Introduction to mixed models with `R`

BCAM - Basque Center for Applied Mathematics, Applied
Statistics Research Line

Dae-Jin Lee < dlee@bcamath.org >

Neiker Courses on Statistical Modelling (2016)

Contents

---

1   Basic concepts

Many common statistical models can be expressed as linear models that incorporate both fixed effects, which are parameters associated with an entire population or with certain repeatable levels of experimental factors, and random effects, which are associated with individual experimental units drawn at random from a population. A model with both fixed effects and random effects is called a *mixed-effects model*.

Linear mixed-effects models (LMM's) are an important class of statistical models that can be used to analyze complex data structures. Such data include correlated and/or clustered observations, repeated measurements, longitudinal measurements, multivariate observations, etc. These models are useful in a wide variety of disciplines in the physical, experimental sciences, medicine, psychology, biological and social sciences. They are also known as mixed-effects models as they contain both fixed and random effects. By associating common random effects to observations sharing the same level of a classification factor, mixed-effects models flexibly represent the covariance structure induced by the grouping of the data.

The increasing popularity of mixed-effects models is explained by the flexibility they offer in modeling the within-group correlation often present in grouped

data, by the handling of balanced and unbalanced data in a unified framework, and by the availability of reliable and efficient software for fitting them.

There are many packages in `R` with functions that allow LMM's in various forms. For instance, `amer`, `nlme`, `MASS`, `glmm`, `lmm`, `MCMCglmm`, etc. . . To facilitate and promote the use of LMM's in practice, we will provide details for few of them. We focus on `nlme` and `lme4` libraries and functions `lme()` and `lmer()` respectively.

1.1   Data types and structures

Mixed-effects models are primarily used to describe relationships between a response variable and some covariates in data that are grouped according to one or more classification factors.

The type of data and its structure determines a first step to know if we need mixed-effects models and how to formulate them.

1.1.1   Hierarchical data (grouped data or multilevel)

- The dependent variable is measured once on each individual (unit of interest) and individuals are grouped (or nested) in more than one level.

- Then the units of analysis are usually individuals (at a lower level) who are nested within contextual/aggregate units (at a higher level).

- **Examples:**

  - student in schools.
  - people in districts.
  - patients in hospitals.
  - plants in a plot.

Multilevel structures can be the result of the experimental design. For example, consider a survey to study the health status: we can consider a design with three levels: 1) we sample regions, then 2) districts and finally 3) individuals.

1.1.2   Repeated measurements and longitudinal data

- **Repeated measurements** if dependent variable is measured more than once on an individual. *Example:* we measure glucose levels on a patient before and after inyecting insuline. Can be considered as multilevel when *Level 2* indicates the individuals and *Level 1* the measurements taken. Given that measurements are taken on the same individual, probably the observations are not independent, hence a linear model may not be appropiate.

- **Longitudinal studies** involves repeated observations of the same variables over long periods of time. Unlike cross-sectional studies, in which different individuals with same characteristics are compared, longitudinal studies track the same people, and therefore the differences observed in those people are less likely to be the result of cultural differences across generations. Because of this benefit, longitudinal studies make observing changes more accurate, and they are applied in various other fields.

From the mixed-effects perspective the distinction between repeated measurements and longitudinal studies is not so crucial.

1.2 Fixed or random effects?

In a mixed-effects model, the key point is to distinguish between **fixed effects** and **random effects**. This is important because the inference and analysis on fixed and random effects is different.

Fixed effects are variables associated with an entire population or with certain repeatable levels of experimental factors (*levels* or *treatments*), while **random effects** are associated with individual *experimental units* drawn at random from a population.

Examples:

- Consider an experiment where the aim is to compare two groups, 1st group with patients where treatment is applied (case) and a 2nd group where not (control). In this situation, the aim of the study is to compare both groups and not generalize the results for other treatments that might be considered.

- Consider a survey where 10 cities were chosen. If we are interested in the results for those 10 cities and we are not interested in the rest of the cities, the variable *city* is a fixed effect. But if the 10 cities are randomly chosen from a large population of cities we are considering the variable *city* as a random effect.

A measure quantity is considered *random* if it changes over the population units. When a variable is chosen as a random effect in a statistical model, we are assuming that we want to extract conclusions about a population from which the observed units are chosen, but not really from the particular units. Then we can think about *interchangeability*, in the sense that we could have changed a single unit with another one from the population with no effect.

Note that the standard methods used to construct significance tests and confidence intervals for the fixed effects are not valid for the random effects. This is because the observed effects are only a sample of all the possible effects.

Then it is important to distinguish (statistically speaking) between *fixed* and *random* effects if:

- The levels of a variable can be interpreted as extracted from a population with a given probability distribution.
- In the case of a fixed effect, usually we are interested in the comparison of the results for the dependent variable for the different levels of the independent variable, i.e. we are interested in the difference between the averages.
- In the case of the random effects, we are not interested in the difference between the averages, but in how the random effect explains the variability in the dependent variable. Then, to consider an effect as random, we have to find the presence of an un-explained variability associated to the units of the random effect.

1.3   Why mixed models?

1.4   Some motivating examples

1.4.1   Single random factor

- Consider an experimental design in which we have J measurements form each of I units (see next Figure)
- Suppose that we are interested in establishing whether there are differences between the units, but are not really interested in the individuals unit effects.
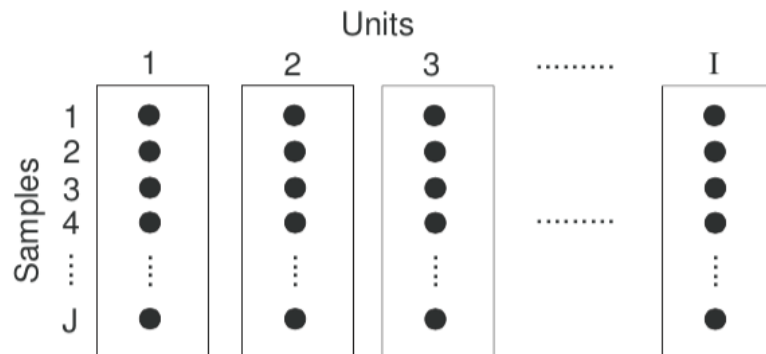

Figure 1: *Schematic illustration of the 1-way or single factor layout. Rectangles are experimental units and •'s are measurements.*

1.4.2 Railway experiment

This example is borrowed from Pinheiro and Bates (2000)

An engineering test for longitudinal stress in rails, involves measuring the time it takes certain ultrasonic waves to travel along the rail. To be a useful test, engineers need to know the average travel time for rails, and the variability to expect between rails, as well as the variability in the measurement process.

Six rails were chosen at random and tested three times each. The only experimental setting that changes between the observations is the rail. We say these observations are arranged in a one-way classification because they are classified according to a single characteristic, the rail on which the observation was made.

```
library(nlme)
data(Rail)
class(Rail)
```

```
## [1] "nffGroupedData" "nfGroupedData"  "groupedData"     "data.frame"
```

```
?Rail
Rail$Rail # ordered factor
```

```
##  [1] 1 1 1 2 2 2 3 3 3 4 4 4 5 5 5 6 6 6
## Levels: 2 < 5 < 1 < 6 < 3 < 4
```

The structure of the data is quite simple each row corresponds to one observation for which the rail and the travel time are recorded.

The quantities the engineers were interested in estimating from this experiment are the average travel time for a *typical* rail (the *expected travel time*), the variation in average travel times among rails (the *between-rail variability*), and the variation in the observed travel times for a single rail (the *within-rail variability*).

Data from a one-way classification like the rails example can be analyzed either with a fixed-effects model or with a random effects model. The distinction between the two models is according to whether we wish to make inferences about those particular levels of the classification factor that were used in the experiment or to make inferences about the population from which these levels were drawn.

Data are grouped (or clustered) which implies:

1. observations from the same rail are more similar between them
2. average travel time varies from rail to rail, but also between measurements.

We can see from next Figure that there is considerable variability in the mean travel time for the different rails.
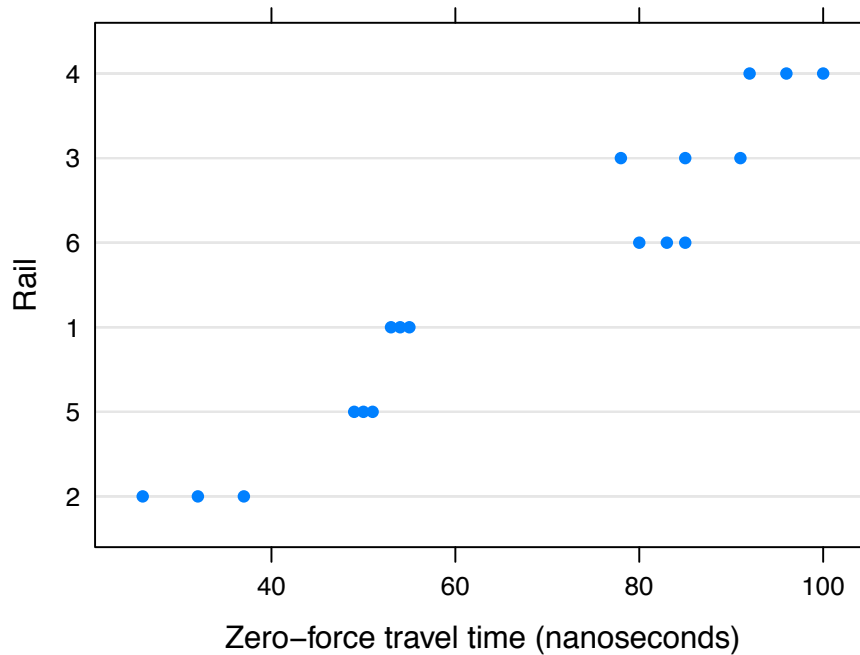
```
plot(Rail)
```


Figure 2: *Travel time in nanoseconds in a sample of six rails.*

It is evident that we should account for the classification factor when modeling grouped data such as the rails example, we initially ignore the grouping structure of the data and assume the simple *single-mean* model:

$$y_{ij} = \mu + \epsilon_{ij}, \qquad i = 1, ..., M, \quad j = 1, ..., n_i, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

where y_{ij} is the observed travel time for observation j on rail i, μ is the mean travel time across the population of rails being sampled. In this case the number of rails $M = 6$ and the number of observation on rail i is n_i, i.e. $n_1 = n_2 = n_3 = 3$. The total number of observations is $N = \sum_{i=1}^{M} n_i = 18$. We know from theory that the maximum likelihood estimator (MLE) is the sample mean $\bar{y}$ and the mean squared error (MSE) is an estimator of the population variance, i.e. $s^2 = \sigma^2$:

```
mean(Rail$travel)
```

```
## [1] 66.5
```

```r
sd(Rail$travel)
```

```
## [1] 23.64505
```

We fit model with `lm` function as:

```r
rail1 <- lm(travel ~ 1,data=Rail) # single-mean model
summary(rail1)
```

```
## 
## Call:
## lm(formula = travel ~ 1, data = Rail)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -40.50 -16.25   0.00  18.50  33.50 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   66.500      5.573   11.93  1.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 23.65 on 17 degrees of freedom
```

The boxplots of the residuals from the 'rail1 fit by rail number is shown as follows. The grouped rail effect is incorporated into the residuals leading to an inflated estimate of the *within-rail* variability.

```r
boxplot(resid(rail1)~Rail,xlab="Rail",ylab="Residuals",data=Rail)
abline(0,0)
```

The *rail effects* may be incorporated into the model for the travel times by allowing the mean of each rail to be represented by a separate parameter. This fixed-effects model for the one-way classification is written as an ANOVA model as:

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \ldots, M, \quad j = 1, \ldots, n_i \quad \epsilon_{ij} \sim N(0, \sigma^2),$$

where μ_i represents the mean travel time of rail i or also can be viewed as $\mu_i = \mu + \alpha_i$ where α_i is the difference (+ or -) of each individual rail to the overall mean.
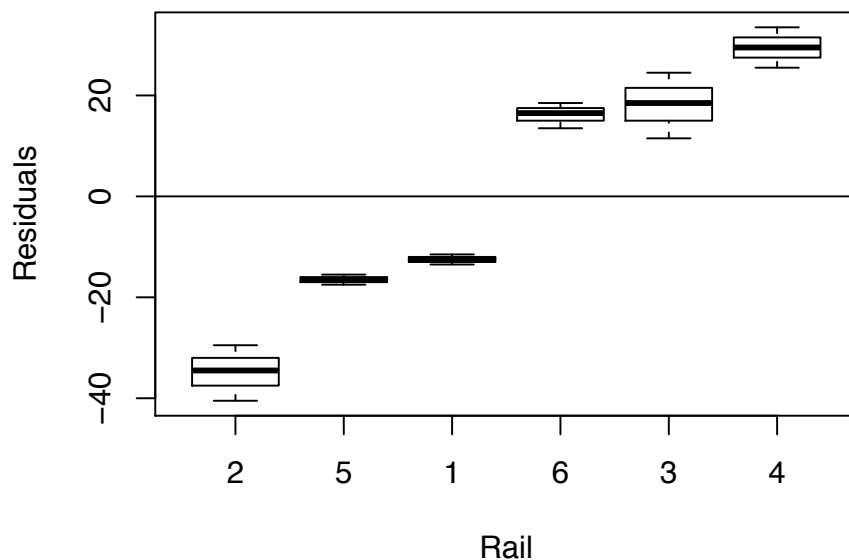
Figure 3: *Boxplot of residuals by number for the `lm` fit the single-mean model.*

```
 rail2 <- lm(travel~Rail-1, data=Rail)
summary(rail2)
```

```
##
## Call:
## lm(formula = travel ~ Rail - 1, data = Rail)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -6.6667 -1.0000  0.1667  1.0000  6.3333
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## Rail2   31.667      2.321   13.64 1.15e-08 ***
## Rail5   50.000      2.321   21.54 5.86e-11 ***
## Rail1   54.000      2.321   23.26 2.37e-11 ***
## Rail6   82.667      2.321   35.61 1.54e-13 ***
## Rail3   84.667      2.321   36.47 1.16e-13 ***
## Rail4   96.000      2.321   41.35 2.59e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.021 on 12 degrees of freedom
## Multiple R-squared:  0.9978, Adjusted R-squared:  0.9967
```

9

```
## F-statistic: 916.6 on 6 and 12 DF,  p-value: 2.971e-15
```

A `-1` is used in the model formula to prevent the default inclusion of an intercept term in the model. As expected, there is considerable variation in the estimated mean travel times per rail. The residual standard error is $\hat{\sigma}^2 = 4.021$ (much less than the single-mean model 23.65).

Looking at the residuals they look centered around zero and have smaller magnitudes, but there are still *some problems*.

```
boxplot(resid(rail2)~Rail,xlab="Rail",ylab="Residuals",data=Rail)
abline(0,0)
```
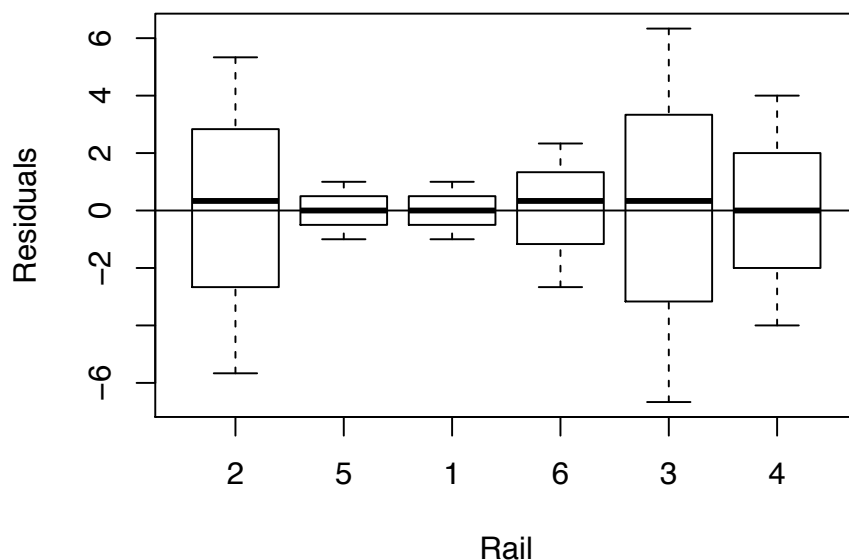

Figure 4: *Boxplot of residuals by rail number for the `lm` fit of the fixed-effects model.*

From `rail2` we see that:

- As expected, there is considerable variation in the estimated mean travel times per rail.
- The residual standard error obtained for the fixed effects model `rail2` is $\hat{\sigma}^2 = 4.0208$, much lower than the one obtained in the single-mean model `rail1`.
- `rail2` succesfully accounts for the rail effects.
- However, it only models the specific sample of rails used in the experiment, while the main interest is in the population of rails from which the sample was drawn.

- It does not provide an estimate of the *between-rail* variability, which is one of the central quantities of interest in the rails experiment.
- Another drawback of this fixed effects model is that the number of parameters in the model increases linearly with the number of rails.

These issues can be solved with a *random effects* model by treating the rail effects as random variations around a population mean. If we reparameterize one-way classification model `rail2` as

$$y_{ij} = \mu + (\mu_i - \mu) + \epsilon_{ij},$$

the random effects model replaces the fixed parameters $\mu_i - \mu = u_i$ with a random effect u_i, which is a random variable specific for the ith rail, with zero mean and unknown variance σ_u^2. Once again they are called random effects because they are associated with the particular experimental units (rails in this case that are selected at random from the population of interest). They are effects because they represent a deviation from an overall mean. That is, the *effect* of choosing rail i is to shift the mean travel time from μ to $\mu + u_i$.

The model would be:

$$y_{ij} = \mu + u_i + \epsilon_{ij}, \quad u_i \sim \mathcal{N}(0, \sigma_u^2) \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2),$$

we also assume that u_i's are *iid* and independent from ϵ_{ij}'s.

Because observations made on the same rail share the same random effect u_i, they are correlated (while in the *fixed effects* model they were independent). The covariance between observations on the same rail is σ_u^2 corresponding to a correlation of $\sigma_u^2/(\sigma_u^2 + \sigma^2)$.

The parameters of the statistical model created by the *random effects* model are μ, σ_u^2 and σ^2. The number of parameters is always three, irrespective of the number of rails in the experiment. Note that, although the random effects u_i are like parameters we are interested in estimating σ_u^2. We will however, form prediction $\hat{u}_i$ given the data we observed.

2 Mixed model general formulation

In general, linear mixed models (LMM) extend the linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$$

to

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon}, \qquad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{G}), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{R}),$$

where $\boldsymbol{X}$ is a $n \times k$ matrix (where k is the number of *fixed effects*), $\boldsymbol{Z}$ is a $n \times p$ matrix (p is the number of *random effects*), $\boldsymbol{G}$ is the variance-covariance matrix of the random effects of dimension $p \times p$. The matrix $\boldsymbol{R}$ is the variance-covariance matrix of the errors usually $\boldsymbol{R} = \sigma^2 \boldsymbol{I}$ but we can assume other forms such as correlation structures.

In our previous rails example, recall that the model for the jth response on the ith rail is

$$y_{ij} = \mu + u_i + \epsilon_{ij},$$

where

$\quad y_{ij}$ is the observed travel time for observation j on rail i
$\quad \mu$ is the mean travel time across the population of rails sampled
$\quad u_i$ is the random effect of the ith rail
$\quad \epsilon_{ij}$ is the error term

with u_i's and ϵ_{ij} mutually independent.

Hence, the model for the ith rail would be:

$$\left(\begin{array}{c} y_{i1} \\ y_{i2} \\ y_{i3} \end{array} \right) = \left(\begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right) \mu + \left(\begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right) u_i + \left(\begin{array}{c} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \end{array} \right)$$

or equivalently:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\,\boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{u}_i + \boldsymbol{\epsilon}_i \quad \boldsymbol{u}_i \sim N(0, \sigma_u^2)$$

and the model for all the observations would have:

$$\boldsymbol{X} = \left(\begin{array}{c} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_6 \end{array} \right), \quad \boldsymbol{Z} = \left(\begin{array}{cccc} \boldsymbol{Z}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 & \cdots & \boldsymbol{0} \\ & & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Z}_6 \end{array} \right), \quad \boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{G}), \quad \boldsymbol{G} = \sigma_u^2 \boldsymbol{I}_6, \quad \boldsymbol{R} = \sigma^2 \boldsymbol{I}_{18}$$

i.e.

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ \vdots \\ y_{61} \\ y_{62} \\ y_{63} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_6 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \vdots \\ \epsilon_{61} \\ \epsilon_{62} \\ \epsilon_{63} \end{pmatrix}$$

2.1 Estimation of linear mixed models

2.2 Estimation of fixed effects

One way to estimate $\boldsymbol{\beta}$ is to use the **marginal model**, i.e. re-write the LMM model as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}, \quad \text{with } \boldsymbol{e} = \boldsymbol{Zu} + \boldsymbol{\epsilon}.$$

This is a linear model with correlated errors with

$$\text{Cov}(\boldsymbol{y}) \ or \ \boldsymbol{V} = \boldsymbol{Z'GZ} + \boldsymbol{R}$$

Given matrix $\boldsymbol{V}$, the estimator of $\boldsymbol{\beta}$ is obtained by *generalized least squares* by minimizing the function

$$\boldsymbol{Q} = (\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}),$$

where

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X'V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X'V}^{-1}\boldsymbol{y}$$

Since linear mixed models are fitted using maximum likelihood estimation, we could, in principle, use generalized likelihood ratio tests for all model comparison/hypothesis testing.

2.3 Prediction of random effects

Since the u_i's are random effects, we do not estimate them (they are not parameters), but we predict them (they are random variables). The usual way to do this is to evaluate the expected value of $\boldsymbol{u}$, given the data $\boldsymbol{y}$.

There are several ways to obtain predictions of $\boldsymbol{u}$ such that they are the *best linear unbiased predictors* (BLUP's). One solution is through *mixed models Henderson equations* obtained maximizing the joint density function of $\boldsymbol{y}$ and $\boldsymbol{u}$:

$$f(y, \boldsymbol{u}) = f(\boldsymbol{y}|\boldsymbol{u})f(\boldsymbol{u}), \quad \boldsymbol{y}|\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{X\beta}, \boldsymbol{R}), \quad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{G})$$

In terms of the *likelihood* we have

$$l \propto -\frac{1}{2}\left[log|\boldsymbol{R}| + log|\boldsymbol{G}| + (\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu}) + \boldsymbol{u'G}^{-1}\boldsymbol{u}\right],$$

deriving w.r.t. $\boldsymbol{\beta}$ and $\boldsymbol{u}$ we obtain the equation:

$$\begin{bmatrix} \boldsymbol{X'R}^{-1}\boldsymbol{X} & \boldsymbol{X'R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z'R}^{-1}\boldsymbol{X} & \boldsymbol{Z'R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z'R}^{-1}\boldsymbol{y} \end{bmatrix}.$$

The solution to the system of equations is:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$$

$$\hat{\boldsymbol{u}} = \boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}),$$

where $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}$. However, $\boldsymbol{V}$ depends on variance parameters of $\boldsymbol{G}$ and $\boldsymbol{R}$.

2.4 Estimation of the variance components

In general, variance components estimation is done by maximum likelihood (ML) or *restricted/residual maximum likelihood* (REML). See Searle, Casella, and McCulloch (1992) for details.

2.4.1 Maximum likelihood

Given $\mathrm{Cov}(\boldsymbol{y}) = \boldsymbol{V}$ defined as

$$\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}$$

The maximum likelihood estimator (MLE) of the variance components in $\boldsymbol{V}$ are based on the model

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V}),$$

then

$$l(\boldsymbol{\beta}, \boldsymbol{V}) \propto \frac{1}{2}\left[(\log(|\boldsymbol{V}|) + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).\right]$$

The MLE of $\boldsymbol{\beta}$ coincides with $\hat{\boldsymbol{\beta}}$. Substituting this expression $l(\boldsymbol{\beta}, \boldsymbol{V})$, we obtain the *profiled* likelihood for $\boldsymbol{V}$:

$$l_p(\boldsymbol{V}) = \frac{1}{2}\left[(\log|\boldsymbol{V}| + \boldsymbol{y}'\boldsymbol{V}^{-1}\left(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\right)\boldsymbol{y}\right].$$

The MLE of the parameters of $\boldsymbol{V}$ are obtained maximizing $l_p(\boldsymbol{V})$. There is no closed form solution to this maximization problem so an iterative numerical method is used to maximize l_p (or more usually $\log l_p$).

2.4.2 Restricted or residual maximum likelihood (REML)

The MLE estimators of variance components tend to become biased, as the number of fixed parameters increases, particularly if sample sizes are not large.

Recall from the classic linear model, the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2}{n}$$

The unbiased estimator is

$$\mathbb{E}(\hat{\sigma}^2) = \frac{n-k}{n}\sigma^2$$

where $k = \dim(\boldsymbol{\beta})$. Then, there is a tendency to underestimate variance components, as k increases. The difficulty is that the estimators take no account of the degrees of freedom lost by estimating the fixed effects.

The REML approach Patterson and Thompson (1971) measures the fit of the variance components not from the joint likelihood of the variance parameters and $\boldsymbol{\beta}$, but rather by the (scaled) average of the likelihood over all the possible values of $\boldsymbol{\beta}$. The maximization is done on:

$$l_R(\boldsymbol{V}) = l_P(\boldsymbol{V}) - \frac{1}{2}log|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}|.$$

In the railway example:

- **Fixed effects model:** $\hat{\mu} = 66,5$ and $\hat{\sigma} = 4,02$
- **Mixed-effects model:** $\hat{\mu} = 66,5$, $\hat{\sigma} = 4,02$ and $\hat{\sigma}_u^2 = 24,80^2$ (with REML and $\hat{\sigma}_u^2 = 22,62^2$ with ML)

We will estimate this variance components with `lme` and control options `method="REML"` and `method="ML"`.

3 Mixed models estimation with `nlme` and `lme4` R packages

There are a large number of R packages that can be used to fit LMMs. Rather than attempting to describe all of these packages, we focus mainly on two of them, namely, `nlme` and `lme4`. In this way, we can provide a more detailed account of the tools offered by the two packages, which include a wide variety of functions for model fitting, diagnostics, inference, etc

3.1 `lme` function

The main model fitting function in `library(nlme)` is `lme`. A call to the `lme` function is similar to `lm`, except that an extra argument specifying the random effects structure must also be supplied to the model.

`lme` assumes that your data are grouped according to the levels of some factor(s), and that the same random effects structure is required for each group, with random effects independent between groups.

```
lme(response ~ predictors,random= ,correlation=, method=,data= )
```

The simplest way to specify the random effects structure is with a one sided formula. Suppose a response variable y with predictors x and z, and grouping factor variable g. For a single random effect:

```
lme(y ~ x+z,random = ~1|g,data= df) # default method = "REML"
```

To specify a random effect that depends on a x, we must specify ~x|g

```
lme(y ~ x+z,random = ~x|g,data= df)
```

A combination of both can be specified as

```
lme(y ~ x+z,random = ~1+x|g,data= df)
```

Equivalently, we can use the `list` command

```
lme(y ~ x+z,random = list(g=~1),data= df)
lme(y ~ x+z,random = list(g=~x),data= df)
lme(y ~ x+z,random = list(g=~1+x),data= df)
```

For more than one random effects we use lists, i.e.:

```
lme(y ~ x+z,random =list(~1|g1,~1|g2),data= df)
```

If random effects are *nested* the `lme` syntax is

```
lme(y ~ x+z,random =~1|g1/g2,data= df)
```

where factor `g2` is nested in `g1`.

Crossed random effects on the other hand, can be dealt with through a combination of `pdBlocked` and `pdIdent` objects.

Example: *Rail data*

Now we can easily fit the random effects model with `lme`

```
library(nlme)
rail3 <- lme(travel~1,random=~1|Rail,data=Rail)
# or equivalently
# rail3 <- lme(travel~1,random=list(Rail=~1),data=Rail)
rail3
```

```
## Linear mixed-effects model fit by REML
##   Data: Rail
##   Log-restricted-likelihood: -61.0885
##   Fixed: travel ~ 1
## (Intercept)
##        66.5
##
## Random effects:
##  Formula: ~1 | Rail
##         (Intercept) Residual
## StdDev:    24.80547 4.020779
##
## Number of Observations: 18
## Number of Groups: 6
```

REML is the default method for estimation, to use ML we have to specify `method="ML"`.

The `summary()` command gives us additional information of the fitted model.

```
summary(rail3)
```

```
## Linear mixed-effects model fit by REML
##  Data: Rail
##        AIC      BIC    logLik
##   128.177 130.6766 -61.0885
##
## Random effects:
##  Formula: ~1 | Rail
##         (Intercept) Residual
## StdDev:    24.80547 4.020779
##
## Fixed effects: travel ~ 1
##              Value Std.Error DF  t-value p-value
## (Intercept)  66.5  10.17104 12 6.538173       0
##
## Standardized Within-Group Residuals:
##         Min         Q1         Med         Q3        Max
## -1.61882658 -0.28217671  0.03569328  0.21955784  1.61437744
##
## Number of Observations: 18
## Number of Groups: 6
```

```
model <- rail3
summ  <- summary(rail3)
```

where $\sigma_u = 24.805$ and $\sigma = 4.020$.

Variance components can be obtained by `VarCorr`:

```r
VarCorr(model)
```

```
## Rail = pdLogChol(1)
##             Variance  StdDev
## (Intercept) 615.31111 24.805465
## Residual     16.16667  4.020779
```

Given this information we can compute the intra-class correlation coefficient as:

$$IIC = \frac{615,31}{615,31 + 16,16} = 0,974$$

i.e. $97,4\%$ of the total variability is due to the heterogeneity between the rails.

We can fit the model by ML

```r
rail3.ml <- update(rail3,method="ML")
VarCorr(rail3.ml)
```

```
## Rail = pdLogChol(1)
##             Variance  StdDev
## (Intercept) 511.86111 22.624348
## Residual     16.16667  4.020779
```

Notice that the ML estimate of σ is $4,0208$, the same as the REML estimate (it occurs in simple models but not in general). The ML estimate of $\sigma_u^2 = 22,624^2$ is smaller than the REML estimate $\sigma_u^2 = 24,805^2$ (this is because ML estimates are not always unbiased, and those for variance components tend to be biased downwards). The estimate of $\hat{\beta}$ are the same $66,5$.

The next table summarizes the R syntax to extract results from an `lme` object

Results	`lme` object
Estimation method	`model$method`
$\hat{\beta}$ coefficients	`fixef(model)`
$\hat{\boldsymbol{\beta}}$ se($\hat{\boldsymbol{\beta}}$) and t-test	`summary(model)$tTable`
$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})$	`vcov(model)`
I.C. 95% for $\boldsymbol{\beta}$	`intervals(model, which=''fixed'')`
$\hat{\sigma}$	`summary(model)$sigma`
I.C. 95% for σ, σ_u	`intervals(model, which=''var-cov'')`
$\hat{\boldsymbol{u}}$	`ranef(model)`

18

Results	`lme` object
$\beta_0 + \boldsymbol{u}$	`coef(model)`
$\hat{\boldsymbol{G}}$	`getVarCov(model)`
$\hat{\boldsymbol{G}}$ and $\hat{\sigma}$	`VarCorr(model)`
$\hat{\boldsymbol{R}}$	`getVarCoc(model,type=''conditional'')`
$\hat{\boldsymbol{V}}$	`getVarCoc(model,type=''marginal'')`
log ML	`logLik(model,REML=FALSE)`
log REML	`logLik(model,REML=TRUE)`
AIC	`AIC(model)`
BIC	`BIC(model)`
Fitted values:	`fitted(model)`
Residuals:	`residuals` or `resid`
— Conditioned	`resid(model,type=''response'')`
— Marginal	`resid(model, type=''response'',level=0)`
Normalized residuals	`resid(model,type=''normalized'')`
Pearson residuals	`resid(model, type=''pearson'')`
Predicted values:	see `?predict.lme`
— Conditioned	`predict(model, newdata)`
— Marginals	`predict(model, newdata,level=0)`

The fitted model should be examined using graphical and numerical summaries. An initial plot is the one of the residuals versus the fitted responses of the model in order to assess the assumption of constant variance of the ϵ_{ij}.

```
plot(model)
```

Note that standardized residuals in the vertical axis are the raw residuals, $\epsilon_{ij} = y_{ij} - \hat{\beta} - \hat{u}_i$, divided by the estimated standard deviation, $\hat{\sigma}$, of the ϵ_{ij}.

We should also examine numerical summaries of the model. A basic summary are the confidence intervals on the parameters:

```
intervals(model)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##               lower est.    upper
## (Intercept) 44.33921 66.5 88.66079
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
```

19
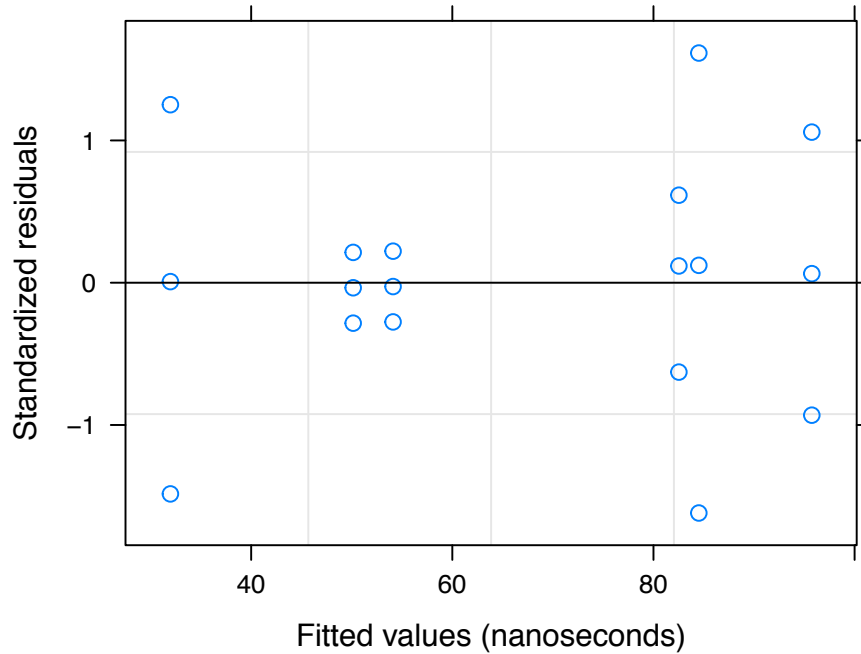
Figure 5: *Standard residuals vs fitted values*

```
##   Level: Rail
##                   lower     est.    upper
## sd((Intercept)) 13.27434 24.80547 46.35341
##
##  Within-group standard error:
##    lower     est.    upper
## 2.695007 4.020779 5.998747
```

We can also assess the significance of the fixed-effects with the `anova` function

```
anova(model)
```

```
##             numDF denDF F-value p-value
## (Intercept)     1    12 42.7477  <.0001
```

3.2 `lmer` function

`lmer` is implemented in the library `lme4`. The mixed-effects model is estimated using a formula with random effects in parenthesis:

20

```
lmer( y ~ x+(1|g), data=df)
```

The main difference with `lme` is the limited flexibility in the structure of the covariance matrix G.

For the rail example:

```
library(lme4)
rail.lmer <- lmer(travel ~1+(1|Rail),data=Rail)
summary(rail.lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: travel ~ 1 + (1 | Rail)
##    Data: Rail
##
## REML criterion at convergence: 122.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.61883 -0.28218  0.03569  0.21956  1.61438
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Rail     (Intercept) 615.31   24.805
##  Residual             16.17     4.021
## Number of obs: 18, groups:  Rail, 6
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    66.50      10.17   6.538
```

3.3 A randomized design block

See Pinheiro and Bates (2000) Section 1.2

In the railway rails example, the observations were classified according to one characteristic only – the rail on which the observation was made. In other experiments we may have more than one classification factor for each observation.

A *randomized block design* is a type of experimentin which there are two classification factors:

- an *experimental factor or treatment*: for which we use fixed effects and
- a *blocking* factor for which we use the random effects.

Example: ergonometrics experiment with stool types

The data consists of an ergonometrics experiment that has a randomized block design. The experimenters recorded the effort required (Borg scale) by each of nine different subjects to arise from each of four types of stools.

```
data(ergoStool)
names(ergoStool)
```

```
## [1] "effort"  "Type"     "Subject"
```

The aim is to compare these four particular types of stools so we used `Type` factor as a fixed effect. The nine different subjects represent a sample from the population about which we wish to make inferences so we use random effects to model the `Subject` factor.
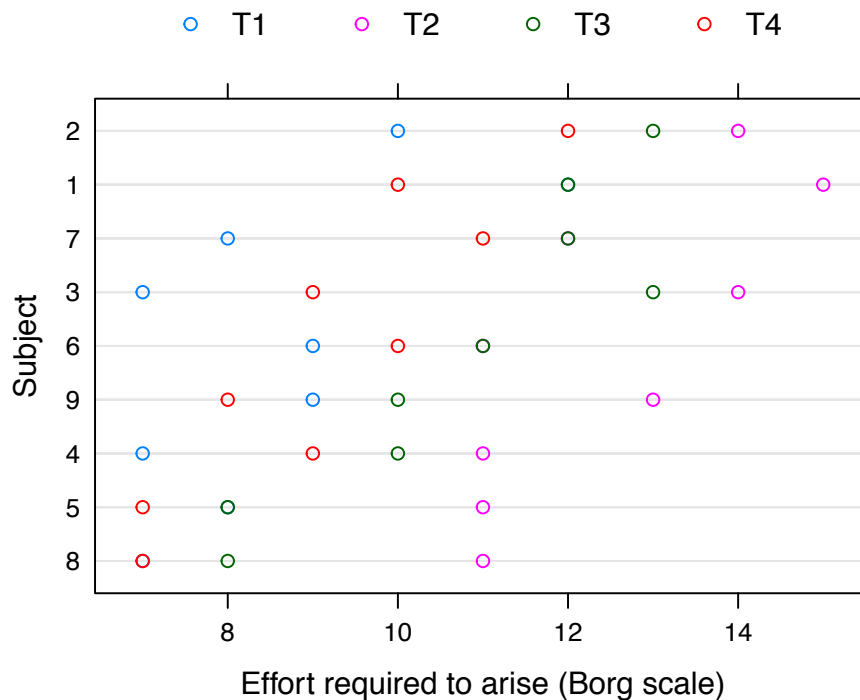
```
plot(ergoStool)
```


Figure 6: *Effort required (Borg scale) to arise from a stool for 9 different subjects each using 4 different types of stools (T1,T2,T3 and T4).*

From the Figure, we observe that there are systematic differences between stool types.

- `T2` required the greatest effort.
- `T1` was one of the lowest effort types.

The subjects also exhibited variability in their scoring of the effort. `Subject` is a *blocking factor* because it represents a known source of variability in the experiment. `Type` is an *experimental factor* because the purpose of the experiment is to determine if there are systematic differences in the level of effort to arise from the different types of stools.

More summaries:

```
# results not shown
plot.design(ergoStool)  # Design plot
ergoDF <- as.data.frame(ergoStool)   # convert to data.frame
by(ergoDF$effort, ergoDF$Type,mean)  # averages by Type
by(ergoDF$effort, ergoDF$Subject,mean) # averages by Subject
```

A simple mixed-effects model would be:

$$y_{ij} = \mu_j + u_i + \epsilon_{ij}, \quad i = 1, 2, 3, ..., 9 \qquad j = 1, 2, ..., 4$$

where μ_j is the fixed effect for the `Type` factor and u_i the random effects for the `Subject` factor.

Adding the term `-1` to the formula for the fixed effects we remove the columns of one's from the model matrices.

```
stool.lme <- lme(effort ~ Type-1, random = ~1|Subject,data=ergoStool)
summary(stool.lme)
```

```
## Linear mixed-effects model fit by REML
##  Data: ergoStool
##        AIC      BIC    logLik
##   133.1308 141.9252 -60.56539
##
## Random effects:
##  Formula: ~1 | Subject
##         (Intercept) Residual
## StdDev:    1.332465 1.100295
##
## Fixed effects: effort ~ Type - 1
##              Value Std.Error DF  t-value p-value
## TypeT1  8.555556 0.5760123 24 14.85308       0
## TypeT2 12.444444 0.5760123 24 21.60448       0
## TypeT3 10.777778 0.5760123 24 18.71102       0
```

```
## TypeT4  9.222222 0.5760123 24 16.01046        0
##  Correlation:
##        TypeT1 TypeT2 TypeT3
## TypeT2 0.595
## TypeT3 0.595  0.595
## TypeT4 0.595  0.595  0.595
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -1.80200345 -0.64316591  0.05783115  0.70099706  1.63142054
##
## Number of Observations: 36
## Number of Groups: 9
```

```r
VarCorr(stool.lme)
```

```
## Subject = pdLogChol(1)
##             Variance StdDev
## (Intercept) 1.775463 1.332465
## Residual    1.210648 1.100295
```

To examine the random effects structure, we can compare two models and apply a *likelihood ratio test*, i.e., the hypothesis test is $H_0 : \sigma_u^2 = 0$. Then:

```r
stool.lm <- lm(effort~Type-1,data=ergoStool) # only fixed effects
test <- -2*logLik(stool.lm) + 2 * logLik(stool.lme,REML=FALSE)
test
```

```
## 'log Lik.' 15.16194 (df=5)
```

```r
mean(pchisq(test,df=c(0,1),lower.tail=FALSE)) # if <0.05 reject H0
```

```
## [1] 4.933627e-05
```

We can also use the `library(RLRsim)`

```r
library(RLRsim)
exactLRT(update(stool.lme,method="ML"),stool.lm)
```

```
##
##  simulated finite sample distribution of LRT. (p-value based on
##  10000 simulated values)
##
## data:
## LRT = 15.162, p-value = 1e-04
```

We can also use AIC (*Akaike Information Criteria*) and select the model with lower AIC

```
AIC(stool.lm)
```

```
## [1] 147.3064
```

```
AIC(stool.lme)
```

```
## [1] 133.1308
```

Using the `anova` function we can test the fixed effects, i.e. the hypothesis being tested is
$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0,$$

which is equivalent to reducing the model to

$$\boldsymbol{y}_i = \boldsymbol{Z}_i u_i + \boldsymbol{\epsilon}_i, i = 1, 2, 3, ..., 9 \qquad u_i \sim \mathcal{N}(0, \sigma_u^2), \boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$$

That is, eliminates the fixed-effects so the mean response across the population would be zero. Note that, this hypothesis is not meaningful in the context of this experiment.

```
anova(stool.lme)
```

```
##      numDF denDF  F-value p-value
## Type     4    24 130.5186  <.0001
```

Another option is to compare nested models by likelihood ratio test and ML estimates.:

```
stool.lme1 <- lme(effort ~ Type,random = ~1|Subject,data=ergoStool,
                  method="ML")
stool.lme0 <- lme(effort ~ 1,random = ~1|Subject,data=ergoStool,
                  method="ML")
anova(stool.lme0,stool.lme1)
```

```
##            Model df      AIC      BIC    logLik   Test L.Ratio p-value
## stool.lme0     1  3 164.1500 168.9006 -79.07502
## stool.lme1     2  6 134.1444 143.6455 -61.07222 1 vs 2 36.0056  <.0001
```

We can examine the parameters estimates

```
intervals(stool.lme)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##            lower       est.      upper
## TypeT1   7.366725   8.555556   9.744386
## TypeT2 11.255614 12.444444 13.633275
## TypeT3   9.588947 10.777778 11.966609
## TypeT4   8.033391   9.222222 10.411053
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##    Level: Subject
##                     lower       est.     upper
## sd((Intercept)) 0.7496297 1.332465 2.368453
##
##  Within-group standard error:
##     lower       est.      upper
## 0.8293707 1.1002946 1.4597190
```

```
plot(stool.lme)
```

```
qqnorm(stool.lme, abline=c(0,1))
```

```
shapiro.test(resid(stool.lme)) # for a formal normality test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(stool.lme)
## W = 0.97783, p-value = 0.6716
```

Important:

- The overall effect of the factor should be assessed with `anova`.
- The interpretation of the parameter estimates for a fixed-effects term depends on the contrasts being used (see `?contrasts`)
- For REML estimation, likelihood ratio tests or comparisons of AIC or BIC require the same fixed-effects structure and the same choice of contrasts in all models.
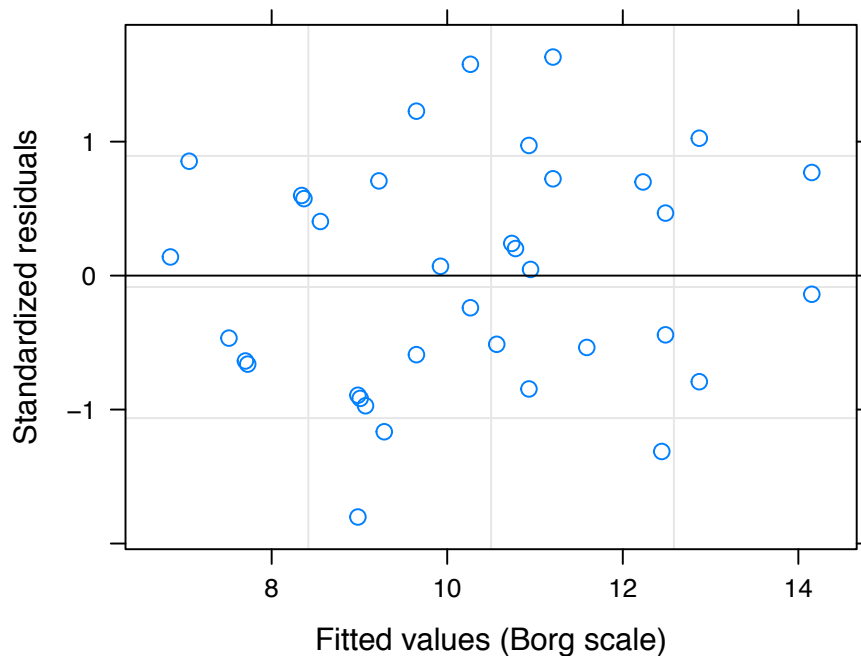
26

Figure 7: *Plot of fitted values vs standardized residuals*

- Adding `-1` to a model formula or not usually give the same results for `anova`.

```
plot(stool.lme,form=resid(.,type="p") ~ fitted(.)|Subject, abline=0)
```

Estimation using `lmer` function in `library(lme4)`

```
library(lme4)
stool.lmer <- lmer(effort ~ Type -1 + (1|Subject), data=ergoStool)
ranef(stool.lmer)
```

```
## $Subject
##     (Intercept)
## 8 -1.708716e+00
## 5 -1.495127e+00
## 4 -8.543581e-01
## 9 -2.135895e-01
## 6 -3.283826e-15
## 3  4.271791e-01
## 7  4.271791e-01
```
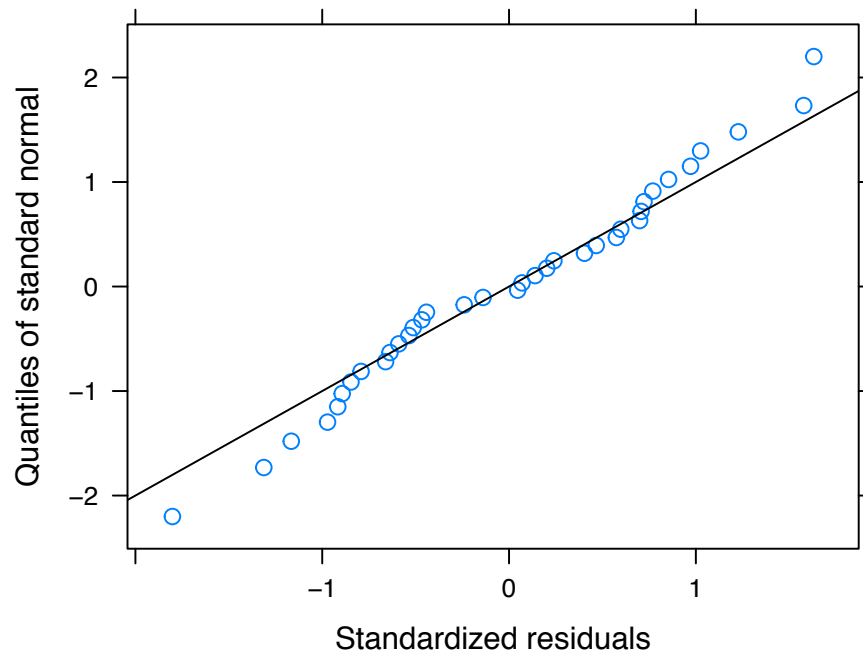
27

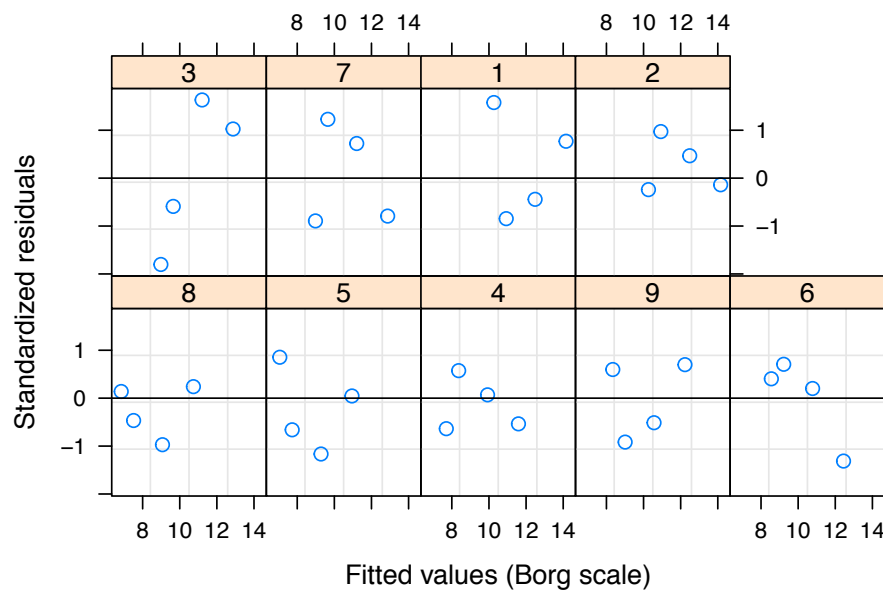Figure 8: *QQ-plot of standardized residuals*

Figure 9: *Standardized residuals vs fitted values by subject*

```
## 1  1.708716e+00
## 2  1.708716e+00
```

```
rr1 <- ranef(stool.lmer, condVar = TRUE)
library(lattice)
dotplot(rr1) # or qqmath(rr1)
```
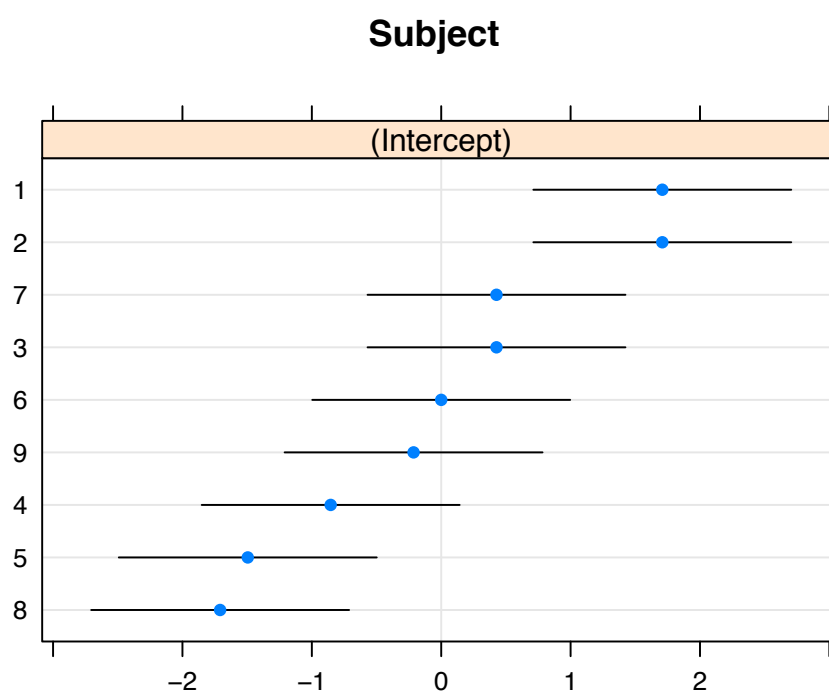
```
## $Subject
```

Subject

Figure 10: *Random effects plot*

4 Multilevel models

Example: High School and Beyond data set

High School & Beyond is a nationally representative survey of U.S. public and Catholic high schools conducted by the National Center for Education Statistics (NCES). The data are a subsample of the 1982 HSB survey with 7,185 students from 160 schools. The average sample size per school is approximately 45 students.

```r
MathAchieve <- read.table("data/MathAchieve.txt",header=TRUE,sep="\t")
# Variables
names(MathAchieve)
```

```
## [1] "School"  "Sex"     "SES"     "MathAch" "Sector"  "CSES"
```

```r
str(MathAchieve)
```

```
## 'data.frame':    7185 obs. of  6 variables:
##  $ School : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Sex    : int  1 1 0 0 0 0 1 0 1 0 ...
##  $ SES    : num  -1.528 -0.588 -0.528 -0.668 -0.158 ...
##  $ MathAch: num  5.88 19.71 20.35 8.78 17.9 ...
##  $ Sector : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CSES   : num  -1.1 -0.16 -0.1 -0.24 0.27 0.45 -0.19 -0.57 -0.46 -0.03 ...
```

```r
MathAchieve$School <- factor(MathAchieve$School,ordered=FALSE)
MathAchieve$Sex <- factor(MathAchieve$Sex,ordered=FALSE)
levels(MathAchieve$Sex) <- c("Male","Female")
MathAchieve$Sector <- factor(MathAchieve$Sector,ordered=FALSE)
levels(MathAchieve$Sector) <- c("Public","Private")
```

- `School`, an ordered factor identifying the school that the studend attends
- `Sex`, `Female` or `Male`
- `SES`, a standardized scale of socio-economic status
- `MathAch`, a measure of mathematics achievement
- `CSES`, centered mean of the SES values for the school

The variable of interest is `MathAch` with mean = 12.75 and std = 6.88. We created a new variable `CSES` the centered individual `SES`, relative standing on the `SES` measure for a student within school. We need this so that intercept is the mean achievement for the school.

A simple model would estimate the difference in `MathAch` based on the socio-economic status, i.e.:

$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j,$$

This model ignores that the students attends different schools (j^{th} subindex corresponds to the individual unit level).

```r
multilev0 <- lm(MathAch~CSES, data=MathAchieve) # fit a lm
multilev0
```

```
##
## Call:
## lm(formula = MathAch ~ CSES, data = MathAchieve)
##
## Coefficients:
## (Intercept)          CSES
##      12.761         2.191
```

The intercept is 12.76 and the slope 2.19.

```
plot(MathAchieve$CSES,MathAchieve$MathAch,cex=.5,col="grey")
abline(multilev0,col=2,lwd=3)
```
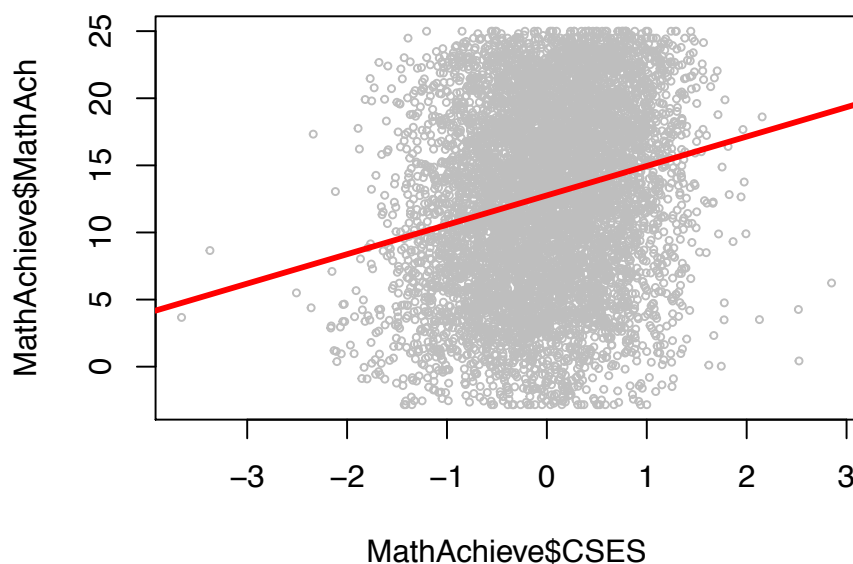

Figure 11: *Linear model fit*

Suppose we are interested in comparing differences between schools.

$$y_{ij} = \beta_{0i} + \beta_1 x_{ij} + \epsilon_{ij}$$

where the subindex i indicates the school of the student, with β_{0i} we specify a separate intercept for each school. Indeed, what we include is a categorical variable with as many categories as schools. Model `multilev1` considers the Schools as a fixed effect, i.e. we implicitly assume that we are interested in the schools of the survey.

```
##
```

```
## Call:
## lm(formula = MathAch ~ CSES + School, data = MathAchieve)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -18.9903  -4.3986   0.1181   4.5536  18.4433
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.72943    0.88731  10.965  < 2e-16 ***
## CSES         2.19117    0.10865  20.168  < 2e-16 ***
## School2      3.79539    1.50582   2.520 0.011742 *
## School3     -2.08142    1.24830  -1.667 0.095478 .
## School4      6.53921    1.62405   4.026 5.72e-05 ***
## School5      3.46067    1.24830   2.772 0.005581 **
## School6      1.48922    1.42155   1.048 0.294859
## School7      0.01140    1.45221   0.008 0.993739
## School8     10.00286    1.35816   7.365 1.97e-13 ***
## School9      8.39553    1.27606   6.579 5.07e-11 ***
## School10     7.12535    1.38155   5.158 2.57e-07 ***
## School11     0.78016    1.19855   0.651 0.515120
## School12     4.51310    1.17651   3.836 0.000126 ***
## School13    -2.05609    1.21882  -1.687 0.091657 .
## School14    -2.69128    1.46896  -1.832 0.066980 .
## School15     6.26771    1.21882   5.142 2.79e-07 ***
## School16     4.70688    1.45221   3.241 0.001196 **
## School17     8.39461    1.43643   5.844 5.32e-09 ***
## School18     3.19204    1.31763   2.423 0.015437 *
## School19     2.36107    1.25485   1.882 0.059938 .
## School20     5.68802    1.18493   4.800 1.62e-06 ***
## School21    -0.41882    1.18066  -0.355 0.722797
## School22     1.42148    1.15742   1.228 0.219437
## School23     6.80235    1.25485   5.421 6.13e-08 ***
## School24     4.26986    1.19855   3.563 0.000370 ***
## School25     0.43064    1.22432   0.352 0.725042
## School26     7.33691    1.19855   6.121 9.77e-10 ***
## School27     3.67997    1.32707   2.773 0.005569 **
## School28     5.19072    1.19855   4.331 1.51e-05 ***
## School29    -3.09987    1.29166  -2.400 0.016425 *
## School30     1.36791    1.32707   1.031 0.302682
## School31     2.62813    1.22432   2.147 0.031858 *
## School32     3.67890    1.26872   2.900 0.003747 **
## School33     6.76013    1.25485   5.387 7.39e-08 ***
## School34     1.16976    1.50582   0.777 0.437289
## School35     2.12882    1.20836   1.762 0.078156 .
## School36     4.15728    1.29166   3.219 0.001294 **
```

```
## School37    -1.73682    1.28370   -1.353 0.176106
## School38     8.73227    1.24830    6.995 2.89e-12 ***
## School39    -0.17084    1.26165   -0.135 0.892289
## School40     2.89504    1.21882    2.375 0.017562 *
## School41     4.67869    1.18934    3.934 8.44e-05 ***
## School42     7.24820    1.59670    4.539 5.73e-06 ***
## School43    -0.57128    1.31763   -0.434 0.664616
## School44     3.49267    1.22432    2.853 0.004347 **
## School45     4.56083    1.32707    3.437 0.000592 ***
## School46     1.74867    1.31763    1.327 0.184509
## School47    -0.52884    1.26872   -0.417 0.676813
## School48     9.99926    1.24198    8.051 9.57e-16 ***
## School49     6.67516    1.21882    5.477 4.48e-08 ***
## School50     3.56047    1.32707    2.683 0.007315 **
## School51     0.69376    1.24830    0.556 0.578390
## School52     5.63839    1.16855    4.825 1.43e-06 ***
## School53    -0.19472    1.23000   -0.158 0.874216
## School54     4.93992    1.28370    3.848 0.000120 ***
## School55     0.61613    1.26872    0.486 0.627241
## School56     0.65168    1.29995    0.501 0.616169
## School57     6.34580    1.21350    5.229 1.75e-07 ***
## School58     2.23379    1.29995    1.718 0.085774 .
## School59     2.31904    1.22432    1.894 0.058245 .
## School60     4.92805    1.21882    4.043 5.33e-05 ***
## School61     1.22785    1.26165    0.973 0.330483
## School62     4.59914    1.16855    3.936 8.37e-05 ***
## School63     3.00837    1.27606    2.358 0.018424 *
## School64     4.90634    1.26872    3.867 0.000111 ***
## School65    -0.30410    1.19387   -0.255 0.798946
## School66     3.14840    1.16474    2.703 0.006886 **
## School67     3.52400    1.21882    2.891 0.003848 **
## School68     2.14000    1.38155    1.549 0.121429
## School69     1.74896    1.50582    1.161 0.245493
## School70     3.75583    1.29995    2.889 0.003874 **
## School71     4.15842    1.39417    2.983 0.002867 **
## School72    -3.90389    1.24830   -3.127 0.001771 **
## School73     3.69305    1.19387    3.093 0.001987 **
## School74    -1.36510    1.25485   -1.088 0.276695
## School75    -0.66083    1.17247   -0.564 0.573030
## School76     4.88178    1.18066    4.135 3.59e-05 ***
## School77     2.59363    1.36956    1.894 0.058296 .
## School78     4.07505    1.19387    3.413 0.000645 ***
## School79     0.69259    1.45221    0.477 0.633434
## School80     5.69873    1.19855    4.755 2.03e-06 ***
## School81     5.69923    1.16103    4.909 9.37e-07 ***
## School82     3.44294    1.19855    2.873 0.004084 **
```

33

```
## School83      4.55627    1.26872    3.591 0.000331 ***
## School84      4.06184    1.18066    3.440 0.000584 ***
## School85      4.56697    1.21882    3.747 0.000180 ***
## School86      1.42177    1.22432    1.161 0.245568
## School87     -5.39154    1.33695   -4.033 5.57e-05 ***
## School88      3.43398    1.43643    2.391 0.016846 *
## School89     -2.44493    1.50582   -1.624 0.104496
## School90      2.42349    1.23589    1.961 0.049926 *
## School91      3.97439    1.40749    2.824 0.004760 **
## School92      7.05976    1.43643    4.915 9.09e-07 ***
## School93      4.06226    1.20338    3.376 0.000740 ***
## School94      5.85342    1.38155    4.237 2.30e-05 ***
## School95     -1.17224    1.28370   -0.913 0.361186
## School96      4.46434    1.59670    2.796 0.005188 **
## School97      0.39040    1.35816    0.287 0.773776
## School98      5.94064    1.19387    4.976 6.65e-07 ***
## School99      3.07872    1.18493    2.598 0.009390 **
## School100     2.14449    1.21350    1.767 0.077241 .
## School101    -0.24075    1.42155   -0.169 0.865519
## School102    -2.62406    1.43643   -1.827 0.067774 .
## School103     8.74047    1.19855    7.293 3.37e-13 ***
## School104     3.19643    1.35816    2.354 0.018625 *
## School105     2.27677    1.20338    1.892 0.058534 .
## School106     1.98659    1.20338    1.651 0.098815 .
## School107    -0.42977    1.27606   -0.337 0.736280
## School108     4.82215    1.20836    3.991 6.66e-05 ***
## School109     5.38233    1.24198    4.334 1.49e-05 ***
## School110    -3.74016    1.21882   -3.069 0.002158 **
## School111     4.09802    1.38155    2.966 0.003025 **
## School112     2.13364    1.45221    1.469 0.141813
## School113    -1.64987    1.27606   -1.293 0.196076
## School114     2.82660    1.22432    2.309 0.020988 *
## School115     2.96261    1.21882    2.431 0.015094 *
## School116     4.91982    1.24830    3.941 8.19e-05 ***
## School117     0.07806    1.23000    0.063 0.949398
## School118     1.44975    1.19387    1.214 0.224664
## School119     1.62172    1.20338    1.348 0.177818
## School120     4.45665    1.27606    3.493 0.000481 ***
## School121     5.34929    1.23000    4.349 1.39e-05 ***
## School122     8.70627    1.21350    7.174 8.00e-13 ***
## School123     6.00495    1.39417    4.307 1.68e-05 ***
## School124     0.84415    1.57141    0.537 0.591154
## School125    -1.37583    1.23000   -1.119 0.263365
## School126     5.13440    1.33695    3.840 0.000124 ***
## School127     4.36928    1.25485    3.482 0.000501 ***
## School128     5.13638    1.27606    4.025 5.75e-05 ***
```

34

```
## School129     6.73507    1.24198    5.423 6.06e-08 ***
## School130     1.98101    1.38155    1.434 0.151643
## School131     3.02468    1.42155    2.128 0.033394 *
## School132     6.51510    1.28370    5.075 3.97e-07 ***
## School133     1.99677    1.35816    1.470 0.141551
## School134     4.66613    1.46896    3.176 0.001497 **
## School135    -5.16256    1.85216   -2.787 0.005329 **
## School136     2.80619    1.33695    2.099 0.035857 *
## School137     3.81218    1.29995    2.933 0.003373 **
## School138     1.16780    1.21882    0.958 0.338024
## School139     6.81166    1.18066    5.769 8.30e-09 ***
## School140     3.16738    1.24830    2.537 0.011191 *
## School141    -0.25002    1.24830   -0.200 0.841263
## School142    -2.37980    1.39417   -1.707 0.087873 .
## School143    -5.47705    1.39417   -3.929 8.63e-05 ***
## School144     5.58031    1.16855    4.775 1.83e-06 ***
## School145     2.33825    1.34730    1.736 0.082695 .
## School146     0.65978    1.19387    0.553 0.580527
## School147     1.27546    1.23000    1.037 0.299792
## School148     4.98116    1.20338    4.139 3.52e-05 ***
## School149     7.11483    1.20836    5.888 4.09e-09 ***
## School150    -1.17079    1.21882   -0.961 0.336792
## School151     9.37600    1.40749    6.662 2.91e-11 ***
## School152     4.95020    1.34730    3.674 0.000240 ***
## School153     0.56404    1.65376    0.341 0.733066
## School154     1.46295    1.43643    1.018 0.308496
## School155     3.82324    1.19855    3.190 0.001430 **
## School156     5.55417    1.21882    4.557 5.28e-06 ***
## School157     0.63890    1.25485    0.509 0.610666
## School158     3.85837    1.35816    2.841 0.004512 **
## School159     1.37278    1.43643    0.956 0.339264
## School160     5.14707    1.18934    4.328 1.53e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.083 on 7024 degrees of freedom
## Multiple R-squared:  0.2353, Adjusted R-squared:  0.2178
## F-statistic:  13.5 on 160 and 7024 DF,  p-value: < 2.2e-16
```

We can complicate more the model adding a different socio-economic status for each school, i.e. 1 unit increase of socio-econonomic status can explain different maths achievements in different schools. A model like:

$$y_{ij} = \beta_{0i} + \beta_{1i}x_{x_{ij}} + \epsilon_{ij}$$

now we have the subindex i for the slope β_{1i}.

We are not interested in these schools in particular, but in the population of schools in order to compare schools with different characteristics.

4.1   Multilevel model for group means

Now we consider a data structure with two levels, students are grouped in *Level 1* and in *Level* 2 by `Schools`. Our first model consist on a simple model with no explicative variables, i.e. our only interest is to test the difference between the maths achievements averages among schools.

Next Figure illustrates the variability by schools and within schools (we took a subsample)

```
attach(MathAchieve)
subMathAchieve=MathAchieve[School==c("91","3","31","52","74"),]
dotplot(reorder(School, MathAch) ~ MathAch,
        subMathAchieve,ylab="School")
```
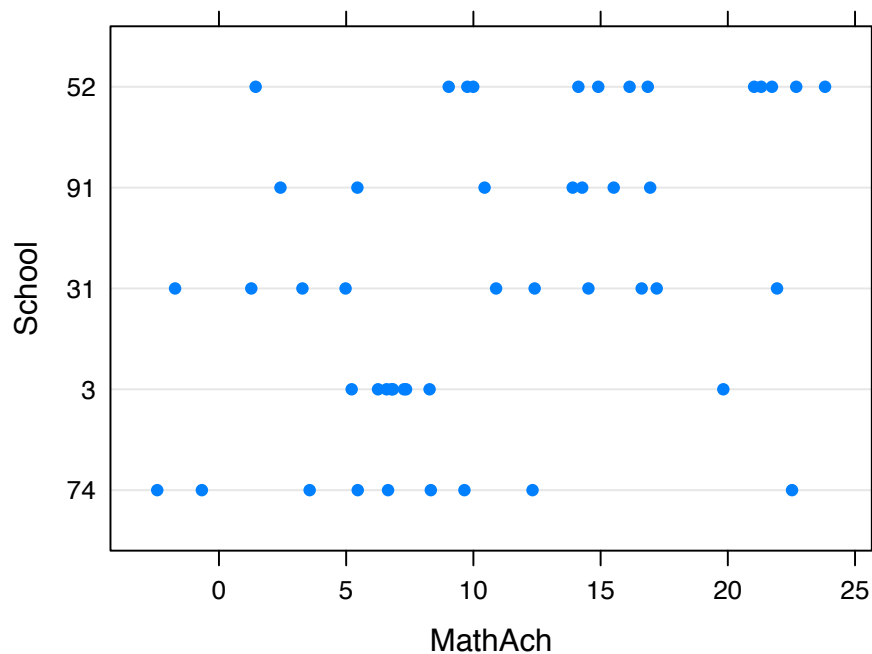

Figure 12: *Subsample of Schools*

We specify two levels in the model:

$$y_{ij} \quad = \quad \mu_i + \epsilon_{ij} \quad \text{(Level 1)}$$

36

where subindex j corresponds to individuals and i to schools. If we consider the schools as a random effect, then μ_i (the average of each school) would be given by:

$$\mu_i = \beta_0 + u_i \quad \text{(Level 2)}$$

where β_0 is the average of all the students and u_i are deviations of the ith school from the total average. Then, the previous equation can be written as:

$$y_{ij} = \beta_0 + u_i + \epsilon_{ij}, \quad i = 1, ..., m \quad j = 1, ..., n_m.$$

Indicating that the maths achievement of the jth student in the ith school is the sum of the overall mean (β_0), plus the deviation of the ith school to the overall mean (u_i), plus the deviation of the jth student to the average of his/her school (ϵ_{ij}).

In matrix notation the model can be written as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\epsilon}$$

$$\boldsymbol{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{mn_m} \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{1_1} \\ \vdots \\ \boldsymbol{1_m} \end{bmatrix} \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{1_1} & 0 & \dots & 0 \\ 0 & \boldsymbol{1_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & \boldsymbol{1_m} \end{bmatrix} \quad \boldsymbol{1_i} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{\boldsymbol{n_i \times 1}}$$

The mean of $\boldsymbol{y}$ for the ith group is given by $\beta_0 + u_i$, then u_i is the deviation of the mean of ith group to the total average. Individual level residuals ϵ_{ij} are the difference between the value of the response variable of the jth individual and the average of the group they belong. Next Figure shows the decomposition of the residuals in this model.
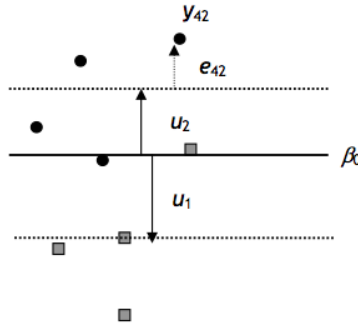

Figure 13: *Illustrative example of residuals decomposition.*

At both levels, we assume $u_i \sim \mathcal{N}(0, \sigma_{u_i}^2)$ and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ (both independent).

We fit the model with `lme`

```
multilev.lme <- lme(MathAch~1,random=~1|School, data=MathAchieve)
multilev.lme
```

```
## Linear mixed-effects model fit by REML
##   Data: MathAchieve
##   Log-restricted-likelihood: -23558.4
##   Fixed: MathAch ~ 1
## (Intercept)
##    12.63697
##
## Random effects:
##  Formula: ~1 | School
##         (Intercept) Residual
## StdDev:    2.934966 6.256862
##
## Number of Observations: 7185
## Number of Groups: 160
```

```
VarCorr(multilev.lme)
```

```
## School = pdLogChol(1)
##             Variance  StdDev
## (Intercept)  8.614025 2.934966
## Residual    39.148322 6.256862
```

The estimated mean is $12,64$ (`intercept`), and the mean for the ith school is $12,64 + \hat{u}_i$ where $\hat{u}_i$ is the predicted random effect for the school.

```
beta0 <-  fixef(multilev.lme)
beta0
```

```
## (Intercept)
##    12.63697
```

```
fixef(multilev.lme)-(predict(multilev.lme,
                        newdat=list(School=c("91","3","31","52","74"))))
```

```
##         91         3        31        52        74
## -0.9180495 4.5684656  0.2683476 -2.5377692 3.9073987
## attr(,"label")
## [1] "Predicted values"
```

38

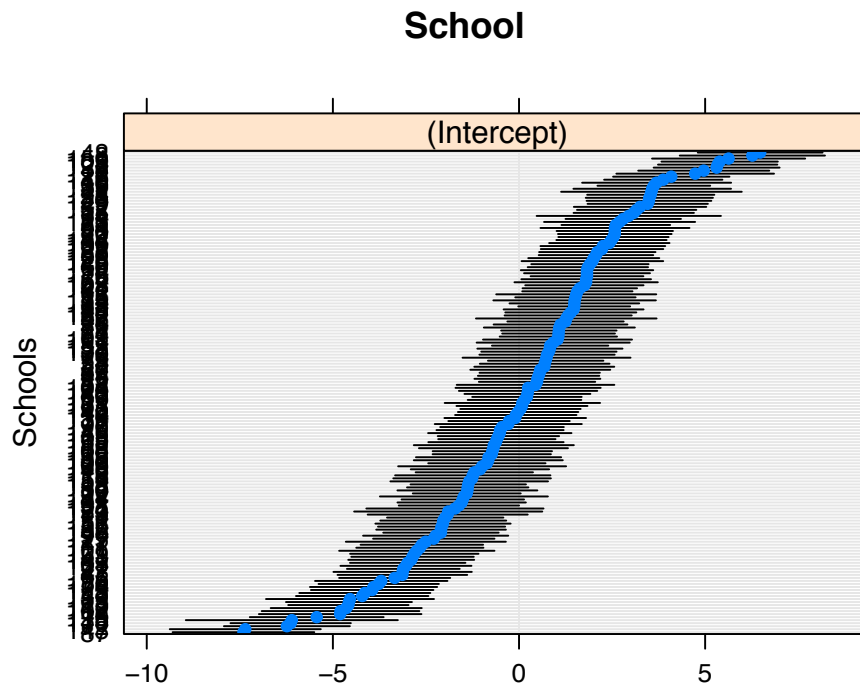We can plot the random effects

$School

School

Figure 14: *Random effects and confidence intervals*

If we had considered a fixed effect model, we would get a single variance $\sigma^2 = 6,88^2$ obtained from `summary(multilev.NULL)$sigma`. Considering a random effects model, we have σ_u^2 and σ^2, and the total variability of the data would be given by:

$$\mathrm{Var}(y_{ij}) = \sigma^2 + \sigma_u^2$$

i.e. the variability within the schools plus the variability among the schools. The intra-class correlation coefficient would be

$$ICC = \frac{2,93^2}{2,93^2 + 6,27^2} = 0,18$$

Then the 18% of the total variability is due to the difference among the averages of the schools.

4.2 Constrast for the group effect

A first attempt to contrast if there are differences among groups would be to compute the confidence intervals for σ_u,

```
intervals(multilev.lme)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                 lower      est.    upper
## (Intercept) 12.15789 12.63697 13.11606
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##    Level: School
##                      lower      est.    upper
## sd((Intercept)) 2.595995 2.934966 3.318198
##
##  Within-group standard error:
##    lower      est.    upper
## 6.154239 6.256862 6.361196
```

where `intervals(multilev.lme)$reStruct` does not contain the zero value.

A more formal way to test its significance is by the hypothesis test

$$
\begin{aligned}
H_0: & \quad y_{ij} = \beta_0 + \epsilon_{ij} \\
H_1: & \quad y_{ij} = \beta_0 + u_i + \epsilon_{ij}
\end{aligned}
$$

which is equivalent to contrast $H_0: \sigma_u^2 = 0$. We use *Likelihood Ratio Test*

$$
LR = -2(\log L_0 - \log l_1)
$$

We can directly estimate the value of the Chi-square statistic or use the function `exactLRT` in `library(RLRsim)`

```
## 'log Lik.' 983.9187 (df=2)
```

```
## [1] 2.810333e-216
```

```
## No restrictions on fixed effects. REML-based inference preferable.
```

```
##
##   simulated finite sample distribution of LRT. (p-value based on
##   10000 simulated values)
##
## data:
## LRT = 983.92, p-value < 2.2e-16
```

So, we reject H_0.

The next step is to try to improve the model by including explanatory variables to reduce the variability among schools.

Model 1: explanatory variable at Level 1

The explanatory variable is measured at Level 1. Hence, we include it at Level 1 as:

$$
\begin{array}{rcll}
y_{ij} & = & \mu_i + \beta_1 x_{ij} + \epsilon_{ij} & \text{Level 1} \\
\mu_i & = & \beta_0 + u_i & \text{Level 2}
\end{array}
$$

If x is a continous variable, this model assumes that the slope is the same (β_1) for all the schools, putting together both equations we have:

$$
y_{ij} = \underbrace{\beta_0 + \beta_1 x_{ij}}_{\text{Fixed effects}} + \underbrace{u_i + \epsilon_{ij}}_{\text{Random effects}}
$$

In matrix notation:

$$
\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\epsilon}
$$

$$
\boldsymbol{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{mn_m} \end{bmatrix}, \qquad
\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} \\ \vdots & \vdots \\ 1 & x_{mn_m} \end{bmatrix}, \qquad
\boldsymbol{\beta} = [\beta_0, \beta_1]'
$$

$$
\boldsymbol{Z} = \begin{bmatrix} \mathbf{1}_1 & 0 & \dots & 0 \\ 0 & \mathbf{1}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{1}_m \end{bmatrix}, \qquad
\mathbf{1}_j = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n_j \times 1}
$$

In this model, the relationship between $\boldsymbol{y}$ and $\boldsymbol{X}$ is represented by a straight line with intercept β_0 and slope β_1. However, the intercept for a particular ith group is $\beta_0 + u_i$, i.e., it is greater or lower than β_0 by a quantity u_i. Although the intercept varies by group the slope is the same for all groups (lines are parallel) (see next Figure).

In the example:

```
ml1 <- lme(MathAch ~ CSES, random = ~1|School, data=MathAchieve)
ml1
```
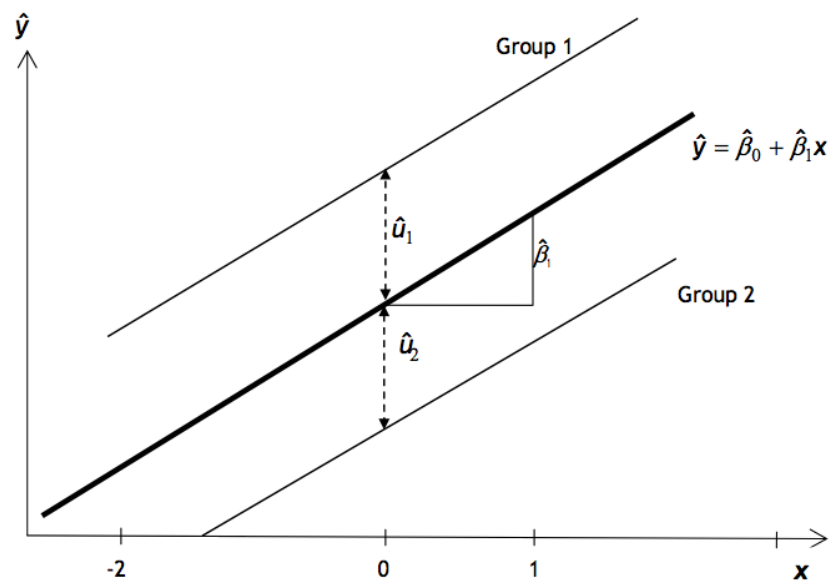
41

Figure 15: *Illustrative example of fitted global straigth line and lines for each group*

```
## Linear mixed-effects model fit by REML
##   Data: MathAchieve
##   Log-restricted-likelihood: -23362
##   Fixed: MathAch ~ CSES
## (Intercept)        CSES
##   12.649286    2.191168
##
## Random effects:
##  Formula: ~1 | School
##         (Intercept) Residual
## StdDev:    2.944893 6.083618
##
## Number of Observations: 7185
## Number of Groups: 160
```

The fixed effects are

- $\hat{\beta}_0 = 12.62$
- $\hat{\beta}_1 = 2.19$

β_0 is the average math achievement for those students with average socio-economic status. The average straight line is given by:

$$12,62 + 2,19 * CSES$$

In order to constrast if the slope is significatively not equal to zero, we need to perform a **Likelihood Ratio Test** for the fixed effects by means of maximum likelihood (ML not REML):

```
 ml0a <-update(multilev.lme, method="ML")
 ml1a <-update(ml1, method="ML")
anova(ml0a,ml1a)
```

```
##       Model df      AIC      BIC    logLik   Test L.Ratio p-value
## ml0a     1  3 47121.81 47142.45 -23557.90
## ml1a     2  4 46728.41 46755.93 -23360.21 1 vs 2 395.3969  <.0001
```

The comparison of both models shows that model `ml1` (which includes the socio-economic status covariable `CSES`) reduces the student level variance to a 5% ($(6.25^2 - 6.08^2)/6.25^2 = 0.05$). The next figure shows the fitted lines for each school:

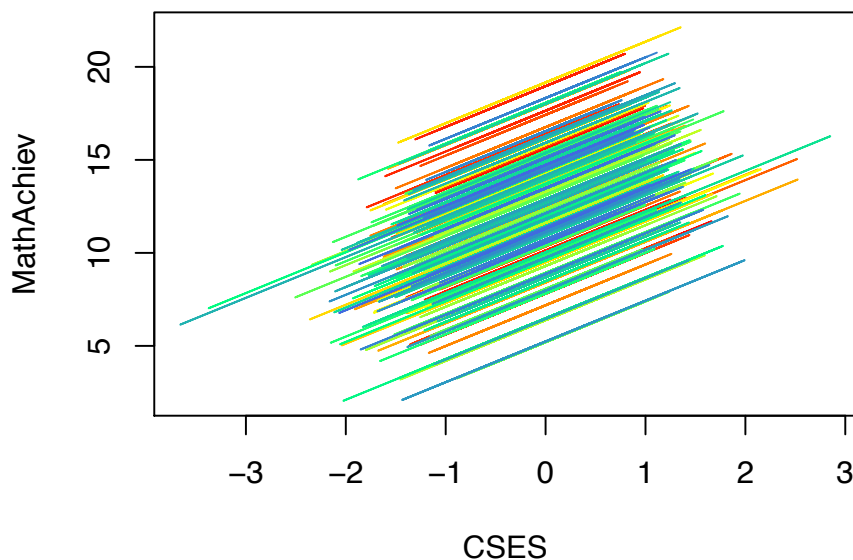**Model 2: explanatory variable at Level 2**

Figure 16: *Fitted lines for each school model*

If the explanatory variable is measured at Level 2:

$$
\begin{aligned}
y_{ij} &= \mu_i + \beta_1 x_{ij} + \epsilon_{ij} && \text{(Level 1)} \\
\mu_i &= \beta_0 + \beta_2 s_i + u_i && \text{(Level 2)} \\
y_{ij} &= \underbrace{\beta_0 + \beta_2 s_i}_{\text{fixed effects}} + \underbrace{u_i + \epsilon_i}_{\text{random effects}}
\end{aligned}
$$

In the example, we include the variable `sector`, i.e.:

```
ml2 <- lme(MathAch ~ Sector, random = ~1|School, data=MathAchieve)
ml2
```

```
## Linear mixed-effects model fit by REML
##   Data: MathAchieve
##   Log-restricted-likelihood: -23540.07
##   Fixed: MathAch ~ Sector
##   (Intercept) SectorPrivate
##     11.393044      2.804887
##
## Random effects:
##  Formula: ~1 | School
##         (Intercept) Residual
## StdDev:    2.583981 6.257108
##
```

```
## Number of Observations: 7185
## Number of Groups: 160
```

The estimated model `ml2` shows that the variance of the random effect at Level 2 (σ_u^2) has decreased ($(2,93^2 - 2,58^2)/2,93^2 = 0,22$), i.e., the unexplained variability among schools has decreased a 22% when we included the variable `sector`. For the fixed effects we have:

- $\hat{y} \mid$ sector="Public" = 11.39
- $\hat{y} \mid$ sector="Private" = 11.39 + 2.8 = 14.19

Interpreted as follows: *The math achievement of a student of a private school is expected to be 2,8 points higher than for a student in a public school.* This result can be generelized as we have considered a random effect for the schools.

In order to check the significance of the variable `Sector` we performed a likelihood ratio test:

```
ml2a <- update(ml2, method="ML")
anova( ml0a,ml2a)
```

```
##       Model df      AIC      BIC   logLik   Test  L.Ratio p-value
## ml0a      1  3 47121.81 47142.45 -23557.90
## ml2a      2  4 47087.11 47114.62 -23539.55 1 vs 2 36.70476  <.0001
```
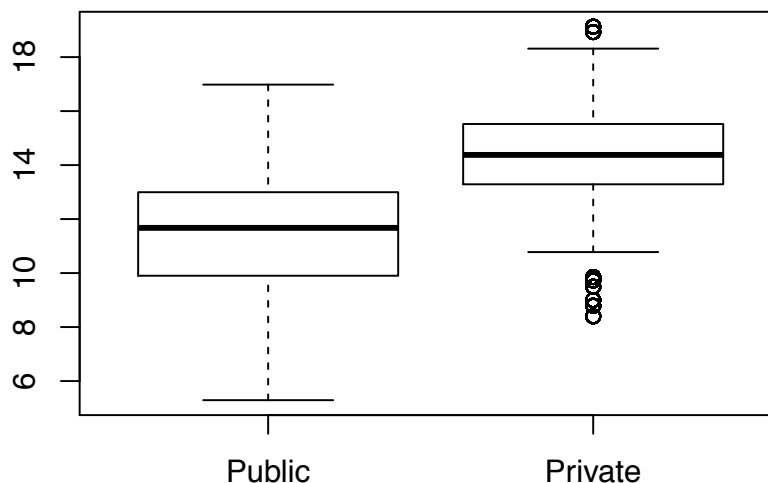

Figure 17: *Boxplot for fitted values by School sector*

4.3 Random slopes models

In this type of models, we assume that the relationship between the response variable and the explanatory variables could be different for each of the units at Level 2, i.e. the relationship may change from school to school. We can think for example that the effect of the socio-economic status in math achievements is different among the schools, and then relax the previous model by allowing for a different random slope for all the groups.

Model 3: explanatory variable at Level 1

$$
\begin{array}{rcll}
y_{ij} & = & \mu_i + \beta_{1i} x_{ij} + \epsilon_{ij} & \text{(Level 1)} \\
\mu_i & = & \beta_0 + u_i & \text{(Level 2)} \\
\beta_{1i} & = & \beta_1 v_i &
\end{array}
$$

$$
y_{ij} = \underbrace{\beta_0 + \beta_1 x_{ij}}_{\text{fixed effects}} + \underbrace{u_i + v_i x_{ij} + \epsilon_{ij}}_{\text{random effects}}, \quad
\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N(\mathbf{0}, \mathbf{G}_i) \quad
\mathbf{G}_i = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}
$$

where σ_{uv} is the covariance between the intercepts of the groups and the slopes. A positive value of the covariance implies that groups with high u_i effect tend to have high values of the v_i effect, or equivalently, school with high intercepts have higher slope. In matrix form the model can be written as:

$$
\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon}
$$

$$
\boldsymbol{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{mn_m} \end{bmatrix}, \qquad
\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_m \end{bmatrix}, \qquad
\boldsymbol{X}_i = \begin{bmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{bmatrix},
$$

$$
\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{X}_1 & 0 & \dots & 0 \\ 0 & \boldsymbol{X}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{X}_m \end{bmatrix}, \qquad
\boldsymbol{1}_i = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n_i \times 1} \qquad
\boldsymbol{\beta} = (\beta_0, \beta_1)'.
$$

The next Figure shows a representation of the models with and without a random intercept, i.e. with and without interaction between student socio-economic status and school.

In our example, this previous figure would mean that a increase of the socio-economic status tends to increase the results in both types of schools (public and private), this improvement is higher in school 1 than in school 2, and both difference incresases as long as the socio-economic status incresases.
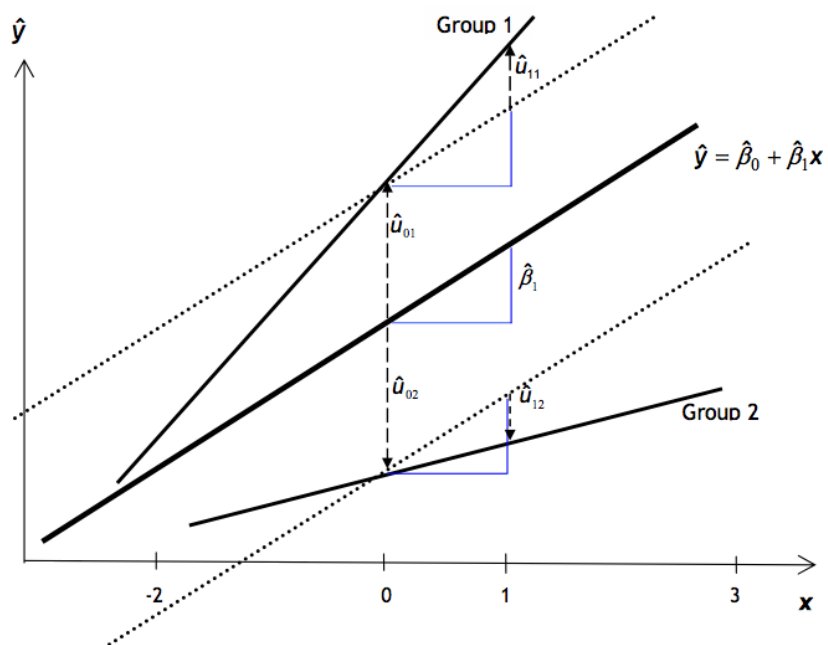
Figure 18: *Illustrative comparison of models with and without random intercept*

```
ml3 <- lme(MathAch ~ CSES, random = ~CSES|School, data=MathAchieve)
ml3
```

```
## Linear mixed-effects model fit by REML
##   Data: MathAchieve
##   Log-restricted-likelihood: -23357.12
##   Fixed: MathAch ~ CSES
## (Intercept)       CSES
##   12.649339    2.193192
##
## Random effects:
##  Formula: ~CSES | School
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 2.9464629 (Intr)
## CSES        0.8330632 0.021
## Residual    6.0580687
##
## Number of Observations: 7185
## Number of Groups: 160
```

The socio-economic status at school i is estimated as $2,19 + \hat{u}_i$ and the variance of the slopes among schools is $0,833^2 = 0,694$. For the **average school**, we predict an increase of $2,19$ units in the math achievement score when CSES increases one unit. The estimated variance are:

$$\hat{\sigma}_u^2 = 2.946^2 = 8.67 \quad \hat{\sigma}_v^2 = 0.833^2 = 0.694 \quad \hat{\sigma}_{uv} = \rho\sigma_u\sigma_v = 0.051 \quad \hat{\sigma}^2 = 6.058^2 = 36.7$$

The estimated variance of the intercepts ($\hat{\sigma}_u^2 = 8.67$) is interpreted as the variability among the schools with an average CSES . The next figure shows a plot of $\hat{u}_i$ versus $\hat{v}_i$, and it is used to show graphically if there exists a linear relationship between the intercepts and the slopes and decide if an additional term σ_{uv} is needed or not.

```
plot(ranef(ml3)[,1],ranef(ml3)[,2],
     xlab="intercepts (u_i)", ylab="slopes (v_i)")
```
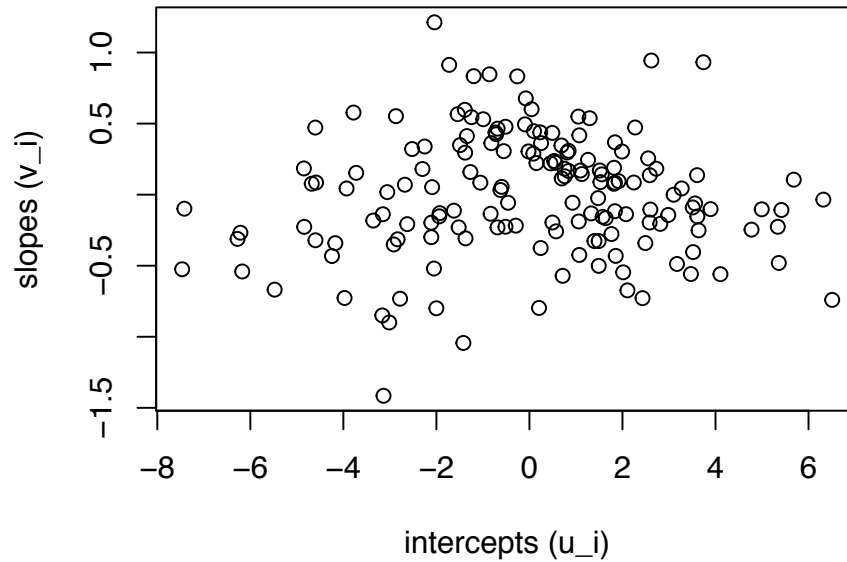
More technically, we are interested in testing the hypothesis H_0: $\sigma_{uv} = 0$ y H_1: $\sigma_{uv} \neq 0$. Then, the exact likelihood ratio test is

```
ml3a <- lme(MathAch~CSES,random = list(School=pdDiag(~CSES)),
            data=MathAchieve)
anova(ml3a,ml3)
```

Figure 19: *Random effects plot*

```
##       Model df      AIC      BIC   logLik   Test    L.Ratio p-value
## ml3a     1  5 46724.25 46758.65 -23357.12
## ml3      2  6 46726.24 46767.51 -23357.12 1 vs 2 0.01559681  0.9006
```

The next step would be testing if the slopes are different, i.e. H_0: $\sigma_v^2 = 0$, H_1: $\sigma_v^2 > 0$. In this case the LRT is not exact and we need to approximate it by:

```
test=-2*logLik(ml1, REML=TRUE) +2*logLik(ml3a, REML=TRUE)
mean(pchisq(test,df=c(0,1),lower.tail=FALSE))
```

```
## [1] 0.0008984918
```

We can also use the Akaike Information Criteria to compare the models (the lower the better):

```
AIC(logLik(ml3))
```

```
## [1] 46726.24
```

```
AIC(logLik(ml3a))
```

```
## [1] 46724.25
```

49

```
AIC(logLik(ml1))
```

```
## [1] 46732
```

The next Figure shows the fitted lines for each school with model `ml3a`
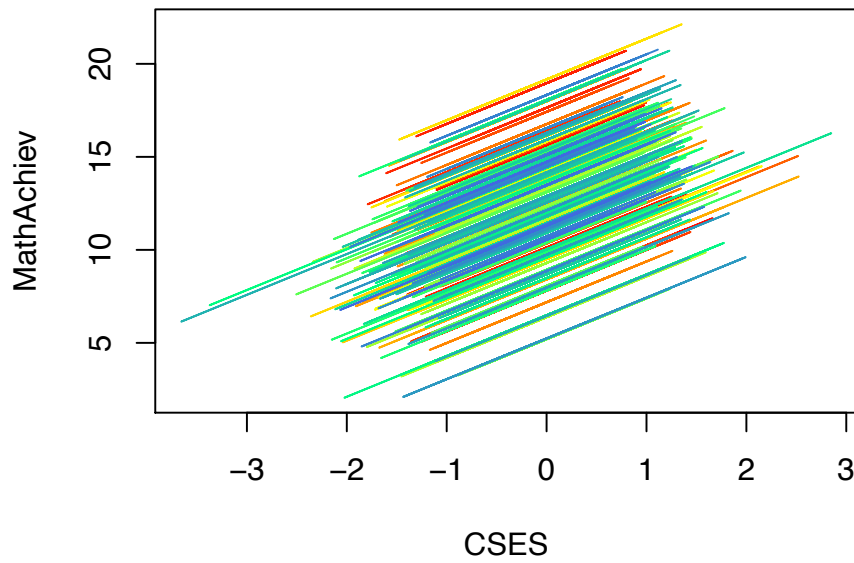
Figure 20: *Fitted lines for each school model*

Model `ml3a` assumes that the variability among schools depends on the socio-economic status. We can compute the intra-school variance as:

$$
\begin{aligned}
Var(\mu_i) &= Var(u_i + v_i x_{ij}) = Var(u_i) + Var(v_i)x_{ij}^2 + 2x_{ij}\underbrace{Cov(u_i,v_i)}_{0} \\
&= 8.68 + 0.694 \times \texttt{CSES}^2
\end{aligned}
$$

Model 4: explanatory variables at Level 1 and 2

$$
\begin{aligned}
y_{ij} &= \mu_i + \beta_{1i}x_{ij} + \epsilon_{ij} \quad \text{(Level 1)} \\
\mu_i &= \beta_0 + \beta_2 s_i + u_i \quad \text{(Level 2)} \\
\beta_{1i} &= \beta_1 + \beta_3 s_i + v_i
\end{aligned}
$$

$$
y_{ij} = \underbrace{\beta_0 + \beta_1 x_{ij} + \beta_2 s_i + \beta_3 x_{ij} : s_i}_{\text{fixed effects}} + \underbrace{u_i + v_i x_{ij} + \epsilon_{ij}}_{\text{random effects}}
$$

50

As we included a variable measured at Level 2, the fixed part is modified (with respect to Model 3), but the random effect structure remains unchanged. However, one may expect a reduction of the variance components of the random effects, as the unexplained variability is captured by the variable at Level 2. We estimated β_2 in order to check if the `Private` schools are significatively different to the `Public` schools in terms of average math scores. β_3 is estimated to check whether `Private` schools differ from `Public` schools in terms of socio-economic status and math scores.

```
ml4 = lme(MathAch~CSES*Sector,random = list(School=pdDiag(~CSES)),
          data=MathAchieve)
summary(ml4)
```

```
## Linear mixed-effects model fit by REML
##  Data: MathAchieve
##        AIC      BIC    logLik
##   46662.88 46711.03 -23324.44
##
## Random effects:
##  Formula: ~CSES | School
##  Structure: Diagonal
##         (Intercept)      CSES  Residual
## StdDev:    2.597427 0.5157795  6.058111
##
## Fixed effects: MathAch ~ CSES * Sector
##                       Value Std.Error   DF  t-value p-value
## (Intercept)       11.409644 0.2929341 7023 38.94952       0
## CSES               2.784446 0.1556796 7023 17.88575       0
## SectorPrivate      2.797268 0.4394370  158  6.36557       0
## CSES:SectorPrivate -1.345710 0.2345135 7023 -5.73830      0
##  Correlation:
##                    (Intr) CSES   SctrPr
## CSES                0.003
## SectorPrivate      -0.667 -0.002
## CSES:SectorPrivate -0.002 -0.664  0.003
##
## Standardized Within-Group Residuals:
##         Min         Q1        Med         Q3        Max
## -3.05497682 -0.73276553  0.01508149  0.75506848  2.99189223
##
## Number of Observations: 7185
## Number of Groups: 160
```

All the fixed effects in model `ml4` are significant. In order to check it we perform a test for the fixed effects:

```
ml4a <- update(ml4,method="ML")
anova(ml4a)
```

```
##             numDF denDF  F-value p-value
## (Intercept)     1  7023 3393.246  <.0001
## CSES            1  7023  358.710  <.0001
## Sector          1   158   41.233  <.0001
## CSES:Sector     1  7023   33.353  <.0001
```

`Private` schools has an average score significantly higher $(2, 79)$ and a much lower slope $(-1, 34)$ than `Public` schools, i.e. in a `Private` school and improvement of the math score is lower than in `Public` schools with respect to the socio-economic status.

The next Figure shows the fitted lines for `Public` and `Private` schoools. It can be noticed that the slopes for `Private` schools are lower that those for `Public` schools.
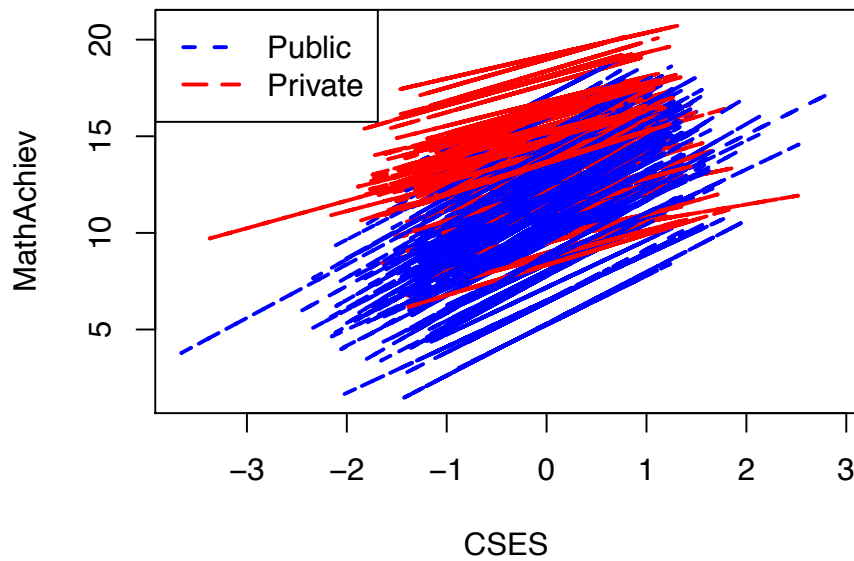
Figure 21: *Fitted lines for each school model*

5 Repeated measurements and longitudinal data

The term repeated measures refers to experimental designs (or observational studies) in which each experimental unit (or subject) is measured at several

points in time. The term *longitudinal data* is also used for this type of data.

Typical Design: Experimental units are randomly allocated to one of g treatments. A short time series is observed for each observation. An example in which there are 3 treatment groups with 3 units per treatment, and each unit is measured at four times is as follows:

Treatment	Unit	Time 1	Time 2	Time 3
1	1			
1	2			
1	3			
2	4			
2	5			
3	7			
3	8			
3	9			

In the example design just given, we have observed nine short time series, one for each experimental unit. The focus of the analysis would be to determine treatment differences, both in mean level and patterns across time. We may also want to characterize the overall time pattern (assuming that pattern doesn't differ substantially for different treatments).

Repeated measurements and longitudinal data can be viewed as multilevel models nested among individuals where for example with 2 levels, the lowest level (Level 1) are the repeated measurements nested within individuals (Level 2).

To illustrate these kind of models, we will use data from Singer and Willet (2003) consisting of weight measurements of asian children un the UK. The children were weighted between 1 and 5 times, we also know the sex and the age of the children (`child.txt`).

```
child <- read.table("data/child.txt",header=TRUE)
```

We can plot the data in order to look the temporal trend.

```
library(lattice)
library(fields)
xyplot(weight ~ age,groups=id,col=tim.colors(length(unique(child$id))),
       lwd=1,t="b",pch=19,data=child, ylim=c(2,20))
```

Looking at the Figure, we can try to:

1. Linear or quadratic trend
2. Specific intercept for each child
3. Specific straight line for each child
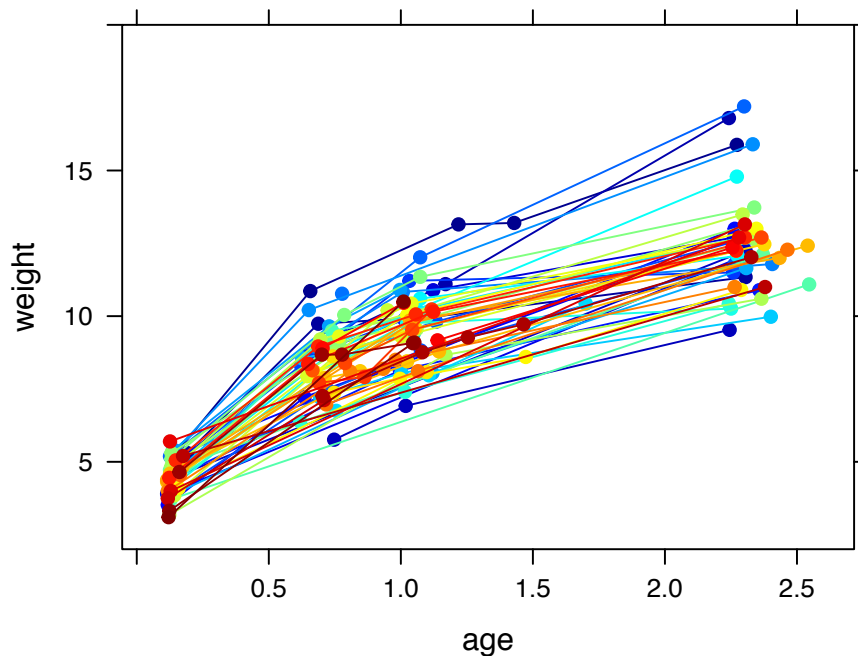
Figure 22: *Plot of children weights and ages*

5.1 Model with random intercept

We start fitting a model with a quadratic global trend and a random effect for each child. We allow for each child to have an individual trajectory that will be parallel to the mean trajectory.

$$\text{weight}_{ij} = \beta_0 + \beta_1 \text{age}_{ij} + \beta_2 \text{age}_{ij}^2 + u_i + \epsilon_{ij}$$

```r
# model with random intercept
child.mod1 <- lme(weight~age+I(age^2),random=~1|id,data=child)
child.mod1
```

```
## Linear mixed-effects model fit by REML
##   Data: child
##   Log-restricted-likelihood: -281.0327
##   Fixed: weight ~ age + I(age^2)
## (Intercept)         age     I(age^2)
##    3.432819    7.818011    -1.705631
##
## Random effects:
```

```
## Formula: ~1 | id
##          (Intercept)  Residual
## StdDev:   0.9258152 0.7401676
##
## Number of Observations: 198
## Number of Groups: 68
```

```
# Equivalently with lmer
child.mod1 <-lmer(weight~age+I(age^2)+(1|id),data=child)
```

We observed a greater variability among individuals than within individuals. A way to explain this is to allow for a difference bewteen the global trajectory and the trajectory for each child would be represented by a line, and the slope of that line varies from child to child.
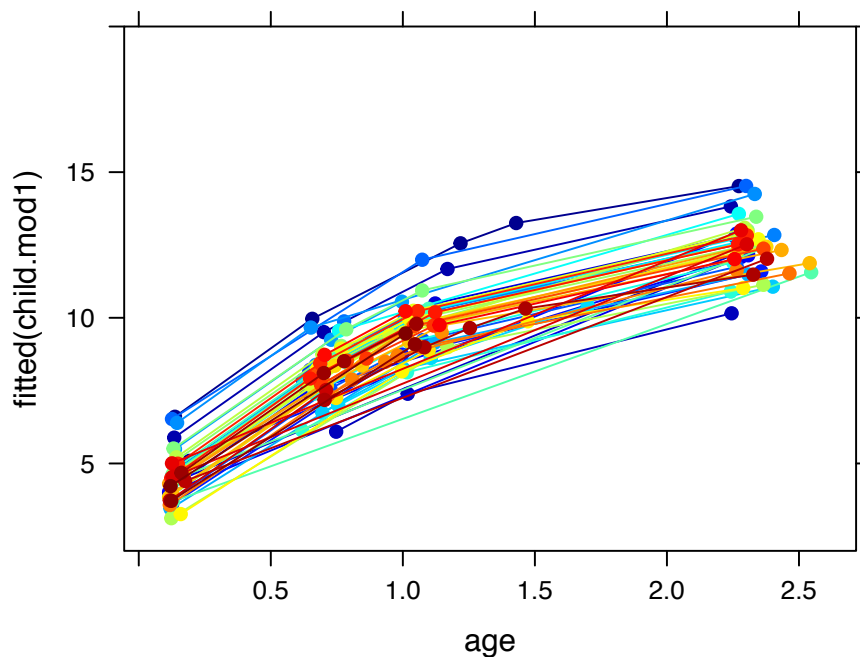
Figure 23: *Fitted model with random intercepts and quadratic trend*

5.2   Model with random slopes

The model for the `weight` of the ith child at time j is:

$$\text{weight}_{ij} = (\beta_0 + u_i) + (\beta_1 + v_i)\text{age}_{ij} + \beta_2\text{age}_i^2 + e_i,$$

$$\text{weight}_{ij} = \underbrace{\beta_0 + \beta_1 \text{age}_{ij} + \beta_2 \text{age}_{ij}^2}_{\text{fixed}} + \underbrace{u_i + v_i \text{age}_{ij} + \epsilon_{ij}}_{\text{random}}$$

```
child.mod2<-lme(weight~age+I(age^2),random=~age|id,
                data=child)
# or with lmer
# child.mod2 <- lmer(weight~age+I(age^2)+(age|id),data=child)
child.mod2
```

```
## Linear mixed-effects model fit by REML
##   Data: child
##   Log-restricted-likelihood: -262.4327
##   Fixed: weight ~ age + I(age^2)
## (Intercept)        age     I(age^2)
##    3.494664   7.703452   -1.660091
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev     Corr
## (Intercept) 0.6459801  (Intr)
## age         0.5116161  0.258
## Residual    0.5780657
##
## Number of Observations: 198
## Number of Groups: 68
```

We can see that the variability within individuals decreased, and there is a positive correlation. We can also test if both parameters are equal to zero or not:

```
child.mod2a<-lme(weight~age+I(age^2),random = list(id=pdDiag(~age)),
                data=child)
anova(child.mod2a, child.mod2)
```

```
##             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## child.mod2a     1  6 537.6541 557.2921 -262.8270
## child.mod2      2  7 538.8654 561.7764 -262.4327 1 vs 2 0.7886937  0.3745
```

The p-value shows that we do not reject the null hypothesis, i.e. correlation is zero. Now, we can contrast if the slopes must be different:

```
test=-2*logLik(child.mod1, REML=TRUE)+2*logLik(child.mod2a,REML=TRUE)
mean(pchisq(test,df=c(0,1),lower.tail=FALSE))
```

```
## [1] 7.988806e-10
```

The *Likelihood ratio test* shows that we need a model with random slopes. The next figure shows the fitted curves

```
library(fields)
xyplot(fitted(child.mod2) ~ age,groups=id,
       col=tim.colors(length(unique(child$id))),
       lwd=1,t="b",pch=19,data=child,ylim=c(2,20))
```
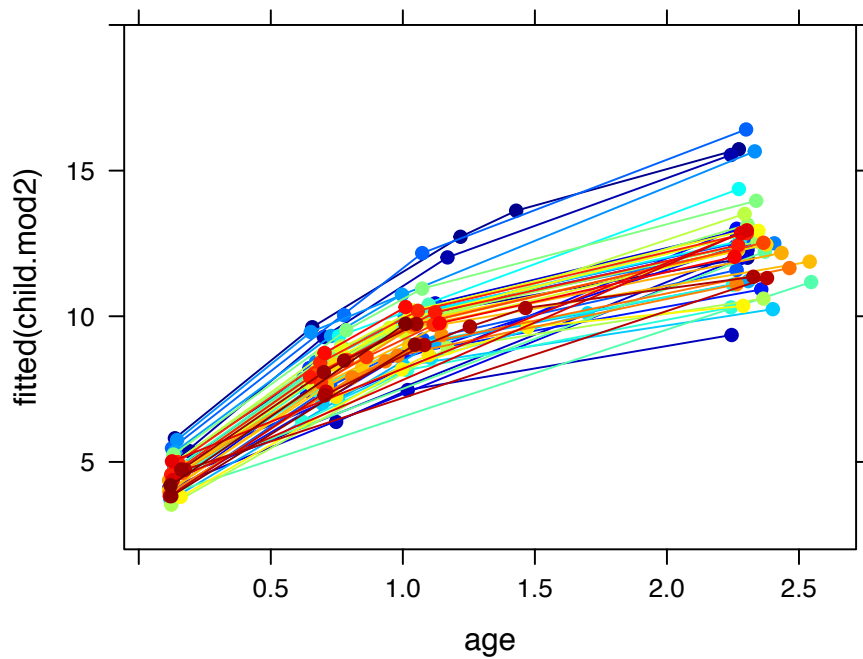

Figure 24: *Fitted model with random intercepts and quadratic trend*

Is there a different trajectory for boys and girls? We can plot the fitted values separately

```
xyplot(fitted(child.mod2) ~ age|sex,groups=id,lwd=1,t="b",pch=19,
       data=child,ylim=c(2,20))
```
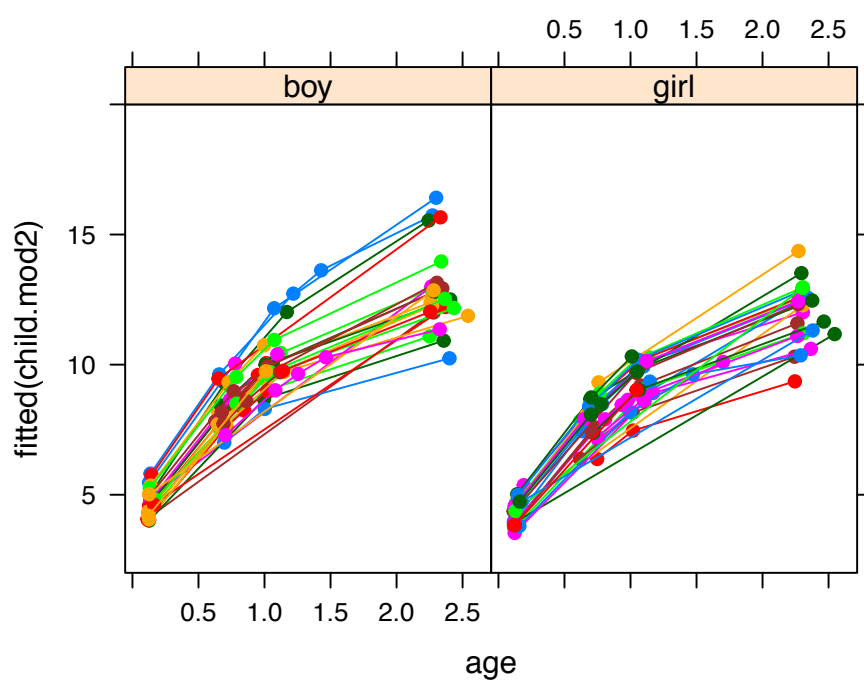
Figure 25: Fitted model with random intercepts and quadratic trend for boys and girls

The Figure shows a higher weight average for boys, and also a higher variability among boys that increases as long as the age increases.

A first step is to include the child sex and its interaction with age as a fixed effect in the model, i.e.:

$$\text{weight}_{ij} = (\beta_0 + \beta_3 \text{child}_i + u_i) + (\beta_1 + \beta_4 \text{child}_i + v_i)\, \text{age}_{ij} + \beta_2 \text{age}_{ij}^2 + \epsilon_{ij},$$

i.e.

$$\text{weight}_{ij} = \beta_0 + \beta_1 \text{age}_{ij} + \beta_2 \text{age}_{ij}^2 + \beta_3 \text{sexgirl}_i + \beta_4(\text{age}_{ij} : \text{child}_i) + u_i + v_j \text{age}_{ij} + \epsilon_{ij}$$

```
child.mod3<-lme(weight~age*sex+I(age^2),random=list(id=pdDiag(~age)),
                data=child)
summary(child.mod3)$tTable
```

```
##                    Value  Std.Error  DF    t-value      p-value
## (Intercept)    3.7543813 0.17495268 127  21.459411 4.574972e-44
## age            7.8068688 0.25507768 127  30.605848 1.838193e-60
## sexgirl       -0.5107904 0.21772204  66  -2.346066 2.198582e-02
## I(age^2)      -1.6541318 0.08795207 127 -18.807196 1.608868e-38
## age:sexgirl   -0.2296062 0.18099354 127  -1.268588 2.069084e-01
```

The variable sex is significant but the interaction not.

Interpretation: On average, boys are heavier than girls, but their average rate of linear growth is not different. How can we check if the interaction is significant?

In the previous model, we included the `sex` effect in the average growth, but we assumed that the variability of the specific deviations of each child were the same between boys and girls. To check if that assumption is reasonable, we introduce the `sex` variable as a random effect to allow the individuals lines to be different from the average for each child sex, i.e.:

$$\text{weight}_{ij} = (\beta_0 + \beta_3\, \text{sexgirl}_i + u_i : \text{sex}_i) + (\beta_1 + \beta_4 \text{sexgirl}_i + v_i : \text{sexgirl}_i)\, \text{age}_{ij} + \beta_2 \text{age}_{ij}^2 + \epsilon_{ij},$$

```
child.mod4<-lme(weight~age*sex+I(age^2),
                random=list(id=pdDiag(~sex-1),id=pdDiag(~sex:age-1)),
                data=child)
print(child.mod4)
```

```
## Linear mixed-effects model fit by REML
##   Data: child
##   Log-restricted-likelihood: -254.9944
##   Fixed: weight ~ age * sex + I(age^2)
## (Intercept)         age      sexgirl     I(age^2) age:sexgirl
##   3.7672164   7.7820066   -0.5150251   -1.6399541   -0.2428170
##
## Random effects:
##  Formula: ~sex - 1 | id
##  Structure: Diagonal
##            sexboy    sexgirl
## StdDev: 0.5763521 0.7727271
##
##  Formula: ~sex:age - 1 | id %in% id
##  Structure: Diagonal
##         sexboy:age sexgirl:age  Residual
## StdDev:  0.7012966   0.2678847 0.5588593
##
## Number of Observations: 198
## Number of Groups:
##          id id.1 %in% id
##          68         68
```

```r
summary(child.mod4)$tTable
```

```
##                   Value Std.Error  DF    t-value      p-value
## (Intercept)   3.7672164 0.16548522 127   22.764670 1.176554e-46
## age           7.7820066 0.26279169 127   29.612834 7.272974e-59
## sexgirl      -0.5150251 0.22134640  66   -2.326783 2.305571e-02
## I(age^2)     -1.6399541 0.08673299 127  -18.908078 9.738796e-39
## age:sexgirl  -0.2428170 0.17769546 127   -1.366478 1.742033e-01
```

For all the previous models in this section, we assumed that the residuals at the lowest level ϵ_{ij} were independents. In the next Section we will see how we can relax that assumption.

6   Extensions of the linear mixed model

6.1   Heteroskedasticity

Until now, we have assumed mixed models with constant variance such that:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\epsilon}, \quad Var(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$$

This assumption does not hold in many situations:

- When the variance increases as long as the magnitude of the response variable increases.
- When the variances are different for each group.
- When the variability depends on the explanatory variable.

In those situations, a better option is to model the variance as a functions of the covariates, a grouping factor or the mean of the response, i.e.:

$$Var(\epsilon_{ij}) = \sigma^2 g^2(\boldsymbol{v}_i, \boldsymbol{\theta})$$

where

$$
\begin{aligned}
v_i &= \text{is a vector of one or more covariates (including } E[y_{ij}]) \\
\boldsymbol{\theta} &= \text{a vector of unknown parameters to be estimated} \\
g^2(.) &= \text{a known functions}
\end{aligned}
$$

The library `lme4` does not include these option, instead we have to use `nlme` library.

Available functions in `library(nlme)`

- `varFixed` The variance function is $g^2(\boldsymbol{v}_i) = \boldsymbol{v}_i$:

$$Var(\epsilon_{ij}) = \sigma^2 \boldsymbol{v}_i$$

the variance is proportional to the values of the covariate (also known as *weighted least squares*).

- `varIdent` It corresponds to specifying different variances for each level of the grouping factor, $\boldsymbol{s}$. The variance function is $g^2(s_{ij}, \boldsymbol{\theta}) = \theta_{s_{ij}}$:

$$Var(\epsilon_{ij}) = \sigma^2 \theta_{s_{ij}}^2$$

- `varPower` It is a generalization of `varFixed`, the error variance is a power of the absolute value of the variance covariate: $g^2(\boldsymbol{v}_i) = |\boldsymbol{v}_i|^{2\theta}$, and

$$Var(\epsilon_{ij}) = \sigma^2 |\boldsymbol{v}_i|^{2\theta}$$

A particular case is when the covariate is the mean, i.e.:

$$Var(\epsilon_{ij}) = \sigma^2 = |\boldsymbol{m}\boldsymbol{u}_i|^{2\theta}$$

- `varComb` It allow the combination of two, or more, variance models by multiplying together the corresponding variance functions.

Example: *Body Weight Growth in Rats*

This example describes data on the body weights of rats measured over 64 days. The body weights of the rats (in grams) are measured on day 1 and every seven days thereafter until day 64, with an extra measurement on day 44. The experiment started several weeks before `day 1`. There are three groups of rats, each on a different diet.

```r
library(nlme)
data(BodyWeight)
head(BodyWeight)
```

```
## Grouped Data: weight ~ Time | Rat
##   weight Time Rat Diet
## 1    240    1   1    1
## 2    250    8   1    1
## 3    255   15   1    1
## 4    260   22   1    1
## 5    262   29   1    1
## 6    258   36   1    1
```

Figure show plots of the body weights versus time by diet. The plots indicate strong differences among the three diet groups. There is one rat with a high initial weight in group 2. The weight increases linearly with time. We can consider a model with different intercepts and different slopes for each diet group and random effects to account for the variability among rats.

An initial model would be:

```r
# create a centered variable cTime
BodyWeight$cTime <- BodyWeight$Time - mean(BodyWeight$Time)
rat1 <- lme(weight~cTime*Diet, data=BodyWeight,random=~cTime|Rat)
rat1
```

```
## Linear mixed-effects model fit by REML
##   Data: BodyWeight
##   Log-restricted-likelihood: -575.8599
##   Fixed: weight ~ cTime * Diet
## (Intercept)        cTime        Diet2        Diet3 cTime:Diet2 cTime:Diet3
## 263.7159091    0.3596391  220.9886364  262.0795455   0.6058392   0.2983375
##
## Random effects:
##  Formula: ~cTime | Rat
##  Structure: General positive-definite, Log-Cholesky parametrization
##            StdDev       Corr
```
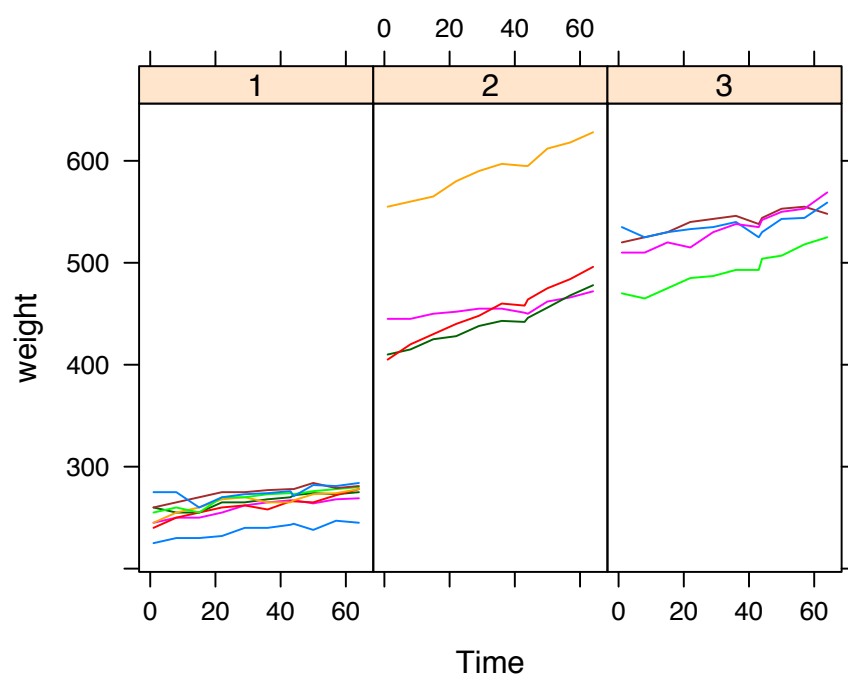
Figure 26: Body weights of rats by Time and Diet

```
## (Intercept) 36.6355112 (Intr)
## cTime         0.2484113 0.077
## Residual      4.4436052
##
## Number of Observations: 176
## Number of Groups: 16
```

Results of the model `rat1` shows that the variability of the intercepts is much higher than the variability of the slopes. A first step is to check whether there exists an interaction between `Diet` and `Time` and if it is necessary to allow for different slopes for each rat.

```
rat1a <- update(rat1,method="ML")
rat1b <- lme(weight~cTime+Diet, data=BodyWeight,random=~cTime|Rat,
             method="ML")
anova(rat1b,rat1a)
```

```
##       Model df      AIC      BIC   logLik   Test  L.Ratio p-value
## rat1b     1  8 1194.219 1219.583 -589.1095
## rat1a     2 10 1185.858 1217.563 -582.9291 1 vs 2 12.36078  0.0021
```

Hence the interaction is significant.

```
rat1.1 <- lme(weight~cTime*Diet, data=BodyWeight,
              random = list(Rat=pdDiag(~cTime)))
anova(rat1.1,rat1)
```

```
##        Model df      AIC      BIC   logLik   Test    L.Ratio p-value
## rat1.1     1  9 1169.792 1198.014 -575.8959
## rat1       2 10 1171.720 1203.078 -575.8599 1 vs 2 0.07195708  0.7885
```

```
# Null model
Model_NULL <- lme(weight~cTime*Diet, data=BodyWeight,
                  random=~1|Rat) # rat1

# LRT
test=-2*logLik(Model_NULL, REML=TRUE) +2*logLik(rat1.1, REML=TRUE)

# p-value
mean(pchisq(test,df=c(0,1),lower.tail=FALSE))
```

```
## [1] 1.488156e-19
```

And hence, we need different slopes for each rat.

In previous models, we have assumed that intercepts and slopes had the same variability per group. Next plot shows the clear within-group heteroscedasticity.
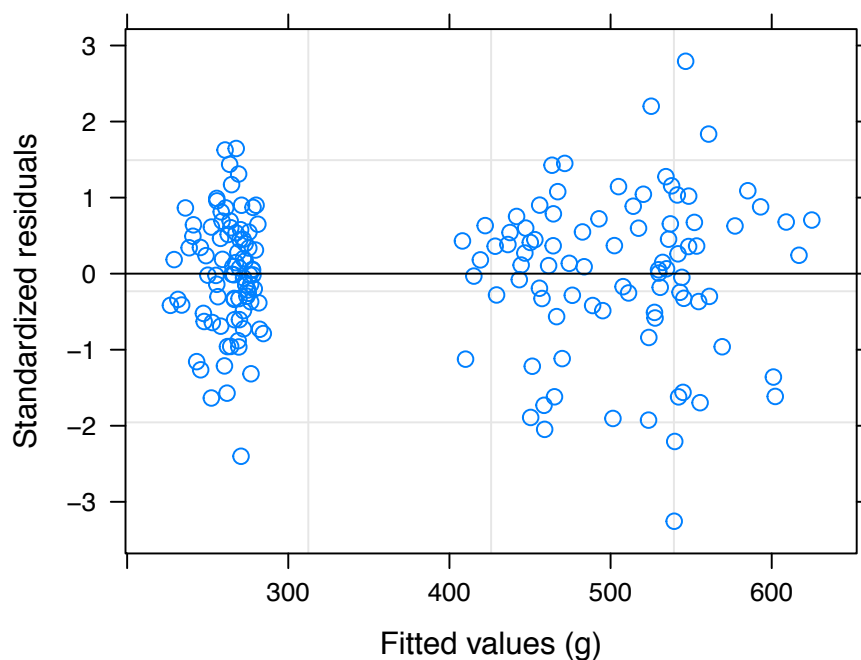
```
plot(rat1.1)
```


Figure 27: *Plot of standardized residuals versus fitted values for the homoscedastic model*

Because the fitted values are bounded away from zero, we can use the `varPower` variance function to model the heteroscedasticity.

```
rat2 <- update(rat1.1,weights=varPower(form=~fitted(.)))
rat2
```

```
## Linear mixed-effects model fit by REML
##   Data: BodyWeight
##   Log-restricted-likelihood: -570.9949
##   Fixed: weight ~ cTime * Diet
## (Intercept)        cTime        Diet2        Diet3 cTime:Diet2 cTime:Diet3
## 263.7150249    0.3610754 220.9652249 262.0741998   0.6014446   0.2952008
```

```
##
## Random effects:
##  Formula: ~cTime | Rat
##  Structure: Diagonal
##          (Intercept)      cTime   Residual
## StdDev:     36.61393 0.2435684 0.1749746
##
## Variance function:
##  Structure: Power of variance covariate
##  Formula: ~fitted(.)
##  Parameter estimates:
##      power
## 0.5430376
## Number of Observations: 176
## Number of Groups: 16
```

The residuals variance has decreased (from 4.443 to 0.175).

The plot of the standardized residuals versus the fitted values for the heteroscedastic fit corresponding to `rat2` indicates that the `varPower` variance function adequately represents the within-group heteroscedasticity.

```
plot(rat2)
```

We can test the significance of the variance parameter ($\theta = 0$)

```
anova(rat1.1,rat2)
```

```
##       Model df     AIC      BIC    logLik  Test  L.Ratio p-value
## rat1.1    1  9 1169.792 1198.014 -575.8959
## rat2      2 10 1161.990 1193.348 -570.9949 1 vs 2 9.801898  0.0017
```

The primary question of interest for the rats data is whether the growth rates differ significantly among diets.

```
summary(rat2)
```

```
## Linear mixed-effects model fit by REML
##  Data: BodyWeight
##      AIC      BIC    logLik
##   1161.99 1193.348 -570.9949
##
## Random effects:
##  Formula: ~cTime | Rat
```
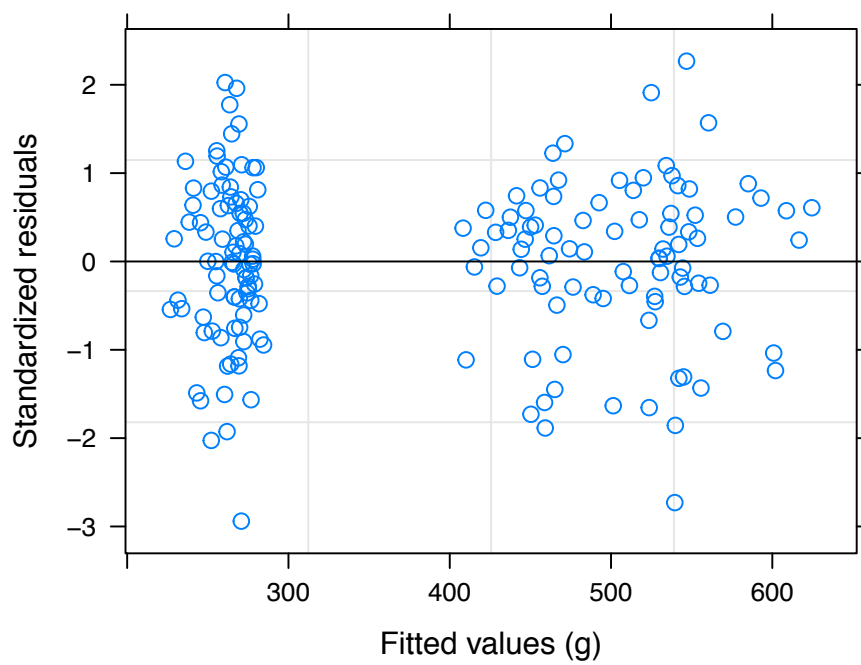
Figure 28: *Plot of standardized residuals versus fitted values for the* `varPower` *model*

```
##  Structure: Diagonal
##         (Intercept)       cTime  Residual
## StdDev:    36.61393 0.2435684 0.1749746
##
## Variance function:
##  Structure: Power of variance covariate
##  Formula: ~fitted(.)
##  Parameter estimates:
##     power
## 0.5430376
## Fixed effects: weight ~ cTime * Diet
##                  Value Std.Error  DF   t-value p-value
## (Intercept) 263.71502 12.950705 157 20.362985  0e+00
## cTime         0.36108  0.088349 157  4.086917  1e-04
## Diet2       220.96522 22.437479  13  9.848042  0e+00
## Diet3       262.07420 22.438653  13 11.679587  0e+00
## cTime:Diet2   0.60144  0.155375 157  3.870915  2e-04
## cTime:Diet3   0.29520  0.155844 157  1.894210  6e-02
##  Correlation:
##             (Intr) cTime  Diet2  Diet3  cTm:D2
## cTime        0.000
## Diet2       -0.577  0.000
## Diet3       -0.577  0.000  0.333
## cTime:Diet2  0.000 -0.569  0.000  0.000
## cTime:Diet3  0.000 -0.567  0.000  0.000  0.322
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.93954908 -0.44396488  0.07709085  0.60043464  2.26745821
##
## Number of Observations: 176
## Number of Groups: 16
```

summary(rat2) only provide tests for differences between Diet 1 (reference group) and Diet 2 and between Diet 1 and Diet 3. We can test the difference in growth rates between Diets 2 and 3 using the anova method.

```
anova(rat2, L=c("cTime:Diet2"=1,"cTime:Diet3"=-1))
```

```
## F-test for linear combination(s)
## cTime:Diet2 cTime:Diet3
##           1          -1
##   numDF denDF  F-value p-value
## 1     1   157 2.857771  0.0929
```

The argument `L` is used to specify constrats of coefficients to be tested as equal to zero. There does not seem to be a significant difference in growth rate between Diets 2 and 3.

Orthodont data

Investigators at the University of North Carolina Dental School followed the growth of 27 children (16 males, 11 females) from age 8 until age 14. Every two years they measured the distance between the pituitary and the pterygomaxillary fissure, two points that are easily identified on x-ray exposures of the side of the head.

```
data(Orthodont)
```

Next Figure shows the data with the individual linear regression fits by sex groups.

```
xyplot(distance~age | Subject, groups=Sex,
       data=Orthodont,type=c("p","r"),
       par.strip.text=list(cex=0.75), layout=c(8,4))
```
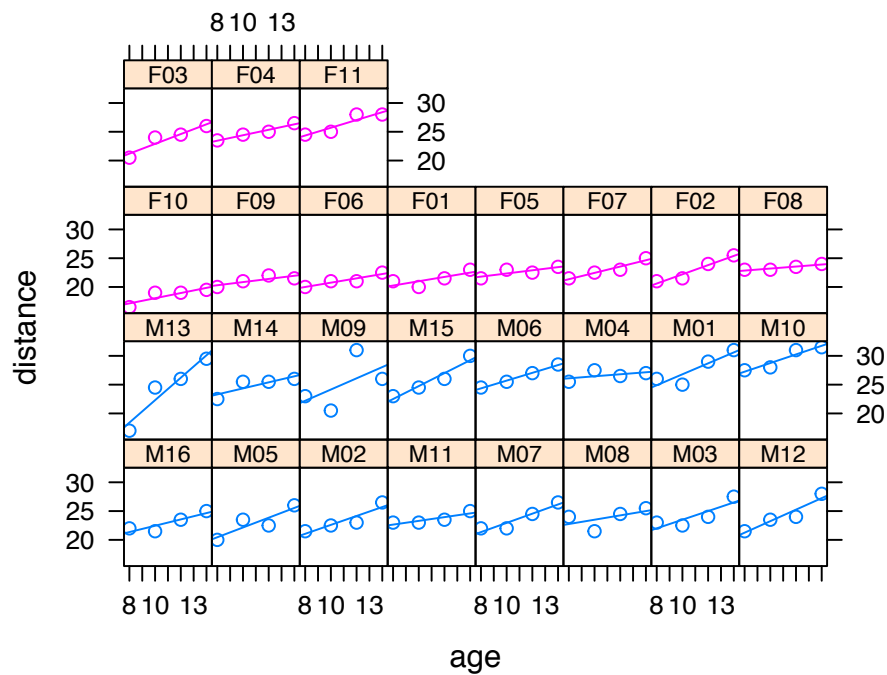

Figure 29: *Orthodont data with individual `lm` fits*

A very simple practical way to check whether the slopes are different is to compare two samples, i.e.:

```r
ab <- t(sapply(split(Orthodont, Orthodont$Subject),
        function(u)coef(lm(distance~age, data=u))))
ab <- cbind(ab, data.frame(Sex=sapply(split(as.character(Orthodont$Sex),
Orthodont$Subject),function(x)x[1])))

names(ab) <- c("ybar", "b", "Sex")

sex <- as.character(ab$Sex)

plot(ab[, 1], ab[, 2], col=c(Female="red", Male="blue")[sex],
     pch=c(Female=1, Male=3)[sex], xlab="Intercept", ylab="Slope")
```
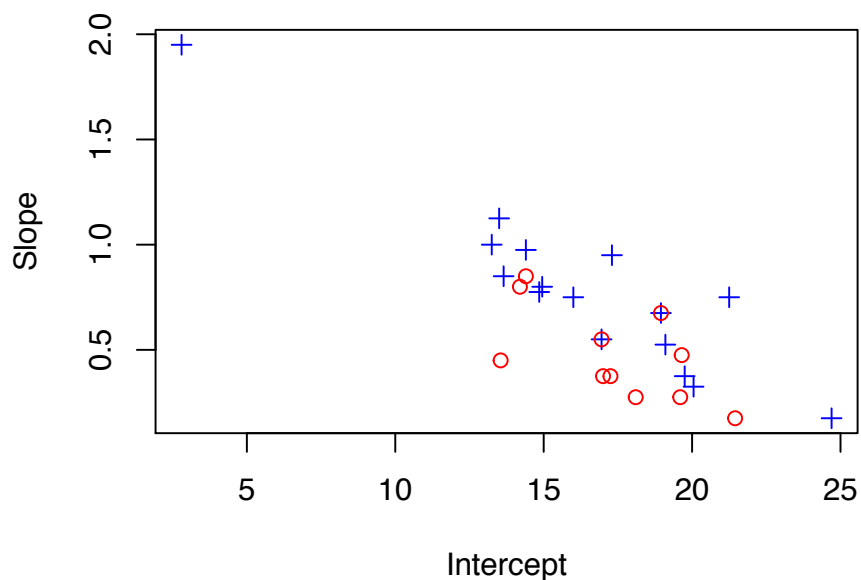

Figure 30: Plot of intercepts VS slopes by sex group

```r
#lmList.obj <- lmList(distance~age|Subject,data=Orthodont)
#coef(lmList.obj)
#plot(intervals(lmList.obj))
#plot(coef(lmList.obj)[,1],(coef(lmList.obj)[,2]),xlab= "Estimated slope",
# ylab="Estimated intercept",pch=c(Female=1, Male=3)[sex], col=c(Female="red", Male="blue")
```

```r
attach(ab)
```

```
## The following object is masked from MathAchieve:
##
##      Sex
```

```r
t.test(b[Sex=="Male"], b[Sex=="Female"], var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  b[Sex == "Male"] and b[Sex == "Female"]
## t = 2.2624, df = 25, p-value = 0.03261
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.02733691 0.58232218
## sample estimates:
## mean of x mean of y
## 0.7843750 0.4795455
```

```r
detach("ab")
```

The p-value is less than 5% and hence we reject H_0 (that the slopes are equal).

Now we consider a random effects model. First, we assume a model with the same slope for both groups and second a model with different slopes for males and females.

Note: `Orthodont` is a `groupedData` object, and by default, the random effects have the same form as the fixed effects:

```r
# Ages are centered at 11 years
ortho1<-lme(distance~I(age-11),random= ~I(age-11) | Subject,data=Orthodont)
```

One of the questions of interest is whether boys and girls have different growth patterns.

```r
ortho2 <- update(ortho1,fixed = distance~Sex*I(age-11))
```

And use the `anova` function to compare both models:

BCAM – Basque Center for Applied Mathematics

```r
ortho1.ml <- update(ortho1,method="ML")
ortho2.ml <- update(ortho2,method="ML")
anova(ortho1.ml,ortho2.ml)
```

```
##           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## ortho1.ml     1  6 451.2116 467.3044 -219.6058
## ortho2.ml     2  8 443.8060 465.2630 -213.9030 1 vs 2 11.40565  0.0033
```

```r
ortho4 <- lme(distance ~ Sex+age, random=pdDiag(~I(age-11)),
              data=Orthodont, method="ML")
anova(ortho2.ml,ortho4)
```

```
##           Model df     AIC      BIC    logLik   Test  L.Ratio p-value
## ortho2.ml     1  8 443.806 465.2630 -213.9030
## ortho4        2  6 445.117 461.2098 -216.5585 1 vs 2 5.311018  0.0703
```

```r
ortho3 <- update(ortho2.ml,method="REML")
summary(ortho3)
```

```
## Linear mixed-effects model fit by REML
##  Data: Orthodont
##        AIC      BIC    logLik
##   448.5817 469.7368 -216.2908
##
## Random effects:
##  Formula: ~I(age - 11) | Subject
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 1.8303267 (Intr)
## I(age - 11) 0.1803453 0.206
## Residual    1.3100397
##
## Fixed effects: distance ~ Sex + I(age - 11) + Sex:I(age - 11)
##                         Value Std.Error DF  t-value p-value
## (Intercept)          24.968750 0.4860007 79 51.37596  0.0000
## SexFemale            -2.321023 0.7614167 25 -3.04829  0.0054
## I(age - 11)           0.784375 0.0859995 79  9.12069  0.0000
## SexFemale:I(age - 11) -0.304830 0.1347353 79 -2.26243  0.0264
##  Correlation:
##                       (Intr) SexFml I(-11)
## SexFemale             -0.638
## I(age - 11)            0.102 -0.065
## SexFemale:I(age - 11) -0.065  0.102 -0.638
```

```
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.16807863 -0.38593916  0.00710392  0.44515472  3.84946317
##
## Number of Observations: 108
## Number of Groups: 27
```

The next Figure shows that the variability in the orthodontic distance measurements is greater among boys than among girls. Within each gender the variability seems to be constant. There are few outlying observations for subjects M09 and M13.

```r
plot(ortho3,resid(.,type="p")~fitted(.)|Sex, id=0.05, adj=-0.3)
```
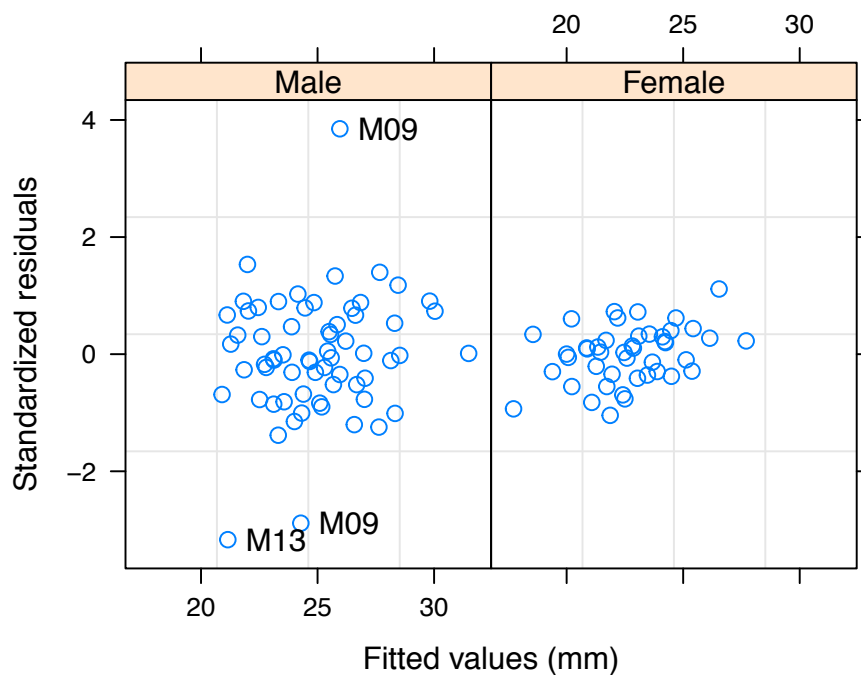

Figure 31: Scatterplots of standardized residuals versus fitted values for `ortho3`

A more general model to represent the orthodontic growth data allows different variances by gender for the within-group error. The `lme` function allow the modelling of heteroscedasticity of the within-error group via `weights` argument.

```
ortho5 = update(ortho3,weights=varIdent(form = ~1|Sex))
ortho5
```

```
## Linear mixed-effects model fit by REML
##   Data: Orthodont
##   Log-restricted-likelihood: -205.7612
##   Fixed: distance ~ Sex + I(age - 11) + Sex:I(age - 11)
##          (Intercept)             SexFemale          I(age - 11)
##           24.9687500            -2.3210227            0.7843750
## SexFemale:I(age - 11)
##           -0.3048295
##
## Random effects:
##  Formula: ~I(age - 11) | Subject
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 1.854979 (Intr)
## I(age - 11) 0.156517 0.394
## Residual    1.629585
##
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | Sex
##  Parameter estimates:
##      Male    Female
## 1.0000000 0.4088464
## Number of Observations: 108
## Number of Groups: 27
```

The parameters for `varIdent` give the ratio of the stratum standard errors to the within-group standard error. To allow identifiability, we assume

$$Var(\epsilon_{ij}) = \sigma^2 \theta^2_{s_{ij}}$$

where $\theta^2_{s_{ij}}$ is the ratio between each variance and the one for the first stratum, i.e.:

$$\hat{\sigma}_{boys} = 1.6296 \quad \hat{\sigma}_{girls} = 1.6296 \times 0.408 = 0.665$$

The next Figure shows the scatter plots of the residuals of the heteroscedastic model `ortho5`.

```
plot(ortho5,resid(.,type="p")~fitted(.)|Sex, id=0.05, adj=-0.3)
```
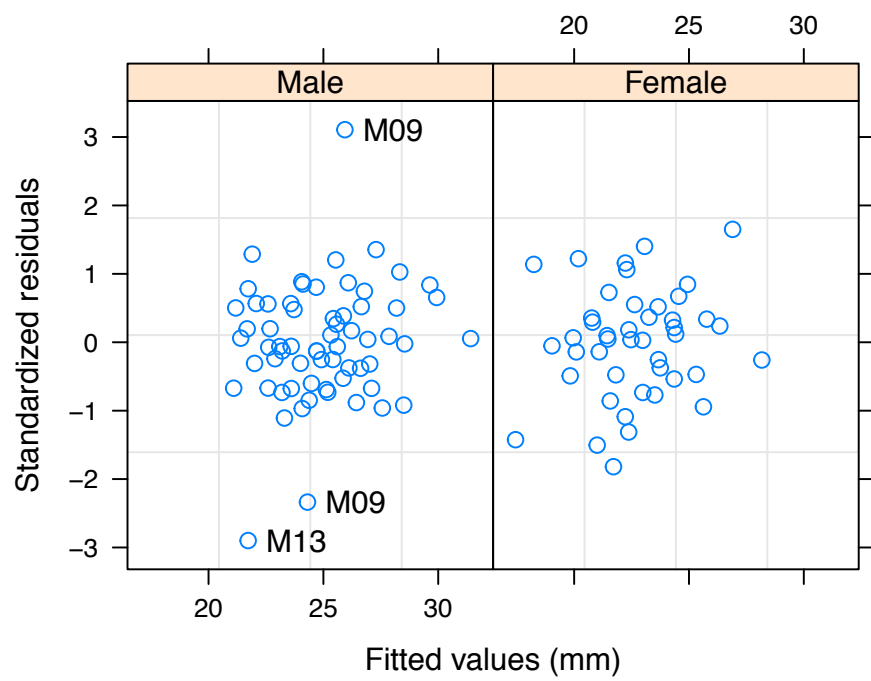
74

Figure 32: *Scatterplots of standardized residuals versus fitted values for* `ortho5`

```
plot(ortho5,distance~fitted(.), id=0.05, adj=-0.3)
```
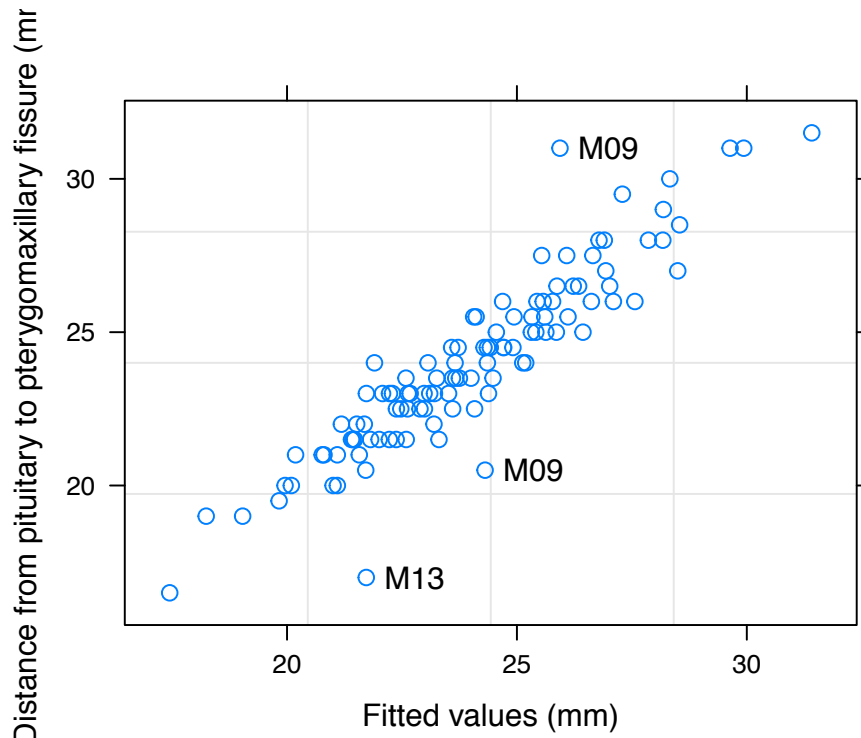

Figure 33: *Observed versus fitted values for `ortho5`*

The need of the heteroscedastic model can be formally tested with the `anova` method:

```
anova(ortho3,ortho5)
```

```
##        Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## ortho3     1  8 448.5817 469.7368 -216.2908
## ortho5     2  9 429.5225 453.3220 -205.7612 1 vs 2 21.05918  <.0001
```

The small p-value of the LRT statistic confirms that the heteroscedastic model explains better the orthodontic growth data. The assumption of normality for the within-group errors can be assessed with the normal probability plot of the residuals:
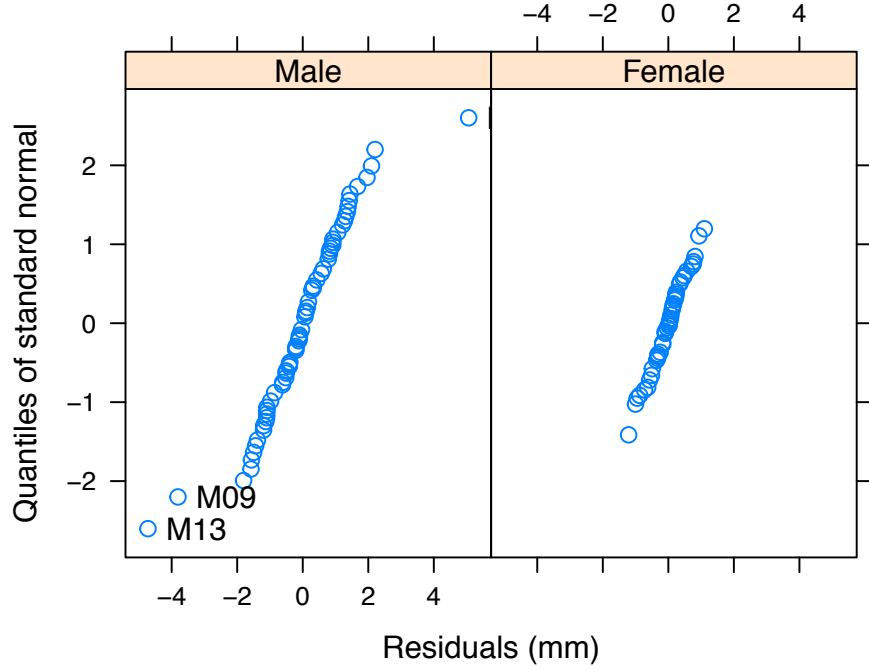
76

```
qqnorm(ortho5,~resid(.)|Sex, id=0.05, adj=-0.3)
```


Figure 34: *Scatterplots of standardized residuals versus fitted values for* `ortho5`

6.2 Correlation

Correlation strcutures are used to model dependence among observations. In the context of mixed-effects models and extended linear models, they are used to model dependence among the within-group errors. To establish a general framework for correlation structures, we assume that the within-group errors ϵ_{ij} are associated with position vectors $\boldsymbol{p}_{ij}, \boldsymbol{p}_{ij\prime}$, only through some distance between them, say $d(\boldsymbol{p}_{ij}, \boldsymbol{p}_{ij\prime})$, and not with the particular values they assume. The general within-group correlation structure for single-level grouping is

$$\text{Corr}(\boldsymbol{\epsilon}_{ij}, \boldsymbol{\epsilon}_{ij\prime}) = h\left[d(\boldsymbol{p}_{ij}, \boldsymbol{p}_{ij\prime}), \boldsymbol{\rho}\right],$$

where

- $\boldsymbol{\rho}$ is a vector of correlation parameters
- $h(\ )$ a known correlation function taking values between -1 and 1, assumed continuous in $\boldsymbol{\rho}$, and such that $h(0, \boldsymbol{\rho}) = 1$, that is, if two observations

have identical position vectors, they are the same observation and therefore have correlation 1.

- $\boldsymbol{p}_{ij}, \boldsymbol{p}_{ij'}$ position of the observations y_{ij}, y_{ik}
- $d(\ )$ a known distance function.

6.2.1 Serial correlation structures

- `corAR1` This class implements an autoregressive correlation structure of orden 1. The argument `value` initializes the single correlation parameter ϕ, which takes values between -1 and 1, and by default is set to 0. The position variable must evaluate to an integer vector, with non-repeated values per group, but its values are not required to be consecutive, so that missing time points are naturally accommodated.

$$Corr(\epsilon_{ij}, \epsilon_{ik}) = \rho^{|i-j|}$$

For instance, for $t = 5$:

$$corr(\boldsymbol{\epsilon}_i) = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ & 1 & \rho & \rho^2 & \rho^3 \\ & & 1 & \rho & \rho^2 \\ & & & 1 & \rho \\ & & & & 1 \end{pmatrix}$$

```
cs1AR1 <- corAR1(0.8,form=~1|Subject)
cs1AR1 <- Initialize(cs1AR1,data=Orthodont)
corMatrix(cs1AR1)
```

- `corCAR1` This class implements the continious time $AR(1)$ structure. Its arguments are defined as in `corAR1`, but the position variable can be any continuous variable with non-repeated values per group and the correlation parameter ϕ can only take positive values.

- `corARMA`. This class is used to specify *Autorregressive-moving average* $ARMA(p, q)$ processes.

```
cs1ARMA <- corARMA(0.4,form=~1|Subject,q=1)
cs1ARMA <- Initialize(cs1ARMA,data=Orthodont)
corMatrix(cs1ARMA)

cs2ARMA <- corARMA(c(0.8,0.4),form=~1|Subject,p=1,q=1)
cs2ARMA <- Initialize(cs2ARMA,data=Orthodont)
corMatrix(cs2ARMA)
```

- `corCompSymm` This class implements the compound symmetry correlation structure. The argument `value` is used to initialize the intraclass correlation coefficient, assuming a default value of 0. Because the compound symmetry correlation model does not depend on the position of the observation, but just on the group to which it belongs, the `form` argument is used only to specify the grouping structure.

$$Corr(\epsilon_{ij}, \epsilon_{ik}) = \left\{ \begin{array}{ll} 1 & \text{if } j = k \\ \rho & \text{if } j \neq k \end{array} \right.$$

For example,

```
cs1CompSymm <- corCompSymm(value=0.3,form=~1|Subject)
cs1CompSymm <- Initialize(cs1CompSymm,data=Orthodont)
corMatrix(cs1CompSymm)
```

specifies a compound symmetry structure with intraclass correlation of $0, 3$ and grouping defined by `Subject`. By default, `form=~1`, implying that all observations belong to the same group.

- `corSymm` This class implements the general correlation structure. The argument `value` is used to initialize the correlation parameters, being given as a numeric vector with the lower diagonal elements of the largest correlation matrix represented by the `corSymm` object stacked columnwise.

$$corr(\boldsymbol{\epsilon}_i) = \begin{pmatrix} 1 & \rho & \rho_2 & \rho_3 & \rho_4 \\ & 1 & \rho_5 & \rho_6 & \rho_7 \\ & & 1 & \rho_8 & \rho_9 \\ & & & 1 & \rho_{10} \\ & & & & 1 \end{pmatrix}$$

For example,

```
cs1Symm <- corSymm(value=c(0.2,0.1,-0.1,0,0.2,0),form=~1|Subject)
cs1Symm <- Initialize(cs1Symm, data=Orthodont)
corMatrix(cs1Symm)
```

- `corExp` This class implements an exponential spatial correlation structure. The argument `value` is used to specify values for the range ρ,

$$Corr(\epsilon_{ij}, \epsilon_{ik}) = exp(-s/\rho) \quad s = d(p_{ij}, p_{ik})$$

79

Example: Ovary data

The data are collected from a study of the number of ovarian follicles larger than 10mm in diameter detected in 11 different mares (*female horses*) at several times in their estrus cycles. The data were recorded daily from three days before ovulation until three days after the next ovulation. The measurement times for each mare are scaled so that the ovulations for each mare occur at times 0 and 1.

```
data(Ovary) # see ?Ovary
```

The figure of the number of follicles versus time per mare, suggest a periodic behavior for the number of follicles over time.
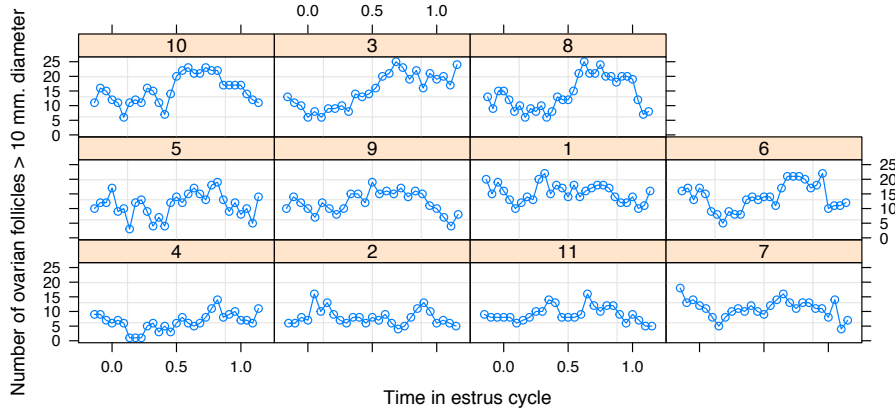
```
plot(Ovary)
```


Figure 35: *Number of ovarian follicles greater than 10mm in diameter detected in mares at various times in their estrus cycles.*

A preliminar reasonable model for the number of follicles y_{ij} for the ith mare at time t_{ij}:

$$y_{ij} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i})\sin(2\pi t_{ij}) + \beta_2 \cos(2\pi t_{ij}) + \epsilon_{ij},$$

where β_0, β_1 and β_2 are the fixed effects, $\boldsymbol{u}_i = \binom{u_{0i}}{u_{1i}} \sim \mathcal{N}(0, diag(\sigma_0^2, \sigma_1^2))$ is the random effects vector, assumed independent for different mares, and ϵ_{ij} is the within-group error, assumed independent for different i, j and independent of the random effects. The random effects u_{0i} and u_{1i} are assumed to be independent with variances σ_0^2 and σ_1^2, respectively.

```
ovary1 <- lme(follicles~sin(2*pi*Time) + cos(2*pi*Time), data=Ovary,
              random = pdDiag(~sin(2*pi*Time)))
ovary1
```

```
## Linear mixed-effects model fit by REML
##   Data: Ovary
##   Log-restricted-likelihood: -813.0409
##   Fixed: follicles ~ sin(2 * pi * Time) + cos(2 * pi * Time)
##       (Intercept) sin(2 * pi * Time) cos(2 * pi * Time)
##         12.1820241        -3.2985373        -0.8623725
##
## Random effects:
##  Formula: ~sin(2 * pi * Time) | Mare
##  Structure: Diagonal
##         (Intercept) sin(2 * pi * Time) Residual
## StdDev:    3.052136          2.079312 3.112854
##
## Number of Observations: 308
## Number of Groups: 11
```

```
ovary2 <- lme(follicles~sin(2*pi*Time) + cos(2*pi*Time), data=Ovary,
              random = pdSymm(~sin(2*pi*Time))) # general positive-def ranef
ovary2
```

```
## Linear mixed-effects model fit by REML
##   Data: Ovary
##   Log-restricted-likelihood: -811.3405
##   Fixed: follicles ~ sin(2 * pi * Time) + cos(2 * pi * Time)
##       (Intercept) sin(2 * pi * Time) cos(2 * pi * Time)
##         12.1802616        -3.2992439        -0.8620191
##
## Random effects:
##  Formula: ~sin(2 * pi * Time) | Mare
##  Structure: General positive-definite
##                     StdDev   Corr
## (Intercept)        3.052574 (Intr)
## sin(2 * pi * Time) 2.082481 -0.592
## Residual           3.112607
##
## Number of Observations: 308
## Number of Groups: 11
```

```r
anova(ovary2,ovary1) # can we assume indepedent random effects for different mares
```

```
##        Model df      AIC      BIC   logLik   Test  L.Ratio p-value
## ovary2     1  7 1636.681 1662.723 -811.3405
## ovary1     2  6 1638.082 1660.404 -813.0409 1 vs 2 3.400713  0.0652
```

```
                        # p-value is >0.05 Do not Reject H0 (correlation =0)
```

The `ACF` method obtains the empirical autocorrelation function from the residuals of an `lme` object:

```r
ACF(ovary1)
```

```
##    lag         ACF
## 1    0  1.000000000
## 2    1  0.379480128
## 3    2  0.179722025
## 4    3  0.035692749
## 5    4  0.059778881
## 6    5  0.002096953
## 7    6  0.064327101
## 8    7  0.071634861
## 9    8  0.048578247
## 10   9  0.027782488
## 11  10 -0.034275979
## 12  11 -0.077203767
## 13  12 -0.161131992
## 14  13 -0.196030197
## 15  14 -0.289337297
```

We can plot the empirical `ACF`, which shows significantly different from 0 at the first two lags, and decreases approximately exponentially for the first four lags.

```r
plot(ACF(ovary1,maxLag=20),alpha=0.01)
```

This suggests an $AR(1)$ model may be suitable for the within-group correlation:

```r
ovary3 <- update(ovary2,correlation=corAR1())
ovary3
```

```
## Linear mixed-effects model fit by REML
##   Data: Ovary
```
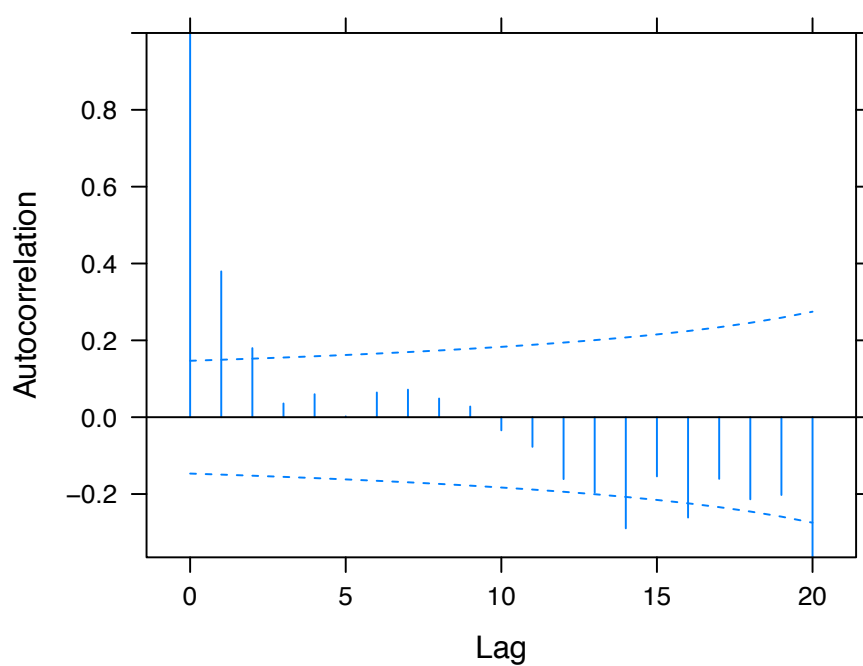
Figure 36: *Empirical autocorrelation function for the standard residuals of* **ovary1** *object.*

```
##   Log-restricted-likelihood: -773.3439
##   Fixed: follicles ~ sin(2 * pi * Time) + cos(2 * pi * Time)
##        (Intercept) sin(2 * pi * Time) cos(2 * pi * Time)
##         12.1813088         -2.9900694        -0.8767559
##
## Random effects:
##  Formula: ~sin(2 * pi * Time) | Mare
##  Structure: General positive-definite
##                      StdDev   Corr
## (Intercept)        2.863130 (Intr)
## sin(2 * pi * Time) 1.278563 -0.837
## Residual           3.490714
##
## Correlation Structure: AR(1)
##  Formula: ~1 | Mare
##  Parameter estimate(s):
##        Phi
## 0.5673788
## Number of Observations: 308
## Number of Groups: 11
```

Because model `ovary2` is nested within `ovary3` (corresponding to $\phi = 0$) we can compare them using a LRT

```
anova(ovary2,ovary3)
```

```
##        Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## ovary2     1  7 1636.681 1662.723 -811.3405
## ovary3     2  8 1562.688 1592.450 -773.3439 1 vs 2 75.99315  <.0001
```

The p-value ≈ 0 indicates that $AR(1)$ model provides better fit than model with independent errors, suggesting that within-group serial correlation is present. Approximate confidence intervals of ϕ can be obtain as:

```
intervals(ovary3)$corStruct
```

```
##        lower     est.    upper
## Phi 0.427705 0.5673788 0.6805904
## attr(,"label")
## [1] "Correlation structure:"
```

which consistently with the LRT results shows that ϕ is different from 0.

The autocorrelation pattern shows in the residuals is also consistent with a $MA(2)$ model, in which only the two first lags have non-zero correlations.

```
ovary4 <- update(ovary2,correlation=corARMA(q=2))
anova(ovary3,ovary4,test=FALSE)
```

```
##        Model df      AIC      BIC     logLik
## ovary3     1  8 1562.688 1592.450 -773.3439
## ovary4     2  9 1570.244 1603.726 -776.1218
```

Note that $AR(1)$ and $MA(2)$ are not nested, and therefore, cannot be compared through a LRT. They can however, be compared via their information criteria (the lower AIC/BIC the better). Hence $AR(1)$ is a better model.

An intermediate model between the $AR(1)$ and the $MA(2)$ models is the $ARMA(1,1)$ model. The $AR(1)$ model is nested within the $ARMA(1,1)$ model (corresponding to $\theta_1 = 0$) and we can use anova to compare both models through a LRT:

```
ovary5 <- update(ovary2,correlation=corARMA(p=1,q=1))
anova(ovary3,ovary5)
```

```
##        Model df      AIC      BIC     logLik    Test  L.Ratio p-value
## ovary3     1  8 1562.688 1592.450 -773.3439
## ovary5     2  9 1559.354 1592.837 -770.6770 1 vs 2 5.333903  0.0209
```

The low p-value for the LRT indicates that the $ARMA(1,1)$ model provides a better fit to the data. The next Figure shows the empirical ACF of the $ARMA(1,1)$ model. No significant autocorrelations are observed, indicating that the normalized residuals behave like uncorrelated noise, as expected under the appropiate correlation model.

```
plot(ACF(ovary5, maxLag=10, resType="n"),alpha=0.01)
```

6.3 Non-gaussian responses: generalized linear mixed models

GLMM's are an extension of generalized linear models (e.g., logistic regression) to include both fixed and random effects (hence mixed models). In a GLM:

$$\boldsymbol{\eta} = \boldsymbol{X\beta} \Rightarrow \boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zu}$$

In a mixed model:

$$\boldsymbol{u} \sim N(0, \boldsymbol{G}) \quad \boldsymbol{y}|\boldsymbol{u} \sim N(\boldsymbol{X\beta} + \boldsymbol{Zu}, \boldsymbol{R}) \Rightarrow \boldsymbol{y}|\boldsymbol{u} \sim B(m, p(\boldsymbol{u})); \ P(\lambda(\boldsymbol{u}))$$

In order to fit the model, we need to compute the likelihood function of $\boldsymbol{y}$ which implies the calculus of an integral with no analytical solution. There are three ways to estimate a GLMM:
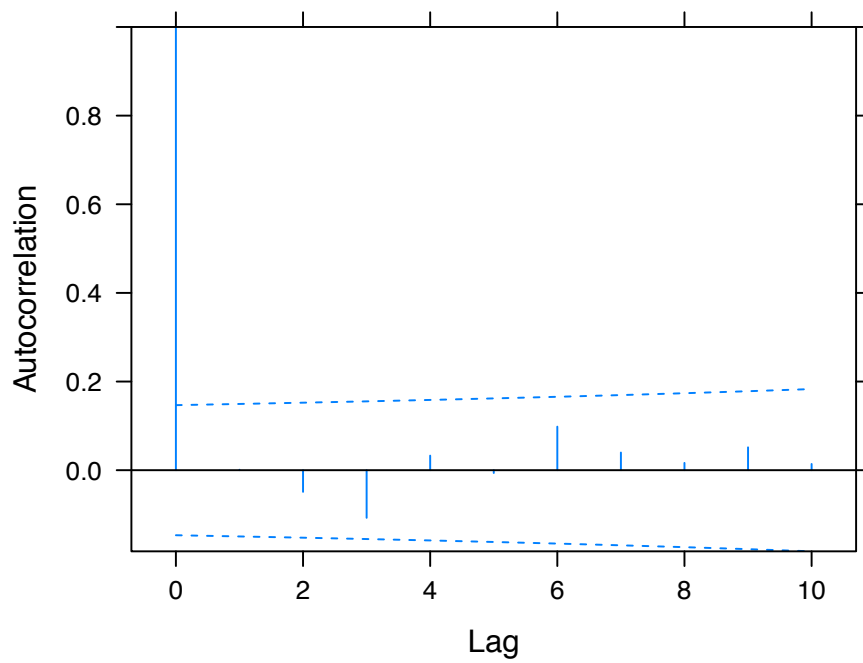
Figure 37: *Empirical autocorrelation function corresponding to the normalized residuals of model* `ovary5`.

- Maximum likelihood (ML)
- Markov chain Montecarlo (MCMC) bayesian methods
- Generalized Estimating Equations (GEE)

We will focus on ML, where a numerical approximation is use to compute the integral (Laplace approximation or Gaussian quadrature).

In R the function to estimate a GLMM is `glmer` in library `lme4`, it works like funciona como `lmer`, but for we need to include the argument `family=` and `link` function, similarly to function `glm`.

GLMMs for binary data: Bangladesh Demographic and Health Survey

The data comes from the Bangladesh Demographic and Health Survey (BDHS), a nationally representative cross-sectional survey of women of reproductive age (13-49 years). Our response variable is a binary indicator of whether a woman received antenatal care from a medically-trained provider (a doctor, nurse or midwife) at least once before her most recent live birth. To minimise recall errors, the question was asked only about children born within five years of the survey. For this reason, our analysis sample is restricted to women wh o had a live birth in the five-year period before the survey. Note that if a woman had more than one live birth during the reference period, we consider only the most recent.

The data have a two-level hierarchical structure with 5366 women at level 1, nested within 361 communities at level 2. In rural areas a community corresponds to a village, while an urban community is a neighbourhood based on census definitions.

At level 1, we consider variables such as a woman's age at the time of the birth and education. Level 2 variables include an indicator of whether the region of residence is classified as urban or rural. Explanatory variables are:

- `comm` Community identifier
- `wonid` Woman identifier
- `antemed` Received antenatal care at least once from a medically-trained provider, e.g. doctor, nurse or midwife (1 = yes, 0 = no)
- `bord` Birth order of child (ranges from 1 to 13)
- `mage` Mother's age at the child's birth (in years)
- `urban` Type of region of residence at survey (1=urban, 0=rural)
- `meduc` Mother's level of education at survey (1=none, 2=primary, 3=secondary or higher)
- `slam` Mother's religion (1=Islam, 0=other)
- `wealth` Household wealth index in quintiles (1=poorest to 5=richest)

In the case of a model with random intercept for Gaussian data, we have:

$$E[y_{ij}|u_i] = \beta_0 + u_i + \epsilon_{ij} \quad u_i \sim N(0, \sigma_u^2) \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

The binary response variable (`antemed`) 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). We will begin by fitting a model with only an intercept and community effects:

$$\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_0 + u_{0j}$$

The intercept β_0 is shared by all communities while the random effect u_{oj} is specific to community j. The random effect is assumed to follow a normal distribution with variance $\sigma_{u_0}^2$.

```r
BDHS <- read.table("data/BDHS.txt", header = TRUE)
library(lme4)
fit <- glmer(antemed ~ (1 | comm), family = binomial("logit"), data = BDHS)
```

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 (`comm`). The `family` option is used to specify a binomial distribution for the response with a logit link function. The data option specifies the dataframe being used to fit the model.

The estimation procedure optimizes a function of the log likelihood using penalized iteratively re-weighted least squares. The log-likelihood is evaluated using the Laplacian approximation.

```r
summary(fit)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: antemed ~ (1 | comm)
##    Data: BDHS
##
##      AIC      BIC   logLik deviance df.resid
##   6639.5   6652.7  -3317.8   6635.5     5364
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7779 -0.7458  0.3423  0.7118  2.6784
##
```

```
## Random effects:
##  Groups Name        Variance Std.Dev.
##  comm   (Intercept) 1.464    1.21
## Number of obs: 5366, groups:  comm, 361
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.14809    0.07178   2.063   0.0391 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the model estimates (using Laplacian approximation), we can say that the log-odds of receiving antenatal care from a medically-trained provider in an *average* community (one with u_{0j}) is estimated as $\hat{\beta}_0 = 0.148$. The intercept for community j is $0.148 + u_{0j}$, where the variance of u_{0j} is estimated as $\hat{\sigma}^2_{0j} = 1.464$.

The likelihood ratio statistic for testing the null hypothesis, that $\hat{\sigma}^2_{0j} = 0$, can be calculated by comparing the two-level model, with the corresponding single-level model without the level 2 random effect.

```
fit0 <- glm(antemed ~ 1, family=binomial("logit"),data=BDHS)
test = -2*logLik(fit0) + 2*logLik(fit)
mean(pchisq(test,df=c(0,1),lower.tail=FALSE))
```

```
## [1] 3.169612e-176
```

We will now examine estimates of the community effects or residuals, $\hat{u}_{0j}$, obtained from the null model. The plot shows the estimated residuals for all 361 communities in the sample. For substantial number of communities, the 95% confidence interval does not overlap the vertical line at zero, indicating that uptake the antenatal care in these communities is significantly above average (crossing the zero line). The confidence intervalss are also quite wide. This is because the sample size with a community is small. leading to large standard errors for the estimated community residuals $\hat{u}_{0j}$.

```
library(lattice)
qqmath(ranef(fit, condVar=TRUE),strip=FALSE)$comm
```
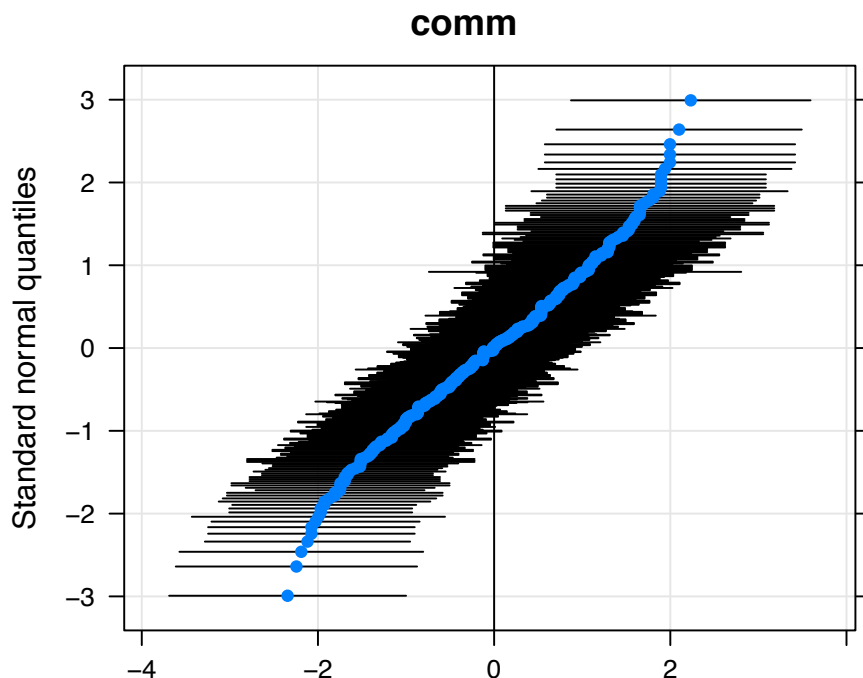
comm

Now we include maternal age (`mage`) as an explanatory variable in the model.
First we center the covariate:

```
BDHS$magec <- BDHS$mage - mean(BDHS$mage)
fit2 <- glmer(antemed~magec+(1|comm),family=binomial("logit"),data=BDHS)
summary(fit2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: antemed ~ magec + (1 | comm)
##    Data: BDHS
##
##      AIC      BIC   logLik deviance df.resid
##   6603.4   6623.2  -3298.7   6597.4     5363
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9757 -0.7431  0.3357  0.7190  3.2357
##
## Random effects:
##  Groups Name        Variance Std.Dev.
```

```
##  comm   (Intercept) 1.462    1.209
## Number of obs: 5366, groups:  comm, 361
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.144604   0.071781   2.015    0.044 *
## magec       -0.032357   0.005235  -6.181 6.37e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr)
## magec 0.008
```

Note that there is little change in the estimate of the between-community variance, suggesting that the distribution of maternal age is similar across communities.

The equation of the average fitted regression line, expressing the relationship between the log-odds of receiving antenatal care and maternal age, is:

$$\log\left(\frac{\hat{p}_{ij}}{1-\hat{p}_{ij}}\right) = 0.144 - 0.032\mathrm{magec}_{ij}$$

The fitted line for a given community will differ from the average line in its intercept, by an amount $\hat{u}_j$ for community j. A plot of the predicted community lines will therefore show a set of parallel lines.

```
predprob <- fitted(fit2)
library(VGAM)
```

```
## Loading required package: stats4
```

```
##
## Attaching package: 'stats4'
```

```
## The following object is masked from 'package:spam':
##
##     mle
```

```
## Loading required package: splines
```

```
predlogit <- logit(predprob)
datapred=unique(data.frame(cbind(predlogit = predlogit, comm = BDHS$comm,mage = BDHS$mage)))
xyplot(predlogit ~ mage, data = datapred, groups = comm, type = "l", col = "blue", xlim = c(
```

91
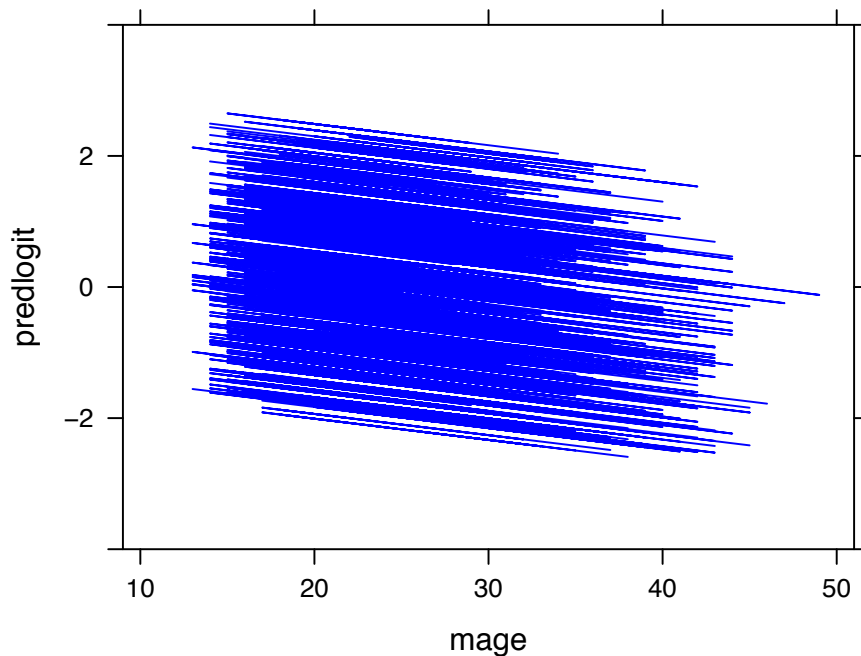
Figure 38: *Fitted values*

Hence, for a woman of 22 years the log-odds of receiving antenatal care ranges from about -2.2 to 2.5 depending on which community she lives in. This translates to a range in probabilities of $exp(-2.2)/[1 + exp(-2.2)] = 0.10$ to $exp(-2.5)/[1 + exp(-2.5)] = 0.92$, so there are strong community effects.

Now, we include additional variables, such as `wealth` and `meduc`.

```
BDHS$meduc <- factor(BDHS$meduc)
BDHS$wealthc <- BDHS$wealth-mean(BDHS$wealth) # we center wealth
fit3 <- glmer(antemed~magec+wealthc+meduc+(1|comm), family=binomial("logit"), data=BDHS)
summary(fit3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: antemed ~ magec + wealthc + meduc + (1 | comm)
##    Data: BDHS
##
##      AIC      BIC   logLik deviance df.resid
##   5993.5   6033.0  -2990.7   5981.5     5360
```

92

```
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4726 -0.6437  0.2581  0.6518  3.7138
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  comm   (Intercept) 0.8667   0.931
## Number of obs: 5366, groups:  comm, 361
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.490324   0.080317  -6.105 1.03e-09 ***
## magec       -0.005141   0.005722  -0.898    0.369
## wealthc      0.402697   0.030012  13.418  < 2e-16 ***
## meduc2       0.544514   0.085843   6.343 2.25e-10 ***
## meduc3       1.305071   0.099406  13.129  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr) magec  welthc meduc2
## magec   -0.195
## wealthc  0.213 -0.105
## meduc2  -0.581  0.215 -0.178
## meduc3  -0.593  0.317 -0.376  0.554
```

Variance Partition Coefficient

The variance partition coefficient (VPC) is calculated as $\frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}$. From theory, the standard logistic distribution has variance $\sigma^2 = \pi^2/3 = 3.29$. Then with $\hat{\sigma}_u^2 = 0.8667$, $VPC = 0,868/(0,868 + 3,29) = 0.208$.

```
# variance of r.e.
as.numeric(summary(fit3)$varcor)
```

```
## [1] 0.8667314
```

Thus 21% of the residual variation in the propensity to use antenatal care services is attributable to unobserved community characteristics. In model `fit2` (with no predictors) the variance was $1,462$, giving a VPC of $1,462/(1,462 + 3,29) = 0,307 \approx 31\%$.

Predict probabilities from a Multilevel model

By substituing the mean of the Level 2 residuals ($u_j = 0$) in the formula for the response probability:

$$p_{ij} = \frac{exp(z_{ij})}{1 + exp(z_{ij})}$$

where

$$z_{ij} = -0.491 - 0.005 \text{magec}_{ij} + 0.403 \text{wealthc}_{ij} + 0.545 \text{meduc2}_{ij} + 1.306 \text{meduc3}_{ij}$$

The resulting predictions represent probabilities for the median community, and are sometimes referred to as *cluster-specific* predictions.

Two-level random slope model

In the previous model we assumed that the maternal age, education and household wealth are assumed to be the same in each community. We will now fit a random slope (coefficient) for wealth to allow its effect to vary across communities.

$$\log\left(\frac{p_{ij}}{1 - pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{meduc}_{ij} + \beta_3 \text{wealthc}_{ij} + u_{0j} + u_{4j} \text{wealthc}_i j$$

Note that the new term u_{4j} has been added to the model, so that the coefficient of `wealthc` has become $\beta_{4j} = \beta_4 + u_{4j}$, and so the community-level variance has been replaced by a matrix with two new parameters, $\sigma^2_{u_4}$ and $\sigma_{u_{04}}$. The two community level residuals σ_{0j} and σ_{4j} are assumed to follow a bivariate normal distribution with mean vector 0 and variance-covariance matrix

$$\begin{pmatrix} \sigma^2_{u_0} & \\ \sigma_{u_{04}} & \sigma^2_{u_4} \end{pmatrix}$$

The model is fitted in `R`

```r
fit4 <- glmer(antemed~magec+meduc+wealthc+(1+wealthc|comm), data = BDHS, family=binomial("lo
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
## eigenvalues
```

```r
summary(fit4)
```

94

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: antemed ~ magec + meduc + wealthc + (1 + wealthc | comm)
##    Data: BDHS
##
##      AIC      BIC   logLik deviance df.resid
##   5987.3   6040.0  -2985.7   5971.3     5358
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2754 -0.6513  0.2928  0.6409  4.3347
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  comm   (Intercept) 0.84641  0.9200
##         wealthc     0.01399  0.1183   -1.00
## Number of obs: 5366, groups:  comm, 361
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.490825   0.079968  -6.138 8.37e-10 ***
## magec       -0.005163   0.005727  -0.902    0.367
## meduc2       0.541494   0.086211   6.281 3.36e-10 ***
## meduc3       1.299956   0.098867  13.149  < 2e-16 ***
## wealthc      0.408347   0.030908  13.212  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr) magec  meduc2 meduc3
## magec   -0.194
## meduc2  -0.584  0.213
## meduc3  -0.592  0.315  0.558
## wealthc  0.042 -0.103 -0.174 -0.362
## convergence code: 0
## unable to evaluate scaled gradient
## Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```

Note, that we specified the random term (`1+wealthc|comm`) which implies that the random intercepts and slopes are correlated. Now we test if the intercept and the slops are independent, i.e. the null hypothesis is that the two new parameters $\sigma_{u_4}^2$ and $\sigma_{u_{04}}$ are simultaneously equal to zero.

```
anova(fit3,fit4) # reject H0
```

```
## Data: BDHS
## Models:
## fit3: antemed ~ magec + wealthc + meduc + (1 | comm)
## fit4: antemed ~ magec + meduc + wealthc + (1 + wealthc | comm)
##      Df    AIC  BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## fit3  6 5993.5 6033 -2990.7   5981.5
## fit4  8 5987.3 6040 -2985.7   5971.3 10.14     2   0.006283 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We therefore conclude that the effect of wealth does indeed vary across communities. The effect of wealth on the log-odds of receiving antenatal care in community j is estimated as $0,408 + \hat{u}_{4j}$, and the between-community variance in the effect of wealth is estimated as 0.014. Because `wealthc` has been centered about is sample mean, the intercept variance $\hat{\sigma}_{u0}^2 = 0,846$ is interpreted as the between-community variance in the log-odds of antenatal care at the mean of the wealth index. The negative intercept-slope covariance estimate $(\hat{\sigma}_{04} = 0,92 \times 0,1183 \times -1,00 = -0,110)$ implies that communities with above-average antenatal care uptake (intercept residual $\hat{u}_{0j} > 0$) tend also to have below-average effects of wealth (slope residual $\hat{u}_{4j} < 0$). In other words, in those communities with a high level of antenatal care, it tends to decrease with wealth.

The equation for the fitted regression for community j, for a woman of mean age (`magec=0`), and no education (`meduc2=meduc3=0`) is:

$$log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = (-0.4908 + \hat{u}_i) + (0.408 + \hat{v}_i)\text{weathc}_{ij}$$

To obtain the fitted line of a woman with different ages or levels of education, only the intercept would change. For a woman with primary education, for example, the intercept would increase from -0.4908 to $-0.4908 + 0.5415 = 0.0507$.

We can include a Level 2 explanatory variable (e.g. `urban`)

```
BDHS$urban <- factor(BDHS$urban)
levels(BDHS$urban) <- c("rural","urban")
fit5 <- glmer(antemed~magec+meduc+wealthc+urban+(1+wealthc|comm),
              data = BDHS, family=binomial("logit"))
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.0129444 (tol =
## 0.001, component 1)
```

96

```
summary(fit5)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: antemed ~ magec + meduc + wealthc + urban + (1 + wealthc | comm)
##    Data: BDHS
##
##      AIC      BIC   logLik deviance df.resid
##   5911.7   5970.9  -2946.8   5893.7     5357
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6969 -0.6555  0.2729  0.6472  4.3840
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  comm   (Intercept) 0.6491   0.8057
##         wealthc     0.0271   0.1646   -1.00
## Number of obs: 5366, groups:  comm, 361
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.863081   0.087340  -9.882  < 2e-16 ***
## magec       -0.004906   0.005731  -0.856    0.392
## meduc2       0.569354   0.086176   6.607 3.93e-11 ***
## meduc3       1.350763   0.098549  13.707  < 2e-16 ***
## wealthc      0.354832   0.031391  11.304  < 2e-16 ***
## urbanurban   1.032676   0.114230   9.040  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) magec  meduc2 meduc3 welthc
## magec      -0.185
## meduc2     -0.569  0.213
## meduc3     -0.593  0.315  0.560
## wealthc     0.105 -0.099 -0.170 -0.355
## urbanurban -0.480  0.020  0.075  0.121 -0.230
## convergence code: 0
## Model failed to converge with max|grad| = 0.0129444 (tol = 0.001, component 1)
```

There exists significative differences between rural and urban regions of residence.
We find that urban women are more likely than rural women to use antenatal care

services. The intercept variance, representing the between-community variance for households has decreased slightly from $0,93$ to $0,80$.

Deer data

Source Zuur et al. (2009) et al. (page 326)

The analysis is focused on the relationship between the presence and absence of larvae in deer and the explanatory variables `length` and `sex` of the deer and the `Farm`. Instead of using farm as a fixed effect with 24 levels, we use it as a random effect and the model becomes

$$y_{ij} \sim Bin(1, p_{ij})$$

$$logit(p_{ij}) = \beta_0 + \beta_1 \text{Length}_{ij} + u_i, \qquad u_i \sim \mathcal{N}(0, \sigma_u^2)$$

y_{ij} is 1 if deer j of Farm i is infected and 0 otherwise. The random intercept u_i is assumed to be normally distributed with mean 0 and variance σ_u^2. Using farm as a random intercept has another major advantage. A random intercept model is implicitly introducing the compound symmetrical correlation structure. This implies that the probability of a deer carrying the parasite is correlated to other deer on the same farm.

There are various functions in `R` that can be used for GLMM. The main ones are `glmmPQL` from the `MASS` package, `glmer` from the `lme4` package, and `glmmML` from the `glmmML` package.

```r
deer <- read.table('data/deer.txt',header=TRUE)
deer$Farm <- factor(deer$Farm)
```

```r
plot(deer$Length,deer$infect,col='#00000011', pch=19)
```

```r
deer.glm <- glm(infect~Length+Farm,data=deer,family=binomial)
library(lme4)
# use glmer() to fit GLMM.
#  Farm is a unique identifier for a cluster; does not need to be a factor
deer.glmm <- glmer(infect~Length+(1|Farm),data=deer,family=binomial)
deer.glmm2 <- update(deer.glmm,.~.-Length,data=deer)
anova(deer.glmm2,deer.glmm)
```

```
## Data: deer
## Models:
## deer.glmm2: infect ~ (1 | Farm)
## deer.glmm: infect ~ Length + (1 | Farm)
```
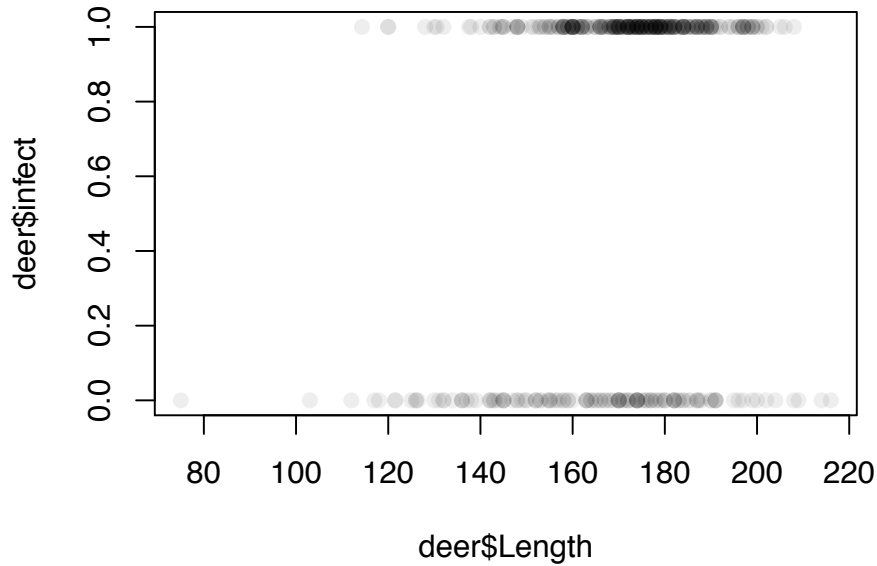
Figure 39: *Deer infection data*

```
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## deer.glmm2  2 464.26 472.46 -230.13   460.26
## deer.glmm   3 436.81 449.12 -215.41   430.81 29.442      1   5.76e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using `glmmPQL` and `glmmML`, note that the results may differ as the estimation procedure is different.

```
library(MASS)
deer2 <- glmmPQL(infect ~ Length, random=~1|Farm,family=binomial("logit"),
                 data=deer)
summary(deer2)

library(glmmML) # similar to glmer
deer3 <- glmmML(infect ~ Length, cluster=Farm,family=binomial("logit"),
                data=deer)
summary(deer3)
```
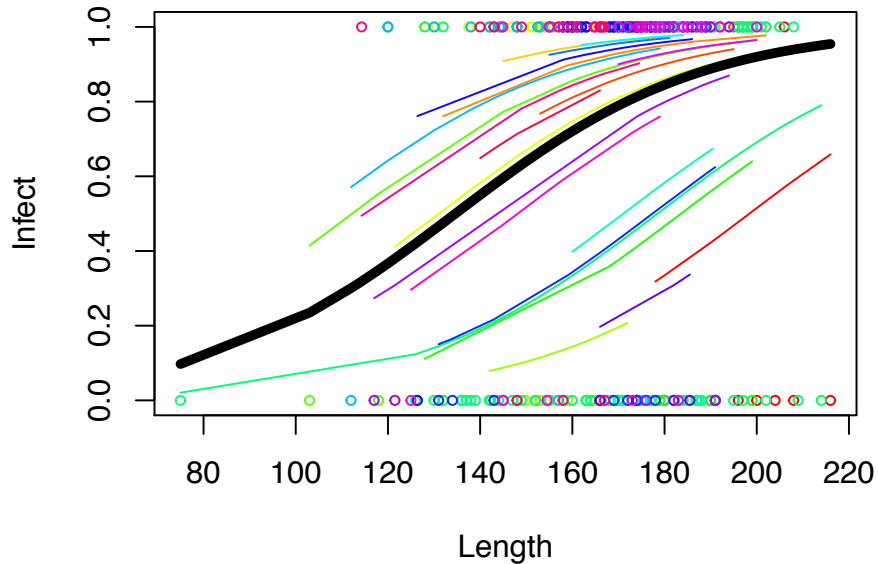
Figure 40: *GLMM predicted probabilities of parasitic infection. The thick line in the middle represents the predicted values for the 'population of farms'*

References

Patterson, H. D., and R. Thompson. 1971. "Recovery of Inter-Block Information When Block Sizes Are Unequal." *Biometrika* 58: 545–54.

Pinheiro, J. C., and D. M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer-Verlag.

Searle, S.R., G. Casella, and C.E. McCulloch. 1992. *Variance Components*. Wiley Series in Probability; Mathematical Statistics.

Singer, J.D., and J.B. Willet. 2003. *Applied Longitudinal Data Analysis*. Oxford University Press.

Zuur, A, E.N. Ieno, N. Walker, A.A. Saveliev, and G.M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R (Statistics for Biology and Health)*. Edited by M. H. Gail, K. Krickeberg, J. M. Samet, A. Tsiatis, and W. Wong. Statistics for Biology and Health. Springer.