Solutions week 6

6.1

(a)

This will not work since X_1 is a confounder of X_2 and X_4 and will create a backdoor path.

(b)

This will work since the backdoor path is now blocked.

(c)

This will not work since the path we want to estimate is now blocked.

(d)

This will not work since the path we want to estimate is now blocked.

6.2

What we need to to is to generate the variables according to the graph, for example $X_i = n_i + \sum_{j:X_j \in PA(X_i)} X_j$ for $n_i \sim \mathcal{N}(0,1)$.

$A \perp\!\!\!\perp B \mid C$ means that there is no dependence between A and B given C, so if we try to find the coefficients as `smf.ols('A ~ B + C - 1', data=data).fit()` and the coefficient for A's dependency on B is equal (or close to) zero we can conclude they are independent.

(a)

Yes, because X_5 blocks the direct path and the backdoor path that could be introduced by including the collider X_5 is blocked since neither X_4 or X_6 is in the set.

(b)

No, now we have a backdoor path since both colliders along the path are included.

(c)

No, X_6 is a descendant of the collider X_4 and will also open up the path when included.

(d)

Yes, X_5 blocks the direct path and even though X_5 and X_6 together will open up the backdoor path this will be blocked by including the confounder X_2.

6.3

The backdoor criterion states that if $Y \notin PA(X)$ then any set of variables Z that does not contain X, Y is a valid adjustment set if it contains no descendant of X and blocks all paths from X to Y that leave X through paths going in to X.

Let's pick $Z = PA(X)$ and show that this fulfills both conditions. Since we are working with a DAG we know that none of the parents of a node can also be an ancestor, this would mean there is a cycle in the graph which would mean it is not a DAG. Since $Z = PA(X)$ we also know that for every possible backdoor path we have a non-collider from the path in Z, and thus we block all backdoor paths.

This shows that if $Y \notin PA(X)$ we have that $Z = PA(X)$ is a valid adjustment set for (X, Y) which is exactly what the theorem we wanted to show states.

6.4

The backdoor criterion states that if Z contains no descendants of X and Z blocks all backdoor paths from X to Y it is a valid adjustment set. This tells us that neither D or E can be in Z since they are descendants of X and either one or both of B and C are in Z since this will block the backdoor path. From this we see that

(a)

does not work since we have Z containing D

(b)

works since we have B blocking and we don't care about A

(c)

works since we have B blocking

(d)

works since we have C blocking

(e)

does not work since we have Z containing E

6.5

(a)

$$
\begin{aligned}
&p(\text{cancer} = 1 \mid \mathbf{do}(\text{smoking} = 0)) = \\
&= \sum_{t=[0,1]} p(\text{tar} = t \mid \text{smoking} = 0) \sum_{s=[0,1]} p(\text{cancer} = 1 \mid \text{smoking} = s, \text{tar} = t) p(\text{smoking} = s) \\
&= p(\text{tar} = 0 \mid \text{smoking} = 0)\big(p(\text{cancer} = 1 \mid \text{smoking} = 0, \text{tar} = 0)p(\text{smoking} = 0) + \\
&\qquad p(\text{cancer} = 1 \mid \text{smoking} = 1, \text{tar} = 0)p(\text{smoking} = 1)\big) + \\
&\quad p(\text{tar} = 1 \mid \text{smoking} = 0)\big(p(\text{cancer} = 1 \mid \text{smoking} = 0, \text{tar} = 1)p(\text{smoking} = 0) + \\
&\qquad p(\text{cancer} = 1 \mid \text{smoking} = 1, \text{tar} = 1)p(\text{smoking} = 1)\big) \\
&= 0.9 \cdot (0.1 \cdot 0.6 + 0.3 \cdot 0.4) + 0.1 \cdot (0.2 \cdot 0.6 + 0.4 \cdot 0.4) = 0.19
\end{aligned}
$$

$$
\begin{aligned}
&p(\text{cancer} = 1 \mid \mathbf{do}(\text{smoking} = 1)) = \\
&= \sum_{t=[0,1]} p(\text{tar} = t \mid \text{smoking} = 1) \sum_{s=[0,1]} p(\text{cancer} = 1 \mid \text{smoking} = s, \text{tar} = t) p(\text{smoking} = s) \\
&= p(\text{tar} = 0 \mid \text{smoking} = 1)\big(p(\text{cancer} = 1 \mid \text{smoking} = 0, \text{tar} = 0)p(\text{smoking} = 0) + \\
&\qquad p(\text{cancer} = 1 \mid \text{smoking} = 1, \text{tar} = 0)p(\text{smoking} = 1)\big) + \\
&\quad p(\text{tar} = 1 \mid \text{smoking} = 1)\big(p(\text{cancer} = 1 \mid \text{smoking} = 0, \text{tar} = 1)p(\text{smoking} = 0) + \\
&\qquad p(\text{cancer} = 1 \mid \text{smoking} = 1, \text{tar} = 1)p(\text{smoking} = 1)\big) \\
&= 0.7 \cdot (0.1 \cdot 0.6 + 0.3 \cdot 0.4) + 0.3 \cdot (0.2 \cdot 0.6 + 0.4 \cdot 0.4) = 0.21
\end{aligned}
$$

(b)

(i)

The first front door criterion says that tar blocks all directed paths from X=smoking to Y=cancer. So adding a directed path from smoking to cancer would mean that there is some other effect from smoking that can give you cancer that is not related to tar.

(ii)

The second front door criterion says that there is no backdoor paths from X=smoke to M=tar. If we assume that some people genetically like the smell of smoke and every chance they get they breathe it in. This could mean that they get tar from other sources than smoking, but also that they could be more likely smokers and we now have a backdoor path through the confounder ”gene”

(iii)

The third front door criterion says that there is no backdoor paths from M=tar to Y=cancer. If we assume there is some gene that both increase your chance of having cancer and making tar get stuck in your lungs easier (just making things up here) this would break the third condition.

6.6

- $X \to Y \to Z$ would imply that night light (X) leads to parent myopia (Y) which does not really work with the order in which things occur. Normally parent myopia is developed before you have kids and use night lights for them.
- $X \leftarrow Y \leftarrow Z$ would imply that child myopia (Z) leads to parent myopia (Y) which is hard to imagine why that would be.