GLM Practical 5: Solutions

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).

The model being fitted is of the form

$$\Pr(Y=1|X=x,Z=z)=\frac{e^{\beta_0+\beta_X x+\beta_Z z}}{1+e^{\beta_0+\beta_X x+\beta_Z z}}.$$

The conditional OR is e^{β_X}, which is the same for $Z=0,1$ under this model. The estimate of this is $e^{\hat{\beta}_X}=1.42$ (95% CI (0.91,2.24))

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.

The model being fitted is of the form

$$\Pr(Y=1|X=x,Z=z)=\frac{e^{\beta_0+\beta_X x+\beta_Z z+\beta_{XZ}xz}}{1+e^{\beta_0+\beta_X x+\beta_Z z+\beta_{XZ}xz}}.$$

The conditional OR given $Z=0$ is e^{β_X} and the conditional OR given $Z=1$ is $e^{\beta_X+\beta_{XZ}}$. The estimates are $e^{\hat{\beta}_X}=2.08$ (95% CI (0.84,5.11) $e^{\hat{\beta}_X+\hat{\beta}_{XZ}}=1.23$ (95% CI (0.71,2.12))

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
```

   The estimates of the conditional probabilities are:
   $\widehat{\Pr}(Y = 1|do(X = 0), Z = 0) = 0.069$
   $\widehat{\Pr}(Y = 1|do(X = 0), Z = 1) = 0.270$
   $\widehat{\Pr}(Y = 1|do(X = 1), Z = 0) = 0.133$
   $\widehat{\Pr}(Y = 1|do(X = 1), Z = 1) = 0.313$

   ii) The conditional risk differences

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

   The estimates of the conditional risk differences are:
   $\widehat{\Pr}(Y = 1|do(X = 1), Z = 0) - \widehat{\Pr}(Y = 1|do(X = 0), Z = 0) = 0.064$
   $\widehat{\Pr}(Y = 1|do(X = 1), Z = 1) - \widehat{\Pr}(Y = 1|do(X = 0), Z = 1) = 0.043$

   iii) The conditional risk ratios

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

   The estimates of the conditional risk ratios are:
   $\widehat{\Pr}(Y = 1|do(X = 1), Z = 0)/\widehat{\Pr}(Y = 1|do(X = 0), Z = 0) = 1.93$
   $\widehat{\Pr}(Y = 1|do(X = 1), Z = 1)/\widehat{\Pr}(Y = 1|do(X = 0), Z = 1) = 1.16$

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'.

The results from the logistic regression model including the interaction between X and Z give the same conditional OR estimates to those seen in the notes(except for some rounding errors in the results given in the notes!). The interpretations of the conditional estimates are as follows.

Conditional ORs: In those with small stone size ($Z = 0$) the odds of a bad outcome ($Y = 1$) under lithotripsy are estimated to be 2.08 times the odds of a bad outcome under surgery. In those with large stone size ($Z = 1$) the odds of a bad outcome ($Y = 1$) under lithotripsy are estimated to be 1.23 times the odds of a bad outcome under surgery. However, the 95% CIs for both estimates include 1. There is in fact no evidence suggest that Z modifies the effect of treatment on outcome, as the OR for the interaction term is 0.59, with 95% CI (0.21,1.70), and corresponding p-value 0.329.

Conditional risk differences: In those with small stone size ($Z = 0$) the absolute risk of a bad outcome ($Y = 1$) under lithotripsy is estimated to 0.064 higher than the risk under surgery. In those with large stone size ($Z = 1$) the absolute risk of a bad outcome ($Y = 1$) under lithotripsy is estimated to 0.043 higher than the risk under surgery.

Conditional risk ratios: In those with small stone size ($Z = 0$) the risk of a bad outcome ($Y = 1$) under lithotripsy is estimated to be 1.93 times the risk of a bad outcome under surgery. In those with large stone size ($Z = 1$) the risk of a bad outcome ($Y = 1$) under lithotripsy are estimated to be 1.16 times the risk of a bad outcome under surgery.

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.

   See estimates in the table below.

   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.

   The marginal risk difference estimate is:
   $\widehat{\Pr}(Y = 1|do(X = 1)) - \widehat{\Pr}(Y = 1|do(X = 0)) = 0.054$

   The marginal risk ratio estimate is:
   $\frac{\widehat{\Pr}(Y=1|do(X=1))}{\widehat{\Pr}(Y=1|do(X=0))} = 1.32$

The marginal odds ratio estimate is:
$\frac{\widehat{\Pr}(Y=1|do(X=1))/\widehat{\Pr}(Y=0|do(X=1))}{\widehat{\Pr}(Y=1|do(X=0))/\widehat{\Pr}(Y=0|do(X=0))} = 1.41$

$\Pr(Y=1|do(X=1))$	$\Pr(Y=1|X=1,Z=0)$	0.133
	$\Pr(Y=1|X=1,Z=1)$	0.313
	$\Pr(Y=1|X=1,Z=0)\times\Pr(Z=0)$	0.068
	$\Pr(Y=1|X=1,Z=1)\times\Pr(Z=1)$	0.153
	$\Pr(Y=1|X=1,Z=0)\times\Pr(Z=0)$ $+\Pr(Y=1|X=1,Z=1)\times\Pr(Z=1)$	0.221
$\Pr(Y=1|do(X=0))$	$\Pr(Y=1|X=0,Z=0)$	0.069
	$\Pr(Y=1|X=0,Z=1)$	0.270
	$\Pr(Y=1|X=0,Z=0)\times\Pr(Z=0)$	0.035
	$\Pr(Y=1|X=0,Z=1)\times\Pr(Z=1)$	0.132
	$\Pr(Y=1|X=0,Z=0)\times\Pr(Z=0)$ $+\Pr(Y=1|X=0,Z=1)\times\Pr(Z=1)$	0.167

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.

Marginal risk difference: If everyone had received lithotripsy we would expect the risk of a bad outcome to be 0.054 higher than if everyone had received surgery.

Marginal risk ratio: If everyone had received lithotripsy we would expect the risk of a bad outcome to be 1.32 times higher (i.e. 32% higher) than if everyone had received surgery.

Marginal odds ratio: If everyone had received lithotripsy we would expect the odds of a bad outcome to be 1.41 times higher (i.e. 41% higher) than if everyone had received surgery.

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.

See Stata do file for i) and ii). Using ‘margins’ gives the marginal risk difference estimate of 0.054, with 95% CI (-0.012,0.119) and p-value 0.110. The interpretation of the estimate is as above. The 95% CI includes 0, so there is no evidence at the 5% level against the null hypothesis that there is no difference in the risk of a bad outcome had everyone received lithotripsy versus had everyone received surgery.

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.

   See Stata do file.

   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 $x, z = 0, 1$?

   The estimated expectations are:
   $\widehat{E}(Y|X = 0, Z = 0) = 0.201$
   $\widehat{E}(Y|X = 0, Z = 1) = -0.848$
   $\widehat{E}(Y|X = 1, Z = 0) = 0.557$
   $\widehat{E}(Y|X = 1, Z = 1) = 0.130$

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?

   The linear regression model is of the form:

   $$E(Y|X = x, Z = z) = \beta_0 + \beta_X x + \beta_Z z + \beta_{XZ} xz.$$

   The above expectations relate to the model parameters as follows:
   $E(Y|X = 0, Z = 0) = \beta_0$
   $E(Y|X = 0, Z = 1) = \beta_0 + \beta_Z$
   $E(Y|X = 1, Z = 0) = \beta_0 + \beta_X$
   $E(Y|X = 1, Z = 1) = \beta_0 + \beta_X + \beta_Z + \beta_{XZ}$

8. Use your results from question 6 and 7 to estimate the conditional effect of X on Y given Z for $Z = 0, 1$, using a conditional mean difference:

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

   The conditional mean difference estimates are

$$\begin{aligned}&\widehat{E}(Y|do(X = 1), Z = 0) - \widehat{E}(Y|do(X = 0), Z = 0)\\&\quad= \widehat{E}(Y|X = 1, Z = 0) - \widehat{E}(Y|X = 0, Z = 0)\\&\quad= \hat{\beta}_X = 0.355\end{aligned}$$

   and

$$\begin{aligned}&\widehat{E}(Y|do(X = 1), Z = 1) - \widehat{E}(Y|do(X = 0), Z = 1)\\&\quad= \widehat{E}(Y|X = 1, Z = 1) - \widehat{E}(Y|X = 0, Z = 1)\\&\quad= \hat{\beta}_X + \hat{\beta}_{XZ} = 0.978\end{aligned}$$

   with 95% CIs (-0.375,1.086) and (0.221,1.735) respectively.

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

   In those with $Z = 0$, if everyone received lithotripsy the expected outcome would be 0.355 higher than if everyone had received surgery. In those with $Z = 1$, if everyone received lithotripsy the expected outcome would be 0.978 higher than if everyone had received surgery.

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

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

   See table below.

$E(Y\mid do(X=1))$	$E(Y\mid X=1, Z=0)$	0.557
	$\Pr(Z=0)$	0.51
	$E(Y\mid X=1, Z=1)$	0.130
	$\Pr(Z=1)$	0.49
$E(Y\mid do(X=1))$		0.347
$E(Y\mid do(X=0))$	$E(Y\mid X=0, Z=0)$	0.201
	$\Pr(Z=0)$	0.51
	$E(Y\mid X=0, Z=1)$	-0.848
	$\Pr(Z=1)$	0.49
$E(Y\mid do(X=0))$		-0.313
$E(Y\mid do(X=1)) - E(Y\mid do(X=0))$		0.661

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)
```

See Stata do file

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.

    See Stata do file

    Discuss: Interpret the marginal treatment effect estimate.

    If everyone had received lithotripsy we would expect the mean outcome to be 0.661 higher than had everyone received surgery.

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)
```

    Fitting a linear regression of Y on X and Z without the interaction gives an estimated conditional mean difference of 0.656, with 95% CI (0.130,1.182). This is the expected mean difference in both the $Z = 0$ and $Z = 1$ groups, under the assumption of no interaction between X and Z. The interpretation is that if everyone in the $Z = 0$ group had received lithotripsy we would expect the mean outcome to be 0.656 higher than had everyone in the $Z = 0$ group received surgery. Similarly, if everyone in the $Z = 1$ group had received lithotripsy we would expect the mean outcome to be 0.656 higher than had everyone in the $Z = 1$ group received surgery. It was not asked for in the question, but in fact there is no evidence that Z modifies the effect of X on Y.

Based on this model, the estimated marginal mean difference is also 0.656. The interpretation is that if everyone had received lithotripsy we would expect the mean outcome to be 0.656 higher than had everyone received surgery.