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

Fiona Steele
Centre for Multilevel Modelling

Pre-requisites box

- Modules 5, 6 and 7

Contents

P9.1 Introduction to the Eurobarometer 2009 Dataset on Interest in EU Elections ... 2

P9.2 Cumulative Logit Model for Single-Level Data ... 4

P9.2.1 Specifying and estimating a cumulative logit model ... 4
P9.2.2 Adding gender ... 8
P9.2.3 Testing the proportional odds assumption ... 10
P9.2.4 Adding further explanatory variables ... 13

P9.3 Continuation Ratio Model ... 15

P9.4 Random Intercept Cumulative Logit Model ... 20

P9.4.1 Specifying and estimating a simple two-level model ... 21
P9.4.2 Interpretation of the null two-level model ... 26
P9.4.3 Adding explanatory variables ... 28

P9.5 Random Slope Cumulative Logit Model ... 30

P9.5.1 Specifying and testing a random slope for age ... 30
P9.5.2 Interpretation of the random slope model ... 34

P9.6 Contextual Effects ... 40

P9.1 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.[1] The analysis sample contains residents of the 29 European Union Member States[2] 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.[3]

The file contains the following variables:

[1] 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.

[2] 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.

[3] 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?"

Variable name	Description and codes
state	EU state identifier
person	Individual identifier
electint	Interest in EU elections (1=very low, 2=low, 3=some, 4=very high)[4]
female	Individual gender (1=female, 0=male)
agecen50	Individual age in years (centred at 50)
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)
cons	A column of ones. This variable will be included as an explanatory variable in all models and its coefficient will be the intercept

To open the worksheet:

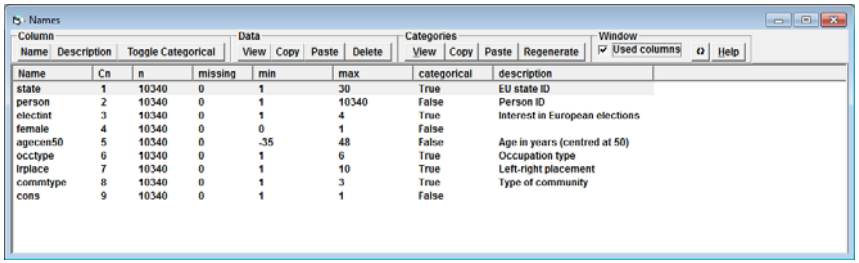
From within the LEMMA Learning Environment

- Go to **Module 9: Single-Level and Multilevel Models for Ordinal Responses**, and scroll down to *MLwiN* **Datafiles**
- If you do not already have MLwiN to open the datafile with, click (**get *MLwiN***).
- Click "9.1.wsz"

The **Names** window will appear.

- Click the check box next to **Used columns** to view only those columns that contain data

[4] 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'.



Name	Cn	n	missing	min	max	categorical	description
state	1	10340	0	1	30	True	EU state ID
person	2	10340	0	1	10340	False	Person ID
electint	3	10340	0	1	4	True	Interest in European elections
female	4	10340	0	0	1	False	
agecen50	5	10340	0	-35	48	False	Age in years (centred at 50)
occtype	6	10340	0	1	6	True	Occupation type
lrplace	7	10340	0	1	10	True	Left-right placement
commtype	8	10340	0	1	3	True	Type of community
cons	9	10340	0	1	1	False	

P9.2 Cumulative Logit Model for Single-Level Data

P9.2.1 Specifying and estimating a cumulative logit model

We will begin by examining the distribution of our response variable, level of interest in EU elections.

- From the **Basic Statistics** menu, select **Tabulate**
- Under **Display**, check **Percentages of row totals**
- From the **Columns** drop-down list, select **electint**
- Click **Tabulate**

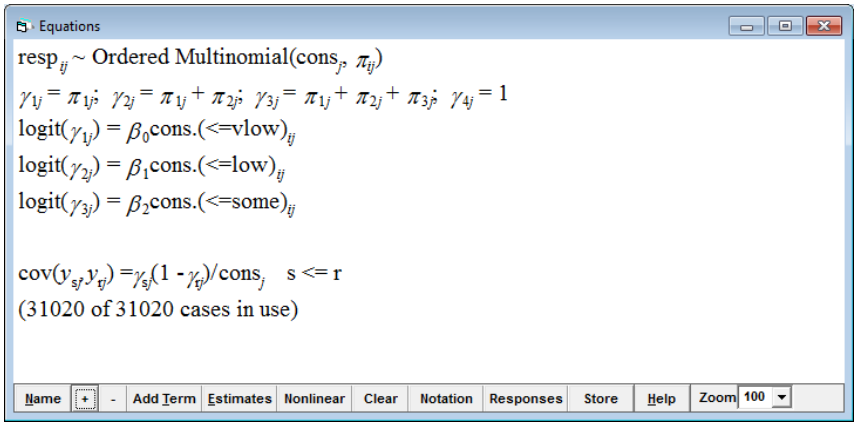
	vlow	low	some	vhigh	TOTALS
N	1773	3255	4144	1168	10340
%	17.1	31.5	40.1	11.3	100.0

The percentage in each of the four response category is shown. The cumulative response percentages, working upwards from the 'very low' category are 17.1%, 48.6%, 88.7%, 100%.

Our first model will simply reproduce the cumulative probabilities, from which we can derive the response probabilities.

- From the **Model** menu, select **Equations**
- Click on either red *y* in the **Equations** window
- From the drop-down list labelled ***y*:** select **electint**
- From the drop-down list labelled **N levels:** select **1-i**
- From the drop-down list labelled **level 1(i):** select **person**
- Click **done**
- Click on $\mathbf{N}(\boldsymbol{XB}, \boldsymbol{\Omega})$ and, under **Select distribution**, check **Multinomial**. Only the logit link function is available. Under **Multinomial options**, check **Ordered proportional odds**. Finally, for **ref category**, switch to **vhigh** (which is the last category). Click **Done**

- We now need to specify the denominator which for an ordered response is always equal to 1, so click on the red n_j and select **cons** from the drop-down list. Check **Done**. If you go to the **Names** window and click the refresh button (Under **Window**) you will see that a new variable called **denom** has been added to the worksheet. This has been created from **cons.** You will also notice that several new variables with 31020 records have been created; this will be explained in a moment.
- In the **Equations** window click on **Add Term** and select **cons** from the **variable** drop-down list. Click **add Separate coefficients**. This adds the threshold parameters which MLwiN names β_0, β_1 and β_2 (corresponding to α_1, α_2 and α_3 in equation (9.4) in C9.1.3 with k = 1, 2, 3).
- Click the **+** button twice to see the full model specification

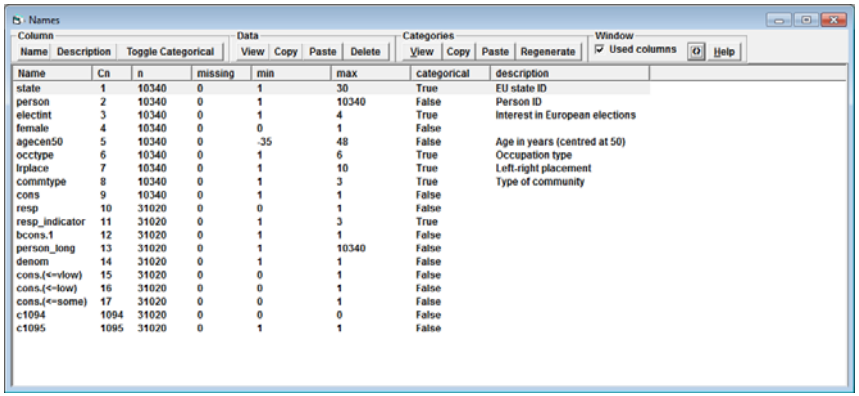


The second line in the **Equations** window reminds us what cumulative probabilities we are modelling: the probability of being in category 1 (very low); categories 1 or 2 (very low or low); and categories 1, 2 or 3 (very low, low or some). The three equations are for the log-odds of each of these cumulative probabilities.

In MLwiN, the subscript i indexes the response category (1, . . ., 4) and j indexes the individual.

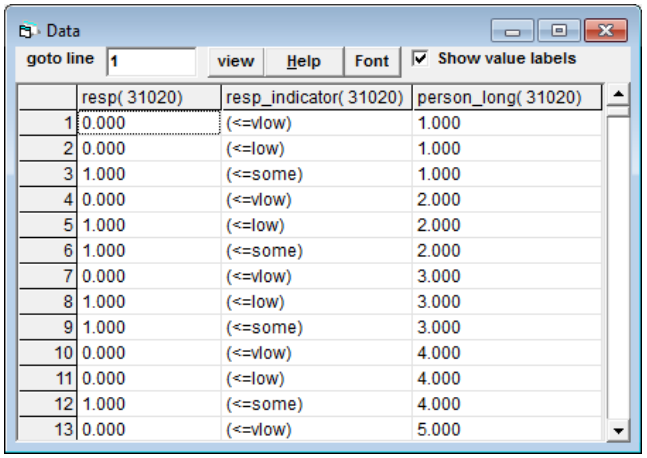
Before estimating this model, we will look at the new variables that specifying the model has created.

- Go to the **Names** window



Name	Cn	n	missing	min	max	categorical	description
state	1	10340	0	1	30	True	EU state ID
person	2	10340	0	1	10340	False	Person ID
electint	3	10340	0	1	4	True	Interest in European elections
female	4	10340	0	0	1	False	
agecen50	5	10340	0	-35	48	False	Age in years (centred at 50)
occtype	6	10340	0	1	6	True	Occupation type
lrplace	7	10340	0	1	10	True	Left-right placement
commtype	8	10340	0	1	3	True	Type of community
cons	9	10340	0	1	1	False	
resp	10	31020	0	0	1	False	
resp_indicator	11	31020	0	1	3	True	
bcons.1	12	31020	0	1	1	False	
person_long	13	31020	0	1	10340	False	
denom	14	31020	0	1	1	False	
cons.(<=vlow)	15	31020	0	0	1	False	
cons.(<=low)	16	31020	0	0	1	False	
cons.(<=some)	17	31020	0	0	1	False	
c1094	1094	31020	0	0	0	False	
c1095	1095	31020	0	1	1	False	

- Highlight **resp**, **resp_indicator** and **person_long** (using ctrl-click) and click **View** under **Data**



	resp(31020)	resp_indicator(31020)	person_long(31020)
1	0.000	(<=vlow)	1.000
2	0.000	(<=low)	1.000
3	1.000	(<=some)	1.000
4	0.000	(<=vlow)	2.000
5	1.000	(<=low)	2.000
6	1.000	(<=some)	2.000
7	0.000	(<=vlow)	3.000
8	1.000	(<=low)	3.000
9	1.000	(<=some)	3.000
10	0.000	(<=vlow)	4.000
11	0.000	(<=low)	4.000
12	1.000	(<=some)	4.000
13	0.000	(<=vlow)	5.000

These variables contain three records for each person, leading to a total of 3 × 10340 = 31020 records, and **person_long** is the person identifier in this expanded dataset. These three records correspond to the first three categories of the ordinal response, **electint** (the last category is omitted because it was taken as the reference category). For each of the first three categories, a binary response has been created indicating whether the response was in that category *or lower*; these have been stacked into a single column called **resp**. The variable **resp_indicator** tells us which category the binary response refers to. For example, we can infer that persons 1 and 4 had **electint**=some because the sequence of binary responses 0, 0 and 1 for **resp_indicator** categories '<=vlow', '<=low' and '<=some'

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.