Module 5: Introduction to Multilevel Modelling

Stata Practical

George Leckie [1]
Centre for Multilevel Modelling

Pre-requisites

- Modules 1-4

Contents

[1] This Stata practical is adapted from the corresponding MLwiN practical: Steele, F. (2008) Module 5: Introduction to Multilevel Modelling. LEMMA VLE, Centre for Multilevel Modelling. Accessed at http://www.cmm.bris.ac.uk/lemma/course/view.php?id=13.

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.[2]

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. 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

[2] 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.

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 33,988 students in 508 schools.

P5.1 Comparing Groups using Multilevel Modelling

Load “5.1.dta” into memory and open the do-file for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 5: Introduction to Multilevel Modelling**, and scroll down to ***Stata* Datasets and Do-files**
- Click “5.1.dta” to open the dataset

and use the `describe` command to produce a summary of the dataset:

```
. describe

Contains data from 5.1.dta
  obs:        33,988
 vars:             9                          3 Sep 2009 09:31
 size:       713,748 (99.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
caseid          float  %9.0g                  Case ID
schoolid        int    %9.0g                  School ID
score           byte   %9.0g                  Score
cohort90        byte   %9.0g                  Cohort
female          byte   %9.0g                  Female
sclass          byte   %9.0g                  Social class
schtype         byte   %9.0g                  School type
schurban        byte   %9.0g                  School urban-rural classification
schdenom        byte   %9.0g                  School denomination
-------------------------------------------------------------------------------
Sorted by:
```

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

$$\textbf{score}_{ij} = \beta_0 + u_{0j} + e_{ij}$$

where $\textbf{score}_{ij}$ is the attainment of student i in school j, β_0 is the overall mean across schools, u_{0j} is the effect of school j on attainment, and e_{ij} is a student-level residual. The school effects u_{0j}, 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^2_{u0}.

Stata's main command for fitting multilevel models for continuous response variables is the `xtmixed` command.[3] To fit the above model using the `xtmixed` command, we type: `xtmixed score || schoolid:, mle variance nostderr`.

The response variable (**score**) follows the command which is then followed by the list of fixed part explanatory variables (excluding the constant as this is included by default[4]). The above model contains only an intercept and so no fixed part explanatory variables are specified. The level 2 random part of the model is specified after two vertical bars `||`. The level 2 identifier (**schoolid**) is specified first followed by a colon and then the list of random part explanatory variables (again excluding the constant as this is included by default). The `mle` option is used to request maximum likelihood estimation (as opposed to the default of restricted maximum likelihood estimation). The `variance` option reports the variances of the random intercept and any random coefficients included in the model (as opposed to the default of standard deviations). The `nostderr` option is specified to avoid calculating standard errors for the random part parameters. This speeds up the time it takes to fit each `xtmixed` model and we can still use likelihood ratio tests to compare nested models with different random part specifications.

[3] Note, two-level random intercept models can equally be fitted with the `xtreg` command (with the `mle` option); see `help xtreg`. We do not discuss the `xtreg` command as it cannot be used to fit more complicated multilevel models while `xtmixed` can. However, we do note that `xtreg` (with the `mle` option) fits models considerably faster than `xtmixed` and is therefore recommended for fitting two-level random intercept models. See Rabe-Hesketh and Skrondal (2008) for examples of two-level random intercept models fitted with both commands.
[4] Note, the `noconstant` option can be used to omit the constant; see `help xtmixed`.

Module 5 (Stata Practical): Introduction to Multilevel Modelling
P5.1 Comparing Groups using Multilevel Modelling

Issuing the `xtmixed` command gives the following output:

```
. xtmixed score || schoolid:, mle variance nostderr

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -143269.53
Iteration 1:   log likelihood = -143269.53

Mixed-effects ML regression                     Number of obs      =     33988
Group variable: schoolid                        Number of groups   =       508

                                                Obs per group: min =         1
                                                               avg =      66.9
                                                               max =       190


                                                Wald chi2(0)       =         .
Log likelihood = -143269.53                     Prob > chi2        =         .

------------------------------------------------------------------------------
       score |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    30.6006   .3694317    82.83   0.000     29.87652    31.32467
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
schoolid: Identity           |
                  var(_cons) |   61.02457          .             .           .
-----------------------------+------------------------------------------------
               var(Residual) |   258.3572          .             .           .
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  3749.78 Prob >= chibar2 = 0.0000
```

Before interpreting the model, we will discuss the estimation procedure that `xtmixed` uses.[5] The default estimation option is to fit the model using the EM (expectation maximisation) algorithm until convergence (or 20 iterations have been reached). At that point, maximization switches to a gradient-based method, unless the `emonly` option is specified, in which case maximization stops.[6] In the analysis which follows we will mainly use this default estimation option.

While the default estimation options are normally the preferred approach, complicated models can be very slow to iterate. The advantage of specifying `emonly` is that EM iterations are typically much faster than those for gradient-based methods. However, the disadvantage is that it can take a large number of EM iterations to converge (if at all).

[5] For further details see `help xtmixed`.
[6] By default, the gradient-based method is Newton-Raphson iterations, but other methods are available by specifying the appropriate maximize options; see `help xtmixed`.

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.)

Before we continue, we store the results using the `estimates store` command:

```
. estimates store nullmodel
```

We can then explore other model specifications with the option of restoring these estimates later (by using the `estimates restore` command) without having to refit this model. This will be particularly helpful when we fit more complex models that are slower to converge. We can even store each model we fit under a different name so that we can restore any previously fitted model at a later point.

Partitioning variance

The between-school (level 2) variance `var(_cons)` in attainment is estimated as $\hat{\sigma}^2_{u0}$ = 61.02, and the within-school between-student (level 1) variance `var(Residual)` is estimated as $\hat{\sigma}^2_e$ = 258.36. Thus the total variance is 61.02 + 258.36 = 319.38.

The variance partition coefficient (VPC) is 61.02/319.38 = 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:

$$\textbf{score}_{ij} = \beta_0 + e_{ij}$$

```
. xtmixed score, mle variance nostderr

Mixed-effects ML regression                     Number of obs      =     33988

                                                Wald chi2(0)       =         .
Log likelihood = -145144.42                     Prob > chi2        =         .

------------------------------------------------------------------------------
       score |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   31.09462   .0939156   331.09   0.000     30.91055    31.27869
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
                var(Residual) |   299.7787          .             .           .
------------------------------------------------------------------------------
```

The likelihood ratio test statistic is calculated as two times the difference in the log likelihood values for the two models:

LR = 2(-143269.53 - -145144.42) = 3750 on 1 d.f. (because there is only one parameter difference between the models, σ^2_{u0}).

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.[7]

Note, the `xtmixed` command automatically compares the specified model with the equivalent single-level model. The likelihood ratio test statistic for this comparison can be seen in the last line of the `xtmixed` output of the first model we fitted: `chibar2(01) = 3749.78`. Note that there is not a corresponding likelihood ratio test statistic for the second model we fitted as this model is a single-level model.

[7] Note that this test statistic has a non-standard sampling distribution as the null hypothesis of a zero variance is on the boundary of the parameter space; we do not envisage a negative variance. In this case the correct p-value is half the one obtained from the tables of chi-squared distribution with 1 degree of freedom. In the output of the `xtmixed` command, Stata automatically reports the correct p-value for this test. See `help j_xtmixedlr` for further details.