## 5.10 Practical 5

| Datasets required |
|---|
| kidney_example.csv |
| kidney_example_continuousY.csv |

### Introduction

This practical is in two parts. Part A uses the kidney stones example data from the Notes, in which the outcome $Y$ is binary. Part B considers data on a similar study but with a continuous outcome $Y$. The other variables are the same: $X$ denotes a binary treatment and $Z$ a binary confounder. Both data sets include 700 individuals.

| Variable | Description |
|---|---|
| `y` | Binary: 0 = success; 1 = failure |
| | Continuous: a score on a quality of life questionnaire |
| `x` | Treatment. 0 = Surgery; 1 = Lithotripsy |
| `z` | Size of kidney stone. 0 = small; 1 = large |

### Aims

1 To estimate conditional treatment effects for both binary and continuous outcomes.

2 To estimate marginal treatment effects for both binary and continuous outcomes using standardization.

3 To interpret estimates of marginal and conditional treatment effects.

### Part A: Binary outcome

The data for this section are in the file "kidney_example.csv". You can load the dataset into Stata using: `import delimited kidney_example.csv , clear`

The data are summarised in the table below.

| | $(X = 0)$ | $(X = 1)$ | Total |
|---|---|---|---|
| Success $(Y = 0)$ | 273 | 289 | 562 |
| Failure$(Y = 1)$ | 77 | 61 | 138 |
| Total | 350 | 350 | 700 |

| | Small stone size $(Z = 0)$ | | Large stone size $(Z = 1)$ | |
|---|---|---|---|---|
| | Surgery $(X = 0)$ | Lithotripsy $(X = 1)$ | Surgery $(X = 0)$ | Lithotripsy $(X = 1)$ |
| Success $(Y = 0)$ | 81 | 234 | 192 | 55 |
| Failure$(Y = 1)$ | 6 | 36 | 71 | 25 |
| Total | 87 | 270 | 263 | 80 |

1 In this question we will use logistic regression to obtain an estimate of the conditional odds ratio

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

for $Z = 0, 1$.

i) First using a logistic regression model for $Y$ with $X$ and $Z$ as explanatory variables (not including their interaction).

ii) Using a logistic regression model for $Y$ with $X$, $Z$ and their interaction $X \times Z$ as explanatory variables. You might find the `lincom` command useful for this.

2 Using your model from Question 1(ii) obtain estimates of:

i) The conditional probabilities $\Pr(Y = 1|do(X = 1), Z = z)$ and $\Pr(Y = 1|do(X = 0), Z = z)$ for $Z = 0, 1$. After fitting the logistic regression model the 'predict' command can be used to obtain estimates of the $\Pr(Y = 1|X = x, Z = z)$ components.

```
predict ypred
```

The estimates of $\Pr(Y = 1|X = 0, Z = 0)$ (for example) can be seen using

```
list ypred if x==0 & z==0
```

ii) The conditional risk differences

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

iii) The conditional risk ratios

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

**Discuss: Compare your results from the logistic regression models with those given in Table 5.2 in the Notes (which were not obtained using regression). Interpret your conditional treatment effect estimates obtained in questions 1 and 2. What assumption do we make to interpret these estimates 'causally'.**

3 We will next estimate marginal treatment effects.

i) Using the results from your model in Question 1(ii), and information about $\Pr(Z = z)$, use the standardization formula in (5.8) to estimate $\Pr(Y = 1|do(X = 1))$ and $\Pr(Y = 1|do(X = 0))$. The standardization formula is:

$$\Pr(Y = 1|do(X = x)) = \sum_{z=0,1} \Pr(Y = 1|X = x, Z = z) \Pr(Z = z)$$

Note that the probabilities of having small and large kidney stone size are $\Pr(Z = 0) = 0.51$ and $\Pr(Z = 1) = 0.49$ (check you agree!).

You may find it useful to use the table below to keep track of your estimates.

ii) Using your results from (i) obtain estimates of the marginal risk difference, marginal risk ratio, and marginal odds ratio, as defined in equations (5.1)-(5.3) in the Notes.

| | | |
|---|---|---|
| $\Pr(Y=1\|do(X=1))$ | $\Pr(Y=1\|X=1,Z=0)$ | |
| | $\Pr(Y=1\|X=1,Z=1)$ | |
| | $\Pr(Y=1\|X=1,Z=0)\times\Pr(Z=0)$ | |
| | $\Pr(Y=1\|X=1,Z=1)\times\Pr(Z=1)$ | |
| | $\Pr(Y=1\|X=1,Z=0)\times\Pr(Z=0)$ $+\Pr(Y=1\|X=1,Z=1)\times\Pr(Z=1)$ | |
| $\Pr(Y=1\|do(X=0))$ | $\Pr(Y=1\|X=0,Z=0)$ | |
| | $\Pr(Y=1\|X=0,Z=1)$ | |
| | $\Pr(Y=1\|X=0,Z=0)\times\Pr(Z=0)$ | |
| | $\Pr(Y=1\|X=0,Z=1)\times\Pr(Z=1)$ | |
| | $\Pr(Y=1\|X=0,Z=0)\times\Pr(Z=0)$ $+\Pr(Y=1\|X=0,Z=1)\times\Pr(Z=1)$ | |

**Discuss: What are the interpretations of the causal marginal treatment effects you have estimated? Compare these with the conditional treatment effects estimated in questions 1 and 2.**

4 We can also use the 'margins' postestimation command in Stata to get to the treatment effect estimates more directly.

i) After running the logistic regression model of $Y$ on $X, Z, X \times Z$ run the following command
`margins x`

This provides estimates of $\Pr(Y=1|do(X=1))$ and $\Pr(Y=1|do(X=0))$ using the same standardization procedure that you used above.

ii) Estimates of the marginal risk difference can be obtained using

`margins x, dydx(x)`

iii) The results from the 'margins' command provide an estimate of the 95% confidence interval for the marginal risk difference. Interpret the point estimate and its 95% confidence interval.

5 Lastly, we will use the empirical standardization method to estimate the marginal probabilities $\Pr(Y=1|do(X=1))$ and $\Pr(Y=1|do(X=0))$. This was outlined in section 5.6 of the Notes (see equation (5.15)). The Stata code for this is a bit tricky so please follow the steps outlined in the accompanying Stata Do file, and make sure you understand what is being done in each step. Check that you get the same estimates as found in Questions 3 and 4.

EXTRA: Extra Stata code is provided to show how we can obtain bootstrap estimates of 95% confidence intervals for treatment effect estimates. Those who have done the Robust Methods module may wish to look through this as an extra exercise.

## Part B: Continuous outcome

In the second part of this practical we will use a modified version of the kidney study data with a continuous outcome $Y$ instead of a binary outcome. For example, $Y$ could be a score on a quality of life questionnaire. The data are available in the file kidney_example_continuousY.csv. Load the data.

6 What are the means of $Y$ in each of the four groups defined by $X$ and $Z$? That is, what are $E(Y|X=x,Z=z)$ for $x,z=0,1$ ?

7 Fit a linear regression of $Y$ on $X$, $Z$ and their interaction $X \times Z$. How do the coefficients relate to the means of each group?

8 Use your results from question 4 and 5 to estimate the conditional effect of $X$ on $Y$ given $Z$ for $Z = 0, 1$, using a conditional mean difference:

$E(Y|do(X = 1), Z = 0) - E(Y|do(X = 0), Z = 0)$
$E(Y|do(X = 1), Z = 1) - E(Y|do(X = 0), Z = 1)$

**Discuss: Interpret the estimates. What assumption do you make to interpret the conditional mean differences as causal treatment effect estimates?**

9 We now estimate the causal marginal treatment effect.

i) By hand, use standardization to complete the table below, and hence to estimate the causal marginal treatment effect:

$$E(Y|do(X = 1)) - E(Y|do(X = 0))$$

Reminder, by standardization $E(Y|do(X = x))$ is given by

$$E\left(Y|X = x, Z = 0\right) \times Pr\left(Z = 0\right) + E\left(Y|X = x, Z = 1\right) \times Pr\left(Z = 1\right)$$

<table>
<tr><td rowspan="4">$E(Y|do(X = 1))$</td><td>$E(Y|X = 1, Z = 0)$</td><td></td></tr>
<tr><td>$\Pr(Z = 0)$</td><td>0.51</td></tr>
<tr><td>$E(Y|X = 1, Z = 1)$</td><td></td></tr>
<tr><td>$\Pr(Z = 1)$</td><td>0.49</td></tr>
<tr><td colspan="2">$E(Y|do(X = 1))$</td><td></td></tr>
<tr><td rowspan="4">$E(Y|do(X = 0))$</td><td>$E(Y|X = 0, Z = 0)$</td><td></td></tr>
<tr><td>$\Pr(Z = 0)$</td><td>0.51</td></tr>
<tr><td>$E(Y|X = 0, Z = 1)$</td><td></td></tr>
<tr><td>$\Pr(Z = 1)$</td><td>0.49</td></tr>
<tr><td colspan="2">$E(Y|do(X = 0))$</td><td></td></tr>
<tr><td colspan="2">$E(Y|do(X = 1)) - E(Y|do(X = 0))$</td><td></td></tr>
</table>

ii) You can also obtain the same estimate using the 'margins' command after fitting your linear regression model as follows:

```
regress y i.x##i.z

margins x, dydx(x)
```

10 Use the empirical standardization approach described in the notes (section 5.6, equation 5.15), and used above in Question 5, to estimate the marginal treatment effect. You can do this by modifying the Stata code used for Question 5.

**Discuss: Interpret the marginal treatment effect estimate.**

11 Fit a linear regression of Y on X and Z, without an interaction term between X and Z. What is the conditional mean difference $E(Y|do(X = 1), Z = z) - E(Y|do(X = 0), Z = z)$ for $Z = 0, 1$? Compare this with the result from using

```
regress y i.x i.z
margins x, dydx(x)
```