Module 5: Introduction to Multilevel Modelling

R Practical

Camille Szmaragd and George Leckie[1]
Centre for Multilevel Modelling

Pre-requisites

- Modules 1-4

Contents

[1] This R 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
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)

[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 and Raffe (2006).

There are 33,988 students in 508 schools.

P5.1 Comparing Groups using Multilevel Modelling

Download the R dataset for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 5: Introduction to Multilevel Modelling**, and scroll down to **R Datasets and R files**
- Right click "5.1.txt" and select **Save Link As …** to save the dataset to your computer

Read the dataset into R using the `read.table` command and create a dataframe object named **mydata**[3]**:**

```
> mydata <- read.table(file = "5.1.txt", sep = ",", header = TRUE)
```

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

```
> str(mydata)
'data.frame':   33988 obs. of  9 variables:
 $ caseid  : int  18 17 19 20 21 13 16 14 15 12 ...
 $ schoolid: int  1 1 1 1 1 1 1 1 1 1 ...
 $ score   : int  0 10 0 40 42 4 0 0 14 27 ...
 $ cohort90: int  -6 -6 -6 -6 -6 -6 -6 -6 -6 -6 ...
 $ female  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ sclass  : int  2 2 4 3 2 2 3 4 3 2 ...
 $ schtype : int  0 0 0 0 0 0 0 0 0 0 ...
 $ schurban: int  1 1 1 1 1 1 1 1 1 1 ...
 $ schdenom: int  0 0 0 0 0 0 0 0 0 0 ...
```

[3] At the beginning of your R session, you will need to set R's working directory to the file location where you saved the dataset. This can be done using the command line and the function `setwd`:

```
> setwd("C:/userdirectory")
```

Or through selecting Change dir… on the File menu.

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:

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

where $\mathbf{score}_{ij}$ 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}.

R's main command for fitting multilevel models is part of the additional `lme4`[4] library which can be installed through the R Packages menu; select Install Package(s) and then select the correct Mirror and package from the scroll-down menus. As you will see, there is a variety of additional packages that can be installed with R. You only need to install a package once to your own computer. If you then want to use the package, you simply need to call it from within R prior to using the command for the first time in each R session. This can be done with the `library()` function and in this case `library(lme4)`.

```
> library(lme4)
Loading required package: Matrix
Loading required package: lattice

Attaching package: 'lme4'


        The following object(s) are masked from package:stats :

         AIC
```

The output informs us that R has loaded two additional packages `Matrix` and `lattice` which are required for the `lme4` package to work. We are also told that the `AIC` object is masked from a third package `stats`. This means that when you call these commands you need to specify from which packages you are calling them from. This is done by using the name of the package followed by two '`:`' and then the name of the command; for instance in this case `stats::AIC`.

We will use the `lmer()` function from the `lme4` library to fit the above model. The syntax for this function is very similar to the syntax used for the `lm()` function for multiple regression which we introduced in Module 3.[5] Below we choose to store the model as a new object called **nullmodel**:

```
> nullmodel <- lmer(score ~ (1 | schoolid), data = mydata, REML = FALSE)
```

[4] `lme4` is a package developed by Douglas Bates and Martin Maechler for fitting linear and generalized linear mixed-effect models.
[5] To obtain details of the different options available for the `lmer()` function, just type `help("lmer")`

The response variable (**score**) follows the command which is then followed by a ~ and then by a list of fixed part explanatory variables (excluding the constant as this is included by default)[6]. 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 in brackets by the list of random part explanatory variables (the constant has to be explicitly specified by `1`, followed by a single vertical bar **|** and then by the level 2 identifier (**schoolid**). The `data` option specifies the dataframe being used to fit the model. The `REML = FALSE` option is used to request maximum likelihood estimation (as opposed to the default of restricted maximum likelihood estimation).

[6] Note, to omit the constant you need to add `-1` to the right-hand side of the "~" sign.

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

We then display the results using the `summary` command, which gives the following output:

```
> summary(nullmodel)
Linear mixed model fit by maximum likelihood

Formula: score ~ (1 | schoolid)
   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 286545 286570 -143270   286539  286539

Random effects:
 Groups   Name        Variance Std.Dev.
 schoolid (Intercept)  61.024   7.8118
 Residual             258.357  16.0735
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept)  30.6006     0.3693   82.85
```

Before interpreting the model, we will discuss the estimation procedure that `lmer` uses.[7] The estimation procedure used by both MLE and REML is based on optimizing a function of the log likelihood using penalized iteratively re-weighted least squares. The log-likelihood is evaluated using an adaptive Gauss-Hermite approximation, which, when using the default value of one, reduces to the Laplacian approximation. This default approximation can be changed by using the `nAGQ = n` option, where `n` is an integer greater than one, representing the number of points used for evaluating the adaptive Gauss-Hermite approximation. The greater the value of `n`, the more accurate the evaluation of the log-likelihood, but the longer it takes to fit the model.

The output of `lmer` consists of three parts. The first part under `Formula`: and `Data:` reports a range of summary statistics (`AIC`, `BIC`, `LogLik`,…). The second part under `Random effects:` summarises the variance and standard deviation of each random effect (including the level 1 model residuals). Underneath the random effects table, the total number of observations is provided along with the number of units (or groups) for each higher level in the model. Here, schools are our only higher level and the output reports that we have 508 different schools. The final part of the output is the `Fixed effects:` table which reports the parameter estimate (`Estimate`) standard error (`Std. Error`) and t-value (`t value`), for each parameter in the model. For models with more than one fixed part explanatory variable (including the intercept), a correlation table between these variables is also provided underneath the table of parameter estimates (see later examples).

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

[7] For further details see the PDF vignettes available on the `lme4` website http://cran.r-project.org/web/packages/lme4, in particular the vignette entitled "Computational Methods" which deals with the statistical theory.

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 `schoolid (Intercept)` in attainment is estimated as $\hat{\sigma}^2_{u0}$ = 61.02, and the within-school between-student (level 1) variance `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:

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

```
> fit <- lm(score ~ 1, data = mydata)

> summary(fit)

Call:
lm(formula = score ~ 1, data = mydata)

Residuals:
    Min      1Q  Median      3Q     Max
-31.095 -12.095   1.905  13.905  43.905

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 31.09462    0.09392   331.1   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.31 on 33987 degrees of freedom
```

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

You can obtained the log likelihood value for a model with the `loglik` command:

```
> logLik(nullmodel)
'log Lik.' -143269.5 (df=3)

> logLik(fit)
'log Lik.' -145144.4 (df=2)
```

LR = 2(-143269.5 - -145144.4) = 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.[8]

P5.1.2 Examining school effects (residuals)

To estimate the school-level residuals $\hat{u}_{0j}$ and their associated standard errors, we use the `ranef` command with the `postVar` option. This create a random effects object, containing the variance-covariance matrix in the `postVar` attribute.

```
> u0 <- ranef(nullmodel, postVar = TRUE)

> u0se <- sqrt(attr(u0[[1]], "postVar")[1, , ])
```

The 508 school level residuals are stored in **u0**, a type of R object called a list. It is actually a list of lists. The first and unique element of the list, u0[1], is the list corresponding to the first set of random effects. We can obtain a description of `u0[1]` by using the `str` command:

```
> str(u0[1])
List of 1
 $ schoolid:'data.frame':       508 obs. of  1 variable:
  ..$ (Intercept): num [1:508] -11.84 3.21 3.4 -7.42 3.43 ...
  ..- attr(*, "postVar")= num [1, 1, 1:508] 5.71 1.7 2.24 4.29 2.66 ...
```

The first line of the output confirms that there are 508 schools in the data. The second line (`$ (Intercept)`) lists the school effects while the third line corresponding to the "`postVar`" attribute lists their associated posterior variances.

In our case there is only one set of random effects and therefore **u0[1]** is a list of only one object, u0[[1]]. u0[[1]] is itself a dataframe containing the school-level residuals and the "posterior variances" of these residuals within the attribute `postVar`: To access the elements of this dataframe, we need to use two sets of square brackets as opposed to one set of square bracket.

[8] 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.

You can see the difference the second pair of square brackets makes by using the `str` command:

```
> str(u0[[1]])
'data.frame':   508 obs. of  1 variable:
 $ (Intercept): num  -11.84 3.21 3.4 -7.42 3.43 ...
 - attr(*, "postVar")= num [1, 1, 1:508] 5.71 1.7 2.24 4.29 2.66 ...
```

The output seems similar to the previous one, except that the name of the higher level random-effect is no longer specified.

R uses lists in particular to associate specific attributes to data objects such as dataframes or vectors. By default, R returns a dataframe for the random effects, even when there is only one set of random effect.

As there is only one set of random effects, the `postVar` attribute only contains the "posterior variance" of each school-level residual. To access this set of variances, we look into the attribute `postVar` of the dataframe **u0[[1]]**. This returns a three-dimensional array with the third dimension referring to each individual residual. To reduce this array into a simple vector containing the "posterior variances" for each residual, we use `attr(u0[[1]], "postVar")[1,,]`. To view the first few elements of this vector, we can use the `head` command:

```
> head(attr(u0[[1]], "postVar")[1, , ])
[1] 5.714615 1.698067 2.243282 4.291562 2.658550 1.969786
```

The school-level residuals and their standard errors have been calculated and stored for each individual school. We can therefore calculate summary statistics and produce graphs based on these data.

Next we create a dataframe containing an identifier, residual and standard error for every school:

```
> schoolid <- as.numeric(rownames(u0[[1]]))

> u0tab <- cbind(schoolid, u0[[1]], u0se)

> colnames(u0tab) <- c("schoolid", "u0", "u0se")
```

We then sort this table by ascending order based on the values of **u0:**

```
> u0tab <- u0tab[order(u0tab$u0), ]
```

and create a new column containing the ranks:

```
> u0tab <- cbind(u0tab, c(1:dim(u0tab)[1]))

> colnames(u0tab)[4] <- "u0rank"
```

We finally reorder the table based on the school identifier:

```
> u0tab <- u0tab[order(u0tab$schoolid), ]
```


To see the school residual, standard error and ranking for a particular school, we can list the data by using the indexing structure of the R dataframe. Here we do this for the first 10 schools in the data.

```
> u0tab[1:10, ]
   schoolid         u0     u0se u0rank
1         1 -11.844059 2.390526     37
2         2   3.207216 1.303099    337
3         3   3.396920 1.497759    344
4         4  -7.416852 2.071609     73
5         5   3.427138 1.630506    345
6         6  12.437109 1.403491    487
7         7  -1.652372 1.460226    199
8         8  20.984041 2.021872    508
9         9  -8.693975 6.438403     59
10       10   1.737830 1.904961    291
```

From these values we can see, for example, that school 1 had an estimated residual of -11.84 which was ranked 37, i.e. 37 places from the bottom. For this school, we estimate a mean score of 30.60 – 11.84 = 18.76. In contrast, the mean for school 8 (ranked 508, the highest) is estimated as 30.60 + 20.98 = 51.58.

Finally, we use the `plot` and `segments` commands to produce a 'caterpillar plot' to show the school effects in rank order together with 95% confidence intervals.

We start by creating the plot but without plotting any data

```
> plot(u0tab$u0rank, u0tab$u0, type = "n", xlab = "u_rank", ylab = "conditional
modes of r.e. for school_id:_cons")
```

By using the `type = "n"` option, we create the axis for the plot but prevent any data from being plotted.

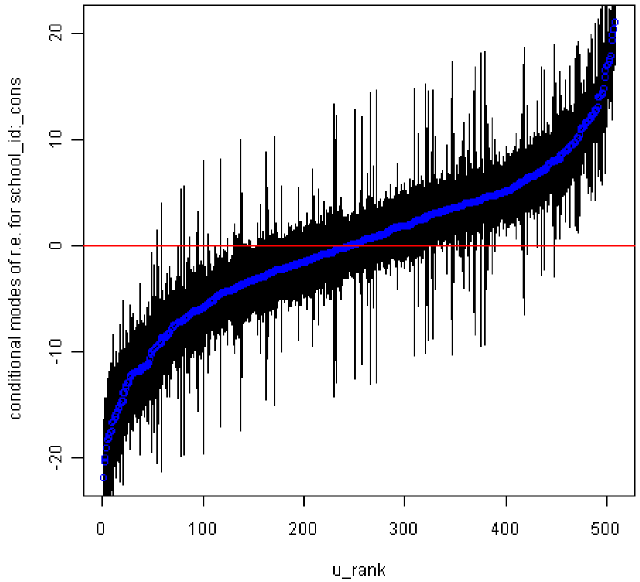
We then add to the plot the 95% confidence intervals by using the `segments` command:

```
> segments(u0tab$u0rank, u0tab$u0 - 1.96*u0tab$u0se, u0tab$u0rank, u0tab$u0 +
1.96*u0tab$u0se)
```

The `segments` command takes at least four arguments corresponding to the pair of (x,y) coordinates, corresponding to the two end points of the segments, or in this case the lower and upper values of the 95% confidence intervals.

We can finally superimpose onto the confidence interval the point estimates, and a horizontal line corresponding to y = 0, the average school.

```
> points(u0tab$u0rank, u0tab$u0, col = "blue")

> abline(h = 0, col = "red")
```



The symbol for the point estimates can be changed using the `pch` option[9], similarly the colour for the confidence intervals can be changed using the `col` option.

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

Note that because we have not accounted for intake ability, we cannot interpret these residuals as "school effects" in the value-added sense that it is used in school effectiveness research. Unfortunately, no measure of prior attainment is available from the Scottish School Leavers Survey. Nevertheless, exam performance at age 16 is an important educational outcome because it is a strong predictor of post-16 educational attainment and entry to university depends on attainment rather than progress. In these exercises, we will study trends in mean attainment and variation in attainment between individuals and between schools.

[9] For more details on the plotting option type `help(points)` or `help(plot)`.

Don't forget to take the online quiz!

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

P5.2 Adding Student-level Explanatory Variables: Random Intercept Models

Download the R dataset for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 5: Introduction to Multilevel Modelling**, and scroll down to **R Datasets and R files**
- Right click "5.2.txt" and select **Save Link As...** to save the dataset to your computer

Read the dataset into R

```
> mydata <- read.table("5.2.txt", sep = ",", header = TRUE)
```

and load the `lme4` library:

```
> library(lme4)
Loading required package: Matrix
Loading required package: lattice

Attaching package: 'lme4'


        The following object(s) are masked from package:stats :

         AIC
```

We begin by allowing for a linear cohort effect:

$$\mathbf{score}_{ij} = \beta_0 + \beta_1 \mathbf{cohort90}_{ij} + u_{0j} + e_{ij}$$

```
> fit <- lmer(score ~ cohort90 + (1 | schoolid), data = mydata, REML = FALSE)

> summary(fit)
Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + (1 | schoolid)
   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 280922 280955 -140457   280914  280921
Random effects:
 Groups   Name        Variance Std.Dev.
 schoolid (Intercept)  45.988   6.7815
 Residual             219.288  14.8084
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 30.55913    0.32250   94.76
cohort90     1.21496    0.01553   78.24

Correlation of Fixed Effects:
         (Intr)
cohort90 -0.002
```

The equation of the average fitted regression line (across schools) is

$$\hat{\mathbf{score}}_{ij} = 30.559 + 1.215\,\mathbf{cohort90}_{ij}$$

The fitted line for a given school will differ from this average line in its intercept, by an amount $\hat{u}_{0j}$ for school j. However, the slope of the school lines is assumed to be fixed at 1.215, i.e. the effect of cohort is assumed the same for all schools. A plot of the predicted school lines will show a set of parallel lines. To produce this plot, we first need to compute **scôre** for each student, based on their cohort and school. We do this using the `fitted` command to create a new variable (**predscore**) which is equal to the average fitted regression line plus the relevant school's intercept:

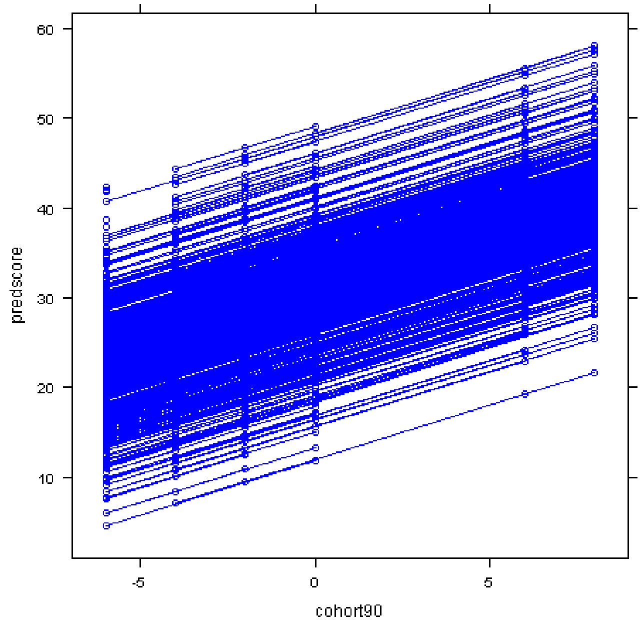
```
> predscore <- fitted(fit)
```

Next we create a variable to pick out the minimum amount of data required to plot the predicted school lines (see P3.1.2 in Module 3 where we explain this approach in detail):.

```
> datapred <- unique(data.frame(cbind(predscore = predscore, cohort90 =
mydata$cohort90, schoolid = mydata$schoolid)))
```


We will use the `xyplot` command in order to display markers (i.e. the data points) in addition to the school lines. This is obtained by using the `type = c("p", "l")` option, where `"p"` stands for points and `"l"` for lines.

```
> xyplot(predscore ~ cohort90, data = datapred, groups = schoolid, type = c("p",
"l"), col = "blue")
```



Careful examination of the top left hand corner of the graph shows that a small number of schools are observed for only one cohort.

We may thus want to reproduce the graph for the subset of schools which are observed for two or more cohorts. To do this, we first sort the dataframe by **schoolid** and **cohort90** using the `order` command, and then we create a new variable **multiplecohorts** and initially set its values equal to 0.

```
> datapred <- datapred[order(datapred$schoolid, datapred$cohort90), ]

> datapred$multiplecohorts <- rep(0, length(datapred$schoolid))
```

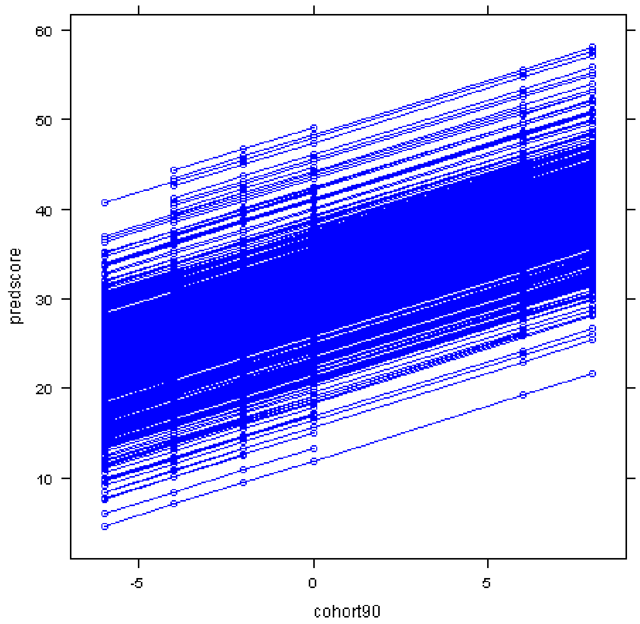
We will then replace **multiplecohorts** with the value 1 for those schools observed for more than one cohort. The relevant command is:

```
> datapred$multiplecohorts[datapred$schoolid %in%
unique(datapred$schoolid[duplicated(datapred$schoolid)])] <- 1
```

In this command, we have used the `unique` and `duplicated` commands. The `duplicated(datapred$schoolid)` command returns the list of indexes of **schoolid** which appear more than once in the dataset. However if the same **schoolid** appears more than twice in the dataset, for example as it does for **schoolid=1** which corresponds to the first, the second and third record in the data, `duplicated(datapred$schoolid)` will return 2 and 3, that is the 2nd and 3rd position in the original dataset. By applying the `unique` command to the value of **schoolid** for the positions returned by `duplicated`, we can obtain the list of unique **schoolid** which appear in more than one cohort. Finally, we replace the value of **multiplecohorts** in the dataset for every record for which the **schoolid** is in the list of schools with more than one cohort.

Now we can simply repeat the previous `xyplot` command, but this time we condition upon **multiplecohorts** taking the value 1.

```
> xyplot(predscore ~ cohort90, data = datapred[datapred$multiplecohorts == 1, ],
groups = schoolid, type = c("p", "l"), col = "blue")
```



The graph now contains only lines for schools which are observed in more than one cohort.

Module 5 (R Practical): Introduction to Multilevel Modelling
P5.2 Adding Student-level Explanatory Variables: Random Intercept Models

Returning to the results and comparing with the results for the null model of P5.1, we can see that the addition of cohort has reduced the amount of variance at both the school and the student level. The between-school variance has reduced from 61.02 to 45.99, and the within-school variance has reduced from 258.36 to 219.29. The decrease in the within-school variance is expected because cohort is a student-level variable. The large reduction in the between-school variance suggests that the distribution of students by cohort differs from school to school (see C5.2.3). In Module 3 (C3.1.1) we found that, pooling across all schools, the proportions in each cohort were:

Table 5.1. Proportion of students in each cohort

Year	**1984**	**1986**	**1988**	**1990**	**1996**	**1998**
% students	19.1	18.6	15.4	12.9	12.5	21.6

One source of the variation in these proportions across schools can be seen from the plot of the predicted lines above. If you look at the top line (corresponding to the school with the highest intercept), you can see that there are only three predicted values, for **cohort90** = -4, -2 and 0 (1986, 1988 and 1990). This is because, in this school, no data were collected for 1984, 1996 and 1998. Clearly, for this school, the proportions for the missing years will be zero. Similarly, for the school with the second *lowest* intercept, there are no data points for the last two cohorts (**cohort90** = 6 and 8).

After accounting for cohort effects, the proportion of unexplained variance that is due to differences between schools decreases slightly to 45.99/(45.99 + 219.29) = 17%.

Don’t forget to take the online quiz!

From within the LEMMA learning environment

- Go down to the section for **Module 5: Introduction to Multilevel Modelling**
- Click " [5.2 Multilevel Regression with a Level 1 Explanatory Variable: Random Intercept Models](#)"
  to open Lesson 5.2
- Click Q 1 to open the first question

P5.3 Allowing for Different Slopes across Schools: Random Slope Models

In the previous exercise, we allowed for school effects on the mean attainment by allowing the *intercept* of the regression of attainment on cohort to vary randomly across schools. We assumed, however, that cohort changes in attainment are the same for all schools, i.e. the *slope* of the regression line was assumed fixed across schools. We will now extend the random intercept model fitted at the end of P5.2 to allow both the intercept and the slope to vary randomly across schools.

Download the R dataset for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 5: Introduction to Multilevel Modelling**, and scroll down to **R Datasets and R files**
- Right click "5.3.txt" and select **Save Link As…** to save the dataset to your computer

Read the dataset into R:

```
> mydata <- read.table("5.3.txt", header = TRUE, sep = ",")
```

Load the `lme4` library:

```
> library(lme4)
Loading required package: Matrix
Loading required package: lattice

Attaching package: 'lme4'


        The following object(s) are masked from package:stats :

         AIC
```

Fit the model:

$$\textbf{score}_{ij} = \beta_0 + \beta_1\textbf{cohort90}_{ij} + u_{0j} + u_{1j}\textbf{cohort90}_{ij} + e_{ij}$$

Note that a new term u_{1j} has been added to the model, so that the coefficient of **cohort90** has become $\beta_{1j} = \beta_1 + u_{1j}$, and so the community-level variance has been replaced by a matrix with two new parameters, σ^2_{u1} and σ_{u01}.

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim MVN(\mathbf{0}, \mathbf{\Omega}_u), \qquad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Omega}_u = \begin{pmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{pmatrix}$$

Note that the slope residual, and associated variance and covariance, have a subscript of '1' because **cohort90** is the 1st explanatory variable in the model (not including the constant).

Fit the model:

```
> fit <- lmer(score ~ cohort90 + (1 + cohort90 | schoolid), data = mydata, REML
= FALSE)

> summary(fit)
Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + (1 + cohort90 | schoolid)
   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 280698 280749 -140343   280686  280692
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  42.85809  6.54661
          cohort90      0.16059  0.40074 -0.390
 Residual             215.73930 14.68807
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 30.60970    0.31345   97.66
cohort90     1.23391    0.02532   48.74

Correlation of Fixed Effects:
         (Intr)
cohort90 -0.267
```

In the output, the estimate of the intercept variance σ^2_{u0} is given in the `Random effects:` table to the right of the `(Intercept)` while the estimate of the slope variance σ^2_{u1} is given to the right of `cohort90`. The estimate of the correlation coefficient ρ_{u01} is given on the same line as the slope variance under the column `Corr.` By defining the random intercepts and slopes together, we indirectly specify that we want the random intercepts and slopes to covary. If we wanted to have independent random intercepts and slopes we would need to replace `(1 + cohort90 | schoolid)` in our R syntax by `(1 | schoolid) + (0 + cohort90 | schoolid)`.

P5.3.1 Testing for random slopes

We can use a likelihood ratio test to test whether the cohort effect varies across schools. The null hypothesis for this test is that the two additional parameters σ_{u01} and σ^2_{u1} are simultaneously equal to zero. The log-likelihood value for the random intercept model was found to be -140457 (P5.2), so the likelihood ratio test statistic is

LR = 2 (-140343 - -140457) = 228 on 2 d.f.

R can actually conduct the likelihood ratio test directly by using the anova command[10]:

```
> fita <- lmer(score ~ cohort90 + (1 | schoolid), data = mydata, REML = FALSE)

> anova(fit, fita)
Data: mydata
Models:
fita: score ~ cohort90 + (1 | schoolid)
fit: score ~ cohort90 + (1 + cohort90 | schoolid)
     Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
fita  4 280922 280955 -140457
fit   6 280698 280749 -140343 227.40      2  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

So there is very strong evidence that the cohort effect differs across schools.

P5.3.2 Interpretation of random cohort effects across schools

The cohort effect for school j is estimated as $1.234 + \hat{u}_{1j}$, and the between-school variance in these slopes is estimated as 0.160. For the ‘average’ school we predict an increase of 1.234 points in the attainment score for each successive cohort. A 95% coverage interval for the school slopes is estimated as $1.234 \pm 1.96\sqrt{0.160}$ = 0.45 to 2.018. Thus, assuming a normal distribution, we would expect the middle 95% of schools to have a slope between 0.45 and 2.018.

The intercept variance of 42.858 is interpreted as the between-school variance when **cohort90** = 0, i.e. for the 1990 cohort.

P5.3.3 Examining intercept and slope residuals for schools

The negative covariance estimate of -1.024 means that schools with a high intercept (above-average attainment in 1990) tend to have a flatter-than-average slope. Similarly, schools with a low intercept (below-average attainment in 1990) tend to have seen a more marked increase in attainment with cohort (above-average slope).

The intercept-slope correlation is estimated as:

$$\hat{\rho}_{u01} = \frac{\hat{\sigma}_{u01}}{\sqrt{\hat{\sigma}^2_{u0}\,\hat{\sigma}^2_{u1}}} = -0.390$$

[10] For more details about the anova command please refer to the end of section 3.4.4 of Module 3.

Module 5 (Practice): Intro to Multilevel Modelling

We can obtain the estimated correlation directly from the table. Alternatively, we can obtain the correlation table for the random effects by using the VarCorr command:

```
> VarCorr(fit)
$schoolid
            (Intercept)   cohort90
(Intercept)   42.858092 -1.0241779
cohort90      -1.024178  0.1605916
attr(,"stddev")
(Intercept)    cohort90
  6.5466092   0.4007388
attr(,"correlation")
            (Intercept)  cohort90
(Intercept)    1.000000 -0.390389
cohort90      -0.390389  1.000000

attr(,"sc")
[1] 14.68807
```

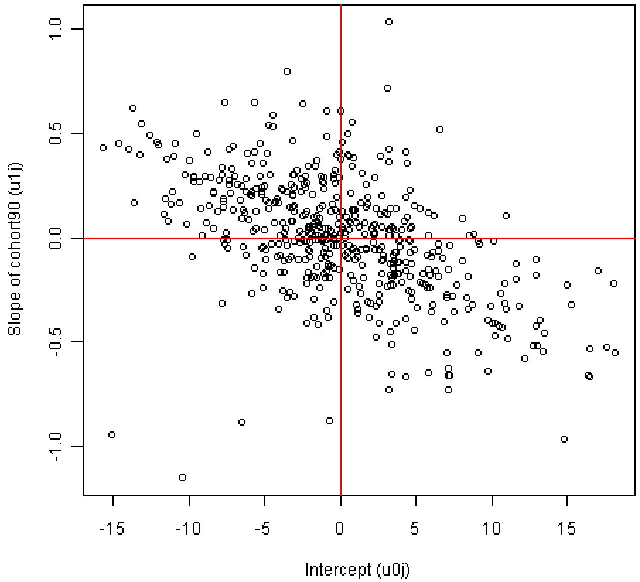
The first table contains the variance/covariance matrix for the random effects. The value of the covariance $\hat{\sigma}_{u01}$ is -1.024. The second table (attr(, ”stddev”)) contains the standard deviation of the variance estimates of the random intercepts and slopes. The third table (attr(, ”correlation”)) is the correlation table between the random intercepts and slopes. Finally the last value (attr(, ”sc”)) returns the standard deviation of the level 1 residuals.

To estimate the school intercepts and slopes we use the raneff command with the postVar option. We specify a new data object **myrandomeff** which will contain the values of the random slopes and intercepts along with their posterior variance:

```
> myrandomeff <- ranef(fit, postVar = TRUE)
```

To obtain a plot of the school slopes versus the school intercepts, $\hat{u}_{1j}$ vs. $\hat{u}_{0j}$:

```
> plot(myrandomeff[[1]], xlab = "Intercept (u0j)", ylab = "Slope of cohort90
(u1j)")

> abline(h = 0, col = "red")

> abline(v = 0, col = "red")
```



where we have used the `xlab` and `ylab` options to add axes titles to the graph.

From this plot, it is possible to identify, for example, those schools which had a lower-than-average attainment in 1990 but a better-than-average year-on-year improvement. Schools in the top-left quadrant are such schools while schools in the bottom-left quadrant also had a below-average mean attainment in 1990, but the below-average slopes for these schools means that they continued at this low level.

The equation for the fitted regression line for school j is

$$\hat{\mathbf{score}}_{ij} = (30.610 + \hat{u}_{0j}) + (1.234 + \hat{u}_{1j})\mathbf{cohort90}_{ij}$$

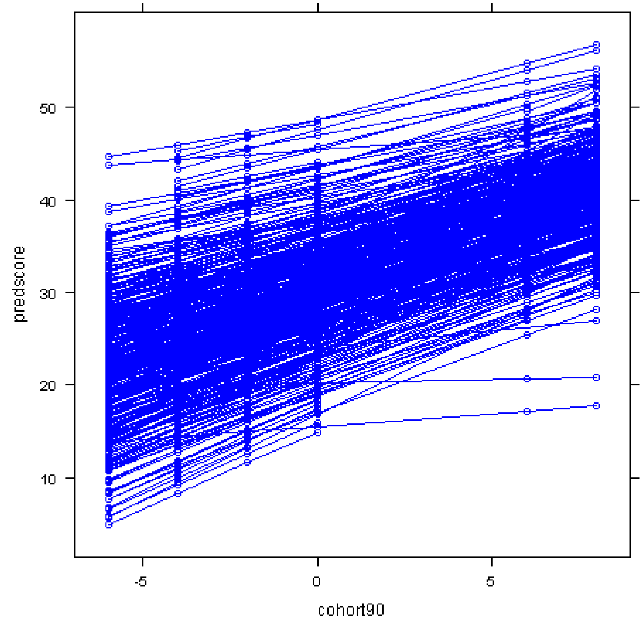
where the values of $\hat{u}_{0j}$ and $\hat{u}_{1j}$ are shown in the pairwise residual plot shown above.

To produce a plot of the predicted school lines, we first need to compute **scôre** for each student, based on their cohort and school:

```
> predscore <- fitted(fit)
```


As in P5.2, we plot the fitted regression lines for the subset of schools for which we have multiple cohorts of data:

```
> datapred <- cbind(predscore = predscore, cohort90 = mydata$cohort90, schoolid
= mydata$schoolid)

> datapred <- data.frame(unique(datapred))

> datapred <- datapred[order(datapred$schoolid, datapred$cohort90), ]

> datapred$multiplecohorts <- rep(0, dim(datapred)[1])

> datapred$multiplecohorts[datapred$schoolid %in%
unique(datapred$schoolid[duplicated(datapred$schoolid)])] <- 1

> xyplot(predscore ~ cohort90, data = datapred[datapred$multiplecohorts == 1, ],
groups = schoolid, type = c("p","l"), col = "blue")
```



P5.3.4 Between-school variance as a function of cohort

The random slope model we have fitted implies that the between-school variance in attainment is a function of cohort; that is, the amount of between-school variance differs across cohorts.

In C5.3.5 (Equation 5.9), we saw that for a model with a random slope for an explanatory variable x_{ij}, the level 2 variance is:

$$\begin{aligned}\text{var}(u_{0j} + u_{1j}x_{ij}) &= \text{var}(u_{0j}) + 2x_{ij}\,\text{cov}(u_{0j}, u_{1j}) + x_{ij}^2\,\text{var}(u_{1j}) \\ &= \sigma_{u0}^2 + 2\sigma_{u01}x_{ij} + \sigma_{u1}^2 x_{ij}^2\end{aligned}$$

Substituting **cohort90** for x, and the estimates for σ_{u0}^2, σ_{u01} and σ_{u1}^2, we obtain:

Between-school variance = 42.859 – 2.048 cohort90 +0.161 cohort90^2

Applying this equation to selected cohorts we obtain the following estimates of level 2 variance.

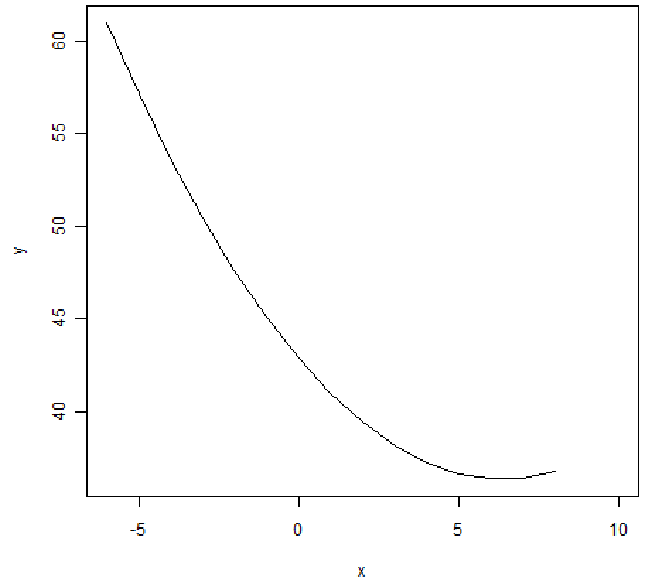
Table 5.2. Estimates of the between-school variance

cohort90	Year	Between-school variance
-6	1984	42.859 – (2.048 × -6) + [0.161 × (-6)^2] = 60.943
0	1990	42.859
6	1996	42.859 – (2.048 × 6) + (0.161 × 6^2] = 36.367

We would therefore conclude that the mean attainment increased with cohort, and the *variation* in mean attainment among schools has decreased.

We can produce a plot of the between-school variance with the `plot` command. We create two new vectors; the first one containing cohort values restricted to the cohorts in the data (1984 to 1998) and the second one corresponding to the equation for the between-school variance.

```
> x <- c(-6:8)

> y <- 42.859 - 2.048*x + 0.161*x^2

> plot(x, y, type = "l", xlim = c(-6, 10))
```



P5.3.5 Adding a random coefficient for gender (dichotomous x)

In Module 3, we found that the mean attainment was higher for girls than for boys. We will now consider whether this gender difference is the same across schools by introducing a random coefficient for gender.[11]

We will start be adding a fixed effect for gender. This will be our comparison model for testing for a random coefficient.

$$\textbf{score}_{ij} = \beta_0 + \beta_1\textbf{cohort90}_{ij} + \beta_2\textbf{female}_{ij} + u_{0j} + u_{1j}\textbf{cohort90}_{ij} + e_{ij}$$

```
> (fit2a <- lmer(score ~ cohort90 + female + (1 + cohort90 | schoolid), data =
mydata, REML = FALSE))

Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + female + (1 + cohort90 | schoolid)
   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 280558 280617 -140272   280544  280552
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  42.57457  6.52492
          cohort90      0.16127  0.40158 -0.393
 Residual             214.83738 14.65733
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 29.58494    0.32405   91.30
cohort90     1.22734    0.02533   48.45
female       1.94453    0.16298   11.93

Correlation of Fixed Effects:
         (Intr) chrt90
cohort90 -0.254
female   -0.265 -0.022
```

Placing brackets around the model declaration fits the model and directly displays the summary of the model results.

[11] As noted in Module 3, we use the more general term ‘coefficient’ rather than ‘slope’ for categorical explanatory variables. The term ‘slope’ is reserved for straight line relationships between y and a continuous x.

Module 5 (R Practical): Introduction to Multilevel Modelling
P5.3 Allowing for Different Slopes across Schools: Random Slope Models

To add a random coefficient for gender:

$$\textbf{score}_{ij} = \beta_0 + \beta_1\textbf{cohort90}_{ij} + \beta_2\textbf{female}_{ij} + u_{0j} + u_{1j}\textbf{cohort90}_{ij} + u_{2j}\textbf{female}_{ij} + e_{ij}$$

```
> (fit2 <- lmer(score ~ cohort90 + female + (1 + cohort90 + female | schoolid),
data = mydata, REML = FALSE))

Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + female + (1 + cohort90 + female | schoolid)
   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 280559 280643 -140269   280539  280547
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  40.55760  6.3685
          cohort90      0.16169  0.4021  -0.394
          female        1.37140  1.1711   0.206 -0.113
 Residual             214.51590 14.6464
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 29.58914    0.31766   93.15
cohort90     1.22777    0.02534   48.44
female       1.93142    0.17391   11.11

Correlation of Fixed Effects:
         (Intr) chrt90
cohort90 -0.253
female   -0.201 -0.046
```

Note that you may have to wait a minute or two for the model to fit. The effect of gender in school j is estimated as 1.931 + $\hat{u}_{2j}$. Allowing for a random effect of gender at the school level has led to the addition of three new random parameters to the model (σ_{u02}, σ_{u12}, σ^2_{u2}). The estimate of the random coefficient variance for **female** σ^2_{u2} is reported in the table `Random effects:` to the right of `female`. The correlation between the intercept and **female** ρ_{u02} is given on the same line as **female** under the column `Corr` and it is the first value while the correlation between **cohort90** and **female** ρ_{u12} is given to the right of ρ_{u02}. A likelihood ratio test, comparing this model with the previous model with a fixed gender effect, is testing the null hypothesis that all three of these parameters are equal to zero.

The likelihood ratio test statistic can be obtained using the `anova` command:

```
> anova(fit2, fit2a)
Data: mydata
Models:
fit2a: score ~ cohort90 + female + (1 + cohort90 | schoolid)
fit2: score ~ cohort90 + female + (1 + cohort90 + female | schoolid)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
fit2a  7 280558 280617 -140272
fit2  10 280559 280643 -140269 5.2362      3     0.1553
```

This is not significant at the 5% level, as indicated by the p-value provided in the table under the column `Pr(>Chisq)`, so we cannot reject the null hypothesis and we conclude that the gender effect is the same for each school. We therefore revert to a model with a fixed coefficient for **female**.

P5.3.6 Adding a random coefficient for social class (categorical *x*)

In Module 3, we found strong social class effects on attainment. We will now explore whether these class effects can be assumed the same across schools.

Before adding social class to the model, we create three dummy variables for when **sclass** is 1, 2 and 4 respectively (taking social class 3 as the reference category).

```
> mydata$sclass1 <- mydata$sclass == 1

> mydata$sclass2 <- mydata$sclass == 2

> mydata$sclass4 <- mydata$sclass == 4
```

We will start by fitting a fixed effect for social class:

$$\mathbf{score}_{ij} = \beta_0 + \beta_1\mathbf{cohort90}_{ij} + \beta_2\mathbf{female}_{ij} + \beta_3\mathbf{sclass1}_{ij} + \beta_4\mathbf{sclass2}_{ij} + \beta_5\mathbf{sclass4}_{ij} + u_{0j} + u_{1j}\mathbf{cohort90}_{ij} + e_{ij}$$

```
> (fit3a <- lmer(score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 | schoolid), data = mydata, REML = FALSE))

Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 | schoolid)
   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 276712 276797 -138346   276692  276705
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  22.51334  4.74482
          cohort90      0.15084  0.38839 -0.317
 Residual             192.94571 13.89049
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 24.60993    0.27962   88.01
cohort90     1.18284    0.02432   48.64
female       1.96135    0.15428   12.71
sclass1TRUE 11.08565    0.20639   53.71
sclass2TRUE  5.87517    0.20405   28.79
sclass4TRUE -3.73775    0.28453  -13.14

Correlation of Fixed Effects:
            (Intr) chrt90 female s1TRUE s2TRUE
cohort90    -0.150
female      -0.296 -0.023
sclass1TRUE -0.395 -0.054  0.008
sclass2TRUE -0.386 -0.020  0.009  0.539
sclass4TRUE -0.271 -0.036  0.013  0.358  0.357
```


Next we add random coefficients for the social class dummy variables:

$$\mathbf{score}_{ij} = \beta_0 + \beta_1\mathbf{cohort90}_{ij} + \beta_2\mathbf{female}_{ij} + \beta_3\mathbf{sclass1}_{ij} + \beta_4\mathbf{sclass2}_{ij} + \beta_5\mathbf{sclass4}_{ij} + u_{0j} + u_{1j}\mathbf{cohort90}_{ij} + u_{3j}\mathbf{sclass1}_{ij} + u_{4j}\mathbf{sclass2}_{ij} + u_{5j}\mathbf{sclass4}_{ij} + e_{ij}$$

where:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{3j} \\ u_{4j} \\ u_{5j} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Omega_u}), \qquad \mathbf{\Omega_u} = \begin{pmatrix} \sigma^2_{u0} & & & & \\ \sigma_{u01} & \sigma^2_{u1} & & & \\ \sigma_{u03} & \sigma_{u13} & \sigma^2_{u3} & & \\ \sigma_{u04} & \sigma_{u14} & \sigma_{u34} & \sigma^2_{u4} & \\ \sigma_{u05} & \sigma_{u15} & \sigma_{u35} & \sigma_{u45} & \sigma^2_{u5} \end{pmatrix}, \qquad e_{ij} \sim N(0, \sigma^2_e)$$

Fit the model:

```
> (fit3 <- lmer(score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 + sclass1 +sclass2 + sclass4 | schoolid), data = mydata, REML = FALSE))

Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 + sclass1 + sclass2 + sclass4 | schoolid)

   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 276657 276843 -138307   276613  276625

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  11.26710  3.35665
          cohort90      0.15599  0.39495 -0.418
          sclass1TRUE   7.13610  2.67135  0.537 -0.105
          sclass2TRUE   3.32133  1.82245  0.827 -0.164  0.857
          sclass4TRUE   7.18489  2.68046  0.898 -0.417  0.439  0.605
 Residual             191.76775 13.84802
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 24.40121    0.23070  105.77
cohort90     1.18452    0.02448   48.38
female       1.96598    0.15408   12.76
sclass1TRUE 11.18175    0.24369   45.88
sclass2TRUE  6.07241    0.22040   27.55
sclass4TRUE -3.18959    0.31406  -10.16

Correlation of Fixed Effects:
            (Intr) chrt90 female s1TRUE s2TRUE
cohort90    -0.176
female      -0.359 -0.023
sclass1TRUE -0.212 -0.074  0.007
sclass2TRUE -0.207 -0.056  0.009  0.585
sclass4TRUE -0.042 -0.158  0.013  0.346  0.377
```

The new model contains a large number of additional random parameters and will again take a few minutes to fit. Note that depending on your computer specifications and the versions of R and lme4, the model may actually fail to reach convergence[12]. If this is the case, then at the end of the output an error message will be displayed warning that convergence may not have been reached. If this happens we would advise to try and fit the model on a different system or use another software package. There are 12 more parameters in this model than in the fixed social class effects model. The likelihood ratio test statistic for a comparison of these models is:

```
> anova(fit3, fit3a)
Data: mydata
Models:
fit3a: score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
fit3a:     cohort90 | schoolid)
fit3: score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
fit3:     cohort90 + sclass1 + sclass2 + sclass4 | schoolid)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
fit3a 10 276712 276797 -138346
fit3  22 276657 276843 -138307 79.122     12  6.069e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

From the very low Pr(>Chisq), we conclude that there is evidence that the effect of social class on attainment differs across schools.

The coefficients of **sclass1**, **sclass2** and **sclass4** have a fixed component, representing contrasts with the reference category 3 (working class) on average, and a school-specific component. For example, after accounting for cohort and gender effects, children with a parent in a professional or managerial occupation (**sclass** = 1) attending school j are expected to have an attainment score that is 11.2 + $\hat{u}_{3j}$ points higher than working class children in the same school.

Due to the large number of parameters in the random part of this model, the simplest way to interpret the random coefficient for class is to compute the between-school variance

$$\text{var}\left(u_{0j} + u_{1j}\textbf{cohort90}_{ij} + u_{3j}\textbf{sclass1}_{ij} + u_{4j}\textbf{sclass2}_{ij} + u_{5j}\textbf{sclass4}_{ij}\right)$$

We will do this for each social class category, holding constant the value of **cohort90** (the other variable with a random coefficient). For convenience, we will fix **cohort90** at zero, so the between-school variances will refer to 1990. This simplifies the expression for the between-school variance to:

$$\text{var}\left(u_{0j} + u_{3j}\textbf{sclass1}_{ij} + u_{4j}\textbf{sclass2}_{ij} + u_{5j}\textbf{sclass4}_{ij}\right)$$

[12] For this document, the model was fitted on a 32 bit Windows PC laptop, dual core 2.10GHz, with 4GB RAM memory.

We output the variance covariance matrix for the random effects:

```
> VarCorr(fit3)$schoolid
            (Intercept)   cohort90 sclass1TRUE sclass2TRUE sclass4TRUE
(Intercept)  11.2670991 -0.5536348   4.8130826   5.0589541   8.0768916
cohort90     -0.5536348  0.1559892  -0.1105532  -0.1179540  -0.4419864
sclass1TRUE   4.8130826 -0.1105532   7.1361021   4.1744010   3.1412509
sclass2TRUE   5.0589541 -0.1179540   4.1744010   3.3213350   2.9574009
sclass4TRUE   8.0768916 -0.4419864   3.1412509   2.9574009   7.1848898

attr(,"stddev")
(Intercept)    cohort90 sclass1TRUE sclass2TRUE sclass4TRUE
  3.3566500   0.3949546   2.6713484   1.8224530   2.6804645

attr(,"correlation")
            (Intercept)   cohort90 sclass1TRUE sclass2TRUE sclass4TRUE
(Intercept)   1.0000000 -0.4176092   0.5367681   0.8269863   0.8976936
cohort90     -0.4176092  1.0000000  -0.1047837  -0.1638737  -0.4174954
sclass1TRUE   0.5367681 -0.1047837   1.0000000   0.8574470   0.4386944
sclass2TRUE   0.8269863 -0.1638737   0.8574470   1.0000000   0.6054019
sclass4TRUE   0.8976936 -0.4174954   0.4386944   0.6054019   1.0000000
```

We can use R to calculate the between-school variance instead of calculating this by hand. For example, the variance for category 1 of **sclass** (**sclass1** = 1, **sclass2** = 0 and **sclass4** = 0) is:

$$\text{var}\left(u_{0j} + u_{3j}\right) = \sigma_{u0}^2 + 2\sigma_{u03} + \sigma_{u3}^2$$

```
> 11.267 + 2*4.813 + 7.136
[1] 28.029
```

Similarly, the variance for category 2 of **sclass** (**sclass1** = 0, **sclass2** = 1 and **sclass4** = 0) is**:**

$$\text{var}\left(u_{0j} + u_{4j}\right) = \sigma_{u0}^2 + 2\sigma_{u04} + \sigma_{u4}^2$$

```
> 11.267 + 2*5.059 + 3.321
[1] 24.706
```

The variance for category 3 of **sclass** (**sclass1** = 0, **sclass2** = 0 and **sclass4** = 0) is simply**:**

$$\text{var}\left(u_{0j}\right) = \sigma_{u0}^2$$

```
> 11.267
[1] 11.267
```

While the variance for category 4 of **sclass** (**sclass1** = 0, **sclass2** = 0 and **sclass4** = 1) is:

$$\text{var}\left(u_{0j} + u_{5j}\right) = \sigma_{u0}^2 + 2\sigma_{u05} + \sigma_{u5}^2$$

```
> 11.267 + 2*8.077 + 7.184
[1] 34.605
```

The between-school variance is similar for the first two categories (professional/managerial and intermediate), highest for the unclassified group (category 4) and lowest for working class children (category 3). This implies that the school attended matters most for the unclassified group (in terms of their age 16 attainment), and least for working class children. For example, the difference between unclassified and working class in school j is estimated as -3.190 + $\hat{u}_{5j}$. The estimated variance of u_{5j} is 7.182, so a 95% coverage interval for the unclassified-working class difference is $-3.190 \pm 1.96\sqrt{7.184}$ = -8.443 to 2.063. Suppose we rank schools according to their unclassified-working class difference, such that schools with the largest difference (in favour of working class children) are ranked lowest. In the bottom 2.5% of schools, unclassified children are expected to have a mean score that is more than 8.443 points *lower* than working class children. In the top 2.5% of schools, however, the difference is estimated to be more than 2.063, in favour of *unclassified* children.

Once again, it is important to note that these school differences should not be interpreted as school effects in the usual sense because we have not accounted for prior attainment. Of most interest is the extent of between-school variance in the *progress* made by children from different social backgrounds.

Don't forget to take the online quiz!

From within the LEMMA learning environment

- Go down to the section for **Module 5: Introduction to Multilevel Modelling**
- Click "5.3 Allowing for Different Slopes Across Groups: Random Slope Models" to open Lesson 5.3
- Click Q 1 to open the first question

P5.4 Adding Level 2 Explanatory Variables

In the last two exercises, you have seen how to add level 1 explanatory variables to the model and interpret the results of random intercept and random slope models. A key motivation for using multilevel modelling, however, is to assess the effects of level 2 explanatory variables on level 1 outcomes and the extent to which they can explain the level 2 variance. In education, for example, we may be interested in the *contextual* effect of prior attainment on students' later academic performance. A student's progress may be affected by the performance of others in their peer group, and this effect may differ according to the student's own prior attainment (a *cross-level interaction*).

Our example dataset contains three school-level variables that are potential predictors of a student's attainment at age 16: **schtype** (independent vs. state schools), **schurban** (urban vs. rural location of school), and **schdenom** (Roman Catholic vs. non-denominational school). In this exercise, we will add these variables to our model and consider whether the effect on attainment of one of them depends on a selected student-level variable.

As in any analysis, we should look at the distribution of our variables before including them in a model.

Download the R dataset for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 5: Introduction to Multilevel Modelling**, and scroll down to **R Datasets and R files**
- Right click "5.4.txt" and select **Save Link As...** to save the dataset to your computer

Read the dataset into R, and load the `lme4` library:

```
> mydata <- read.table("5.4.txt", header = TRUE, sep = ",")

> library(lme4)
Loading required package: Matrix
Loading required package: lattice

Attaching package: 'lme4'


        The following object(s) are masked from package:stats :

         AIC
```

Each school-level variable is binary, so we will simply look at the proportion in each category. This can be done by using the `table`, `prop.table` and `cumsum` commands to produce respectively the table of frequencies, percentages and cumulative percentages. We first restrict the scope of the command to one record per school by using the `unique` command restricted to the columns of interest:

```
> mydata_un <- unique(mydata[, c(2, 7, 8, 9)])

> cbind(Freq = table(mydata_un$schtype), Perc =
prop.table(table(mydata_un$schtype)), Cum =
cumsum(prop.table(table(mydata_un$schtype))))

  Freq      Perc       Cum
0  456 0.8976378 0.8976378
1   52 0.1023622 1.0000000

> cbind(Freq = table(mydata_un$schurban), Perc =
prop.table(table(mydata_un$schurban)), Cum =
cumsum(prop.table(table(mydata_un$schurban))))

  Freq      Perc       Cum
0  163 0.3208661 0.3208661
1  345 0.6791339 1.0000000

> cbind(Freq = table(mydata_un$schdenom), Perc =
prop.table(table(mydata_un$schdenom)), Cum =
cumsum(prop.table(table(mydata_un$schdenom))))

  Freq      Perc       Cum
0  425 0.8366142 0.8366142
1   83 0.1633858 1.0000000
```

You should obtain the following proportion of schools in category 1 of each variable: **schtype** (10% independent), **schurban** (68% urban), and **schdenom** (16% Catholic).

We will add these variables, one at a time, to a simplified version of the model fitted at the end of P5.3. Although we found evidence that the effect of social class on attainment differs across schools, we will work with a simpler model by removing the random coefficients on the class dummy variables.

$$\mathbf{score}_{ij} = \beta_0 + \beta_1\mathbf{cohort90}_{ij} + \beta_2\mathbf{female}_{ij} + \beta_3\mathbf{sclass1}_{ij} + \beta_4\mathbf{sclass2}_{ij} + \beta_5\mathbf{sclass4}_{ij} + u_{0j} + u_{1j}\mathbf{cohort90}_{ij} + e_{ij}$$

```
> (fit1a <- lmer(score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 | schoolid), data = mydata, REML = FALSE))

Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 | schoolid)

   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 276712 276797 -138346   276692  276705

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  22.51334  4.74482
          cohort90      0.15084  0.38839 -0.317
 Residual             192.94571 13.89049
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 24.60993    0.27962   88.01
cohort90     1.18284    0.02432   48.64
female       1.96135    0.15428   12.71
sclass1TRUE 11.08565    0.20639   53.71
sclass2TRUE  5.87517    0.20405   28.79
sclass4TRUE -3.73775    0.28453  -13.14

Correlation of Fixed Effects:
            (Intr) chrt90 female s1TRUE s2TRUE
cohort90    -0.150
female      -0.296 -0.023
sclass1TRUE -0.395 -0.054  0.008
sclass2TRUE -0.386 -0.020  0.009  0.539
sclass4TRUE -0.271 -0.036  0.013  0.358  0.357
```

P5.4.1 Contextual effects

We will begin by adding school type (independent vs. state) to the model.

$$\begin{aligned}\textbf{score}_{ij} = \beta_0 + \beta_1\textbf{cohort90}_{ij} + \beta_2\textbf{female}_{ij} + \beta_3\textbf{sclass1}_{ij} + \beta_4\textbf{sclass2}_{ij} + \beta_5\textbf{sclass4}_{ij} \\ +\beta_6\textbf{schtype}_j \\ +u_{0j} + u_{1j}\textbf{cohort90}_{ij} + e_{ij}\end{aligned}$$

This can be done simply by updating the previous model:

```
> (fit2 <- update(fit1a, . ~ . + schtype))

Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 | schoolid) + schtype

   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 276689 276782 -138333   276667  276678

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  20.57103  4.53553
          cohort90      0.14812  0.38486 -0.263
 Residual             192.99407 13.89223
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 24.27939    0.27872   87.11
cohort90     1.18404    0.02421   48.90
female       1.96377    0.15426   12.73
sclass1TRUE 11.03061    0.20695   53.30
sclass2TRUE  5.85641    0.20414   28.69
sclass4TRUE -3.75024    0.28454  -13.18
schtype      4.24675    0.81687    5.20

Correlation of Fixed Effects:
            (Intr) chrt90 female s1TRUE s2TRUE s4TRUE
cohort90    -0.112
female      -0.297 -0.023
sclass1TRUE -0.378 -0.055  0.008
sclass2TRUE -0.379 -0.021  0.009  0.540
sclass4TRUE -0.270 -0.036  0.013  0.357  0.357
schtype     -0.214  0.007  0.001 -0.079 -0.034 -0.005
```

A child in an independent school would be expected to have a score that is 4.25 points higher than a child in a state school (from the same cohort, and of the same sex and social background). We can see that this effect is strongly statistically significant because the estimated coefficient is more than 5 times its standard error. There has also been a slight reduction in the school-level variance. After accounting for school type, the between-school variance for the 1990 cohort (the intercept variance) reduces from 22.5 to 20.6. However, there remains a large amount of unexplained between-school variance.

Module 5 (R Practical): Introduction to Multilevel Modelling
P5.4 Adding Level 2 Explanatory Variables

We will now add in the urban-rural indicator of school location:

$$\begin{aligned}\textbf{score}_{ij} = \beta_0 + \beta_1\textbf{cohort90}_{ij} + \beta_2\textbf{female}_{ij} + \beta_3\textbf{sclass1}_{ij} + \beta_4\textbf{sclass2}_{ij} + \beta_5\textbf{sclass4}_{ij} \\ +\beta_6\textbf{schtype}_j + \beta_7\textbf{schurban}_j \\ +u_{0j} + u_{1j}\textbf{cohort90}_{ij} + e_{ij}\end{aligned}$$

```
> (fit3 <- update(fit2, . ~ . + schurban))

Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 | schoolid) + schtype + schurban
   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 276682 276783 -138329   276658  276669

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  19.95405  4.46700
          cohort90      0.14825  0.38503 -0.263
 Residual             193.01104 13.89284
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 25.25993    0.42720   59.13
cohort90     1.18191    0.02423   48.78
female       1.96668    0.15425   12.75
sclass1TRUE 11.03299    0.20694   53.32
sclass2TRUE  5.84714    0.20418   28.64
sclass4TRUE -3.73999    0.28457  -13.14
schtype      4.39193    0.80922    5.43
schurban    -1.43718    0.47632   -3.02

Correlation of Fixed Effects:
            (Intr) chrt90 female s1TRUE s2TRUE s4TRUE schtyp
cohort90    -0.099
female      -0.190 -0.023
sclass1TRUE -0.252 -0.055  0.008
sclass2TRUE -0.264 -0.020  0.009  0.540
sclass4TRUE -0.165 -0.037  0.013  0.357  0.356
schtype     -0.096  0.005  0.002 -0.080 -0.036 -0.004
schurban    -0.763  0.035 -0.004  0.007  0.022 -0.015 -0.054
```

On average, a student in an urban school has a score that is 1.44 points lower than a student attending a school in a town or rural area. This difference is adjusted for the effects of school type, and student cohort, gender and social class. The between-school variance in 1990 has decreased further but by a very small amount (from 20.6 to 20.0).

Finally, we will test for differences in attainment by school denomination.

$$\begin{aligned}\textbf{score}_{ij} &= \beta_0 + \beta_1\textbf{cohort90}_{ij} + \beta_2\textbf{female}_{ij} + \beta_3\textbf{sclass1}_{ij} + \beta_4\textbf{sclass2}_{ij} + \beta_5\textbf{sclass4}_{ij} \\ &\quad + \beta_6\textbf{schtype}_j + \beta_7\textbf{schurban}_j + \beta_8\textbf{schdenom}_j \\ &\quad + u_{0j} + u_{1j}\textbf{cohort90}_{ij} + e_{ij}\end{aligned}$$

```
> (fit4 <- update(fit3, . ~ . + schdenom))
Linear mixed model fit by maximum likelihood

Formula: score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 | schoolid) + schtype + schurban + schdenom
   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 276684 276793 -138329   276658  276668

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  19.96576  4.46831
          cohort90      0.14818  0.38494 -0.267
 Residual             193.01279 13.89290
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
            Estimate Std. Error t value
(Intercept) 25.24843    0.42934   58.81
cohort90     1.18203    0.02422   48.80
female       1.96666    0.15425   12.75
sclass1TRUE 11.03348    0.20696   53.31
sclass2TRUE  5.84725    0.20419   28.64
sclass4TRUE -3.74055    0.28458  -13.14
schtype      4.39730    0.81079    5.42
schurban    -1.46221    0.48490   -3.02
schdenom     0.17054    0.60149    0.28

Correlation of Fixed Effects:
            (Intr) chrt90 female s1TRUE s2TRUE s4TRUE schtyp schrbn
cohort90    -0.100
female      -0.189 -0.023
sclass1TRUE -0.252 -0.055  0.008
sclass2TRUE -0.263 -0.020  0.009  0.540
sclass4TRUE -0.164 -0.037  0.013  0.357  0.356
schtype     -0.102  0.005  0.001 -0.079 -0.035 -0.005
schurban    -0.726  0.034 -0.004  0.004  0.020 -0.013 -0.065
schdenom    -0.101  0.001 -0.002  0.015  0.008 -0.007  0.066 -0.189
```

The ratio of the estimated coefficient of **schdenom** to its standard error is less than 0.3, so there is little evidence of a difference between Catholic and non-denominational schools. We will therefore remove this variable from our model.[13]

[13] It is a possible that a variable with a non-significant main effect could be involved in a significant interaction effect. To illustrate how this might arise, suppose we have a binary student-level variable z (coded 0 and 1). Suppose also that attending a Catholic school is associated with higher attainment among students with z = 0, but lower attainment among students with z = 1. This would be an example of an interaction between school denomination and z (actually a cross-level interaction because the two variables are defined at different levels). If the categories of z are of a similar size, ignoring the interaction with z and allowing only for an overall main effect of school denomination is likely to lead to an apparently non-significant effect. We will not pursue this possibility here.

P5.4.2 Cross-level interactions

Our analysis thus far has revealed that student attainment at age 16 is significantly related to the year in which the exams were taken (cohort), and student gender and parental social class. At the school level, there are differences in student attainment between independent and state schools, and between urban and rural schools. However, we have considered only *main effects* of these variables. In practice, the relationship between y and an explanatory variable x_1 may depend on the value of another variable x_2, i.e. an *interaction effect* between x_1 and x_2. In a multilevel model, x_1 and x_2 may be defined at the same or different levels. If they are at different levels, the interaction is referred to as a *cross-level interaction*.

To illustrate cross-level interactions and their interpretation, we will test for an interaction between cohort (level 1) and school type (level 2). We will also explore whether a cohort-school type interaction can explain between-school differences in attainment trends (i.e. whether such an interaction reduces some of the variance of the random part of the slope for cohort). First we create this new interaction variable:

```
> mydata$cohort90Xschtype <- mydata$cohort90*mydata$schtype
```

We then add this interaction to the model:

$$\begin{aligned}\textbf{score}_{ij} &= \beta_0 + \beta_1\textbf{cohort90}_{ij} + \beta_2\textbf{female}_{ij} + \beta_3\textbf{sclass1}_{ij} + \beta_4\textbf{sclass2}_{ij} + \beta_5\textbf{sclass4}_{ij} \\ &+\beta_6\textbf{schtype}_j + \beta_7\textbf{schurban}_j + \beta_8\textbf{cohort90Xschtype}_{ij} \\ &+u_{0j} + u_{1j}\textbf{cohort90}_{ij} + e_{ij}\end{aligned}$$

```
> (fit5 <- update(fit3, . ~ . + cohort90Xschtype))

Linear mixed model fit by maximum likelihood
Formula: score ~ cohort90 + female + sclass1 + sclass2 + sclass4 + (1 +
cohort90 | schoolid) + schtype + schurban + cohort90Xschtype

   Data: mydata
    AIC    BIC  logLik deviance REMLdev
 276651 276761 -138313   276625  276638

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 schoolid (Intercept)  20.41431  4.5182
          cohort90      0.13801  0.3715  -0.233
 Residual             192.85131 13.8871
Number of obs: 33988, groups: schoolid, 508

Fixed effects:
                 Estimate Std. Error t value
(Intercept)      25.18705    0.43201   58.30
cohort90          1.21354    0.02442   49.69
female            1.97026    0.15418   12.78
sclass1TRUE      11.01937    0.20687   53.27
sclass2TRUE       5.83072    0.20411   28.57
sclass4TRUE      -3.74284    0.28444  -13.16
schtype           5.29101    0.83071    6.37
schurban         -1.40365    0.48284   -2.91
cohort90Xschtype -0.59937    0.10380   -5.77

Correlation of Fixed Effects:
            (Intr) chrt90 female s1TRUE s2TRUE s4TRUE schtyp schrbn
cohort90    -0.090
female      -0.188 -0.022
sclass1TRUE -0.249 -0.056  0.008
sclass2TRUE -0.260 -0.023  0.009  0.540
sclass4TRUE -0.163 -0.037  0.013  0.357  0.356
schtype     -0.100  0.044  0.002 -0.080 -0.037 -0.005
schurban    -0.765  0.038 -0.004  0.007  0.021 -0.014 -0.050
chrt90Xscht  0.027 -0.235 -0.005  0.009  0.013  0.002 -0.175 -0.014
```

The estimated coefficient of the interaction variable **cohort90Xschtype** is almost 6 times its standard error, so this is strong evidence that the effect of cohort differs for independent and state schools. (Equivalently, we can say that the difference between independent and state schools differs across cohorts.) However, the addition of this interaction effect does little to explain between-school differences in attainment trends: the school-level variance in the **cohort90** coefficient has reduced only slightly from 0.148 to 0.138.

To see the nature of the interaction effect, consider the fixed part of the model that contains **cohort90** and **schtype**:

$$1.213\,\textbf{cohort90} + 5.291\textbf{schtype} - 0.599\,\textbf{cohort90Xschtype}$$

For **schtype** = 0 (state schools), this equation reduces to:

$$1.213\,\textbf{cohort90}$$

So in the *average* state school (i.e. with $u_{1j} = 0$),[14] we would expect a year-on-year increase in attainment of 1.213 points.

For **schtype** = 1 (independent schools), this equation reduces to:

$$\begin{aligned}1.213\,\textbf{cohort90} + 5.291 - 0.599\,\textbf{cohort90} &= (1.213 - 0.599)\,\textbf{cohort90} + 5.291 \\ &= 0.614\,\textbf{cohort90} + 5.291\end{aligned}$$

So in the average independent school, we would expect a year-on-year increase in attainment of 0.614 points.

The coefficient of **schtype** (estimated as 5.291) is the expected difference in attainment between independent and state schools in 1990 (i.e. when **cohort90** = 0).

Our overall conclusion is that the mean attainment is higher in independent schools than in state schools, but independent schools experienced a smaller increase in attainment with cohort. As in our earlier analyses, it would be interesting to investigate whether the trends in *progress* are different for independent and state schools.

Don't forget to take the online quiz!

From within the LEMMA learning environment

- Go down to the section for **Module 5: Introduction to Multilevel Modelling**
- Click "5.4 Adding Level 2 Explanatory Variables"
  to open Lesson 5.4
- Click Q1(i) to open the first question

[14] The effect of cohort varies randomly across schools, so we fix the school cohort residual at its mean of zero to examine the cohort-school type interaction effect.

P5.5 Complex Level 1 Variation

In a random slope (coefficient) model, the level 2 variance is a function of the explanatory variable(s) with a random coefficient. For example, in P5.3, we allowed the effects of cohort and social class to vary randomly across schools, which implies that the between-school variance depends on cohort and class. Up to this point, however, we have assumed that the level 1 (*within*-school) variance is constant. Some statistical software packages (e.g. MLwiN, Stata) allow the within-school variance to depend on explanatory variables in a complex level 1 variance model. Unfortunately, it is not possible at present to fit this kind of model in R.

P5.6 References

Bates, D. (2010). *Computational methods for mixed models.* *http://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf*

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.