EXERCISES ARE FROM ARVID SJOLANDER: https://staff.ki.se/people/arvid-sjolander

1 Exercises

1.1 Association vs causation

1. Table 1 shows the data collected in a study of 12 individuals. The goal was to estimate the effect of a treatment ($A = 1$) versus no treatment ($A = 0$) on the risk of fatality ($Y = 1$). The table also shows the values of the potential outcomes that would have been observed under treatment (Y_1) and under no treatment (Y_0).

ID	A	Y	Y_1	Y_0
1	0	0	0	0
2	0	0	0	0
3	0	1	0	1
4	0	1	0	1
5	0	1	0	1
6	1	1	1	1
7	1	1	1	1
8	1	0	0	1
9	1	0	0	0
10	1	1	1	0
11	1	1	1	1
12	1	1	1	1

Table 1: Potential outcome data

1. Consider the data in Table 1. Are treated ($A = 1$) and untreated ($A = 0$) exchangeable?

No, because P(Y1=1|A=1) != P(Y1=1|A=0)

2. Suppose that you are given the first three columns of Table 1, i.e. ID, A, and Y, but not the potential outcomes Y_0 and Y_1.

Given the first three columns, which are the possible values for the causal effect? Given no data at all, which are the possible values for the causal effect?

The causal effect is maximized when P(Y1=1) is as large as possible and P(Y0=1) is as small as possible. P(Y1=1) is as large as possible when all of those with A = 0 are assigned Y1 = 1. P(Y0=1) is as small as possible when all of those with A = 1 are assigned Y0 = 0.

This makes P(Y1) = 5/6 and P(Y0) = 1/4

So the maximum causal effect is 5/6 - 1/4 (or 7/12)

Directed Acyclic Graphs

3. (*Intrumental variables*). Many observational studies suffer from confounding. In this exercise we investigate a method of 'confounding adjustment' which, under certain assumptions, has the remarkable property of producing causal inference even in the presence of unmeasured confounding. Let A be the exposure of interest, let Y be the outcome of interest, and let U be all unmeasured variables (confounders) that affect both A and Y. Let Z be a measured variable which has the following properties: a) U does not affect Z, b) Z does not affect U, c) Z and U don't have common causes, d) Z affects A, e) Z has no effect on Y, apart from an indirect effect mediated through A. A variable Z which has properties a)-e) is called an *instrumental variable*.

   (a) Draw a DAG that connects A, Y, U, and Z.

$$\begin{tikzcd} & & U \arrow[dl] \arrow[dr] & \\ Z \arrow[r] & A \arrow[rr] & & Y \end{tikzcd}$$

   (b) Show that an observed association between Z and Y implies that A has a causal effect on Y (that is, we can test whether A has a causal effect on Y by testing whether Z an Y are associated).

   A: If the arrow from A to Y is missing, then there would be no association between Z and Y, since the path $Z \rightarrow A \leftarrow U \rightarrow Y$ is blocked at A. Hence, if we observe an association between Z and Y, then we can say that the arrow from A to Y exists.