR > 1; (2.b) RR $> 1 \Rightarrow$ OR > 1 and (2.c) OR $> 1 \Rightarrow$ RD > 0.
(2.a) If RD > 0, then

$$\begin{aligned}\mathrm{RR} - 1 =& \frac{\mathrm{pr}(Y = 1 \mid Z = 1) - 1}{\mathrm{pr}(Y = 1 \mid Z = 0)} \\ =& \frac{\mathrm{RD}}{\mathrm{pr}(Y = 1 \mid Z = 0)} > 0.\end{aligned}$$

(2.b) If RR > 1, then

$$\mathrm{OR} - 1 = \frac{(\mathrm{RR} - 1)\,\mathrm{pr}(Y = 1|Z = 0)}{\mathrm{pr}(Y = 1|Z = 0)(1 - \mathrm{pr}(Y = 1|Z = 1))} > 0$$

(2.c) Observe that

$$\mathrm{OR} - 1 = \frac{\mathrm{RD}}{\mathrm{pr}(Y = 1|Z = 0)(1 - \mathrm{pr}(Y = 1|Z = 1))}.$$

Then

$$\mathrm{RD} = (\mathrm{OR} - 1)(\mathrm{pr}(Y = 1|Z = 0)(1 - \mathrm{pr}(Y = 1|Z = 1))) > 0.$$

□

Q2. Justify the identifiable assumption for statistician B in Lord's paradox (page 17 in Chapter 1 of lecture notes); i.e. show that $\beta_g = E(Y(1) - Y(0)|G = 1) - E(Y(1) - Y(0)|G = 0)$ when $Y_i(0) = a + bX_i$;

Answer

Proof. Observe that

$$\begin{aligned}E[Y(1)|G = g, X] =& a_g + bX, \\ E[Y(0)|G = g, X] =& a + bX.\end{aligned}$$

□

Therefore,

$$
\begin{aligned}
&E(Y(1)-Y(0)|G=1)-E(Y(1)-Y(0)|G=0)\\
=&E\left(E(Y(1)-Y(0)|G=1)-E(Y(1)-Y(0)|G=0)|X\right)\\
=&E(E[Y(1)|G=1,X]-E[Y(1)|G=1,X]|X)-E(E[Y(0)|G=1,X]-E[Y(0)|G=0,X]|X)\\
=&E(a_1+bX-a_0-bX|X)-E(0|X\\
=&a_1-a_0:=\beta_g.
\end{aligned}
$$

Q3. Show that in randomized trials; i.e. $(Y_i(1), Y_i(0), X_i) \perp Z_i$, (1) $E(Y_i(1)) = E(Y_i|Z_i = 1)$ and (2) $(Y_i(1), Y_i(0)) \perp Z_i \mid X_i$;

Answer

Proof. (1) Observe that

$$
\begin{aligned}
E(Y|Z=1)=&E(ZY(1)+(1-Z)Y(0)|Z=1)\\
=&E(Y(1)|Z=1)=E(Y(1)).
\end{aligned}
$$

(2) Observe that (subscript i is omitted)

$$
\begin{aligned}
&\mathrm{pr}(Y(1)\le y_1, Y(0)\le y_0, Z=z|X\le x)\\
=&\frac{\mathrm{pr}(Y(1)\le y_1, Y(0)\le y_0, Z=z, X\le x)}{\mathrm{pr}(X\le x)}\\
=&\frac{\mathrm{pr}(Y(1)\le y_1, Y(0)\le y_0, X\le x)\,\mathrm{pr}(Z=z)}{\mathrm{pr}(X\le x)}\\
=&\frac{\mathrm{pr}(Y(1)\le y_1, Y(0)\le y_0, X\le x)}{\mathrm{pr}(X\le x)}\,\frac{\mathrm{pr}(X\le x)\,\mathrm{pr}(Z=z)}{\mathrm{pr}(X\le x)}\\
=&\,\mathrm{pr}(Y(1)\le y_1, Y(0)\le y_0|X\le x)\frac{\mathrm{pr}(X\le x, Z=z)}{\mathrm{pr}(X\le x)}\\
=&\,\mathrm{pr}(Y(1)\le y_1, Y(0)\le y_0|X\le x)\,\mathrm{pr}(Z=z|X\le x).
\end{aligned}
$$

Hence we have $(Y_i(1), Y_i(0)) \perp Z_i \mid X_i$. □

Q4. Show that in randomized trials; i.e. $(Y_i(1), Y_i(0), X_i) \perp Z_i$, $E(Z_iY_i) = E(Z_iY_i(1))$ and $\sum Z_i(Y_i - \bar{Y}_1)^2/(n_1-1)$ is an unbiased estimator for $var(Y(1))$ where $\bar{Y}_1 = \sum Z_iY_i/n_1$ and $n_1 = \sum Z_i$.

Answer

Proof. (1) $E(Z_iY_i) = E(Z_iY_i(1))$.
Observe that

$$E(Z[ZY(1)+(1-Z)Y(0)]=E(Z^2Y(1))=E(ZY(1)).$$

(2) Unbiasedness for $\sum Z_i(Y_i - \bar{Y}_1)^2/(n_1 - 1)$.

$$
\begin{aligned}
& E\left(\sum Z_i(Y_i - \bar{Y}_1)^2/(n_1 - 1)\right) \\
=& E\left(\sum Z_i(Y_i - \bar{Y}_1)^2/(-1 + \sum Z_i)\right) \\
=& E_{\mathcal{A}}\left(E\left(\sum Z_i(Y_i - \bar{Y}_1)^2/(-1 + \sum Z_i)|Z_i = 1 \text{ iff. } i \in \mathcal{A} \subset [n]\right)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
& E\left(\sum Z_i(Y_i - \bar{Y}_1)^2/(-1 + \sum Z_i)|Z_i = 1 \text{ iff. } i \in \mathcal{A} \subset [n]\right) \\
=& E(\sum_{i \in \mathcal{A}}(Y_i - \frac{\sum_{i \in \mathcal{A}} Y_i}{|\mathcal{A}|})^2/(|\mathcal{A}| - 1)) \\
=& \operatorname{Var}(Y|Z = 1) \\
=& E(Y^2|Z = 1) - E^2(Y|Z = 1) \\
=& E([ZY(1) + (1 - Z)Y(0)]^2|Z = 1) - E^2([ZY(1) + (1 - Z)Y(0)]|Z = 1) \\
=& E(Y(1)^2|Z = 1) - E^2(Y(1)|Z = 1) \\
=& E(Y(1)^2) - E^2(Y(1)) \\
=& \operatorname{Var}(Y(1)).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& E\left(\sum Z_i(Y_i - \bar{Y}_1)^2/(n_1 - 1)\right) \\
=& E_{\mathcal{A}}(\operatorname{Var}(Y(1))) \\
=& \operatorname{Var}(Y(1)).
\end{aligned}
$$

□