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

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 States[2] 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

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**
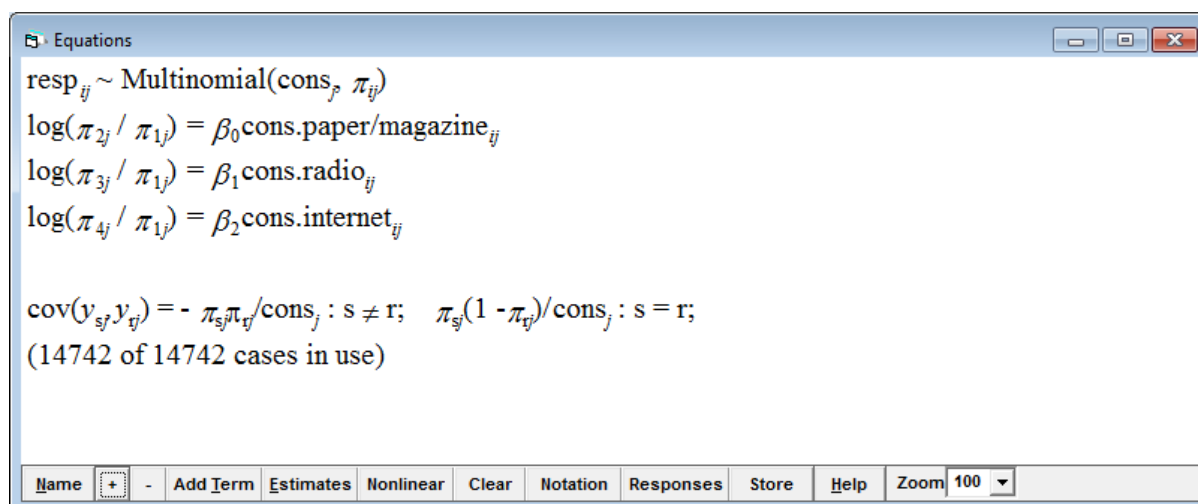- Click **Tabulate**

	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

Equations

$$\text{resp}_{ij} \sim \text{Multinomial}(\text{cons}_j, \pi_{ij})$$

$$\log(\pi_{2j} / \pi_{1j}) = \beta_0 \text{cons.paper/magazine}_{ij}$$

$$\log(\pi_{3j} / \pi_{1j}) = \beta_1 \text{cons.radio}_{ij}$$

$$\log(\pi_{4j} / \pi_{1j}) = \beta_2 \text{cons.internet}_{ij}$$

$$\text{cov}(y_{sj}, y_{rj}) = - \pi_{sj}\pi_{rj}/\text{cons}_j : s \neq r; \quad \pi_{sj}(1 - \pi_{rj})/\text{cons}_j : s = r;$$

(14742 of 14742 cases in use)

Name + - Add Term Estimates Nonlinear Clear Notation Responses Store Help Zoom 100 ▼
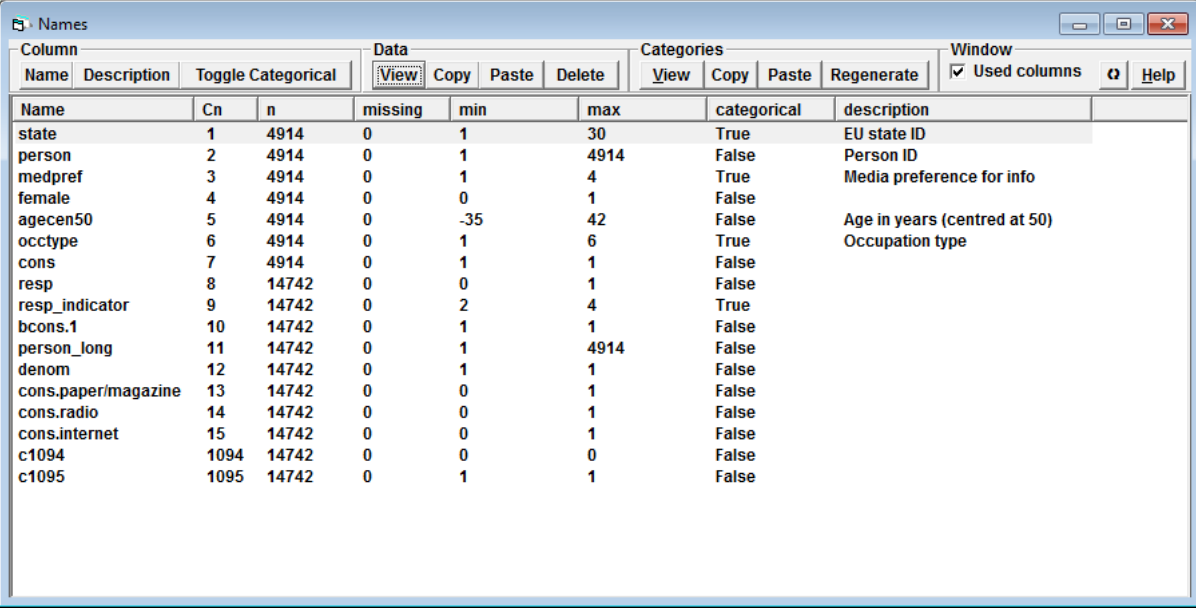
The model consists of three equations or 'contrasts', one for each of categories 2, 3 and 4 with the reference category 1. The intercepts are denoted in MLwiN by β_0, β_1 and β_2, which correspond to β_{02}, β_{03} and β_{04} of equation (10.1) in C10.1.1 (with $C = 4$ and no explanatory variable x). 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:
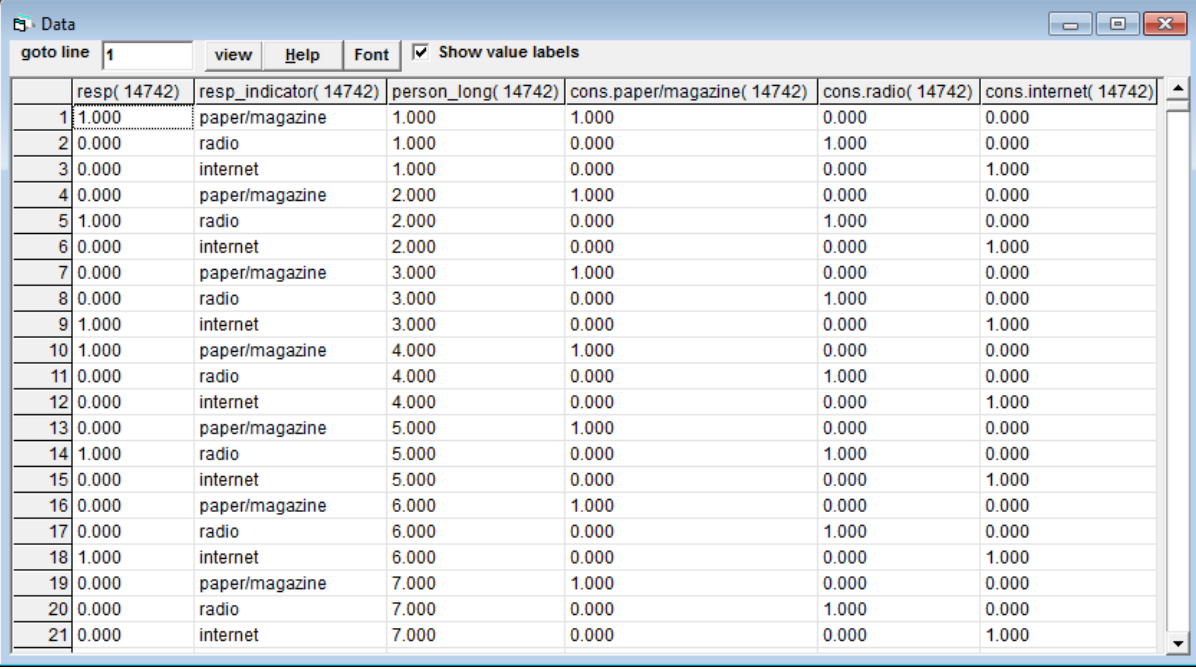
- Go to the **Names** window

> ▪ 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.

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 × 3 = 14742 records. (Note that the bottom line of the **Equations** window reports the number of records in the

Module 11: Three-Level Multilevel Models

George Leckie
Centre for Multilevel Modelling

Contents

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

Leckie, G. (2013). Three-Level Multilevel Models - Concepts. LEMMA VLE Module 11, 1-47.
http://www.bristol.ac.uk/cmm/learning/course.html

Address for correspondence:

George Leckie
Centre for Multilevel Modelling
University of Bristol
2 Priory Road
Bristol, BS8 1TX
UK

g.leckie@bristol.ac.uk

What are Three-Level Multilevel Models?

In the previous modules we illustrated two-level multilevel models for analysing two-level data structures where units (level 1) are nested within groups (or clusters) (level 2). When the groups are themselves nested within supergroups (or superclusters) (level 3), the data form a three-level hierarchy and three-level models can be fitted to account for the additional level. Examples of three-level data structures include: students (level 1) nested within classes (level 2) nested within schools (level 3); voters nested within counties nested within states; and patients nested within doctors nested within clinics. In this module, we describe three-level data structures and multilevel models which can be used to analyse them. Of course, there is nothing to stop data structures being even more complex and having four or more levels and we shall also consider examples of such data structures in this module. Many further examples of three- and four-level data structures are described in C4.2 and C4.3 of Module 4, respectively.

It is important to incorporate three-level structures in to our models when they arise in the data and lead the higher level clusters to differ substantially from one another on the response variable under study. Naively fitting two-level models to three-level data will lead us to misattribute response variation to the two included levels (van Landeghem et al., 2005; Moerbeek, 2004; van den Noortgate et al., 2005; Tranmer and Steele, 2001). This in turn may lead us to draw misleading conclusions about the relative importance of different sources of influence on the response. For example, fitting a students-within-classes two-level model of student attainment and ignoring the fact that classrooms are further nested within schools will likely lead us to overstate the importance of classrooms as a source of variation in student attainment. That is, much of the variation that we attribute to classrooms may be driven by school-to-school differences in attainment. Our naïve analysis would therefore overstate the importance of classrooms on student attainment and would ignore the role of schools (i.e. school policies, practices, context and compositional effects). Furthermore, by incorrectly modelling the dependency in the data we will likely obtain biased standard errors for the predictor variables, particularly those measured at higher levels. We therefore run the risk of making incorrect inferences and drawing misleading conclusions about the relationships being studied. For example, including school-level predictor variables in our students-within-classes two-level model, but ignoring school as a level in the model will typically lead us to severely underestimate the standard errors on these school-level variables. When we then go on to test the significance of these variables, we will run the risk of making type 1 errors of inference.

Introduction to the Example Dataset

In educational research, there is considerable interest in measuring the effects that schools have on students' educational achievements. Measuring the effects that schools have on their students is after all a necessary first step to learning how schools' policies and practices combine to generate differences between schools. Governments are also often interested in measuring school effects, typically for school accountability purposes, but often to also provide parents with information to help guide school choice. However, in nearly all education systems, there are substantial differences between schools in their students' attainments at intake (i.e. when students first arrive at their schools). For the purposes of researching the effects of schools' policies and practices, holding schools accountable, or informing school choice, schools should not be compared simply in terms of their average exam results as these differences will, at least in part, be driven by these initial differences.

Traditional studies of school effects attempt to measure the 'true' effects that schools have on their students by fitting two-level students-within-schools multilevel models to students' exam scores where covariate adjustments are made for students' initial scores, and typically for a range of other student background characteristics. The school-level residuals from these models are then argued to measure the effects that schools have on their students having adjusted for the non random selection of students into schools. These effects are interpreted as measuring the influences schools have on their students' academic progress (improvement or change in attainment) while they attend their schools. In school effectiveness research these influences are referred to as 'value-added' effects.

In terms of studying students' academic progress, there are many other potential sources of clustering or influence which may also be important determinants of student progress. For example, where data contain multiple academic cohorts of students, we can think of schools as potentially having different effects in different academic cohorts. This leads students from the same school-cohort to appear more alike than students from different school-cohorts. The data are then three-level with students (level 1) nested within school-cohorts (level 2) nested within schools (level 3). In this module, we shall introduce three-level multilevel models to explore such data. In particular, we shall focus on the stability of school effects over time by examining the extent to which school effects change from cohort to cohort.

We shall then go on to consider the further nesting of schools within administrative educational regions referred to as local authorities (LAs) (level 4).[1] In England, secondary schools are organised into 150 LAs. Traditionally, LAs controlled the distribution of government funds across schools, co-ordinated school admissions, and were the direct employers of all teachers and staff in many schools. While over the last few years there has been a reduction of LAs' powers, one might still expect to identify LA effects in the data. If nothing else, we would expect LA

---

[1] LAs correspond to school districts in the U.S.

effects to pick up geographic variation in student attainment that exists across England.

We shall use data from England's National Pupil Database (NPD), a census of all students in state (i.e. government funded) schools in England. The data are provided by the Department for Education (http://www.education.gov.uk). The NPD records students' academic attainments and a limited number of background characteristics. We focus on three consecutive academic cohorts of students who sat their General Certificate of Secondary Education (GCSE) examinations (age 16 years) in London schools in 2008, 2009 and 2010, respectively. These students sat their Key Stage 2 (KS2) examinations (age 11 years) five years earlier in 2003, 2004 and 2005, respectively.[2] [3]

Table 11.1 presents the number of units at each level of this data hierarchy.

Table 11.1 Number of units at each level of the data hierarchy

Level number	Level	Number of units
4	LAs	32
3	Schools	427
2	School-cohorts	1,232
1	Students	189,940

Thus, there are 32 LAs at level 4, 427 schools at level 3, 1,232 school-cohorts at level 2 and 189,940 students at level 1 of the data hierarchy. At this point it is helpful to explicitly define 'cohort' and 'school-cohort'. When we say 'cohort' we are referring to the three academic cohorts in the data: 2008, 2009 and 2010. When we refer to 'school-cohorts' we are referring to the 1,232 groups, or school-by-cohort combinations of students, in the data which are formed by crossing the 427 schools by the three cohorts. The number of schools and students present in the data for each cohort are as follows. In 2008 there were 412 schools and 63,208 students. In 2009 there were 410 schools and 63,072 students. In 2010 there were 410 schools and 63,660 students.[4] Three hundred and ninety five schools had all three cohorts represented in the data, 15 schools had only two of the three cohorts, while a further 17 schools had only one of the cohorts present. The 32 schools which were not present for one or more cohorts reflect the opening of new schools and the closing of old schools.

The response variable for all our analyses is a continuous point score summarising students' overall attainment in their GCSE examinations.[5] To ease the

---

[2] GCSE examinations are taken in the last year of secondary schooling. Successful GCSE results are often a requirement for taking A-level examinations (age 18 years) which in turn are a common type of university entrance determinant. For those who leave school at 16 years of age, GCSE results are their main job market qualification.

[3] KS2 examinations are taken in the last year of primary schooling.

[4] The 2010 cohort of 410 schools and 63,660 students will provide the example dataset in Modules 11 and 12.

[5] Specifically, the response variable is the student's capped 'best 8' total point score at GCSE with an additional bonus for attainment in each of English and Mathematics, and is the same measure as

interpretation of this variable, and so that the residuals at each level better approximate the normality assumptions of the models, we transform it to a standard normal score which has the property of being more normally distributed with mean zero and variance one.[6] This transformation allows the effects of the covariates in our multilevel models to be interpreted in terms of standard deviation units of the response. As our focus is on the stability of school effects across cohorts and not on any overall, London-wide, trend in student attainment over time, we carry out this transformation separately for each cohort. Put differently, in this analysis we are interested in the relative performance of schools to one another; we are not interested in the average absolute performance of schools.

We consider eight student-level predictor variables

- Attainment at age 11 (average point score across English, maths and science) (transformed to a standard normal score)

- Female (0 = male; 1 = female)

- Age (ranges from 0 to 1 where higher values correspond to older children; specifically, 0 corresponds to the youngest child in the data, born on the last day of the academic year, while 1 corresponds to the oldest child in the data, born on the first day of the academic year)

- Eligible for free school meals (FSM) (0 = no FSM; 1 = FSM)

- Special education needs (SEN) (0 = no SEN; 1 = SEN)

- English as an additional language (EAL) (0 = no EAL; 1 = EAL)

- Ethnicity (1 = White; 2 = Mixed; 3 = Asian; 4 = Black; 5 = Chinese; 6 = Other)

- Index of deprivation affecting children index (IDACI) a measure of residential neighbourhood social deprivation (transformed to a standard normal score)

and one school-cohort-level variable

- Cohort (1 = 2008; 2 = 2009; 3 = 2010)

---

that published in Government school performance tables (see http://www.education.gov.uk/performancetables).

[6] The transformation is carried out by first ranking the N students by their original scores. The standard normal score for the ith ranked student in the data is then $\Phi^{-1}\{(i - 0.5)/N\}$, where Φ^{-1} denotes the inverse of the standard normal cumulative distribution function. This transformation is order preserving and students with the same original scores will also be tied in terms of their standard normal scores.

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.