Module 5: Introduction to Multilevel Modelling MLwiN Practicals

Fiona Steele
Centre for Multilevel Modelling

Pre-requisites

- Modules 1-4

Contents

Some of the sections within this module have online quizzes for you to test your understanding. To find the quizzes:

EXAMPLE
From within the LEMMA learning environment

- Go down to the section for **Module 5: Introduction to Multilevel Modelling**
- Click " 5.1 Comparing Groups Using Multilevel Modelling" to open Lesson 5.1
- Click Q 1 to open the first question

Introduction to the Scottish Youth Cohort Trends Dataset

You will be analysing data from the Scottish School Leavers Survey (SSLS), a nationally representative survey of young people. We use data from seven cohorts of young people collected in the first sweep of the study, carried out at the end of the final year of compulsory schooling (aged 16-17) when most sample members had taken Standard grades[1].

In the practical for Module 3 on multiple regression, we considered the predictors of attainment in Standard grades (subject-based examinations, typically taken in up to eight subjects). In this practical, we extend the (previously single-level) multiple regression analysis to allow for dependency of exam scores within schools and to examine the extent of between-school variation in attainment. We also consider the effects on attainment of several school-level predictors.

The dependent variable is a total attainment score. Each subject is graded on a scale from 1 (highest) to 7 (lowest) and, after recoding so that a high numeric value denotes a high grade, the total is taken across subjects.

---

[1] We are grateful to Linda Croxford (Centre for Educational Sociology, University of Edinburgh) for providing us with these data. The dataset was constructed as part of an ESRC-funded project on Education and Youth Transitions in England, Wales and Scotland 1984-2002.

Further analyses of the data can be found in Croxford, L. and Raffe, D. (2006) "Education Markets and Social Class Inequality: A Comparison of Trends in England, Scotland and Wales". In R. Teese (Ed.) *Inequality Revisited*. Berlin: Springer.

The analysis dataset contains the student-level variables considered in Module 3 together with a school identifier and three school-level variables:

Variable name	Description and codes
CASEID	Anonymised student identifier
SCHOOLID	Anonymised school identifier
SCORE	Point score calculated from awards in Standard grades taken at age 16. Scores range from 0 to 75, with a higher score indicating a higher attainment
COHORT90	The sample includes the following cohorts: 1984, 1986, 1988, 1990, 1996 and 1998. The COHORT90 variable is calculated by subtracting 1990 from each value. Thus values range from -6 (corresponding to 1984) to 8 (1998), with 1990 coded as zero
FEMALE	Sex of student (1=female, 0=male)
SCLASS	Social class, defined as the higher class of mother or father (1=managerial and professional, 2=intermediate, 3=working, 4=unclassified).
SCHTYPE	School type, distinguishing independent schools from state-funded schools (1=independent, 0=state-funded)
SCHURBAN	Urban-rural classification of school (1=urban, 0=town or rural)
SCHDENOM	School denomination (1=Roman Catholic, 0=non-denominational)

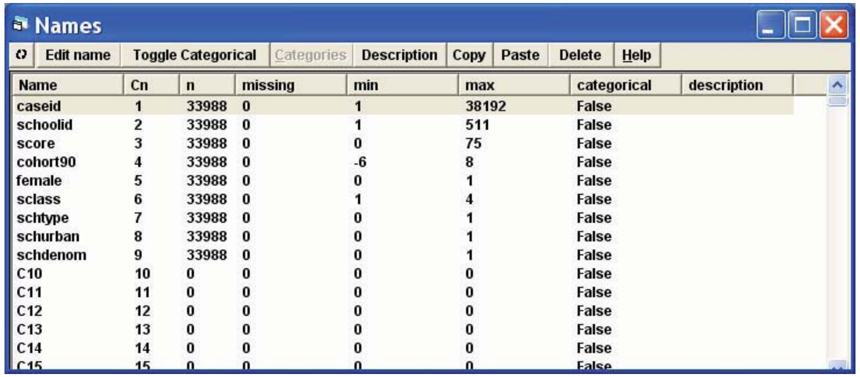
There are 33988 students in 508 schools.

Open the worksheet to

From within the LEMMA Learning Environment

- Go to **Module 5: Introduction to Multilevel Modelling**, and scroll down to ***MLwiN* Datafiles**
- If you do not already have MLwiN to open the datafile with, click (**get *MLwiN***).
- Click "5.1.wsz"

You will see the **Names** window:



Name	Cn	n	missing	min	max	categorical	description
caseid	1	33988	0	1	38192	False	
schoolid	2	33988	0	1	511	False	
score	3	33988	0	0	75	False	
cohort90	4	33988	0	-6	8	False	
female	5	33988	0	0	1	False	
sclass	6	33988	0	1	4	False	
schtype	7	33988	0	0	1	False	
schurban	8	33988	0	0	1	False	
schdenom	9	33988	0	0	1	False	
C10	10	0	0	0	0	False	
C11	11	0	0	0	0	False	
C12	12	0	0	0	0	False	
C13	13	0	0	0	0	False	
C14	14	0	0	0	0	False	
C15	15	0	0	0	0	False	

P5.1 Comparing Groups using Multilevel Modelling

P5.1.1 A multilevel model of attainment with school effects

We will start with the simplest multilevel model which allows for school effects on attainment, but without explanatory variables. This 'null' model may be written

$$y_{ij} = \beta_0 + u_j + e_{ij} \qquad (5.1)$$

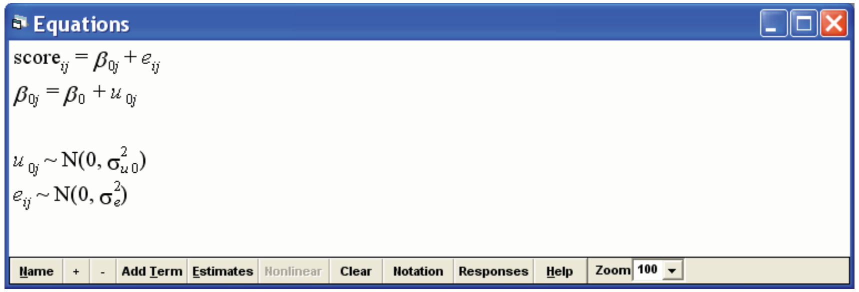
where y_{ij} is the attainment of student i in school j, β_0 is the overall mean across schools, u_j is the effect of school j on attainment, and e_{ij} is a student-level residual. The school effects u_j, which we will also refer to as school (or level 2) residuals, are assumed to follow a normal distribution with mean zero and variance σ_u^2.

To set up this model in MLwiN:

- From the **Model** menu, select **Equations**
- Click **Notation** at the bottom of the **Equations** window, clear the **general** tick box, and click **Done**
- Click on **y** and select **SCORE** from the drop-down list
- Click on **N Levels** and select **2-ij**

- For **level 2(j)**, select **SCHOOLID**
- For **level 1(i)**, select **CASEID**
- Click **Done**
- Click on β_0 , and check **j(schoolid)** to introduce a random school effect, and click **Done**. Click **+** and notice that this step leads to the addition of u_{0j} to the model
- Click **+** again to see the full model specification

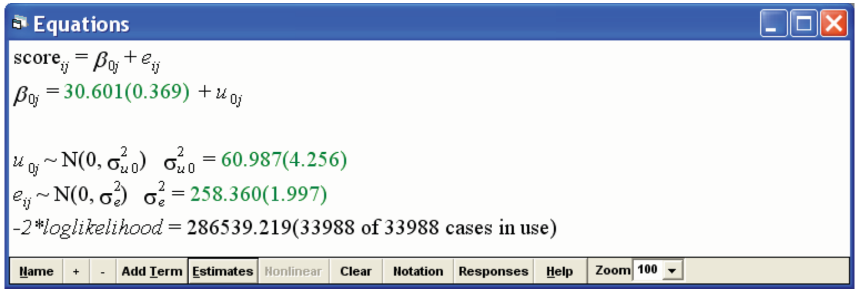
The model should look like this:



If the second equation ($\beta_{0j} = \beta_0 + u_{0j}$) is substituted for β_{0j} in the first equation, we obtain a equation that has the same form as (5.1).

Notice that a '0' subscript has been added to the school effect u_j and its variance σ^2_u, in anticipation of adding further random effects at the school level (see P5.3).

- Click **Start** to fit the model
- Click **Estimates** twice to see the parameter estimates



The overall mean attainment (across schools) is estimated as 30.60. The mean for school j is estimated as 30.60 + $\hat{u}_{0j}$, where $\hat{u}_{0j}$ is the school residual which we will estimate in a moment. A school with $\hat{u}_{0j}$>0 has a mean that is higher than average, while $\hat{u}_{0j}$<0 for a below-average school. (We will obtain confidence intervals for residuals to determine whether differences from the overall mean can be considered 'real' or due to chance.)

Partitioning variance

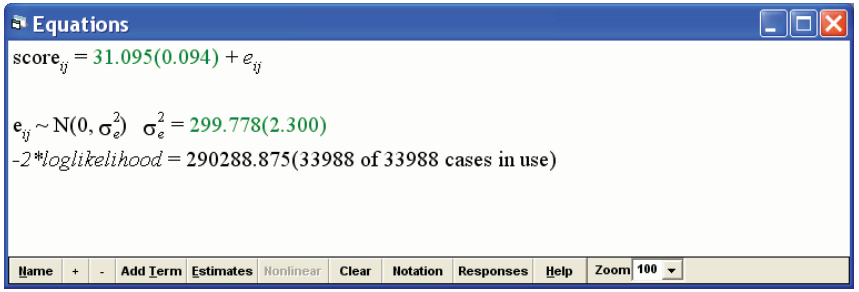
The between-school (level 2) variance in attainment is estimated as $\hat{\sigma}^2_{u0}$=60.99, and the within-school between-student (level 1) variance is estimated as $\hat{\sigma}^2_e$=258.36. Thus the total variance is 60.99+258.36=319.35.

The variance partition coefficient (VPC) is 60.99/319.35 = 0.19, which indicates that 19% of the variance in attainment can be attributed to differences between schools. Note, however, that we have not accounted for intake ability (measured by exams taken on entry to secondary school) so the school effects are not value-added. Previous studies have found that between-school variance in *progress*, i.e. after accounting for intake attainment, is close to 10%.

Testing for school effects

To test the significance of school effects, we can carry out a likelihood ratio test comparing the null multilevel model with a null single-level model. To fit the null single-level model, we need to remove the random school effect:

- In the **Equations** window, click on β_{0j}
- Click on the check box next to **j(schoolid)** to uncheck it, and click **Done**
- The u_{0j} should be removed from the model
- Click **Start** to fit the model



The likelihood ratio test statistic is calculated as the difference in the -2*loglikelihood values for the two models:

LR = 290289 – 286539 = 3750 on 1 d.f. (because there is only parameter difference between the models, σ_{u0}^2).

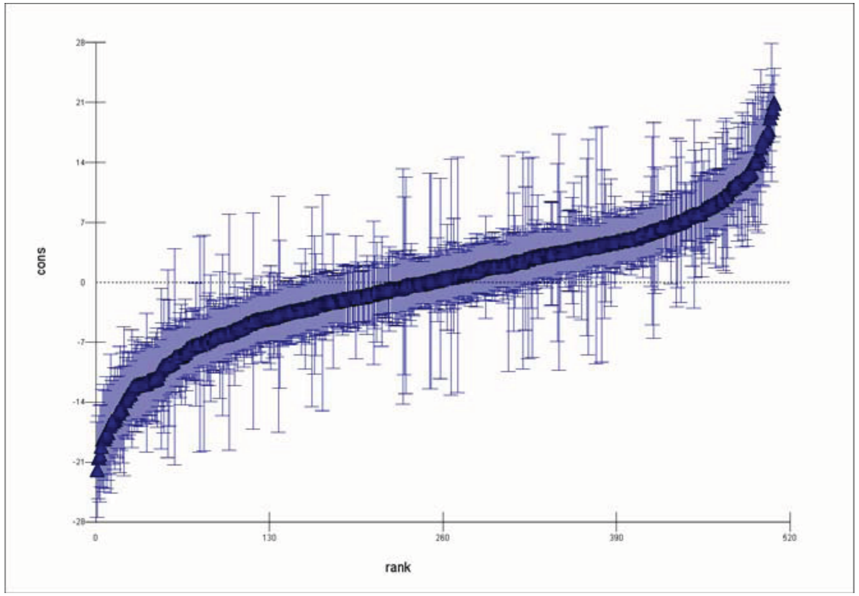
Bearing in mind that the 5% point of a chi-squared distribution on 1 d.f. is 3.84, there is overwhelming evidence of school effects on attainment. We will therefore revert to the multilevel model with school effects.

- In the **Equations** window, click on β_0 (or its estimate 31.095)
- Check **j(schoolid)** and click **Done**
- u_{0j} should be returned to the model
- Click **Start** to fit the model

P5.1.2 Examining school effects (residuals)

To estimate the school-level residuals and produce a caterpillar plot:

- From the **Model** menu, select **Residuals**
- Select the **Settings** tab of the **Residuals** window
- Next to **level:** change from 1:caseid to **2:schoolid**
- In the text box next to **SD(comparative) of residual to** edit 1.0 to **1.96**, so that we obtain 95% confidence limits
- Click **Calc**
- Select the **Plots** tab
- Under **single**, check **residual +/-1.96 sd x rank**
- Click **Apply**



Notice that the confidence intervals around the residual estimates vary greatly in their width; smaller schools will have wider intervals than larger schools.

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.