

Module 10: Single-level and Multilevel Models for Nominal Responses

MLwiN Practical

Fiona Steele
Centre for Multilevel Modelling

Pre-requisites box

- Modules 5, 6 and 7

Contents

P10.1	Specifying a Multinomial Logit Model for Media Preferences	1
P10.1.1	Introduction to the Eurobarometer 2009 Dataset on EU Media Preferences for Political Information	1
P10.1.2	Specifying an intercept-only model and MLwiN notation	3
P10.1.3	Data structure for a multinomial logit model.....	4
P10.2	Multinomial Logit Modelling of Media Preferences	7
P10.2.1	Predicted probabilities from a multinomial logit model	7
P10.2.2	Adding gender	8
P10.2.3	Adding further explanatory variables	10
P10.2.4	Obtaining maximum likelihood estimates in MLwiN	13
P10.3	Random Intercept Multinomial Logit Model.....	16
P10.3.1	Specifying and estimating a simple two-level model	16
P10.3.2	Interpretation of the null two-level model	22
P10.3.3	Adding explanatory variables	26
P10.4	Contextual Effects.....	30
P10.5	Conditional Logit Models: Incorporating Characteristics of Response Alternatives.....	35
P10.5.1	Data structure	35
P10.5.2	Estimation of a conditional logit model as a Poisson model in MLwiN	36
P10.5.3	Combining the multinomial and conditional logit models.....	38
P10.5.4	Note on estimation of multilevel conditional logit model	40

P10.1 Specifying a Multinomial Logit Model for Media Preferences

P10.1.1 Introduction to the Eurobarometer 2009 Dataset on EU Media Preferences for Political Information

In these practical exercises 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 States² who were aged 15 years and over, selected using a multi-stage probability design.

Our response variable is a nominal indicator of the preferred type of media for obtaining information about political matters. Respondents were asked:

If you had to choose between the following means of getting information about politics and political matters, what would you prefer?

They were presented with the following response alternatives: Newspapers, Written magazines, Television, Radio, the Internet, and Don't know.

Some respondents said they preferred a type of media other than those listed above, while others said they had no preference. These 'other' and 'no preference' responses are referred to as 'spontaneous' in the survey because they were not included among the alternatives read out to respondents. For simplicity, we exclude these cases and "don't knows" in our analysis, together with a small percentage of individuals with missing values on any of the explanatory variables considered. We also combine newspapers and written magazines in a single response category. Finally, we take a 25% sample to speed up model estimation. The analysis sample contains 4914 individuals with the sample size for each state ranging from 49 to 239. 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, and occupation type.

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 (preferred type of media) was not available for respondents in these countries.

Variable name	Description and codes
state	EU state identifier
person	Individual identifier
medpref	Media preference for political information (1=paper/magazine, 2=TV, 3=radio, 4=internet)
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)
cons	A column of ones. This variable will be included as an explanatory variable in all models and its coefficient will be the intercept

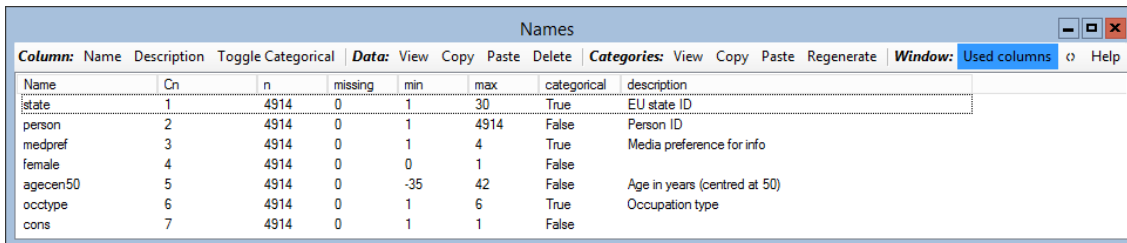
To open the worksheet:

From within the LEMMA Learning Environment

- Go to **Module 10: Single-Level and Multilevel Models for Nominal Responses**, and scroll down to **MLwiN Datafiles**
- If you do not already have MLwiN to open the datafile with, click ([get MLwiN](#)).
- Click “ [10.1.wsz](#)”

The **Names** window will appear.

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



Name	Cn	n	missing	min	max	categorical	description
state	1	4914	0	1	30	True	EU state ID
person	2	4914	0	1	4914	False	Person ID
medpref	3	4914	0	1	4	True	Media preference for info
female	4	4914	0	0	1	False	
agecen50	5	4914	0	-35	42	False	Age in years (centred at 50)
occtype	6	4914	0	1	6	True	Occupation type
cons	7	4914	0	1	1	False	

P10.1.2 Specifying an intercept-only model and MLwiN notation

In this exercise, we will specify a simple multinomial logit model for the four-category nominal response **medpref**. The focus at this stage will be to set up the model in MLwiN, compare the notation with that used in C10.1, and examine the data structure of the analysis file. We will defer estimation of the model, and interpretation of the results, to P10.2.

We begin by examining the distribution of our response variable, preferred media for information on political issues.

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

	TV	paper/magazine	radio	internet	TOTALS
N	2314	1214	541	845	4914
%	47.1	24.7	11.0	17.2	100.0

The percentage in each of the four response categories is shown and we can see, for example, that most people prefer to obtain information from television.

Our first model will include only intercept terms, one for each contrast with the reference category. We will take the largest category 'TV' (category 1) as the reference. When we examine the parameter estimates from this model in the next exercise, we will see that this model simply reproduces the sample response probabilities shown above.

- From the **Model** menu, select **Equations**
- Click on either red **y** in the **Equations** window
- From the drop-down list labelled **y**: select **medpref**
- 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 $N(XB, \Omega)$ and, under **Select distribution**, check **Multinomial**. Only the logit link function is available. Under **Multinomial options**, retain the default of **Unordered**. Finally, for **ref category**, retain **TV** (which is selected by default as the first category). Click **Done**
- We now need to specify the denominator, which for a nominal response is always equal to 1. Click on the red n_j and select **cons** from the drop-down list. Check **Done**. If you go to the **Names** window (from the **Data Manipulation** menu) 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 other new variables have been created, all with 14,742 records; 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 intercept parameters which MLwiN names β_0 , β_1 and β_2
- Click the + button twice to see the full model specification

The screenshot shows the 'Equations' window in MLwiN. The window title is 'Equations'. The content displays the following model specification:

```

respij ~ Multinomial(consj, πij)
log(π2j / π1j) = β0cons.paper/magazineij
log(π3j / π1j) = β1cons.radioij
log(π4j / π1j) = β2cons.internetij

cov(ysjysj) = - πsjπsj/consj : s ≠ r;   πsj(1 - πsj)/consj : s = r;
(14742 of 14742 cases in use)

```

At the bottom of the window, there is a menu bar with the following items: Name + - Add Term Estimates Nonlinear Clear Notation Responses Store Help Zoom 100

The model consists of three equations or ‘contrasts’, one for each of categories 2, 3 and 4 with the reference category 1. The intercepts are denoted in MLwiN by β_0 , β_1 and β_2 , which correspond to β_{02} , β_{03} and β_{04} of equation (10.1) in C10.1.1 (with $C = 4$ and no explanatory variable x). MLwiN uses the labels given to each category of the response variable **medpref** in naming the intercept terms. This is helpful for reminding us which category is being contrasted with the reference in each equation. When we add explanatory variables to the model, these will also be given the response category labels as suffixes.

In MLwiN, the subscript i indexes the response category (2, 3 and 4) and j indexes the individual. The response probabilities are therefore denoted by π_{ij} (rather than π_{ki} as in C10.1.1). MLwiN uses i and j subscripts because the data have been expanded to a type of two-level structure (response categories within individuals). The notation used by MLwiN for nominal response models is consistent with that used for ordinal response models (Module 9). As we shall see, the data structure is also the same for the two types of model.

P10.1.3 Data structure for a multinomial logit model

Specifying the multinomial model has led to a restructuring of the dataset and the creation of several new variables. To look at these new variables:

- Go to the **Names** window

Column:	Name	Description	Toggle	Categorical	Data:	View	Copy	Paste	Delete	Categories:	View	Copy	Paste	Regenerate	Window:	Used columns	Help
	Name	Cn	n	missing	min	max	categorical	description									
	state	1	4914	0	1	30	True	EU state ID									
	person	2	4914	0	1	4914	False	Person ID									
	medpref	3	4914	0	1	4	True	Media preference for info									
	female	4	4914	0	0	1	False										
	agecen50	5	4914	0	-35	42	False	Age in years (centred at 50)									
	occtype	6	4914	0	1	6	True	Occupation type									
	cons	7	4914	0	1	1	False										
	resp	8	14742	0	0	1	False										
	resp_indicator	9	14742	0	2	4	True										
	bcons.1	10	14742	0	1	1	False										
	person_long	11	14742	0	1	4914	False										
	denom	12	14742	0	1	1	False										
	cons.paper/magazine	13	14742	0	0	1	False										
	cons.radio	14	14742	0	0	1	False										
	cons.internet	15	14742	0	0	1	False										
	c1094	1094	14742	0	0	0	False										
	c1095	1095	14742	0	1	1	False										

- Highlight **resp**, **resp_indicator**, **person_long**, **cons.paper/magazine**, **cons.radio** and **cons.internet** (using ctrl-click) and click **View** under **Data**

The column widths have been altered in the screenshot below so that the full variable names are visible.

goto line	1	view	Show value labels	Font	Help	
	resp(14742)	resp_indicator(14742)	person_long(14742)	cons.paper/magazine(14742)	cons.radio(14742)	cons.internet(14742)
1	1.000	paper/magazine	1.000	1.000	0.000	0.000
2	0.000	radio	1.000	0.000	1.000	0.000
3	0.000	internet	1.000	0.000	0.000	1.000
4	0.000	paper/magazine	2.000	1.000	0.000	0.000
5	1.000	radio	2.000	0.000	1.000	0.000
6	0.000	internet	2.000	0.000	0.000	1.000
7	0.000	paper/magazine	3.000	1.000	0.000	0.000
8	0.000	radio	3.000	0.000	1.000	0.000
9	1.000	internet	3.000	0.000	0.000	1.000
10	1.000	paper/magazine	4.000	1.000	0.000	0.000
11	0.000	radio	4.000	0.000	1.000	0.000
12	0.000	internet	4.000	0.000	0.000	1.000
13	0.000	paper/magazine	5.000	1.000	0.000	0.000
14	1.000	radio	5.000	0.000	1.000	0.000
15	0.000	internet	5.000	0.000	0.000	1.000
16	0.000	paper/magazine	6.000	1.000	0.000	0.000
17	0.000	radio	6.000	0.000	1.000	0.000
18	1.000	internet	6.000	0.000	0.000	1.000
19	0.000	paper/magazine	7.000	1.000	0.000	0.000
20	0.000	radio	7.000	0.000	1.000	0.000
21	0.000	internet	7.000	0.000	0.000	1.000

The original dataset of 4914 records, one for each individual, has been expanded to give 3 records for each person (where 3 is the number of response categories minus 1). The expanded dataset therefore has $4914 \times 3 = 14742$ records. (Note that the bottom line of the **Equations** window reports the number of records in the expanded file.) This ‘multivariate’ data structure is the same as for the cumulative logit model for ordinal responses (see C9.2).

person_long is the person identifier in the expanded dataset. For each person, we have a record corresponding to categories 2, 3 and 4 of the nominal response, **medpref** (the first category is omitted because it was taken as the reference

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.