

SCI 2025: Homework 4

Setup

The first three homework problems this week are from Chapter 6 of the textbook.

Question 1: 6M1

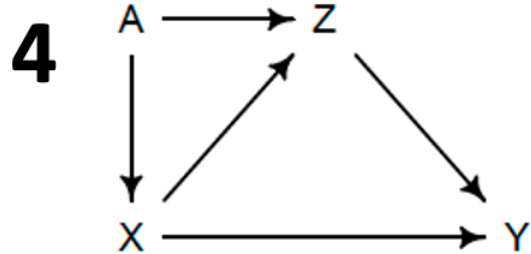
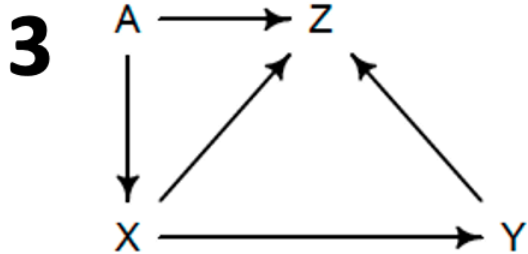
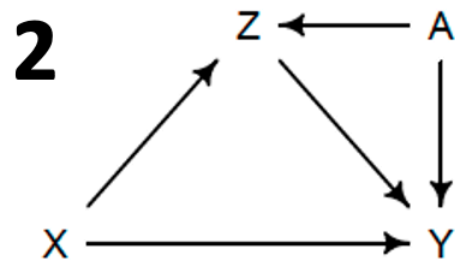
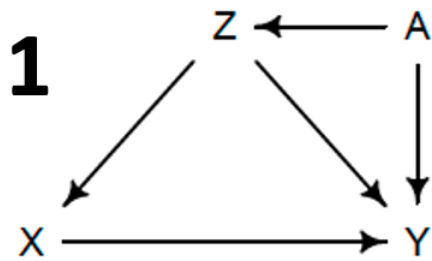
Modify the DAG on page 186 to include the variable V , an unobserved cause of C and Y : $C \leftarrow V \rightarrow Y$. Reanalyze the DAG. How many paths connect X to Y ? Which must be closed? Which variables should you condition on now?

Question 2: 6M2

Sometimes, in order to avoid multicollinearity, people inspect pairwise correlations among predictors before including them in a model. This is a bad procedure, because what matters is the conditional association, not the association before the variables are included in the model. To highlight this, consider the DAG $X \rightarrow Z \rightarrow Y$. Simulate data from this DAG so that the correlation between X and Z is very large. Then include both in a model predicting Y . Do you observe any multicollinearity? Why or why not? What is different from the legs example in the chapter?

Question 3: 6M3

Learning to analyze DAGs requires practice. For each of the four DAGs below, state which variables, if any, you must adjust for (condition on) to estimate the total causal influence of X on Y .



Question 4

Can you think of potential examples of collider bias from your own field or a related literature? Do you think the bias is likely to be positive or negative (with respect to the causal effect of interest)? Positive bias would mean that the estimated relationship between the focal variable and the outcome is too strong, while negative bias would mean that the estimated relationship is too weak.