Modelling Rates

26/12/2023

Contents

1 Modelling Counts **5**

1.1 Introduction . . . 6

1.2 Poisson Regression . . . 6

1.2.1 Introduction . . . 6

1.2.2 Example . . . 7

1.2.3 Constraints . . . 9

1.2.4 Other considerations . . . 10

1.3 Negative Binomial Regression . . . 10

1.3.1 Overdispersion Example . . . 11

2 Modelling Counts: Practical **13**

2.1 Practical For Session 9: Counts . . . 14

2.1.1 Poisson Regression . . . 14

2.1.2 Negative Binomial Regression . . . 15

2.1.3 Using constraints . . . 16

2.1.4 Constraints in Multinomial Logistic Regression . . . 18

1 Modelling Counts

1.1 Introduction

We can use logistic regression to model the prevalence of a condition, i.e. the proportion of people who have that condition. However, we have not yet a way to model *incidence*, i.e. the *rate* at which new cases are occurring.

Incidence is not measured as a proportion, but as a rate: the number of events that happen over a fixed amount of time. If events are happening at a fixed rate λ over a time T, then the expected number of events to occur is λT. The *observed* number of events will follow a Poisson distribution with parameter λT.

1.2 Poisson Regression

1.2.1 Introduction

Poisson regression models the rate at which events occur as a function of the covariates. A rate can never be negative (that is, the number of events that have occurred can never decrease, events can't "unhappen". We commonly model the logarithm of the rate, since as this value goes from $-\infty$ to ∞, the rate goes from 0 to ∞.

So the expected number of events, C for a given observation is

$$E[C] = \lambda T$$

where

C is the number of events

λ is the rate at which events happen

T is the duration of followup for that observation.

So if we model $log(\lambda)$ as a linear function of our covariates, we get

$$\begin{aligned}
\log(\hat{\lambda}) &= \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \\
\hat{\lambda} &= e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p} \\
E[C] &= T\lambda \\
&= T \times e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p} \\
&= e^{\log(T) + \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p} \\
log(E[C]) &= \log(T) + \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p
\end{aligned}$$

So if we are going to model the rate, but what we observe is the number of events, we need to include the log of the exposure time, *with a coefficient fixed as 1*, in our linear predictor. This is referred to as the *offset*.

Since we are modelling the log of the rate, an increase of 1 in a x_p corresponds to and increase of β_p in the log rate. This in turn corresponds to *multiplying* the rate by e^{β_p}. So just like logistic regression, where e^{β} is more meaningful than β itself, so with Poisson regression. In this case, e^{β} is a *Rate Ratio*.

For each observation in our dataset, we have an observed number of events C, and an expected number of events $e^{\log(T)+\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}$. We can calculated a χ squared statistic to test whethere the observed values are further from the expected values that we would expect by chance if the expected values are modelled correctly. This statistic will follow a χ^2 distribution on $N - p - 1$ degrees of freedom, where N is our sample size and p is the number of covariates in our model.

Poisson Regression in Stata The basic command for performing Poisson regression in stata is `poisson`. The first variable after the command will be the outcome variable, any subsequent variables will be predictors. You will also almost always want to include an `exposure(`*`varname`*`)` option, where *`varname`* is the name of a variable giving te duration of exposure for each observation. In order to get Rate Ratios rather than coefficients in the output, use the option `irr` (short for Incidence Rate Ratio).

If you use the `predict` command after a Poisson regression, the following options are available:

`n`	(default)	expected number of events (rate × duration of exposure)
`ir`		incidence rate
`xb`		linear predictor, log of the incidence rate

1.2.2 Example

The data in Table 1.1 shows the mortality by age-group and smoking status for a cohort of British male doctors. The study was set up in 1951, which explains why nearly 80% of the exposure time was in the smokers.

	Smokers		Non-smokers	
Age	Deaths	Person-Years	Deaths	Person-Years
35–44	32	52,407	2	18,790
45–54	104	43,248	12	10,673
55–64	206	28,612	28	5,710
65–74	186	12,663	28	2,585
75–84	102	5,317	31	1,462

Table 1.1: Mortality by Age-Group and Smoking Status among Male British Doctors

If we had a stata dataset containing 10 observations, with a variable `agecat` containing the age group and `smokes` containing the smoking status for that observation, `pyears` containing the Person-Years of followup and `deaths` containing the number of deaths, then we could model that data with the command

```
poisson deaths i.agecat i.smokes, exp(pyears) irr
```

and get the following output:

```
Poisson regression                                Number of obs   =         10
                                                  LR chi2(5)      =     922.93
                                                  Prob > chi2     =     0.0000
Log likelihood = -33.600153                       Pseudo R2       =     0.9321

------------------------------------------------------------------------------
      deaths |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      agecat |
      45-54  |   4.410584   .8605197     7.61   0.000     3.009011    6.464997
      55-64  |    13.8392   2.542638    14.30   0.000     9.654328    19.83809
      65-74  |   28.51678   5.269878    18.13   0.000     19.85177    40.96395
      75-84  |   40.45121   7.775511    19.25   0.000     27.75326    58.95885
             |
      smokes |
        Yes  |   1.425519   .1530638     3.30   0.001     1.154984    1.759421
       _cons |   .0003636   .0000697   -41.30   0.000     .0002497    .0005296
  ln(pyears) |          1  (exposure)
------------------------------------------------------------------------------
```

This output shows that mortality increases with increasing age, and is nearly 43% higher in smokers than it is in non-smokers. However, if we check the goodness of fit using the command `estat gof`, we find that the fit is poor: the observed values are significantly further from the expected values than we would expect if the model were correct.

```
estat gof

         Deviance goodness-of-fit =  12.13244
         Prob > chi2(4)           =    0.0164

         Pearson goodness-of-fit  =  11.15533
         Prob > chi2(4)           =    0.0249
```

This lack of fit can happen if we have not specified the linear predictor correctly. It could be because we have modelled continuous variables incorrectly, for example assuming that the log of the rate increases linearly with the variable, when the increase is really quadratic. However, that cannot be the explanation in this instance, since we have not continuous variables in our model. With categorical variables, this happens when we are missing interaction terms. In our example, we are assuming that the rate ratio is the same for all age groups: if this is not the case, our model will not fit well.

We can use `predict` with the `n` option to get the expected number of events, and see how the expected and predicted numbers differ: these are shown in Table 1.2.

	Smokers		Non-smokers	
Age	Deaths	`pred_n`	Deaths	`pred_n`
35–44	32	27.2	2	6.8
45–54	104	98.9	12	17.1
55–64	206	205.3	28	28.7
65–74	186	187.2	28	26.8
75–84	102	111.5	31	21.5

Table 1.2: Expected and Observed Numbers of Deaths in Doctors Study

The expected numbers of deaths are lower than the observed numbers in smokers in the lowest age groups and in non-smokers in the highest age group. This suggests that the rate ratio is changing with age, and we need to incorporate that into our model. Including the interaction between age and smoking in our model gives the following output:

```
. poisson deaths i.agecat##i.smokes, exp(pyears) irr

Poisson regression                                Number of obs   =         10
                                                  LR chi2(9)      =     935.07
                                                  Prob > chi2     =     0.0000
Log likelihood =  -27.53397                       Pseudo R2       =     0.9444

------------------------------------------------------------------------------
      deaths |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      agecat |
      45-54  |    10.5631   8.067701     3.09   0.002     2.364153    47.19623
      55-64  |   46.07004   33.71981     5.23   0.000     10.97496    193.3901
      65-74  |    101.764   74.48361     6.32   0.000     24.24256    427.1789
      75-84  |   199.2099   145.3356     7.26   0.000     47.67693    832.3648
             |
      smokes |
        Yes  |   5.736637   4.181256     2.40   0.017     1.374811    23.93711
             |
agecat#smokes |
   45-54#Yes  |   .3728337   .2945619    -1.25   0.212     .0792525    1.753951
   55-64#Yes  |   .2559409   .1935392    -1.80   0.072     .0581396    1.126697
   65-74#Yes  |   .2363859   .1788334    -1.91   0.057     .0536612    1.041316
   75-84#Yes  |   .1577109   .1194146    -2.44   0.015     .0357565    .6956154
              |
        _cons |   .0001064   .0000753   -12.94   0.000     .0000266    .0004256
  ln(pyears) |          1  (exposure)
------------------------------------------------------------------------------
```

The rate ratio in the baseline 35-44 category is 5.74, much higher than the overall estimate of 1.43 we got previously. The rate ratios in the other age categories are much lower, although the actual values are not given directly in this output. The estimate in the 45-54 is 5.74 × 0.373 = 2.14, lower than the estimate in the youngest age group, but still higher than the overall estimate.

Rather than work out the rate ratios in the various age groups by hand, we can use the `lincom` command: that way we get confidence intervals and hypothesis tests as well. To get the rate ratio in the 75-84 age group, the stata command would be

```
lincom 1.smokes + 5.age#1.smokes, eform
```

and the corresponding outcome would be

```
( 1)  [deaths]1.smokes + [deaths]5.agecat#1.smokes = 0

------------------------------------------------------------------------------
      deaths |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .9047304   .1855513    -0.49   0.625     .6052658     1.35236
------------------------------------------------------------------------------
```

In this age-group, the rate ratio for smoking is slightly, but not significantly, less than 1.

1.2.3 Constraints

The last Poisson model we fitted, with different rate ratios for each age group, is called a "staturated" model. This means that there are as many parameters in the model (1 for smoking, 4 for age group, 4 for interactions between smoking and age group and 1 for the constant term) as there are observations in the dataset. This means that it is possible to fit the data perfectly, and the observed numbers of deaths in each group will be exactly equal to the expected numbers of deaths.

It also means that it is not possible to perform a goodness of fit test. The χ^2 statistic is 0, since the observed and expected values are all equal. And the number of degrees of freedom for the test is also 0, since we have 10 observations and 9 variables in the model.

However, the interaction terms for the 55–64 age group and the 65–74 age group look very similar. What would happen if we were to force them to be exactly the same ? That would reduce the number of parameters in our model and enable us to perform a goodness of fit test. It would also simplify the presentation of our model: we would only need to give 4 rate ratios for smoking, not 5. Simplifying the presentation of a model is a much more importand reason for using constraints than enabling a goodness of fit test.

Parameters may be constrained to either equal other paramaters, or to equal a particular value. The stata command to define a constraint is `constraint define` *`n parameter`* `=` *`expression`*, where `n` is an integer that will be used later to identify the constraint, *`parameter`* can be the name of a variable, or a way of identifying a particular level or combination of levels for categorical variables, and *`expression`* can be either another parameter, or a numerical value.

For example, the command to force the rate ratio for age 55–64 to be equal to the rate ratio for age 65–74 would be

```
constraint define 1 3.agecat#1.smokes = 4.agecat#1.smokes
```

We can than fit this constrained model to the data by using the `constraint()` option of the `poisson` command:

```
. poisson deaths i.agecat##i.smokes, exp(pyears) irr constr(1)
```

```
Poisson regression                                Number of obs   =         10
                                                  Wald chi2(8)    =     632.14
Log likelihood = -27.572645                       Prob > chi2     =     0.0000

 ( 1)  [deaths]3.agecat#1.smokes - [deaths]4.agecat#1.smokes = 0
------------------------------------------------------------------------------
      deaths |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      agecat |
      45-54  |    10.5631   8.067701     3.09   0.002     2.364153    47.19623
      55-64  |     47.671   34.37409     5.36   0.000     11.60056    195.8978
      65-74  |   98.22765   70.85012     6.36   0.000     23.89324    403.8244
      75-84  |   199.2099   145.3356     7.26   0.000     47.67693    832.3648
             |
      smokes |
        Yes  |   5.736637   4.181256     2.40   0.017     1.374811    23.93711
             |
agecat#smokes |
   45-54#Yes |   .3728337   .2945619    -1.25   0.212     .0792525    1.753951
   55-64#Yes |   .2461772    .182845    -1.89   0.059     .0574155    1.055521
   65-74#Yes |   .2461772    .182845    -1.89   0.059     .0574155    1.055521
   75-84#Yes |   .1577109   .1194146    -2.44   0.015     .0357565    .6956154
             |
       _cons |   .0001064   .0000753   -12.94   0.000     .0000266    .0004256
  ln(pyears) |          1  (exposure)
------------------------------------------------------------------------------
```

You will see that the table of coefficients now has two identical lines: the interaction term between age group and smoking is identical for the 55–64 and 65–74 age groups.

1.2.4 Other considerations

1.3 Negative Binomial Regression

Although Poisson regression can be very useful for modelling count variables, I would not recommend it's use in general. This is because the variance of the Poisson distribution is equal to its mean, but this is not the only kind of distribution that a count variable can follow. If you are modelling a count variable for which the variance is greater than its mean, the variable is said to be "overdispersed".

If you use Poisson regression for a variable that is overdispersed, the standard errors for the model parameters will be too small. This means that hypothesis tests will produce statistically significant results more than 5% of the time that the null hypothesis is true, and confidence intervals will be narrower than they should be. It is therefore essential to test for overdispersion before fitting a Poisson regression model.

Life is made easier by the fact that there is an alternative model for count data which specifically models the overdispersion. This is the negative binomial regression model. There are in fact two types of negative binomial regression model, which differ in the way that they model the overdispersion. They model the variance of the outcome variable Y as either $\mathrm{Var}(Y) = \mu(1+\delta)$ or $\mathrm{Var}(Y) = \mu(1+\alpha\mu)$. I.e. the overdispersion is either constant (first model) or proportional to the mean of Y (model 2). Both models reduce to the Poisson model if α or δ are 0. So by fitting one of these models, you not only test whether fitting a Poisson model would be appropriate, but you also fit it if it is.

The command for fitting negative binomial models in stata is `nbreg`. Almost all of the options, and commands that can be run after `nbreg` are the same as for the command `poisson`. The only difference is that the `nbreg`command has an `overdispersion()` option: by default it uses $\mathrm{Var}(Y) = \mu(1+\alpha\mu)$, but with the option `overdispersion(constant)` it uses $\mathrm{Var}(Y) = \mu(1+\delta)$.

1.3.1 Overdispersion Example

To see the difference between Poisson regression and negative binomial regression, consider the output below.

```
. poisson deaths i.cohort, exposure(exposure) irr

Poisson regression                                Number of obs   =         21
                                                  LR chi2(2)      =      49.16
                                                  Prob > chi2     =     0.0000
Log likelihood = -2159.5158                       Pseudo R2       =     0.0113

------------------------------------------------------------------------------
      deaths |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |
  1960-1967  |   .7393079   .0423859    -5.27   0.000     .6607305      .82723
  1968-1976  |   1.077037   .0635156     1.26   0.208      .959474    1.209005
             |
       _cons |   .0202523   .0008331   -94.80   0.000     .0186836    .0219527
ln(exposure) |          1  (exposure)
------------------------------------------------------------------------------

. nbreg deaths i.cohort, exposure(exposure) irr

Negative binomial regression                      Number of obs   =         21
                                                  LR chi2(2)      =       0.40
Dispersion     = mean                             Prob > chi2     =     0.8171
Log likelihood = -131.3799                        Pseudo R2       =     0.0015

------------------------------------------------------------------------------
      deaths |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |
  1960-1967  |   .7651995   .5537904    -0.37   0.712     .1852434    3.160869
  1968-1976  |   .6329298   .4580292    -0.63   0.527     .1532395    2.614209
             |
       _cons |   .1240922   .0635173    -4.08   0.000     .0455042    .3384052
ln(exposure) |          1  (exposure)
-------------+----------------------------------------------------------------
    /lnalpha |   .5939963   .2583615                       .087617    1.100376
-------------+----------------------------------------------------------------
       alpha |   1.811212   .4679475                       1.09157    3.005294
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) = 4056.27 Prob>=chibar2 = 0.000
```

The Poisson model suggests that the rate is significantly lower in the 1960–1967 cohort than in the baseline cohort. However, the negative binomial model shows that there is highly significant overdispersion (LR test of alpha = 0 has a P-value given as 0.000). Furthermore, there a no longer any significant differences between the cohorts once overdispersion is taken into account.

2 Modelling Counts: Practical

2.1 Practical For Session 9: Counts

Datasets

The datasets that you will use in this practical can be accessed via http from within stata. However, the directory in which they are residing has a very long name, so you can save yourself some typing if you create a global macro for this directory. You can do this by entering

```
global basedir http://personalpages.manchester.ac.uk/staff/mark.lunt
global datadir $basedir/stats/9_Counts/data
```

(In theory, the global variable `datadir` could have been set with a single command, but fitting the necessary command on the page would have been tricky. Far easier to use two separate commands as shown above). If you wish to run the practical on a computer without internet access, you would need to:

1. Obtain copies of the necessary datasets
2. Place them in a directory on your computer
3. Define the global macro `$datadir` to point to this directory.

2.1.1 Poisson Regression

In this section you will be analysing the dataset `$datadir/ships`. This is data from Lloyds of London concerning the rate at which damage occured at different times to different types of ship. There are 5 types of ship (labelled "A" to "E"), which could have been built in any one of 4 time periods, and sailed during one of two time periods. The aggregate duration of operation of each type of ship is given by `months`, and the number of incidents of damage is given by `damage`.

1.1 Familiarise yourself with the the meanings of each of the variables with the command

```
label list
```

Set the reference categories for type and time built to E and 1975-1979 respectively with the commands

```
fvset base 5 type
fvset base 4 built
```

1.2 Are there any differences in the rates at which damage occurs according to the type of ship ? The command to test this is

```
poisson damage i.type, exposure(months) irr
```

1.3 Are there any differences in the rates at which damage occurs according to the time at which the ship was built ? The command to test this is

```
poisson damage i.built, exposure(months) irr
```

1.4 Are there any differences in the rates at which damage occurs accoding to the time in which the ship was operated ? (You can work out this command for yourself).

1.5 Now add all three variables into a multivariate poisson model. Use

```
testparm i.type
```

to test if type is still significant after adjusting for the other predictors.

1.6 Use

```
predict pred_n
```

to obtain predicted numbers of damage incidents. Compare the observed and predicted numbers of incidents with

```
list type built sailed damage pred_n
```

For which type of ship and which time periods are predicted values furthest from the observed values ?

1.7 Use `estat gof` to test whether the model is adequate.

1.8 Add a term for the interaction between ship type and year of construction (`i.type#i.built`). Use `testparm` to determine whether this term is statistically significant.

1.9 Does this term affect the adequacy of the model as determined by `estat gof` ?

2.1.2 Negative Binomial Regression

This section used data concerning childhood mortality in three cohorts, from the dataset `$datadir/nbreg`. The children were divided into 7 age-bands, and the number of deaths, and the persons-months of exposure are recorded in `deaths` and `exposure` respectively. For some reason, one model that converged perfectly well using `xi:` to define indicators failed when `xi:` was not used, which is why `ltol(0.000001)` has been added to one command below: the model converges with a less severe tolerance criterion.

1.10 Fit a poisson regression model using only cohort as a predictor:

```
poisson deaths i.cohort, exposure(exposure) irr
```

Are there differences in mortality rate between the cohorts ?

1.11 Use `estat gof` to test whether the poisson model was appropriate

1.12 Fit a negative binomial regression model to test the same hypothesis:

```
nbreg deaths i.cohort, exposure(exposure) irr
```

Do you reach the same conclusion about the role of `cohort` ?

1.13 What is the value of the parameter α, and its 95% confidence interval ?

1.14 Fit a constant dispersion negative binomial regression model with

```
nbreg deaths i.cohort, exposure(exposure) dispersion(constant) irr
```

Is δ significantly greater than 0 in this model ?

1.15 Does this model suggest any different conclusions as to whether the mortality rate differs between cohorts ?

1.16 One possible source of the extra variation is a change in mortality with age. Fit a model to test whether mortality varies with age with

```
nbreg deaths i.age_gp, exposure(exposure) irr
```

Is age a significant predictor of mortality ?

1.17 Would it be appropriate to use Poisson regression to fit this model ?

1.18 Now fit a negative binomial regression model with both age and cohort as predictors (you will need to add the option `ltol(0.000001)` to get this model to converge). Use `testparm` to determine whether both age and cohort are independently significant predictors of mortality.

1.19 Is α significantly greater than 0 in this model ?

1.20 Fit the same model using `poisson`. Does this model agree with the negative binomial model ?

1.21 Use `estat gof` to test the adequacy of this model. Is using a Poisson regression model appropriate in this case ?

2.1.3 Using constraints

This section uses the data on damage to ships from the dataset `$datadir/ships` again.

1.22 Refit the final Poisson regression model we considered with

```
poisson damage i.type i.built i.sailed, irr exposure(months)
```

Which of the incidence rate ratios are not significantly different from 1 ?

1.23 Create predicted numbers of damage incidents with the command

```
predict pred_n
```

1.24 Define a constraint to force the incidence rate ratio for ships of type D to be equal to 1 with

```
constraint define 1 4.type = 0
```

(Note that the constraints are defined on the *coefficients* of the model, rather than the incidence rate ratios. If the coefficient is 0, the incidence rate ratio is 1.)

1.25 Fit this model with the command

```
poisson damage i.type i.built i.sailed, irr exposure(months) constr(1)
```

How does the output of this command differ from that of the previous Poisson regression command ?

1.26 Use `estat gof` to test the adequacy of this model. How does the constrained model compare to the unconstrained model ?

1.27 Define a second constraint to force the incidence rate ratio for ships of type E to be equal to 1 with

```
constraint define 2 5.type = 0
```

1.28 Fit a Poisson regression model with both of these constraints using the command

```
poisson damage i.type i.built i.sailed, irr exposure(months) constr(1 2)
```

(The above command should be entered on one line.)

1.29 How does the adequacy of this model compare to that of the previous one ?

1.30 It appears that the incidence rate ratio for being built in 1965-1969 is very similar to the incidence rate ratio for being built in 1970-1974. Define a new constraint to force these parameters to be equal with

```
constraint define 3 2.built = 3.built
```

Fit a Poisson regression model with all three constraints using the command

```
poisson damage i.type i.built i.sailed, irr exposure(months) constr(1 2
3)
```

(The above command should be entered on one line.) Notice that the lines for 2.built and 3.built are now identical. In what way do these two lines differ from the lines for the other constrained values ?

1.31 What do you think is the reason for the difference you have just observed ?

1.32 Use `estat gof` to test the adequacy of this constrained model. Have the constraints that you have applied to the model had a serious detrimental effect on the fit of the model.

1.33 Obtain predicted counts from this constrained model with the command

```
predict pred_cn
```

1.34 Compare the predictions from the constrained model and the unconstrained model to each other and to the observed values with

```
corr damage pred_n pred_cn
```

How has the fit of the model been affected by the constraints ?

1.35 If you wish, you can examine the observed and predicted values directly with

```
list type built sailed damage pred_n pred_cn
```

Does this list confirm your answer to the previous question ?

2.1.4 Constraints in Multinomial Logistic Regression

Constraints can be applied to many different types of regression model. However, applying constraints when using `mlogit` can be tricky because there are several equations. The syntax is then similar to the syntax we saw last week for `lincom`. For this part of the practical, we are using the same `$datadir/alligators` dataset that we saw last week.

1.36 Use

```
label list
```

to remind yourself of what the variables mean.

1.37 Fit a multinomial logistic regression model to predict food choice from lake with the command

```
mlogit food i.lake, rrr
```

Are there significant differences between lakes in the primary food choice ?

1.38 What are the odds ratios for preferring invertebrates to fish in Lakes Oklawaha, Trafford and George ?

1.39 It appears that for the choice of invertebrates rather than fish, there is no significant difference between Lake Oklawaha and Lake Trafford. Define the constraint that corresponds to this with

```
constraint define 1 [Invertebrate]2.lake = [Invertebrate]3.lake
```

Fit the model again with this constraint using

```
mlogit food i.lake, rrr const(1)
```

1.40 Even Lake George does not appear to be significantly different from Lake Oklawaha and Lake Trafford. Define a new constraint with

```
constraint define 2 [Invertebrate]4.lake = [Invertebrate]3.lake
```

Fit a multinomial logistic regression model with both of these constraints with

```
mlogit food i.lake, rrr const(1 2)
```

How does the common odds ratio for all three lakes compare to the 3 separate odds ratios you calculated previously ?

Bibliography