MATH464/564: Practical 2b: Causal diagrams and conditioning – Solutions

Tom Palmer

19/20

Introduction

In this practical you need to try and draw a DAG for 3 scenarios.

We will use the same data for each scenario. The data is given in the table below.

	D=1			D=0	
	E*=1	E*=0		E*=1	E*=0
E=1	180	200	E=1	600	200
E=0	20	200	E=0	200	600

The data is in the file `confounding-dags-practical.dta`. The `.dta` extension means it is in Stata format, and so read it into R using the code below.

```
library(foreign)
dat <- read.dta("confounding-dags-practical.dta")
```

Question 1

First perform some estimation so that you know the associations between the 3 variables E, E^*, and D.

a) In R recreate the table above. You may need to produce two smaller tables to do this.

b) What is the marginal odds ratio for the effect of E on D (i.e. use D as the outcome/dependent variable and E as the covariate)?

c) What is the conditional odds ratio for the association between E and D adjusting for E^*?

d) What is the marginal odds ratio for the association between E^* and D?

Question 2: main task

Next you are given 3 scenarios from which the data could have been obtained. For each scenario we wish to estimate the effect of E on D.

- Draw a DAG for each scenario.
- Once you have drawn your DAG check that it conforms to the conditional independencies which you can estimate in the data.
- What model would you fit to estimate the effect of E on D in each scenario?

Scenario 1

- The data come from a case-control study.
- The aetiological question of interest is whether exposure to a particular nonsteroidal anti-inflammatory drug during the first trimester of pregnancy causes a congenital defect (D) arising in the second trimester.
- $D = 1$ for cases, $D = 0$ for controls without the defect.
- The sampling fraction for controls is unknown.
- E^* is use of the drug of interest during the first trimester, as self-reported by the mother 1 month postpartum.
- E is use of the drug of interest as recorded in comprehensive, accurate medical records of 1st trimester medications.
- You can ignore including any other possible confounders or other drug exposures.

Scenario 2

- The data come from a prospective cohort study.
- D is all-cause mortality in a cohort of healthy male miners, all aged 25, all of whom worked underground in a variety of different mine shafts for 6 months in 1967.
- 40 year follow-up is complete. The aetiologic question is whether pulmonary exposure to doses of radon above a certain level causes increased mortality.
- For each miner, the air level of radon in his mine was measured (E^*).
- A subject's actual exposure depends on the level of radon in the mine and the physical demands of the job and this was measured by lung dosimetry (E: 0 = below threshold of interest, 1= above).
- It is known that 6 months of physical exertion at age 25 has no independent effect on subsequent mortality.

Scenario 3

- The data come from a randomized controlled trial.
- D is death over a 15 year period.
- Study subjects were randomly assigned to an educational intervention to encourage them to eat a low fat diet ($E^* = 1$ for intervention, $E^* = 0$ for control).
- Investigators subsequently measured diet accurately in all trial participants ($E = 1$ for low fat diet, $E = 0$ for non-low fat diet).
- Assume the intervention has no effect on D other than through its effect on actual fat consumption E.