Module 9: Single-level and Multilevel Models for Ordinal Responses

Stata Practical[1]

George Leckie, Tim Morris & Fiona Steele
Centre for Multilevel Modelling

Pre-requisites

- Modules 5, 6 and 7

If you find this module helpful and wish to cite it in your research, please use the following citation:

Leckie, G., Morris, T., Steele, F. (2016). Single-level and Multilevel Models for Ordinal Responses - Stata Practical. LEMMA VLE Module 9, 1-45. URL: http://www.bristol.ac.uk/cmm/learning/online-course/.

Contents

[1] This Stata practical is adapted from the corresponding MLwiN practical: Steele, F. (2011). Single-level and Multilevel Models for Ordinal Responses - Stata Practical. LEMMA VLE Module 9, 1-48. Accessed at http://www.bristol.ac.uk/cmm/learning/course.html

Introduction to the Eurobarometer 2009 Dataset on Interest in EU Elections

You will be analysing data from the Eurobarometer Opinion and Social Questionnaire from spring 2009.[2] The analysis sample contains residents of the 29 European Union Member States[3] who were aged 15 years and over, selected using a multi-stage probability design.

Our response variable is an ordinal indicator of the level of interest in European elections. Respondents were asked:

The next European elections will be held in June 2009. How interested or disinterested would you say you are in these elections?

and presented with the following response alternatives: Very interested, Somewhat interested, Somewhat disinterested, Very disinterested, and Don't know.

After excluding the small number of "don't knows" and respondents from candidate EU states who were not asked this question, the sample size is 26,126. For purposes of illustration, and to speed up model estimation, we take a 50% sample and exclude a small percentage of individuals with missing values on any of the explanatory variable considered. The analysis sample contains 10,340 individuals with the sample size for each state ranging from 98 to 509. The data therefore have a two-level hierarchical structure with individuals at level 1, nested within states at level 2.

We consider several predictor variables. The dataset contains only individual-level variables, but we will derive state-level aggregates for consideration as level 2 predictors. The individual-level variables are gender, age, occupation type, and an index of left-right political attitudes.[4]

The file contains the following variables:

Variable name	Description and codes
state	EU state identifier

[2] Eurobarometer 71.1: European Parliament and Elections, Economic Crisis, Climate Change, and Chemical Products, January-February 2009 (Study No. ZA4971). Go to http://www.gesis.org/en/eurobarometer-data-service/ for further information on the Eurobarometer series and to download datasets.

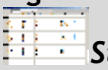
[3] The survey was also conducted in the three candidate countries (Croatia, Turkey and Macedonia) and in the Turkish Cypriot Community, but they are not included in our analysis file because the response variable (interest in EU elections) was not available for respondents in these countries.

[4] Respondents were asked to rate their political views on a 10-point scale in response to the question: "In political matters people talk of 'the left' and 'the right'. How would you place your views on this scale?"

person	Individual identifier
electint	Interest in EU elections (1=very low, 2=low, 3=some, 4=very high)[5]
female	Individual gender (1=female, 0=male)
agecen50	Individual age in years (centred at 50)
agecen50sq	Individual age in years (centred at 50) squared
occtype	Occupation type (1=manager, 2=other employed, 3=looking after home/family, 4=unemployed, 5=retired, 6=student)
lrplace	Placement on scale of left-right political attitudes (a 10-point scale with high values indicating more right wing views)
commtype	Type of community of residence (1=rural, 2=mid-sized town, 3=large town or city)

Load "9.1.dta" into memory and open the do-file "9.do" for this lesson.

From within the LEMMA Learning Environment

- Go to **Module 9: Single-Level and Multilevel Models for Ordinal Responses**, and scroll down to  ***Stata* datasets and dofiles**
- Click "9.1.dta" to open the dataset

Use the `summarize` command to view the variables in the dataset:

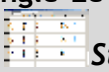
```
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       state |     10,340    14.81576    8.942563          1         30
      person |     10,340      5170.5    2985.045          1      10340
    electint |     10,340    2.455222    .9038433          1          4
      female |     10,340    .5267892     .499306          0          1
    agecen50 |     10,340   -.6407157    17.60089        -35         48
-------------+---------------------------------------------------------
  agecen50sq |     10,340    310.1721    323.4805          0       2304
     occtype |     10,340    3.234816    1.620245          1          6
     lrplace |     10,340    5.292843    2.307556          1         10
    commtype |     10,340    1.904449     .794336          1          3
```

[5] The coding of the original variable was reversed so that high values indicate greater interest. 'Very high' corresponds to 'very interested', 'some' to 'somewhat interested', 'low' to 'somewhat disinterested', and 'very low' to 'very disinterested'.

P9.1 Cumulative Logit Model for Single-Level Data

Load "9.1.dta" into memory, and if it is not already in use open the do-file "9.do" for this lesson.

From within the LEMMA Learning Environment

- Go to **Module 9: Single-Level and Multilevel Models for Ordinal Responses**, and scroll down to  ***Stata*** **datasets and dofiles** Click "9.1.dta" to open the dataset

P9.1.1 Specifying and estimating and cumulative logit model

We will begin by examining the distribution of our response variable, level of interest in EU elections. Use the `tabulate` command to view the number (`Freq.`) and percentage (`Percent`) of respondents in each response category

```
. tabulate electint

Interest in |
   European |
  elections |      Freq.     Percent        Cum.
------------+-----------------------------------
       vlow |      1,773       17.15       17.15
        low |      3,255       31.48       48.63
       some |      4,144       40.08       88.70
      vhigh |      1,168       11.30      100.00
------------+-----------------------------------
      Total |     10,340      100.00
```

The percentage in each of the four response category is shown. The cumulative response percentages, working upwards from the 'very low' category are 17.2%, 48.6%, 88.7%, 100%[6].

Our first model will simply reproduce the cumulative probabilities, from which we can derive the response probabilities. The model is a single-level ordered logistic regression with no covariates. Let $y_i = s$ denote the ordinal response for respondent i ($i = 1, \ldots, n$) where $s = 1,2,3,4$ denotes the four response categories "vlow", "low", "some" and "vhigh".The model can then be written as

$$\text{logit}\{\Pr(y_i > s|x_{1i})\} \equiv \log\left\{\frac{\Pr(y_i > s)}{1 - \Pr(y_i > s)}\right\} = -\kappa_s, \qquad s = 1,2,3$$

[6] Note that the `ologit` and `meologit` estimation commands for fitting single-level and multilevel ordinal response models cumulate the response category probabilities the other way around.

where the only parameters to be estimated are the three cut points κ_1, κ_2 and κ_3.

We fit the above model using the `ologit` command. The model converges after one iteration:

```
. ologit electint

Iteration 0:   log likelihood = -13224.823
Iteration 1:   log likelihood = -13224.823

Ordered logistic regression                     Number of obs     =     10,340
                                                LR chi2(0)        =       0.00
                                                Prob > chi2       =          .
Log likelihood = -13224.823                     Pseudo R2         =     0.0000

------------------------------------------------------------------------------
    electint |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       /cut1 |  -1.575245    .026091                      -1.626382   -1.524107
       /cut2 |  -.0549461   .0196759                      -.0935101   -.0163822
       /cut3 |   2.060862   .0310675                       1.999971    2.121754
------------------------------------------------------------------------------
```

The first cut point `/cut1` is estimated to be -1.575 and tells us that the log-odds of having low, some or very high interest in EU elections ($s > 1$) relative to very low interest ($s = 1$) is 1.575. This corresponds to a probability of having low, some or very high interest in EU elections of exp(1.575)/[1+exp(1.575)] = 0.828. It follows that the probability of having instead very low interest in EU elections is simply 1 – 0.828 or 0.172.

The second cut point `/cut2` is estimated to be -0.055 and so the the log-odds of having some or very high interest in EU elections is 0.055, which corresponds to a probability of 0.514. The probability of having instead very low or low interest in EU elections is 1 – 0.514 or 0.486.

Finally, the third cut point `/cut3` is estimated to be 2.061 and so the log-odds of having very high interest in EU elections is -2.061 which corresponds to a probability of 0.113. The probability of having instead very low, low or some interest in EU elections is 1 – 0.113 or 0.887.

Reassuringly, these probabilities all agree with the cumulative percentages from our earlier tabulation of **electint**.

We could have carried out these calculations using Stata's post estimation `predict` command to calculate the predicted probability for each category of **electint**.

```
. predict p*

(option pr assumed; predicted probabilities)
```

Stata generates four new variables **p1**, **p2**, **p3** and **p4** which store, for each respondent, the predicted probability of each response category. We can use the `summarize` command to display summary statistics of the predictions:

```
. summarize p1-p4
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          p1 |     10,340      .17147           0     .17147     .17147
          p2 |     10,340    .3147969           0   .3147969   .3147969
          p3 |     10,340    .4007737           0   .4007737   .4007737
          p4 |     10,340    .1129594           0   .1129594   .1129594
```

The model includes no covariates and so the predicted probabilities are the same for all 10,340 respondents. The predicted probabilities from the model match the response category percentages reported in the earlier one-way tabulation of **electint**. We can also obtain the cumulative probabilities presented in that tabulation by summing the category-specific probabilities appropriately. We do this by generating a new variable for each cumulative probability using the `generate` command:

```
. generate p12 = p1 + p2

. generate p123 = p1 + p2 + p3

. generate p1234 = p1 + p2 + p3 + p4
```

Summarizing these new variables gives the cumulative response probabilities:

```
. summarize p1 p12 p123 p1234

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          p1 |     10,340      .17147           0     .17147     .17147
         p12 |     10,340    .4862669           0   .4862669   .4862669
        p123 |     10,340    .8870406           0   .8870406   .8870406
       p1234 |     10,340           1           0          1          1
```

These values 0.171, 0.486 and 0.887 agree with our earlier one-way tabulation of **electint**. Finally, we remove all these newly generated variables from the dataset using the `drop` command:

```
. drop p1-p1234
```

P9.1.2 Adding gender

We will next allow for gender differences in election interest, but before including gender in our model we look at a tabulation of **electint** by **female**. Use the `tabulate` command with the option `row` to display row percentages alongside cell and row and column total frequencies:

```
. tabulate female electint, row

+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

           |       Interest in European elections
    female |      vlow        low       some      vhigh |     Total
```

This document is only the first few pages of the full version.
To see the complete document please go to learning materials and register:
http://www.cmm.bris.ac.uk/lemma
The course is completely free. We ask for a few details about yourself for our research purposes only. We will not give any details to any other organisation unless it is with your express permission.