The linear mixed model: modeling hierarchical and longitudinal data

Yves Rosseel
Department of Data Analysis
Ghent University

Summer School – Using R for personality research
August 23–28, 2014
Bertinoro, Italy
Contents

1 Modeling Hierarchical Data 3
   1.1 Overview ................................................................. 3
   1.2 Example: the “High School and Beyond” data ................. 4
   1.3 Further issues ......................................................... 39

2 Modeling Longitudinal Data 40
   2.1 Example: Reaction times in a sleep deprivation study ....... 41
1 Modeling Hierarchical Data

1.1 Overview

• especially in education (think ‘students in schools’), applications of mixed models to hierarchical data have become very popular

• when researchers in the social sciences talk about ‘multilevel analysis’, they typically mean the application of mixed models to hierarchical data

• typical for this type of data is the sharp distinction between ‘levels’: the ‘student level’, and the ‘school level’, where one level (student) is nested in the other (school)

• data often contains both student characteristics and school characteristics; in a multilevel analysis, both types of variables can be included in one analysis simultaneously
1.2 Example: the “High School and Beyond” data

- the following example is borrowed from Raudenbusch and Bryk (2001, chapter 4).

- the data are from the 1982 “High School and Beyond” survey, and pertain to 7185 U.S. high-school students from 160 schools (70 catholic, 90 public)

- these are the variables in the dataset:
  
  - **school** an ordered factor designating the school that the student attends.
  - **minrty** a factor with 2 levels (not used)
  - **sx** a factor with levels Male and Female (not used)
  - **ses** a numeric vector of socio-economic scores
  - **mAch** a numeric vector of Mathematics achievement scores
  - **meanses** a numeric vector of mean ses for the school
  - **sector** a factor with levels Public and Catholic
  - **ceses** a numeric vector of centered ses values where the centering is with respect to the meanses for the school
• the aim of the analysis is to determine how students’ math achievement scores are related to their family socioeconomic status

• but this relationship may very well vary among schools

• if there is indeed variation among schools, can we find any school characteristics that ‘explain’ this variation? the two school characteristics that we will use are:

  – **sector**: public school or Catholic school
  – **meanses**: the average SES of students in the school
### Exploring the data

```r
> library(mlmRev)
> summary(Hsb82)
```

<table>
<thead>
<tr>
<th></th>
<th>school</th>
<th>minrty</th>
<th>sx</th>
<th>ses</th>
<th>mAch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2305</td>
<td>No</td>
<td>67</td>
<td>-3.758000</td>
<td>-2.832</td>
</tr>
<tr>
<td></td>
<td>5619</td>
<td>Yes</td>
<td>66</td>
<td>-0.538000</td>
<td>7.275</td>
</tr>
<tr>
<td></td>
<td>4292</td>
<td>Male</td>
<td>65</td>
<td>0.000143</td>
<td>12.748</td>
</tr>
<tr>
<td></td>
<td>8857</td>
<td>Female</td>
<td>64</td>
<td>0.602000</td>
<td>18.317</td>
</tr>
<tr>
<td>(Other)</td>
<td>4042</td>
<td>No</td>
<td>64</td>
<td>2.692000</td>
<td>24.993</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>meanses</th>
<th>sector</th>
<th>cses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>-1.1939459</td>
<td>Public</td>
<td>-3.6507</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-0.3230000</td>
<td>Catholic</td>
<td>-0.4479</td>
</tr>
<tr>
<td>Median</td>
<td>0.0320000</td>
<td>Median</td>
<td>0.0160</td>
</tr>
<tr>
<td>Mean</td>
<td>0.000143</td>
<td>Mean</td>
<td>0.0000</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.3269123</td>
<td>3rd Qu.</td>
<td>0.4694</td>
</tr>
<tr>
<td>Max.</td>
<td>0.8249825</td>
<td>Max.</td>
<td>2.8561</td>
</tr>
</tbody>
</table>

```r
> head(Hsb82)
```

<table>
<thead>
<tr>
<th></th>
<th>school</th>
<th>minrty</th>
<th>sx</th>
<th>ses</th>
<th>mAch</th>
<th>meanses</th>
<th>sector</th>
<th>cses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1224</td>
<td>No</td>
<td>-1.528</td>
<td>5.876</td>
<td>-0.434383</td>
<td>Public</td>
<td>-1.09361702</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1224</td>
<td>No</td>
<td>-0.588</td>
<td>19.708</td>
<td>-0.434383</td>
<td>Public</td>
<td>-0.15361702</td>
<td></td>
</tr>
</tbody>
</table>
> tail(Hsb82)

<table>
<thead>
<tr>
<th>school</th>
<th>minority</th>
<th>sx</th>
<th>ses</th>
<th>mAch</th>
<th>meanses</th>
<th>sector</th>
<th>cses</th>
</tr>
</thead>
<tbody>
<tr>
<td>7180</td>
<td>Yes</td>
<td>Female</td>
<td>1.612</td>
<td>20.967</td>
<td>0.6211525</td>
<td>Catholic</td>
<td>0.9908475</td>
</tr>
<tr>
<td>7181</td>
<td>No</td>
<td>Female</td>
<td>1.512</td>
<td>20.402</td>
<td>0.6211525</td>
<td>Catholic</td>
<td>0.8908475</td>
</tr>
<tr>
<td>7182</td>
<td>No</td>
<td>Female</td>
<td>-0.038</td>
<td>14.794</td>
<td>0.6211525</td>
<td>Catholic</td>
<td>-0.6591525</td>
</tr>
<tr>
<td>7183</td>
<td>No</td>
<td>Female</td>
<td>1.332</td>
<td>19.641</td>
<td>0.6211525</td>
<td>Catholic</td>
<td>0.7108475</td>
</tr>
<tr>
<td>7184</td>
<td>No</td>
<td>Female</td>
<td>-0.008</td>
<td>16.241</td>
<td>0.6211525</td>
<td>Catholic</td>
<td>-0.6291525</td>
</tr>
<tr>
<td>7185</td>
<td>No</td>
<td>Female</td>
<td>0.792</td>
<td>22.733</td>
<td>0.6211525</td>
<td>Catholic</td>
<td>0.1708475</td>
</tr>
</tbody>
</table>
Exploring the data: math achievement versus ses for several schools

```r
> set.seed(1234)
> library(lattice)
> cat  <- sample(unique(Hsb82$school[Hsb82$sector == "Catholic"]), 18)
> Cat.18 <- Hsb82[is.element(Hsb82$school, cat),]
> plot1 <- xyplot(mAch ~ ses | school, data=Cat.18, main="Catholic",
+     xlab="student SES", ylab="Math Achievement", layout=c(6, 3),
+     panel=function(x, y){
+         panel.xyplot(x, y)
+         #panel.loess(x, y, span=1)
+         panel.lmline(x, y)
+     })

> pub  <- sample(unique(Hsb82$school[Hsb82$sector == "Public"]), 18)
> Pub.18 <- Hsb82[is.element(Hsb82$school, pub),]
> plot2 <- xyplot(mAch ~ ses | school, data=Pub.18, main="Public",
+     xlab="student SES", ylab="Math Achievement", layout=c(6, 3),
+     panel=function(x, y){
+         panel.xyplot(x, y)
+         #panel.loess(x, y, span=1)
+         panel.lmline(x, y)
+     })
```
Fitting a simple regression model for each school separately

```r
> cat.list <- lmList(mAch ~ ses | school, subset = sector=='Catholic',
+ data=Hsb82)
> head(coef(cat.list))

(Intercept)     ses
7172   8.355037 0.9944805
4868  11.845609 1.2864712
2305  10.646595 -0.7821112
8800   9.157380 2.5681254
5192  10.511666 1.6034950
4523   8.479699 2.3807892

> pub.list <- lmList(mAch ~ ses | school, subset = sector=='Public',
+ data=Hsb82)
> head(coef(pub.list))

(Intercept)     ses
8367   4.546383 0.2503748
8854   5.707002 1.9388446
4458   6.999212 1.1318372
5762   3.114085 -1.0140992
6990   6.440875 0.9476903
5815   9.323601 3.0180018
```
Relationship intercepts and slopes and mean school SES

![Scatter plots showing the relationship between intercepts, slopes, and mean SES school. The plots illustrate the distribution of data points and the trend lines associated with the linear mixed model.](image)
Model 1: a random-effects one-way ANOVA

• this is often called the ‘empty’ model, since it contains no predictors, but simply reflects the nested structure

• no level-1 predictors, no level-2 predictors

• in the ‘multilevel’ notation we specify a model for each level

• model for the first (student) level:

\[ y_{ij} = \alpha_{0i} + \epsilon_{ij} \]

• model for the second (school) level:

\[ \alpha_{0i} = \gamma_{00} + u_{0i} \]

• the combined model and the Laird-Ware form:

\[ y_{ij} = \gamma_{00} + u_{0i} + \epsilon_{ij} \]

\[ = \beta_0 + b_{0i} + \epsilon_{ij} \]
• this is an example of a random-effects one-way ANOVA model with one fixed effect (the intercept, $\beta_0$) representing the general population mean of math achievement, and two random effects:

  – $b_{0i}$ representing the deviation of math achievement in school $i$ from the general mean
  – $\epsilon_{ij}$ representing the deviation of individual $j$’s math achievement in school $i$ from the school mean

• there are two variance components for this model:

  – $\text{Var}(b_{0i}) = d^2$: the variance among school means
  – $\text{Var}(\epsilon_{ij}) = \sigma^2$: the variance among individuals in the same school

• since $b_{0i}$ and $\epsilon_{ij}$ are assumed to be independent, the variation in math scores among individuals can be decomposed into these two variance components:

  $$\text{Var}(y_{ij}) = d^2 + \sigma^2$$
R code

> fit.model1 <- lmer(mAch ~ 1 + (1 | school), data=Hsb82)
> summary(fit.model1)

Linear mixed model fit by REML ['lmerMod']
Formula: mAch ~ 1 + (1 | school)
  Data: Hsb82

REML criterion at convergence: 47116.79

Random effects:
  Groups   Name       Variance Std.Dev.
         school (Intercept)  8.614    2.935
         Residual            39.148   6.257
Number of obs: 7185, groups: school, 160

Fixed effects:
  Estimate Std. Error  t value
(Intercept)    12.6370     0.2444   51.71
Intra-class correlation

- The *intra-class correlation coefficient* is the proportion of variation in individuals’ scores due to differences among schools:

\[
\frac{d^2}{\text{Var}(y_{ij})} = \frac{d^2}{d^2 + \sigma^2} = \rho
\]

- \( \rho \) may also be interpreted as the correlation between the math scores of two individuals from the same school:

\[
\text{Cor}(y_{ij}, y_{ij'}) = \frac{\text{Cov}(y_{ij}, y_{ij'})}{\sqrt{\text{Var}(y_{ij}) \times \text{Var}(y_{ij'})}} = \frac{d^2}{d^2 + \sigma^2} = \rho
\]

\[
\begin{aligned}
> d2 &\leftarrow \text{as.numeric(VarCorr(fit.model1)$school)} \\
> s2 &\leftarrow \text{attr(VarCorr(fit.model1), "sc")}^2 \\
> \text{rho} &\leftarrow d2/(d2 + s2) \\
> \text{rho}
\end{aligned}
\]

[1] 0.1803472

- About 18 percent of the variation in students’ match-achievement scores is “attributable” to differences among schools.
• this is often called ‘the cluster effect’
Model 2: a random-effects one-way ANCOVA

- 1 level-1 predictor (SES, centered within school), no level-2 predictors
- random intercept, no random slopes
- model for the first (student) level:

\[ y_{ij} = \alpha_{0i} + \alpha_{1i} \text{cses}_{ij} + \epsilon_{ij} \]

- model for the second (school) level:

\[ \alpha_{0i} = \gamma_{00} + u_{0i} \quad \text{(the random intercept)} \]
\[ \alpha_{1i} = \gamma_{10} \quad \text{(the constant slope)} \]

- the combined model and the Laird-Ware form:

\[ y_{ij} = (\gamma_{00} + u_{0i}) + \gamma_{10} \text{cses}_{ij} + \epsilon_{ij} \]
\[ = \gamma_{00} + \gamma_{10} \text{cses}_{ij} + u_{0i} + \epsilon_{ij} \]
\[ = \beta_0 + \beta_1 x_{1ij} + b_{0i} + \epsilon_{ij} \]
• the fixed-effect coefficients $\beta_0$ and $\beta_1$ represent the average within-schools population intercept and slope respectively
  
  – note: because SES is centered within schools, the intercept $\beta_0$ represents the ‘average’ level of math achievement in the population

• the model has two variance-covariance components:
  
  – $\text{Var}(b_{0i}) = d^2$: the variance among school intercepts
  – $\text{Var}(\epsilon_{ij}) = \sigma^2$: the error variance around the within-school regressions
R code

```r
> fit.model2 <- lmer(mAch ~ 1 + cses + (1 | school), data=Hsb82)
> summary(fit.model2)

Linear mixed model fit by REML ['lmerMod']
Formula: mAch ~ 1 + cses + (1 | school)
  Data: Hsb82

REML criterion at convergence: 46724

Random effects:
  Groups   Name   Variance  Std.Dev.
          school (Intercept)  8.672   2.945
          Residual           37.010   6.084
Number of obs: 7185, groups: school, 160

Fixed effects:
  Estimate   Std. Error   t value
(Intercept) 12.6361    0.2445     51.68
  cses       2.1912    0.1087     20.17

Correlation of Fixed Effects:
                      (Intr)  
cses  0.000
```

The linear mixed model: modeling hierarchical and longitudinal data
Model 3: a random-coefficients regression model

- 1 level-1 predictor (SES, centered within school), no level-2 predictors
- random intercept and random slopes
- model for the first (student) level:

\[ y_{ij} = \alpha_0 + \alpha_1 c\text{ses}_{ij} + \epsilon_{ij} \]

- model for the second (school) level:

\[
\begin{align*}
\alpha_0 &= \gamma_0 + u_0 \quad \text{(the random intercept)} \\
\alpha_1 &= \gamma_1 + u_1 \quad \text{(the random slope)}
\end{align*}
\]

- the combined model and the Laird-Ware form:

\[
\begin{align*}
y_{ij} &= (\gamma_0 + u_0) + (\gamma_1 + u_1) \text{cses}_{ij} + \epsilon_{ij} \\
&= \beta_0 + \beta_1 x_{1ij} + b_0 + b_0 z_{1ij} + \epsilon_{ij}
\end{align*}
\]
the fixed-effect coefficients $\beta_0$ and $\beta_1$ again represent the average within-schools population intercept and slope respectively

the model has four variance-covariance components:

- $\text{Var}(b_{0i}) = d_0^2$: the variance among school intercepts
- $\text{Var}(b_{1i}) = d_1^2$: the variance among school slopes
- $\text{Cov}(b_{0i}, b_{1i}) = d_{01}$: the covariance between within-school intercepts and slopes
- $\text{Var}(\epsilon_{ij}) = \sigma^2$: the error variance around the within-school regressions
R code

```r
> fit.model3 <- lmer(mAch ~ 1 + cses + (1 + cses | school), data=Hsb82)
> summary(fit.model3)

Linear mixed model fit by REML ['lmerMod']
Formula: mAch ~ 1 + cses + (1 + cses | school)
    Data: Hsb82

REML criterion at convergence: 46714.23

Random effects:
  Groups   Name   Variance  Std.Dev.  Corr
  school   (Intercept)  8.6812    2.946
       cses   0.6939    0.833    0.02
  Residual                  36.7002   6.058
Number of obs: 7185, groups: school, 160

Fixed effects:
  Estimate Std. Error t value
(Intercept)   12.6362    0.2445  51.68
       cses     2.1932    0.1283  17.10

Correlation of Fixed Effects:
   (Intr)
       cses  0.009
```

AED The linear mixed model: modeling hierarchical and longitudinal data 26 of 45
Is the random slope necessary?

R code

```r
> anova(fit.model2, fit.model3)
```

Data: Hsb82

Models:

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fit.model2</td>
<td>mAch ~ 1 + cses + (1</td>
<td>4</td>
<td>46728</td>
<td>46756</td>
<td>-23360</td>
<td>46720</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fit.model3</td>
<td>mAch ~ 1 + cses + (1 + cses</td>
<td>6</td>
<td>46723</td>
<td>46764</td>
<td>-23356</td>
<td>46711</td>
<td>9.433</td>
<td>2</td>
<td>0.008946 **</td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

- note: you can only compare two models using likelihood-ratio tests if the two models have the *same* fixed effects!
Model 4: random intercept + level-2 predictor

- we drop the level-1 predictor (cses), but add a level-2 predictor (meanses)
- random intercept, no random slopes
- model for the first (student) level:
  \[ y_{ij} = \alpha_0i + \epsilon_{ij} \]

- model for the second (school) level:
  \[ \alpha_0i = \gamma_{00} + \gamma_{01} \text{meanses}_i + u_{0i} \]

- the combined model and the Laird-Ware form:
  \[ y_{ij} = \gamma_{00} + \gamma_{01} \text{meanses}_i + u_{0i} + \epsilon_{ij} \]
  \[ = \beta_0 + \beta_1 x_{1ij} + b_{0i} + \epsilon_{ij} \]

- note that in the Laird-Ware notation, we use a double index for the fixed-effects \( x_{ij} \), even if the variable (meanses) does not change over students
this model has two fixed effects ($\beta_0$ and $\beta_1$) and two random effects ($b_{0i}$ and $\epsilon_{ij}$)

the interpretation of $b_{0i}$ has changed: whereas in the previous model it had been the deviation of school $i$’s mean from the grand mean, it now represents the residual ($\alpha_{0i} - \gamma_{00} - \gamma_{10}$ meanses$_j$); correspondingly, the variance component $d^2$ is now a conditional variance (conditional on the school mean SES)

in Raudenbush and Bryk, this is called a ‘Regression with means-as-outcomes’ model, because the school’s mean ($\alpha_{0i}$) is predicted by the means SES of the school

note in the following output that the residual variance between schools (2.64) is substantially smaller than the original (8.61)
R code

> fit <- lmer(mAch ~ 1 + meanses + (1 | school), data=Hsb82)
> summary(fit)

Linear mixed model fit by REML ['lmerMod']
Formula: mAch ~ 1 + meanses + (1 | school)
   Data: Hsb82

REML criterion at convergence: 46961.29

Random effects:
  Groups   Name   Variance  Std.Dev.
school   (Intercept)  2.639   1.624
  Residual                  39.157  6.258
Number of obs: 7185, groups: school, 160

Fixed effects:
     Estimate   Std. Error    t value
(Intercept) 12.6846     0.1493    84.97
meanses      5.8635     0.3615   16.22

Correlation of Fixed Effects:
     (Intr)
meanses  0.010
Model 5: random intercept + level-1 predictor + level-2 predictor

- one level-1 predictor (cses), and a level-2 predictor (meanses)
- random intercept, nonrandomly varying slopes
- model for the first (student) level:

\[ y_{ij} = \alpha_{0i} + \alpha_{1i} \text{cses}_{ij} + \epsilon_{ij} \]

- model for the second (school) level:

\[ \alpha_{0i} = \gamma_{00} + \gamma_{01} \text{meanses}_i + u_{0i} \]
\[ \alpha_{1i} = \gamma_{10} + \gamma_{11} \text{meanses}_i \]

- the combined model and the Laird-Ware form:

\[ y_{ij} = (\gamma_{00} + \gamma_{01} \text{meanses}_i + u_{0i}) + (\gamma_{10} + \gamma_{11} \text{meanses}_i) \text{cses}_{ij} + \epsilon_{ij} \]
\[ = \gamma_{00} + \gamma_{01} \text{meanses}_i + \gamma_{10} \text{cses}_{ij} + \gamma_{11} \text{meanses}_i \text{cses}_{ij} + u_{0i} + \epsilon_{ij} \]
\[ = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 (x_{1ij} x_{2ij}) + b_0i + \epsilon_{ij} \]
• this model illustrates that in a mixed model, we combine predictors from both levels to construct an interaction term

• no random slopes are used in this model; the idea is that once we control for the level-2 predictor (meanses), little or no variance in the slopes remains to be explained; the slopes vary across schools, but as a strict function of meanses

• this results in a random-intercept only model
R code

```r
fit <- lmer(mAch ~ 1 + meanses*cses + (1 | school), data=Hsb82)
summary(fit)
```

Linear mixed model fit by REML ['lmerMod']
Formula: mAch ~ 1 + meanses * cses + (1 | school)
Data: Hsb82

REML criterion at convergence: 46567.93

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>school</td>
<td>(Intercept)</td>
<td>2.692</td>
<td>1.641</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>37.017</td>
<td>6.084</td>
</tr>
</tbody>
</table>

Number of obs: 7185, groups: school, 160

Fixed effects:

| (Intercept) | 12.6833 | 0.1494 | 84.91 |
| meannses    | 5.8662  | 0.3617 | 16.22 |
| cses        | 2.2009  | 0.1090 | 20.20 |
| meannses:cses| 0.3248  | 0.2735 | 1.19  |

Correlation of Fixed Effects:

<table>
<thead>
<tr>
<th>(Intr)</th>
<th>meannses</th>
<th>cses</th>
</tr>
</thead>
<tbody>
<tr>
<td>meannses</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>cses</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
meanses:css 0.000 0.000 0.075
Model 6: intercepts-and-slopes-as-outcomes model

• we expand the model by including two level-2 predictors: meanses and sector; the slopes are again allowed to vary randomly

• model for the first (student) level:

\[ y_{ij} = \alpha_{0i} + \alpha_{1i} \text{cse}_{ij} + \epsilon_{ij} \]

• model for the second (school) level:

\[ \alpha_{0i} = \gamma_{00} + \gamma_{01} \text{meanses}_i + \gamma_{02} \text{sector}_i + u_{0i} \]
\[ \alpha_{1i} = \gamma_{10} + \gamma_{11} \text{meanses}_i + \gamma_{12} \text{sector}_i + u_{1i} \]
• the combined model and the Laird-Ware form:

\[
y_{ij} = (\gamma_{00} + \gamma_{01} \text{means}_{i} + \gamma_{02} \text{sector}_{i} + u_{0i}) + \\
(\gamma_{10} + \gamma_{11} \text{means}_{i} + \gamma_{12} \text{sector}_{i} + u_{1i}) \text{cses}_{ij} + \epsilon_{ij} \\
= \gamma_{00} + \gamma_{01} \text{means}_{i} + \gamma_{02} \text{sector}_{i} + \gamma_{10} \text{cses}_{ij} + \\
\gamma_{11} \text{means}_{i} \text{cses}_{ij} + \gamma_{12} \text{sector}_{i} \text{cses}_{ij} + u_{0i} + u_{1i} \text{cses}_{ij} + \epsilon_{ij} \\
= \beta_{0} + \beta_{1}x_{1ij} + \beta_{2}x_{2ij} + \beta_{3}x_{3ij} + \beta_{4}(x_{1ij}x_{3ij}) + \beta_{5}(x_{2ij}x_{3ij}) + \\
b_{0i} + b_{1i}z_{1ij} + \epsilon_{ij}
\]
R code

> fit.model6 <- lmer(mAch ~ 1 + meanses*cses + sector*cses + (1 + cses | school), data=Hsb82)
> summary(fit.model6)

Linear mixed model fit by REML ['lmerMod']
Formula: mAch ~ 1 + meanses * cses + sector * cses + (1 + cses | school)
Data: Hsb82

REML criterion at convergence: 46503.66

Random effects:
Groups   Name        Variance Std.Dev. Corr
school   (Intercept) 2.3796  1.5426  
           cses       0.1012  0.3181  0.39
Residual             36.7212  6.0598  
Number of obs: 7185, groups: school, 160

Fixed effects:
                      Estimate Std. Error  t value     
(Intercept)            12.1279    0.1993  60.86   
meanses                5.3329    0.3692  14.45   
cses                   2.9450    0.1556  18.93   
sectorCatholic         1.2266    0.3063   4.00   
meanses:cses           1.0392    0.2989   3.48   
cses:sectorCatholic   -1.6427    0.2398  -6.85   

Correlation of Fixed Effects:
(Intr) meanss cses  sctrCt mnss:c
meanss 0.256
kses 0.075 0.019
sectorCthlc -0.699 -0.356 -0.053
meanss:css 0.019 0.074 0.293 -0.026
css:sctrCth -0.052 -0.027 -0.696 0.077 -0.351

Do we need the random slopes?

> fit.model6bis <- lmer(mAch ~ 1 + meanss*cses + sector*cses + (1 | school), data=Hsb82)
> anova(fit.model6bis, fit.model6)

Data: Hsb82
Models:
fit.model6bis: mAch ~ 1 + meanss * cses + sector * cses + (1 | school)
fit.model6: mAch ~ 1 + meanss * cses + sector * cses + (1 + cses | school)

Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
fit.model6bis  8 46513 46568 -23249  46497
fit.model6  10 46516 46585 -23248  46496 1.0016 2 0.6061

• no: apparently, the level-2 predictors do a sufficiently good job of accounting for differences in slopes that the variance component for slopes is no longer needed
1.3 Further issues

- modern software (for example \texttt{lmer} in R) can also handle partially nested data structures (for example, some students belong to several schools)

- you should always ‘center’ level-1 covariates

- the standard hierarchical model can be extended to include heterogeneous level-1 variance terms
  
  – in this case: each individual/cluster has a different variance
  
  – for example: math achievement is more variable among boys than girls
  
  – heterogeneity of the level-1 variance may be modelled as a function of predictors at both levels
  
  – in practice, the log of the variance is modelled (to ensure positivity)
  
  – the presence of heterogeneity of variance at level-1 is often due to misspecification of the level-1 model
2 Modeling Longitudinal Data

- longitudinal data is very similar to hierarchical data: you typically have two levels: the timepoints (level-1) and the individuals/units (level-2)

- the focus of the analysis is often to find an appropriate polynomial (linear, quadratic, cubic, ... ) that describes the ‘evolution’ or ‘growth curve’ of the data over time

- the components of that polynomial (intercept, slope, quadratic component, ...) may be constant (fixed) or varying (random) across individuals/units

- a complication in longitudinal data is that it may not be longer reasonable to assume that the ‘level-1’ errors $\epsilon_{ij}$ are independent, since observations taken close in time on the same individual/units may well be more similar than observations farther apart in time; this is called *auto-correlation*

- in the world of linear mixed models, there is a delicate trade-off between adding more random effects to the model (implying more complicated patterns in the covariance matrix) or using a more complicated form for $\Sigma_i$
2.1 Example: Reaction times in a sleep deprivation study

• Dependent variable: the average reaction time per day (Reaction) for subjects in a sleep deprivation study.

• On day 0 the subjects had their normal amount of sleep. Starting that night they were restricted to 3 hours of sleep per night. The observations represent the average reaction time on a series of tests given each day to each subject.

• **Days** Number of days of sleep deprivation

• **Subject** Subject number on which the observation was made

• dataset is available in the lme4 package
Exploring the data: the spaghetti plot

Days of sleep deprivation
Average reaction time (ms)

0 2 4 6 8
Exploring the data: linear regression for each subject
A simple linear growth model

```r
> fit.linear <- lmer(Reaction ~ Days + (Days | Subject), data=sleepstudy)
> summary(fit.linear)

Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (Days | Subject)
   Data: sleepstudy

REML criterion at convergence: 1743.628

Random effects:
  Groups   Name    Variance  Std.Dev.  Corr
     Subject (Intercept)  612.10     24.741
        Days             35.07      5.922   0.07
   Residual              654.94     25.592
Number of obs: 180, groups: Subject, 18

Fixed effects:
   Estimate      Std. Error  t value
(Intercept)    251.405       6.825   36.84
Days            10.467       1.546    6.77

Correlation of Fixed Effects:
    (Intr) Days
(Intercept)   -0.138
Days          -0.078
```
> fit.quad <- lmer(Reaction ~ Days + I(Days^2) + (Days | Subject), data=sleepstudy)
> summary(fit.quad)

Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + I(Days^2) + (Days | Subject)
   Data: sleepstudy

REML criterion at convergence: 1742.816

Random effects:
  Groups     Name       Variance   Std.Dev.   Corr
    Subject (Intercept) 613.11     24.761
           Days       35.11      5.925   0.06
    Residual            651.97     25.534
Number of obs: 180, groups: Subject, 18

Fixed effects:
   Estimate Std. Error  t value
(Intercept) 255.4494     7.5135  34.00
   Days       7.4341      2.8189  2.64
 I(Days^2)   0.3370      0.2619  1.29

Correlation of Fixed Effects:
   (Intr)  Days
   Days  -0.418
 I(Days^2)  0.418 -0.836