Preface

This book is about applied multilevel and longitudinal modeling. Other terms for multilevel models include hierarchical models, random-effects or random-coefficient models, mixed-effects models, or simply mixed models. Longitudinal data are also referred to as panel data, repeated measures, or cross-sectional time series. A popular type of multilevel model for longitudinal data is the growth-curve model.

The common theme of this book is regression modeling when data are clustered in some way. In cross-sectional settings, students may be nested in schools, people in neighborhoods, employees in firms, or twins in twin-pairs. Longitudinal data are by definition clustered because multiple observations over time are nested within units, typically subjects.

Such clustered designs often provide rich information on processes operating at different levels, for instance, people's characteristics interacting with institutional characteristics. Importantly, the standard assumption of independent observations is likely to be violated because of dependence among observations within the same cluster. The multilevel and longitudinal methods discussed in this book extend conventional regression to handle such dependence and exploit the richness of the data.

Volume 1 is on multilevel and longitudinal modeling of continuous responses using linear models. The volume consists of four parts: I. Preliminaries (a review of linear regression modeling, preparing the reader for the rest of the book), II. Two-level models, III. Models for longitudinal and panel data, and IV. Models with nested and crossed random effects. For readers who are new to multilevel and longitudinal modeling, the chapters in part II should be read sequentially and can form the basis of an introductory course on this topic. A one-semester course on multilevel and longitudinal modeling can be based on most of the chapters in volume 1 plus chapter 10 on binary or dichotomous responses from volume 2. For this purpose, we have made chapter 10 freely downloadable from https://www.stata-press.com/books/mlmus4_ch10.pdf.

Volume 2 is on multilevel and longitudinal modeling of categorical responses, counts, and survival data. This volume also consists of four parts: I. Categorical responses (binary or dichotomous responses, ordinal responses, and nominal responses or discrete choice), II. Counts, III. Survival (in both discrete and continuous time), and IV. Models with nested and crossed random effects. Each chapter starts by introducing models for nonclustered data (for example, logistic and Poisson regression) and then extends the models for clustered data by introducing random effects, leading to generalized linear mixed models. Subsequently, alternatives such as generalized estimating equations (GEE) and fixed-effects approaches are discussed. Chapter 10 on binary or dichotomous responses is a core chapter of this volume and should be read before embarking on the other chapters. It is also a good idea to read chapter 14 on discrete-time survival before reading chapter 15 on continuous-time survival.

Our emphasis is on explaining the models and their assumptions, applying the methods to real data, and interpreting results. Many of the issues are conceptually demanding but do not require that you understand complex mathematics. Therefore, wherever possible, we introduce ideas through examples and graphical illustrations, keeping the technical descriptions as simple as possible. Some sections that go beyond an introductory course on multilevel and longitudinal modeling are tagged with the ❖ symbol. Derivations that can be skipped by the reader are given in displays. For an advanced treatment, placing multilevel modeling within a general latent-variable framework, we refer the reader to Skrondal and Rabe-Hesketh (2004), which uses the same notation as this book.

This book shows how all the analyses described can be performed using Stata. There are many advantages of using a general-purpose statistical package such as Stata. First, for those already familiar with Stata, it is convenient not having to learn a new standalone package. Second, conducting multilevel analysis within a powerful package has the advantage that it allows complex data manipulation to be performed, alternative estimation methods to be used, and publication-quality graphics to be produced, all without having to switch packages. Finally, Stata is a natural choice for multilevel and longitudinal modeling because it has gradually become perhaps the most powerful general-purpose statistics package for such models.

Each chapter is based on one or more research problems and real datasets. After describing the models, we walk through the analysis using Stata, pausing to address statistical issues that need further explanation. Do-files for each chapter can be downloaded from https://www.stata-press.com/data/mlmus4.html. Some readers may find it useful to perform the analyses while reading the book.

Stata can be used either via a graphical user interface (GUI) or through commands. We recommend using commands interactively—or preferably in do-files—for serious analysis in Stata. For this reason, and because the GUI is fairly self-explanatory, we use commands exclusively in this book. However, the GUI can be useful for learning the Stata syntax. Generally, we use the `typewriter font` to refer to Stata commands, syntax, and variables. A "dot" prompt followed by a command indicates that you can type verbatim what is displayed after the dot (in context) to replicate the results in the book. Some readers may find it useful to intersperse reading with running these commands. We encourage readers to write do-files for solving the data analysis exercises because this is standard practice for professional data analysis.

The commands used for data manipulation and graphics are explained to some extent, but the purpose of this book is not to teach Stata from scratch. For a basic introduction to Stata, we refer the reader to Acock (2018). Other books and further resources for learning Stata are listed at the Stata website.

If you are new to Stata, we recommend running all the commands given in chapter 1 of volume 1. A list of commands that are particularly useful for manipulating, describing, and plotting multilevel and longitudinal data is given in the appendix of volume 1. Examples using these and other commands can easily be found by referring to the "commands" entry in the subject index.

We have included applications from a wide range of disciplines, including medicine, economics, education, sociology, and psychology. The interdisciplinary nature of this book is also reflected in the choice of models and topics covered. If a chapter is primarily based on an application from one discipline, we try to balance this by including exercises with real data from other disciplines. The two volumes contain over 140 exercises based on over 100 different real datasets. Exercises for which solutions are available to readers are marked with Solutions, and the solutions can be downloaded from https://www.stata-press.com/books/mlmus4-answers.html. Instructors can obtain solutions to all exercises from Stata Press.

All datasets used in this book are freely available for download; for details, see https://www.stata-press.com/data/mlmus4.html. These datasets can be downloaded into a local directory on your computer. Alternatively, individual datasets can be loaded directly into net-aware Stata by specifying the complete URL. For example,

```
. use https://www.stata-press.com/data/mlmus4/pefr
```

If you have stored the datasets in the working directory, omit the path and just type

```
. use pefr
```

We will generally describe all Stata commands that can be used to fit a given model, discussing their advantages and disadvantages. An exception to this rule is that we do not discuss our own `gllamm` command in volume 1 (see the `gllamm` companion, downloadable from http://www.gllamm.org, for how to fit the models of volume 1 in `gllamm`). In volume 1, we extensively use the Stata commands `xtreg` and `mixed`, and we introduce several more specialized commands for longitudinal modeling, such as `xthtaylor`, `xtivreg`, and `xtabond`. The `sem` command for structural equation modeling is used for growth-curve modeling.

In volume 2, we use Stata's `xt` and `me` commands for different response types. For example, we use `xtlogit` and `melogit` for binary responses, `meologit` for ordinal responses, `xtpoisson` and `mepoisson` for counts, and `mestreg` for multilevel continuous-time survival modeling with shared frailties. In chapter 12 on nominal responses, we use Stata's new `cm` (for "choice model") suite of commands, such as `cmxtmixlogit`. `gllamm` is also used throughout volume 2. We also discuss commands for marginal models and fixed-effects models, such as `xtgee` and `clogit`. The online reference manuals available through the `help` command within Stata provide detailed information on all the official Stata commands for multilevel and longitudinal modeling.

The `nolog` option has been used to suppress the iteration logs showing the progress of the log likelihood. This option is not shown in the command line because we do not recommend it to users; we are using it only to save space.

We assume that readers have a good knowledge of linear regression modeling, in particular, the use and interpretation of dummy variables and interactions. However, the first chapter in volume 1 reviews linear regression and can serve as a refresher.

Errata for different editions and printings of the book can be downloaded from https://www.stata-press.com/books/errata/mlmus4.html, and answers to exercises can be downloaded from https://www.stata-press.com/books/mlmus4-answers.html.

In this fourth edition, we have thoroughly revised all chapters and updated the Stata syntax for release 17. Major additions in volume 1 include the Kenward–Roger degree-of-freedom correction for improved inference with a small number of clusters, difference-in-difference estimation for quasi experiments, and instrumental-variables estimation to handle level-1 endogeneity. In volume 2, we now introduce Bayesian estimation for crossed-effects models and extensively use several new commands (since the third edition), including `meologit`, `cmxtmixlogit`, `mestreg`, and `menbreg`.

Berkeley and Oslo Sophia Rabe-Hesketh
August 2021 Anders Skrondal

Contents

Contents

4 Random-coefficient models

4.1 Introduction

In the previous chapter, we considered linear random-intercept models where the overall level of the response was allowed to vary between clusters after controlling for covariates. In this chapter, we include random coefficients or random slopes in addition to random intercepts, thus also allowing the effects of covariates to vary between clusters. Such models involving both random intercepts and random slopes are often called *random-coefficient models*. In longitudinal settings, where the level-1 units are occasions and the clusters are typically subjects, models with a random-coefficient of time are also referred to as growth-curve models. Such models are the topic of chapter 7.

4.2 How effective are different schools?

Here we analyze a dataset on inner-London schools that accompanies the MLwiN software (Rasbash et al. 2019) and is part of the data analyzed by Goldstein et al. (1993).

At age 16, students took their Graduate Certificate of Secondary Education (GCSE) exams in a number of subjects. A score was derived from the individual exam results. Such scores often form the basis for school comparisons, for instance, to allow parents to choose the best school for their child. However, schools can differ considerably in their intake achievement levels. It may be argued that what should be compared is the "value added"; that is, the difference in mean GCSE score between schools after controlling for the students' achievement before entering the school. One such measure of prior achievement is the London Reading Test (LRT) taken by these students at age 11.

The dataset `gcse.dta` has the following variables:

- `school`: school identifier
- `student`: student identifier
- `gcse`: Graduate Certificate of Secondary Education (GCSE) score (z score, multiplied by 10)
- `lrt`: London Reading Test (LRT) score (z score, multiplied by 10)
- `girl`: dummy variable for student being a girl (1: girl; 0: boy)
- `schgend`: type of school (1: mixed gender; 2: boys only; 3: girls only)

One purpose of the analysis is to investigate the relationship between GCSE and LRT and how this relationship varies between schools. The model can then be used to address the question of which schools appear to be most effective, taking prior achievement into account.

We read in the data by using

```
. use https://www.stata-press.com/data/mlmus4/gcse
```

4.3 Separate linear regressions for each school

Before developing a model for all 65 schools combined, we consider a separate model for each school. For school j, an obvious model for the relationship between GCSE and LRT is a simple regression model,

$$y_{ij} \; = \; \beta_{1j} + \beta_{2j} x_{ij} + \epsilon_{ij}$$

where y_{ij} is the GCSE score for the ith student in school j, x_{ij} is the corresponding LRT score, β_{1j} is the school-specific intercept, β_{2j} is the school-specific slope, and ϵ_{ij} is a residual error term with school-specific variance θ_j.

For school 1, OLS estimates of the intercept $\widehat{\beta}_{11}$ and the slope $\widehat{\beta}_{21}$ can be obtained using `regress`,

```
. regress gcse lrt if school==1

      Source |       SS           df       MS      Number of obs   =        73
-------------+----------------------------------   F(1, 71)        =     59.44
       Model |  4084.89189         1  4084.89189   Prob > F        =    0.0000
    Residual |  4879.35759        71  68.7233463   R-squared       =    0.4557
-------------+----------------------------------   Adj R-squared   =    0.4480
       Total |  8964.24948        72  124.503465   Root MSE        =      8.29

------------------------------------------------------------------------------
        gcse | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         lrt |   .7093406   .0920061     7.71   0.000     .5258856    .8927955
       _cons |   3.833302   .9822377     3.90   0.000     1.874776    5.791828
------------------------------------------------------------------------------
```

where we have selected school 1 by specifying the condition `if school==1`.

To assess whether this is a reasonable model for school 1, we can obtain the predicted (ordinary least squares) regression line for this school (with $j = 1$),

$$\widehat{y}_{i1} = \widehat{\beta}_{11} + \widehat{\beta}_{21} x_{i1}$$

by using the `predict` command with the `xb` option:

```
. predict p_gcse, xb
```

We superimpose this line on the scatterplot of the data for the school, as shown in figure 4.1.

```
. twoway (scatter gcse lrt) (line p_gcse lrt, sort) if school==1,
> xtitle(LRT) ytitle(GCSE)
```

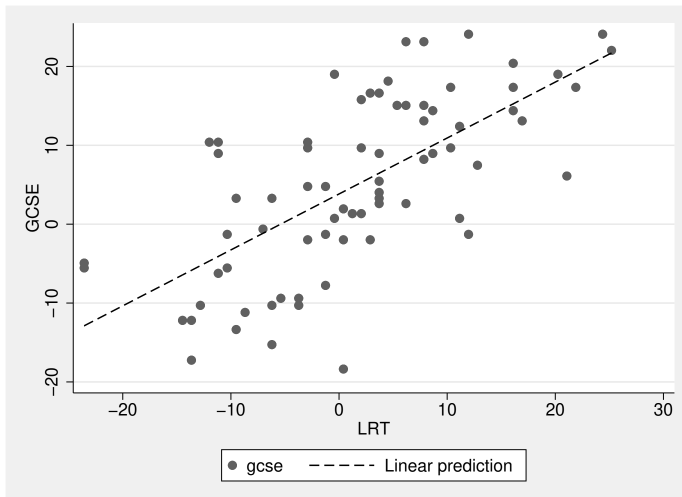


Figure 4.1: Scatterplot of `gcse` versus `lrt` for school 1 with ordinary least-squares regression line

We can also produce a *trellis graph* containing such plots for all 65 schools by using

```
. twoway (scatter gcse lrt) (lfit gcse lrt, sort lpatt(solid)),
> by(school, compact legend(off) cols(5))
> xtitle(LRT) ytitle(GCSE) ysize(3) xsize(2)
```

where `lfit` plots regression lines estimated by OLS. The resulting graph, shown in figure 4.2, suggests that the model assumptions are reasonably met. The schools appear to vary in both their intercepts and slopes.

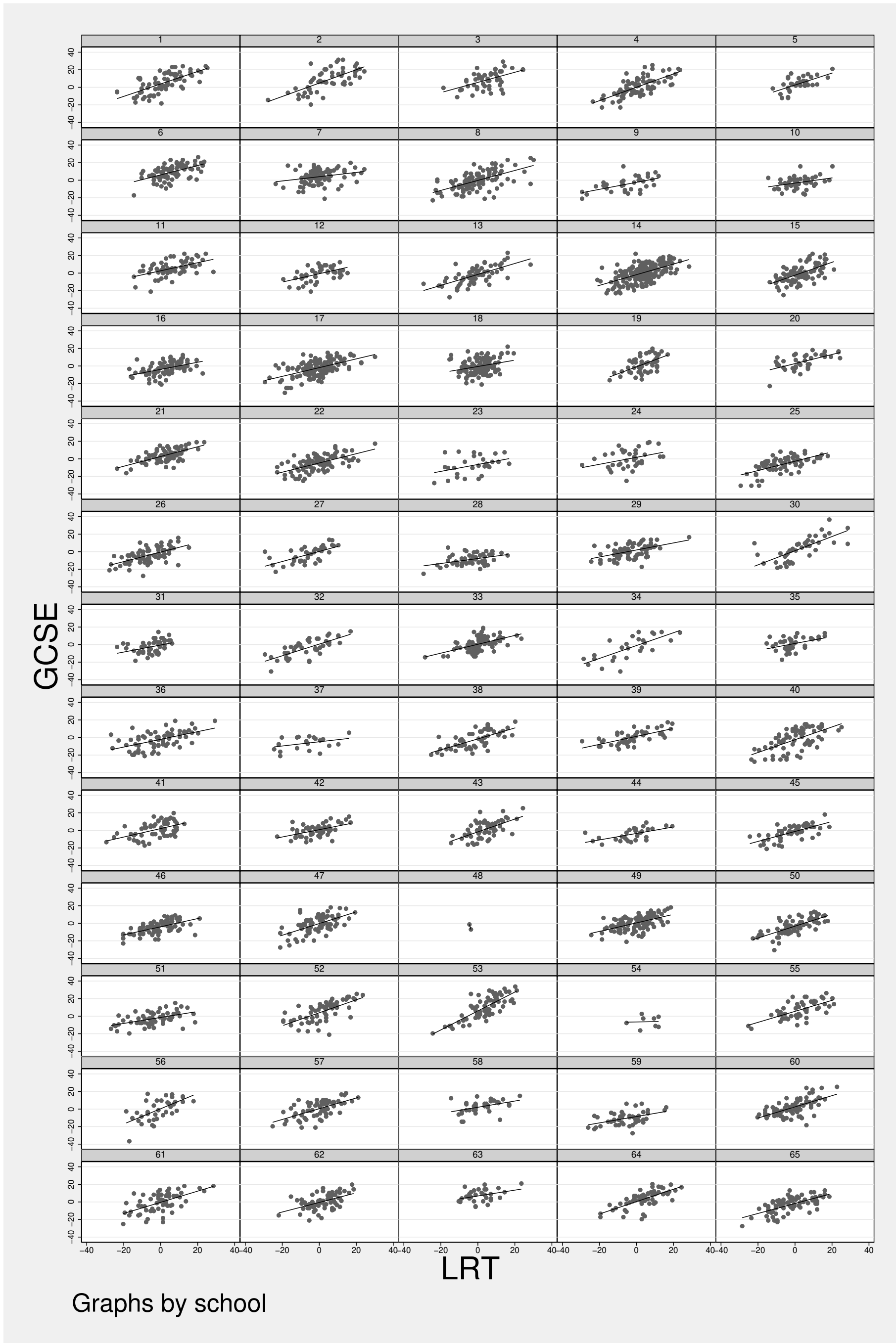


Figure 4.2: Trellis of scatterplots of `gcse` versus `lrt` with fitted regression lines for all 65 schools

We will now fit a simple linear regression model for each school, which is easily done using Stata's prefix command `statsby`. Then we will examine the variability in the estimated intercepts and slopes.

First, calculate the number of students per school by using `egen` with the `count()` function to preclude fitting lines to schools with fewer than five students below:

```
. egen num = count(gcse), by(school)
```

Then, use `statsby` to create a new dataset, `ols.dta`, in the working directory with the variables `inter` and `slope` containing OLS estimates of the intercepts (`_b[_cons]`) and slopes (`_b[lrt]`) from the command `regress gcse lrt if num>4` applied to each school:

```
. statsby inter=_b[_cons] slope=_b[lrt], by(school) saving(ols):
> regress gcse lrt if num>4
(running regress on estimation sample)
      Command: regress gcse lrt if num>4
        inter: _b[_cons]
        slope: _b[lrt]
           By: school
Statsby groups
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
.............
```

The new dataset also contains the variable `school` and is sorted by `school`, making it easy to merge it into the original dataset (the "master data") after sorting the latter by `school`:

```
. sort school
. merge m:1 school using ols
    Result                           Number of obs
    -----------------------------------------
    Not matched                              2
        from master                          2  (_merge==1)
        from using                           0  (_merge==2)
    Matched                              4,057  (_merge==3)
    -----------------------------------------
. drop _merge
```

Here we have specified `m:1` in the `merge` command, which stands for "many-to-one merging" (observations for several students per school in the master data, but only one observation per school in the "using data"). We see that two of the students in the master data did not have matches in the using data (because their school, school 48, had fewer than 5 students in the data, so we did not compute OLS estimates for that school). We have deleted the variable `_merge` produced by the `merge` command to avoid error messages when we run the `merge` command in the future.

A scatterplot of the OLS estimates of the intercept and slope is produced using the following command and given in figure 4.3:

```
. twoway scatter slope inter, xtitle(Intercept) ytitle(Slope)
```

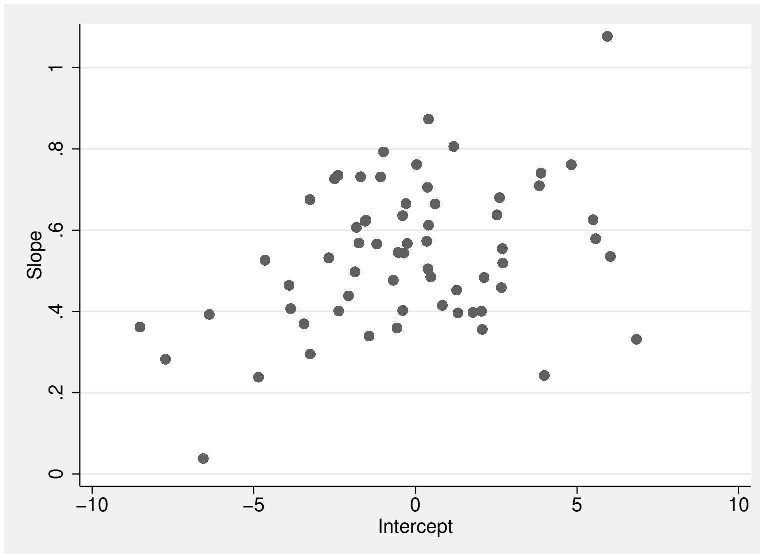


Figure 4.3: Scatterplot of estimated intercepts and slopes for all schools with at least five students

We see that there is considerable variability between the estimated intercepts and slopes of different schools. To investigate this further, we first create a dummy variable to pick out one observation per school,

```
. egen pickone = tag(school)
```

and then produce summary statistics for the schools by using the `summarize` command:

```
. summarize inter slope if pickone==1
```

Variable	Obs	Mean	Std. dev.	Min	Max
inter	64	-.1805974	3.291357	-8.519253	6.838716
slope	64	.5390514	.1766135	.0380965	1.076979

To allow comparison with the parameter estimates obtained from the random-coefficient model considered later on, we also obtain the covariance matrix of the estimated intercepts and slopes:

```
. correlate inter slope if pickone==1, covariance
(obs=64)
```

	inter	slope
inter	10.833	
slope	.208622	.031192

The diagonal elements, 10.83 and 0.03, are the sample variances of the intercepts and slopes, respectively. The off-diagonal element, 0.21, is the sample covariance between the intercepts and slopes, equal to the correlation times the product of the intercept and slope standard deviations.

We can also obtain a *spaghetti plot* of the predicted school-specific regression lines for all schools. We first calculate the fitted values $\widehat{y}_{ij} = \widehat{\beta}_{1j} + \widehat{\beta}_{2j} x_{ij}$,

```
. generate pred = inter + slope*lrt
(2 missing values generated)
```

and sort the data so that `lrt` increases within a given school and then jumps to its lowest value for the next school in the dataset:

```
. sort school lrt
```

We then produce the plot by typing

```
. twoway (line pred lrt, connect(ascending)), xtitle(LRT)
> ytitle(Fitted regression lines)
```

The `connect(ascending)` option is used to connect points only as long as `lrt` is increasing and ensures that only data for the same school are connected. The resulting graph is shown in figure 4.4.

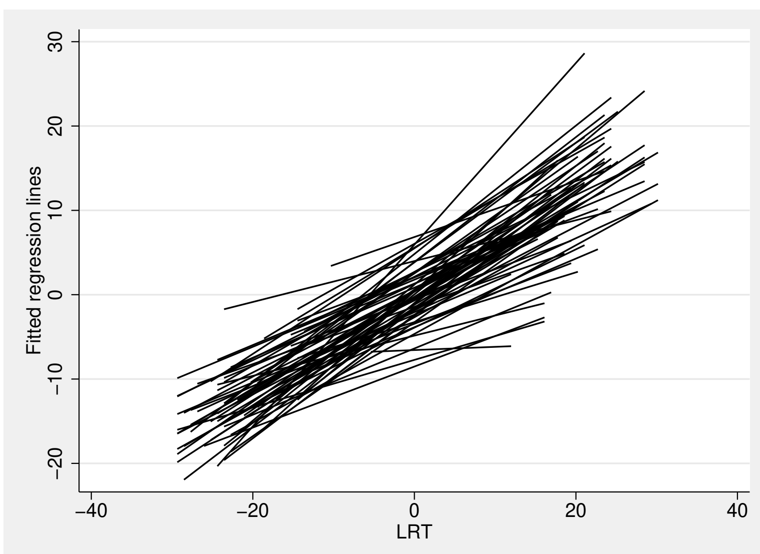


Figure 4.4: Spaghetti plot of ordinary least-squares regression lines for all schools with at least five students

4.4 Specification and interpretation of a random-coefficient model

4.4.1 Specification of a random-coefficient model

How can we develop a joint model for the relationships between `gcse` and `lrt` in all schools that allows intercepts and slopes to differ between schools?

One way would be to use dummy variables for all schools (omitting the overall intercept) and interactions between these dummy variables and `lrt` (omitting the overall slope of `lrt`). The school-specific intercepts are then the coefficients of the dummy variables and the school-specific slopes are the interaction coefficients. The only difference between the resulting model and separate regressions is that a common residual error variance $\theta_j = \theta$ is assumed. However, this model has 130 regression coefficients! Furthermore, if the schools are viewed as a (random) sample of schools from a population of schools, we are not interested in the individual coefficients characterizing each school's regression line. Rather, we would like to estimate the mean intercept and slope as well as the (co)variability of the intercepts and slopes in the population of schools.

A parsimonious model for the relationships between `gcse` and `lrt` can be obtained by specifying a school-specific random intercept ζ_{1j} and a school-specific random slope ζ_{2j} for `lrt` (x_{ij}):

$$\begin{aligned} y_{ij} &= \beta_1 + \beta_2 x_{ij} + \zeta_{1j} + \zeta_{2j} x_{ij} + \epsilon_{ij} \\ &= (\beta_1 + \zeta_{1j}) + (\beta_2 + \zeta_{2j}) x_{ij} + \epsilon_{ij} \end{aligned} \tag{4.1}$$

Here ζ_{1j} represents the deviation of school j's intercept from the mean intercept β_1, and ζ_{2j} represents the deviation of school j's slope from the mean slope β_2.

Given all covariates $\mathbf{X}_j$ in cluster j, it is assumed that the random effects ζ_{1j} and ζ_{2j} have zero expectations:

$$E(\zeta_{1j}|\mathbf{X}_j) = 0$$

$$E(\zeta_{2j}|\mathbf{X}_j) = 0$$

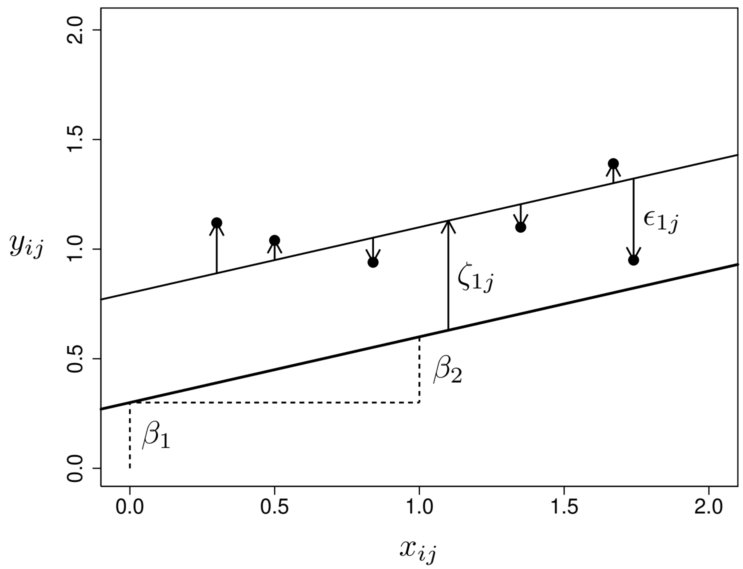
It is also assumed that the level-1 residual ϵ_{ij} has zero expectation, given the covariates and the random effects:

$$E(\epsilon_{ij}|\mathbf{X}_j, \zeta_{1j}, \zeta_{2j}) = 0$$

It follows from these mean-independence assumptions that the random terms ζ_{1j}, ζ_{2j}, and ϵ_{ij} are all uncorrelated with the covariate x_{ij} and with $\overline{x}_{\cdot j}$ and that ϵ_{ij} is uncorrelated with both ζ_{1j} and ζ_{2j}. Both the intercepts ζ_{1j} and slopes ζ_{2j} are assumed to be uncorrelated across schools, and the level-1 residuals ϵ_{ij} are assumed to be uncorrelated across schools and students.

An illustration of this random-coefficient model with one covariate x_{ij} for one cluster j is shown in the bottom panel of figure 4.5. A random-intercept model is shown for comparison in the top panel.

Random-intercept model



Random-coefficient model

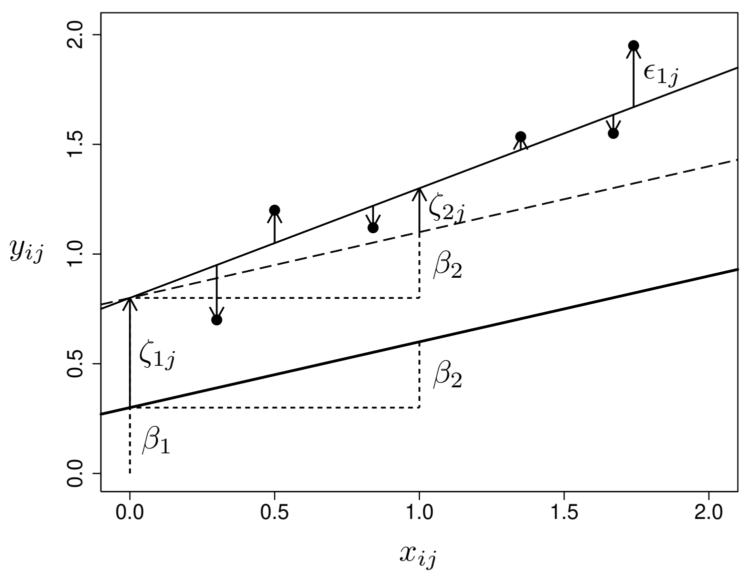


Figure 4.5: Illustration of random-intercept and random-coefficient models

In each panel, the lower bold and solid line represents the population-averaged or marginal regression line

$$E(y_{ij}|x_{ij}) = \beta_1 + \beta_2 x_{ij}$$

across all clusters. The higher and thinner solid line represents the cluster-specific regression line for cluster j. The arrows from the cluster-specific regression lines to the responses y_{ij} are the within-cluster residual error terms ϵ_{ij} (with variance θ).

For the random-intercept model, the cluster-specific line is

$$E(y_{ij}|x_{ij}, \zeta_{1j}) = (\beta_1 + \zeta_{1j}) + \beta_2 x_{ij}$$

which is parallel to the population-averaged line with vertical displacement given by the random intercept ζ_{1j}. In contrast, in the random-coefficient model, the cluster-specific or conditional regression line

$$E(y_{ij}|x_{ij}, \zeta_{1j}, \zeta_{2j}) = (\beta_1 + \zeta_{1j}) + (\beta_2 + \zeta_{2j}) x_{ij}$$

is not parallel to the population-averaged line but has a greater slope because the random slope ζ_{2j} is positive in the illustration. Here the dashed line is parallel to the population-averaged regression line and has the same intercept as cluster j. The vertical deviation between this dashed line and the line for cluster j is $\zeta_{2j}x_{ij}$, as shown in the diagram for $x_{ij} = 1$. The bottom panel illustrates that the total intercept for cluster j is $\beta_1 + \zeta_{1j}$ and the total slope is $\beta_2 + \zeta_{2j}$. It is clear that $\zeta_{2j}x_{ij}$ represents an *interaction* between the clusters, treated as random, and the covariate x_{ij}.

Given $\mathbf{X}_j$, the random intercept and random slope have a bivariate distribution assumed to have 0 means and covariance matrix $\mathbf{\Psi}$:

$$\mathbf{\Psi} = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \equiv \begin{bmatrix} \text{Var}(\zeta_{1j}|\mathbf{X}_j) & \text{Cov}(\zeta_{1j}, \zeta_{2j}|\mathbf{X}_j) \\ \text{Cov}(\zeta_{2j}, \zeta_{1j}|\mathbf{X}_j) & \text{Var}(\zeta_{2j}|\mathbf{X}_j) \end{bmatrix}, \qquad \psi_{21} = \psi_{12}$$

Hence, given the covariates, the variance of the random intercept is ψ_{11}, the variance of the random slope is ψ_{22}, and the covariance between the random intercept and the random slope is ψ_{21}. The correlation between the random intercept and random slope given the covariates becomes

$$\rho_{21} \equiv \text{Cor}(\zeta_{1j}, \zeta_{2j}|\mathbf{X}_j) \;=\; \frac{\psi_{21}}{\sqrt{\psi_{11}\psi_{22}}}$$

For maximum likelihood (ML) and restricted maximum likelihood (REML) estimation a normal distribution is specified for the level-1 error ϵ_{ij} and a bivariate normal distribution for the random intercept and random slope, given $\mathbf{X}_j$. An example of a bivariate normal distribution with $\psi_{11} = \psi_{22} = 4$ and $\psi_{21} = \psi_{12} = 1$ is shown as a perspective plot in figure 4.6. Specifying a bivariate normal distribution implies that the (marginal) univariate distributions of the intercept and slope are also normal.

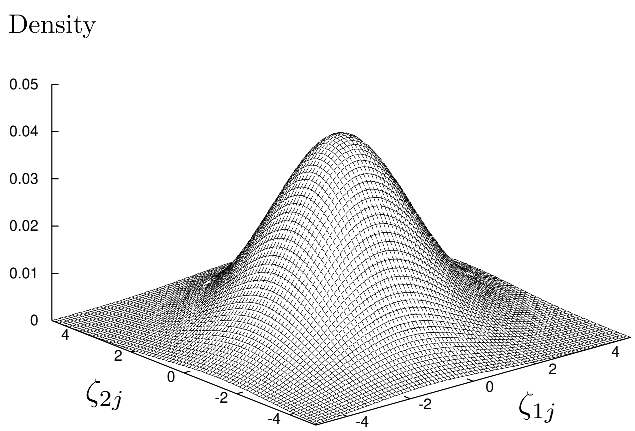


Figure 4.6: Perspective plot of bivariate normal distribution

4.4.2 Interpretation of the random-effects variances and covariances

Interpreting the covariance matrix $\mathbf{\Psi}$ of the random effects (given the covariates $\mathbf{X}_j$) is not completely straightforward.

First, the random-slope variance ψ_{22} and the covariance between random slope and intercept ψ_{21} depend not just on the scale of the response variable but also on the scale of the covariate, here `lrt`. Let the units of the response and covariate be denoted as u_y and u_x, respectively. For instance, in an application in chapter 7 that considers children's increase in weight over time, u_y is kilograms and u_x is years. The units of ψ_{11} are u_y^2, the units of ψ_{21} are u_y^2/u_x, and the units of ψ_{22} are u_y^2/u_x^2. It therefore does not make sense to compare the magnitude of random-intercept and random-slope variances.

Another issue is that the total residual variance is no longer constant as in random-intercept models. The total residual is now

$$\xi_{ij} \equiv \zeta_{1j} + \zeta_{2j}x_{ij} + \epsilon_{ij}$$

and the conditional variance of the responses given the covariate, or the conditional variance of the total residual, is

$$\text{Var}(y_{ij}|\mathbf{X}_j) = \text{Var}(\xi_{ij}|\mathbf{X}_j) = \psi_{11} + 2\psi_{21}x_{ij} + \psi_{22}x_{ij}^2 + \theta \tag{4.2}$$

This variance is a (quadratic) function of the covariate x_{ij}, and the total residual is therefore *heteroskedastic*. The conditional covariance for two students i and i' with covariate values x_{ij} and $x_{i'j}$ in the same school j is

$$\begin{aligned} \text{Cov}(y_{ij}, y_{i'j}|\mathbf{X}_j) &= \text{Cov}(\xi_{ij}, \xi_{i'j}|\mathbf{X}_j) \\ &= \psi_{11} + \psi_{21}x_{ij} + \psi_{21}x_{i'j} + \psi_{22}x_{ij}x_{i'j} \end{aligned} \tag{4.3}$$

and the conditional intraclass correlation becomes

$$\text{Cor}(y_{ij}, y_{i'j}|\mathbf{X}_j) = \frac{\text{Cov}(\xi_{ij}, \xi_{i'j}|\mathbf{X}_j)}{\sqrt{\text{Var}(\xi_{ij}|\mathbf{X}_j)\text{Var}(\xi_{i'j}|\mathbf{X}_j)}}$$

where we can plug in the covariance from (4.3) and the variances from (4.2). When $x_{ij} = x_{i'j} = 0$, the expression for the intraclass correlation is the same as for the random-intercept model and represents the correlation of the total residuals (from the overall mean regression line) for two students in the same school who both have `lrt` scores equal to 0 (the mean in this case). However, for pairs of students i and i' in the same school j with other values of `lrt`, the intraclass correlation is a complicated function of `lrt` (x_{ij} and $x_{i'j}$).

Due to the heteroskedastic total residual variance, it is not straightforward to define coefficients of determination—such as R^2, R_2^2, and R_1^2, discussed in section 3.5—for random-coefficient models. Snijders and Bosker (2012, 114) suggest removing the random coefficient(s) for the purpose of calculating the coefficient of determination because this will usually yield values that are close to correct (see their section 7.2.2 for how to obtain the correct version).

Finally, interpreting the parameters ψ_{11} and ψ_{21} can be difficult because their values depend on the translation of the covariate or, in other words, on how much we add or subtract from the covariate. Adding a constant to `lrt` and refitting the model would result in different estimates of ψ_{11} and ψ_{21} (see also exercise 4.9). This is because the intercept variance is the variability in the vertical positions of school-specific regression lines where `lrt`=0 (the position where `lrt`=0 changes when `lrt` is translated) and the covariance or correlation is the tendency for regression lines that are higher up where `lrt`=0 to have higher slopes. This lack of invariance of ψ_{11} and ψ_{21} to translation of the covariate x_{ij} is illustrated in figure 4.7. Here identical cluster-specific regression lines are shown in the two panels, but the covariate $x'_{ij} = x_{ij} - 3.5$ in the lower panel is translated relative to the covariate x_{ij} in the upper panel. The intercepts are the intersections of the regression lines with the vertical lines at 0. Clearly these intercepts vary more in the upper panel than the lower panel, whereas the correlation between intercepts and slopes is negative in the upper panel and positive in the lower panel.

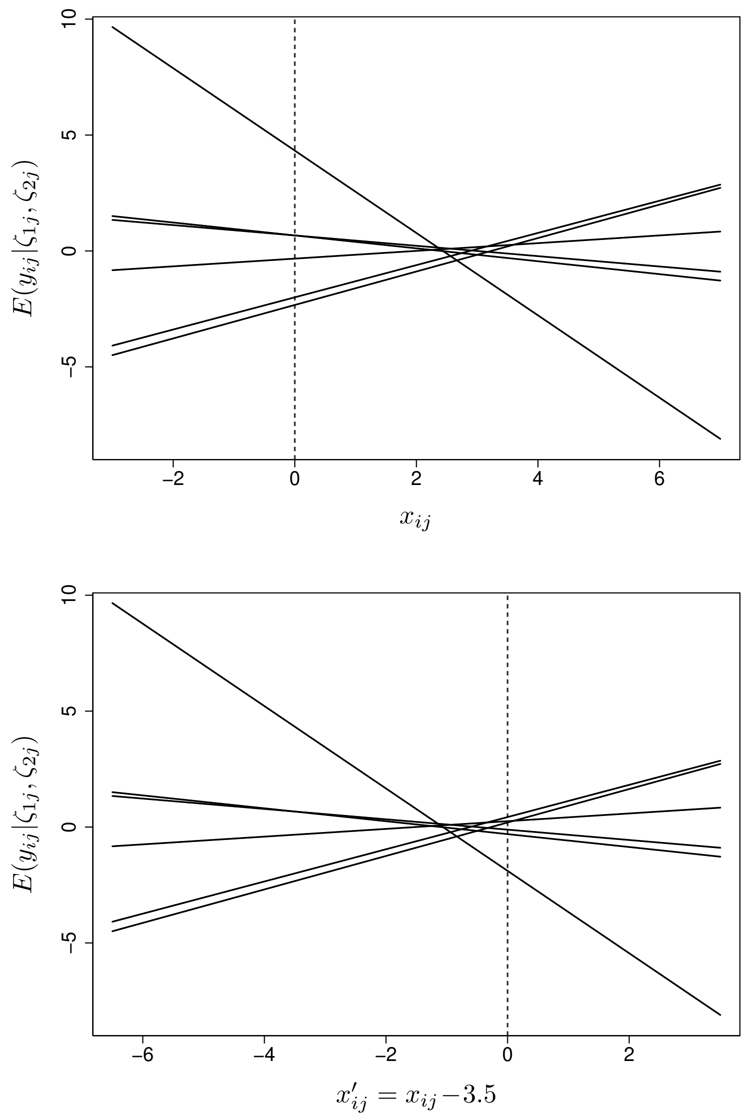


Figure 4.7: Cluster-specific regression lines for random-coefficient model, illustrating lack of invariance under translation of covariate (*Source:* Skrondal and Rabe-Hesketh 2004)

To make ψ_{11} and ψ_{21} interpretable, it makes sense to translate x_{ij} so that the value $x_{ij} = 0$ is a useful reference point in some way. Typical choices are either mean centering (as for `lrt`) or, if x_{ij} is time, as in growth-curve models, defining 0 to be the initial time in some sense. Because the magnitude and interpretation of ψ_{21} depend on the location (or translation) of x_{ij}, which is often arbitrary, it generally does not make sense to set ψ_{21} to 0 by specifying uncorrelated intercepts and slopes.

A useful way of interpreting the magnitudes of the estimated variances $\widehat{\psi}_{11}$ and $\widehat{\psi}_{22}$ is by constructing intervals that contain the intercepts and slopes of 95% of clusters in the population (treating estimates as known parameters). Assuming that the intercepts and slopes are normally distributed with means $\widehat{\beta}_1$ and $\widehat{\beta}_2$ and variances $\widehat{\psi}_{11}$ and $\widehat{\psi}_{22}$, these intervals are $\widehat{\beta}_1 \pm 1.96\sqrt{\widehat{\psi}_{11}}$ and $\widehat{\beta}_2 \pm 1.96\sqrt{\widehat{\psi}_{22}}$. To aid interpretation of the random part of the model, it is also useful to produce plots of predicted school-specific regression lines, as discussed in section 4.8.3.

4.5 Estimation using mixed

The `mixed` command can be used to fit linear random-coefficient models by ML or REML. (`xtreg` can only fit two-level random-intercept models.)

4.5.1 Random-intercept model

We first consider a random-intercept model discussed in the previous chapter:

$$y_{ij} = (\beta_1 + \zeta_{1j}) + \beta_2 x_{ij} + \epsilon_{ij}$$

This model is a special case of the random-coefficient model in (4.1) with $\zeta_{2j} = 0$ or, equivalently, with zero random-slope variance and zero random-intercept and random-slope covariance, $\psi_{22} = \psi_{21} = 0$.

ML estimates for the random-intercept model can be obtained using `mixed` with the `mle` option (the default), and we also use the `vce(robust)` option for robust standard errors:

```
. mixed gcse lrt || school:, mle stddeviations vce(robust)
Mixed-effects regression                        Number of obs     =      4,059
Group variable: school                          Number of groups  =         65
                                                Obs per group:
                                                              min =          2
                                                              avg =       62.4
                                                              max =        198
                                                Wald chi2(1)      =     852.73
Log pseudolikelihood = -14024.799               Prob > chi2       =     0.0000
                                (Std. err. adjusted for 65 clusters in school)
```

gcse	Coefficient	Robust std. err.	z	P>|z|	[95% conf. interval]	
lrt	.5633697	.0192925	29.20	0.000	.5255572	.6011823
_cons	.0238706	.4050143	0.06	0.953	-.7699428	.8176841

Random-effects parameters	Estimate	Robust std. err.	[95% conf. interval]	
school: Identity				
sd(_cons)	3.035269	.3154741	2.475863	3.72107
sd(Residual)	7.521481	.1306016	7.269813	7.781861

To allow later comparison with random-coefficient models via likelihood-ratio tests, we store these estimates by using

```
. estimates store ri
```

The random-intercept model assumes that the school-specific regression lines are parallel. The common coefficient or slope β_2 of `lrt`, shared by all schools, is estimated as 0.56 and the mean intercept as 0.02. Schools vary in their intercepts with an estimated standard deviation of 3.04. Within the schools, the estimated residual standard deviation around the school-specific regression lines is 7.52. The within-school correlation, after controlling for `lrt`, is therefore estimated as

$$\widehat{\rho} \;=\; \frac{\widehat{\psi}_{11}}{\widehat{\psi}_{11}+\widehat{\theta}} \;=\; \frac{3.035^2}{3.035^2+7.521^2} = 0.14$$

We could obtain this within-school correlation by typing `estat icc`.

The ML estimates for the random-intercept model are also given under "Random intercept" in table 4.1.

Table 4.1: Maximum likelihood estimates for inner-London-schools data with robust standard errors

Parameter	Random intercept Est	(SE)	Random coefficient Est	(SE)	Rand. coefficient & level-2 covariates Est	(SE)	γ_{xx}
Fixed part							
β_1 [_cons]	0.02	(0.41)	−0.12	(0.40)	−1.00	(0.55)	γ_{11}
β_2 [lrt]	0.56	(0.02)	0.56	(0.02)	0.57	(0.02)	γ_{21}
β_3 [boys]					0.85	(0.96)	γ_{12}
β_4 [girls]					2.43	(0.84)	γ_{13}
β_5 [boys_lrt]					−0.02	(0.05)	γ_{22}
β_6 [girls_lrt]					−0.03	(0.05)	γ_{23}
Random part							
$\sqrt{\psi_{11}}$	3.04		3.01		2.80		
$\sqrt{\psi_{22}}$			0.12		0.12		
ρ_{21}			0.50		0.60		
$\sqrt{\theta}$	7.52		7.44		7.44		
Log likelihood	$-14,024.80$		$-14,004.61$		$-13,998.83$		

4.5.2 Random-coefficient model

We now relax the assumption that the school-specific regression lines are parallel by introducing random school-specific slopes $\beta_2 + \zeta_{2j}$ of `lrt`:

$$y_{ij} \;=\; (\beta_1 + \zeta_{1j}) + (\beta_2 + \zeta_{2j})x_{ij} + \epsilon_{ij}$$

To introduce a random slope for `lrt` using `mixed`, we simply add that variable name in the specification of the random part, replacing `school:` with `school: lrt`. We must also specify the `covariance(unstructured)` option because `mixed` will otherwise set the covariance ψ_{21} (and the corresponding correlation) to 0 by default. ML estimates for the random-coefficient model are then obtained using

```
. mixed gcse lrt || school: lrt, covariance(unstructured) mle stddeviations
> vce(robust)
Mixed-effects regression                        Number of obs     =      4,059
Group variable: school                          Number of groups  =         65
                                                Obs per group:
                                                              min =          2
                                                              avg =       62.4
                                                              max =        198
                                                Wald chi2(1)      =     767.80
Log pseudolikelihood = -14004.613               Prob > chi2       =     0.0000
                                (Std. err. adjusted for 65 clusters in school)
------------------------------------------------------------------------------
             |               Robust
        gcse | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         lrt |    .556729   .0200919    27.71   0.000     .5173496    .5961084
       _cons |   -.115085   .4009294    -0.29   0.774    -.9008922    .6707222
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects parameters  |   Estimate   std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
                     sd(lrt) |   .1205646   .0236268      .0821128    .1770224
                   sd(_cons) |   3.007444   .3134589      2.451765    3.689065
              corr(lrt,_cons) |   .4975415   .1751841      .0894796    .7625783
-----------------------------+------------------------------------------------
                sd(Residual) |   7.440787   .1251535       7.19949    7.690172
------------------------------------------------------------------------------
```

Because the `stddeviations` option was used, the output shows the standard deviations, `sd(lrt)`, of the slope and `sd(_cons)` of the intercept instead of variances. It also shows the correlation between intercepts and slopes, `corr(lrt,_cons)`, instead of the covariance. We can obtain the estimated covariance matrix either by replaying the estimation results without the `stddeviations` option (or with the `variance` option),

```
mixed, variance
```

or by using the postestimation command `estat recovariance`:

```
. estat recovariance
Random-effects covariance matrix for level school
             |       lrt      _cons
-------------+----------------------
         lrt |   .0145358
       _cons |   .1804042    9.04472
```

The ML estimates for the random-coefficient model were also given under "Random coefficient" in table 4.1. We store the estimates under the name `rc` for later use:

```
. estimates store rc
```

We can also obtain the model-implied residual standard deviations and correlations among the GCSE scores for students in a particular school by using the `estat` wcorrelation command. For schools that have many students in the data, the correlation matrix is too large to display without wrapping, so we choose school 54, which has 8 students in the data, for illustration. First, we sort the data in ascending order of lrt within school and list the values of lrt because they will affect both the standard deviations and correlations, as shown in equations (4.3) and (4.2):

```
. sort school lrt
. list school lrt if school==54, clean noobs
   school       lrt
       54   -5.3806
       54     2.058
       54    2.8845
       54     3.711
       54    9.4967
       54    10.323
       54    11.976
       54    11.976
```

Now, we obtain the estimated residual standard deviations and correlations for school 54:

```
. estat wcorrelation, at(school=54)
Standard deviations and correlations for school = 54:
Standard deviations:
```

obs	1	2	3	4	5	6	7	8
sd	7.930	8.076	8.098	8.121	8.315	8.348	8.415	8.415

```
Correlations:
```

obs	1	2	3	4	5	6	7	8
1	1.000							
2	0.129	1.000						
3	0.130	0.153	1.000					
4	0.131	0.155	0.158	1.000				
5	0.137	0.170	0.173	0.177	1.000			
6	0.138	0.172	0.175	0.179	0.202	1.000		
7	0.139	0.176	0.179	0.183	0.208	0.212	1.000	
8	0.139	0.176	0.179	0.183	0.208	0.212	0.218	1.000

The standard deviations increase with increasing lrt. To interpret the pattern of the correlations, we can look down the columns, which corresponds to holding the lrt for one student constant and looking at the correlations as the lrt of the other student increases. We see that the corresponding correlations increase.

Here we used ML estimation. REML estimation should be used instead when the number of clusters is small ($J - q < 42$, see display 2.1) and this method is requested by specifying the reml option.