6.10 Practical 6

Datasets required
kidney_example.csv
simdata_binary.dta
simdata_binary2.dta

Introduction

This practical is in three parts. In Part A we return the the kidney stones example used in session 5. Parts B and C use simulated data.

Aims

1 To explore relationships between marginal and conditional treatment effect estimates, considering both collapsible and non-collapsible quantities.

2 To gain further practice at interpreting marginal and conditional treatment effect estimates.

3 To extend the scenarios considered in the notes to include a continuous treatment-outcome confounder.

Part A: Kidney stones example

In session 5 and practical 5 we estimated the marginal and conditional treatment effects (of X on Y) using risk differences, risk ratios and odds ratios. See Tables 5.2 and 5.3.

1 In the notes we saw that the marginal risk difference is equal to a weighted average of the conditional risk differences (conditional on stone size Z). What are the weights? Show that the marginal risk difference is equal to a weighted average of the conditional risk differences.

2 In the notes we saw that the marginal risk ratio is equal to a weighted average of the conditional risk ratios (Conditional on stone size Z). What are the weights? Show that the marginal risk ratio is equal to a weighted average of the conditional risk ratios.

3 Can the marginal odds ratio be expressed as a weighted average of the conditional odds ratios? Explain your answer.

Discuss: What are the interpretations of the marginal and conditional risk differences, risk ratios and odds ratios.

Discuss: For a patient with a large kidney stone, which treatment effect estimate(s) are most relevant to inform their treatment?

Part B: A simulated example with binary outcome

In this part we some simulated data (simdata_binary.dta) on a binary treatment X, binary outcome Y and binary covariate Z. The covariate Z is known to be measured temporally prior to X, meaning that we can be sure it does not lie on the causal pathway between X and Y.

4 Quantify the relationship between X and Z.

5 Using logistic regression (using a saturated model) estimate the conditional probabilities $\Pr(Y = 1|X = x, Z = z)$ for $x = 0, 1$ and $z = 0, 1$, and hence the conditional risk differences

$$\Pr(Y = 1|X = 1, Z = z) - \Pr(Y = 1|X = 0, Z = z), \qquad z = 0, 1$$

6 Without performing any further calculations, what do you expect the marginal risk difference to be?

7 From your logistic regression in question 5, what are the estimates of the conditional odds ratios

$$\frac{\Pr(Y = 1|do(X = 1), Z = z)/\Pr(Y = 0|do(X = 1), Z = z)}{\Pr(Y = 1|do(X = 0), Z = z)/\Pr(Y = 0|do(X = 0), Z = z)}, \qquad z = 0, 1$$

8 Fit another logistic regression to obtain an estimate of the marginal odds ratio

$$\frac{\Pr(Y = 1|do(X = 1))/\Pr(Y = 0|do(X = 1))}{\Pr(Y = 1|do(X = 0))/\Pr(Y = 0|do(X = 0))}$$

What assumption do you make? Explain the difference between this estimate and your estimates in question 7.

Part C: A simulated example with binary outcome and continuous confounder

In this part we some simulated data (simdata_binary2.dta) on a binary treatment X, binary outcome Y and continuous variable Z, where Z confounds the association between X and Y.

9 Fit a logistic regression of Y on X and Z and their interaction. What is the conditional odds ratio for an individual with (a) Z equal to its median value in the data, (b) Z equal to its 10th percentile in the data, (c) Z equal to its 90th percentile in the data.

10 Using the empirical standardization method introduced in session 5, obtain an estimate of the effect of X on Y using a marginal odds ratio.

11 Compare this with the odds ratio from a regression of Y on X alone. What is the reason for the difference between this estimate and that in question 10?

Discuss: What are the interpretations of the marginal and conditional odds ratios?