Multilevel Models:

An Introduction based on R

Herwig Friedl

Institute of Statistics
Graz University of Technology, Austria

Introduction

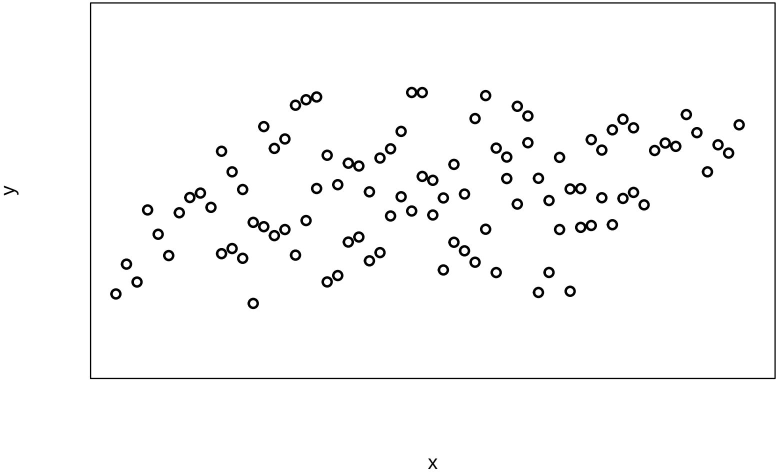
- Based on material from Tom Snijders and Roel Bosker: Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (2nd ed.), SAGE (2012).
- Associated website: http://www.stats.ox.ac.uk/~snijders/
- Interest in **Random Intercept** and **Random Coefficient** Models (Generalized Linear Mixed Models) to relate to **Hierarchical Structures** in the data.
- All models are handled using R functions like `lmer` or `glmer`.

Plan

- Motivation
- Random Intercept Model
- Within-group and Between-group Effects
- Empirical Bayes Estimates
- Random Intercept and Random Slope Model
- Hierarchical Linear Models
- Generalized Linear Mixed Models
- Connections to Social Network Analysis

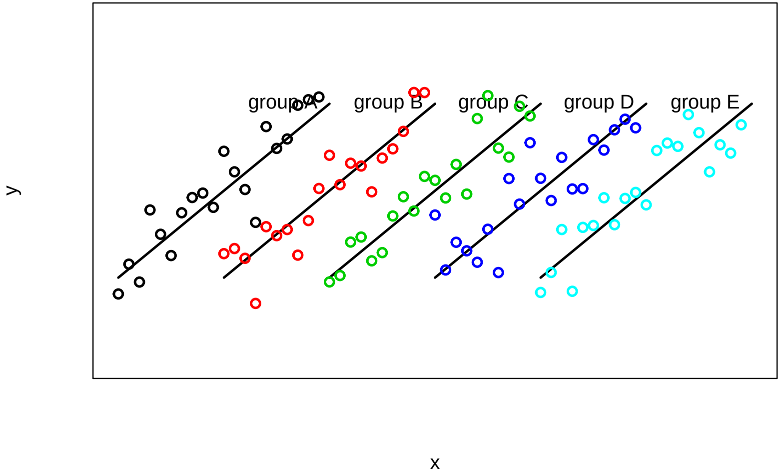
Motivation

Q: Is there any relevant functional relationship between y and x?



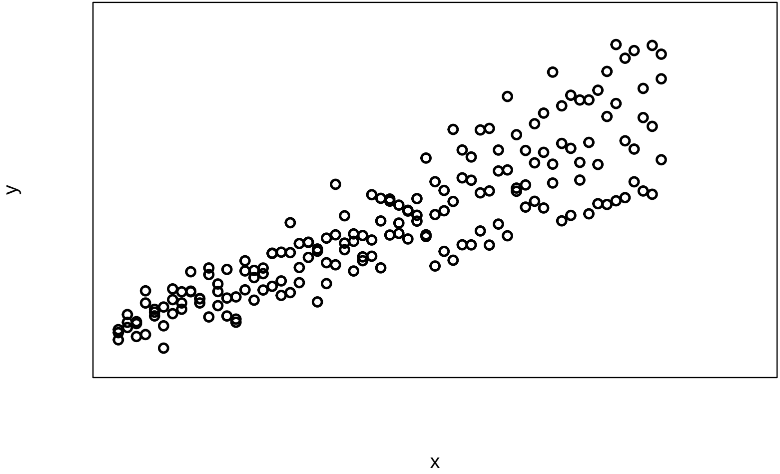
Motivation

A: Yes! There are 5 linear models, one for each group in the data.



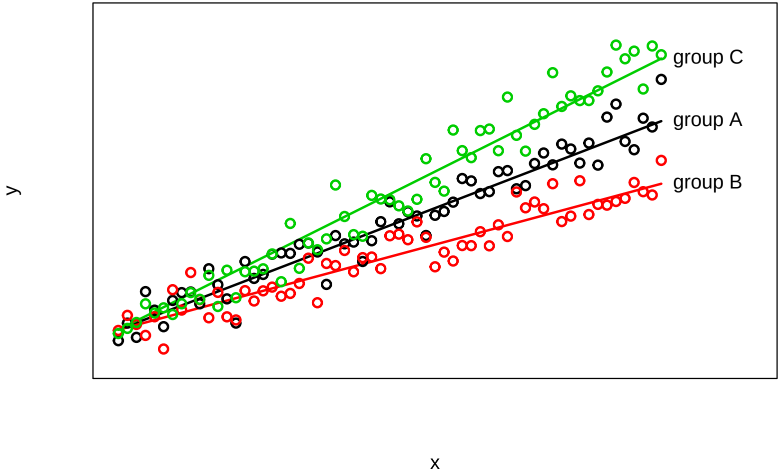
Motivation

Q: Is there constant variance in y?



Motivation

A: Yes! There are 3 homoscedastic groups in the data.



Motivation

How to account for groups or clusters in the data?

- Multilevel analysis is a suitable approach to base the model on social contexts as also on characteristics of individual respondents.
- In a hierarchical (generalized) linear model the dependent variable represents the lowest level (level one).
- Aggregates of level-one variables can serve as explanatory aspects for the second level.
- Explanatory variables could be available at any level.
- Repeated measurements, time series or longitudinal data also form such homogeneous groups.
- Especially, groups, and individuals in these groups, of Social Networks can be compared and modeled utilizing multilevel analysis.

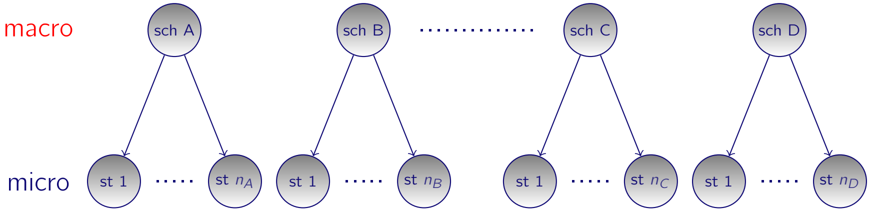
Motivation

Some examples of units at the **macro** and at the **micro** level:

macro-level	micro-level
schools	teachers
classes	pupils
neighborhoods	families
districts	voters
firms	departments
departments	employees
families	children
doctors	patients
interviewers	respondents
judges	suspects
subjects	measurements

Motivation

Two-level models (e.g. students from various schools):
micro-level **student** (level 1) and macro-level **school** (level 2):



Motivation

Arguments in favor of multilevel models (and not to use ordinary least squares regression) in case of multilevel data:

- Relevant effects are often not recognized because they seem to be irrelevant.
- Standard errors and tests conclusions could be sometimes wrong.

Random Intercept Model

- Let i indicate the level-one unit (e.g. individual) and j the level-two unit (e.g. group).
- For individual i in group j, let y_{ij} be the response variable and $\mathbf{x}_{ij}$ the associated vector of explanatory variables at level one.
- For group j, let $\mathbf{z}_j$ be the vector of explanatory variables at level two and denote the size of group j by n_j.

An overall SLR that fully ignores the group structure would be:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$$

Group-dependent SLRs

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

Random Intercept Model

Thus, there are two kinds of **fixed effects** regression models:

1. models in which the group structure is fully ignored,
2. models with fixed effects for the groups, i.e. β_{0j} and β_{1j} are fixed group-specific parameters.

In a **random intercept** model, the intercepts β_{0j} are random variables and represent random differences between the groups

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

where β_{0j} denotes the average intercept γ_{00} plus the group-dependent deviation u_{0j}, i.e.

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Here, we only have one slope β_1, that is common to all groups.

Random Intercept Model

Denote the constant slope parameter β_1 by γ_{10}, then we get

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + \epsilon_{ij}$$

In this **random intercept model**, we additionally assume that

- u_{0j} are independent random variables,
- $\mathrm{E}(u_{0j}) = 0$ and $\mathrm{var}(u_{0j}) = \tau_0^2$,
- they are a simple random sample from a normal population, i.e
$$u_{0j} \overset{iid}{\sim} \mathrm{Normal}(0, \tau_0^2)$$

We are not interested in all individual values of these random effects, but only in their variance τ_0^2.

Random Intercept Model

Arguments for choosing fixed (F) or random (R) intercepts (group indicators):

- If groups are unique entities and inference should focus on these groups: F.
  This often is the case with a small number of groups.
- If groups are regarded as a random sample from a (perhaps hypothetical) population and inference should focus on this population: R.
  This often is the case with a large number of groups.
- If group effects u_{0j} (etc.) are not normally distributed, R could be risky (or use more complicated multilevel models).

Random Intercept Model

Let us start with a random intercept model without explanatory variables:

$$y_{ij} = \gamma_{00} + u_{0j} + \epsilon_{ij}$$

Variance decomposition (u_{0j} and ϵ_{ij} are independent):

$$\text{var}(y_{ij}) = \text{var}(u_{0j}) + \text{var}(\epsilon_{ij}) = \tau_0^2 + \sigma^2$$

Covariance between two responses ($i \neq i'$) in the same group j is

$$\text{cov}(y_{ij}, y_{i'j}) = \text{var}(u_{0j}) = \tau_0^2$$

giving the **intraclass correlation** coefficient

$$\rho(y_{ij}, y_{i'j}) = \frac{\text{cov}(y_{ij}, y_{i'j})}{\sqrt{\text{var}(y_{ij})\,\text{var}(y_{i'j})}} = \frac{\tau_0^2}{\tau_0^2 + \sigma^2}$$

Random Intercept Model

Example: 211 schools in the Netherlands with 3758 pupils (age about 11 years) in elementary schools. Pupils from 1 class are considered. The nesting structure is **students within classes**. The response variable is the pupils result in a language test.

```
> library(lme4)
> summary(lmer(langPOST~(1|schoolnr),data=mlbook_red,REML=FALSE)

Random effects:
 Groups   Name        Variance Std.Dev.
 schoolnr (Intercept) 18.13    4.257
 Residual             62.85    7.928
Number of obs: 3758, groups:  schoolnr, 211

Fixed effects:
            Estimate Std. Error t value
(Intercept)  41.0046     0.3249   126.2
```

Random Intercept Model

Interpretation of these results:

- The (fixed average) intercept is estimated by $\hat{\gamma}_{00} = 41.0$ with standard error $se(\hat{\gamma}_{00}) = 0.3$. Thus, the population from which the y_{ij} are from is normal with mean 41 and standard deviation $\sqrt{18.13 + 62.85} = 9.0$
- The level-two variance (schools variability) is estimated by $\hat{\tau}_0^2 = 18.1$ (or the standard deviation is $\hat{\tau}_0 = 4.3$). Thus, the population from which the random intercepts are drawn is a Normal$(41.0, 18.1)$.
- The level-one variance (students language test scores variability) is estimated by $\hat{\sigma}^2 = 62.85$ (or the standard deviation is $\hat{\sigma} = 7.9$).
- We estimate the intraclass correlation as

$$\hat{\rho} = \frac{18.13}{18.13 + 62.85} = 0.22$$

Random Intercept Model

In a next step we extend this model and also allow for fixed effects of explanatory variables, i.e.

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + \epsilon_{ij}$$

In what follows, x relates to the centered verbal IQ score.

```
> summary(lmer(langPOST~IQ_verb+(1|schoolnr), data=mlbook_red,
+              REML=FALSE)

Random effects:
 Groups   Name        Variance Std.Dev.
 schoolnr (Intercept)  9.845   3.138
 Residual             40.469   6.362
Number of obs: 3758, groups:  schoolnr, 211

Fixed effects:
            Estimate Std. Error t value
(Intercept) 41.05488    0.24339  168.68
IQ_verb      2.50744    0.05438   46.11
```

Random Intercept Model

How does this compare with a SLR not accounting for the multilevel structure induced by schools, i.e.

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \epsilon_{ij}$$

We fit this model and get

```
> summary(lm(langPOST ~ IQ_verb, data = mlbook_red))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 41.29584    0.11517  358.56   <2e-16 ***
IQ_verb      2.65126    0.05643   46.98   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.059 on 3756 degrees of freedom
Multiple R-squared:  0.3702,    Adjusted R-squared:   0.37
F-statistic:  2207 on 1 and 3756 DF,  p-value: < 2.2e-16
```

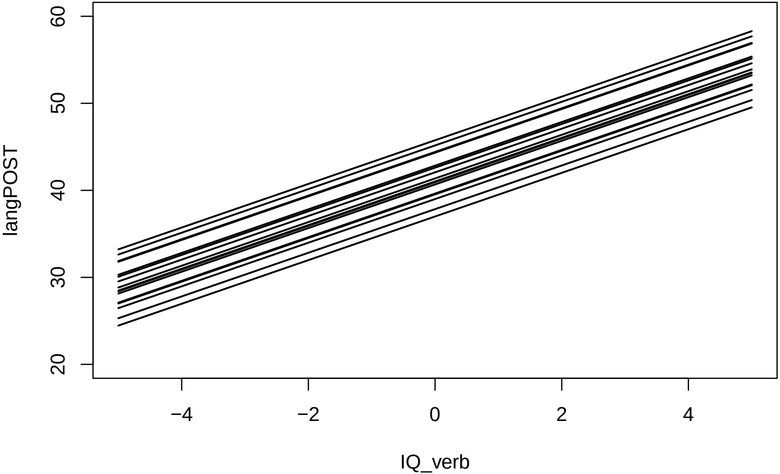
Random Intercept Model

Comparing the results from the random intercept model and from the SLR:

- The random intercept model contains the fixed effects γ_{00} and γ_{10} (as also the SLR) and the variance components σ^2 and τ_0^2 from the random effects. The SLR assumes that $\tau_0^2 = 0$.
- The multilevel model has more structure and accounts for the dependence of responses from the same school.
- The numerical results are surprisingly very similar.

Random Intercept Model

15 randomly chosen models with $u_{0j} \overset{iid}{\sim} \text{Normal}(0, 9.8)$:



Random Intercept Model

Several explanatory variables:

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{1,ij} + \cdots + \gamma_{p0}x_{p,ij} + \gamma_{01}z_{1j} + \cdots + \gamma_{0q}z_{qj} + u_{0j} + \epsilon_{ij}$$

Included are

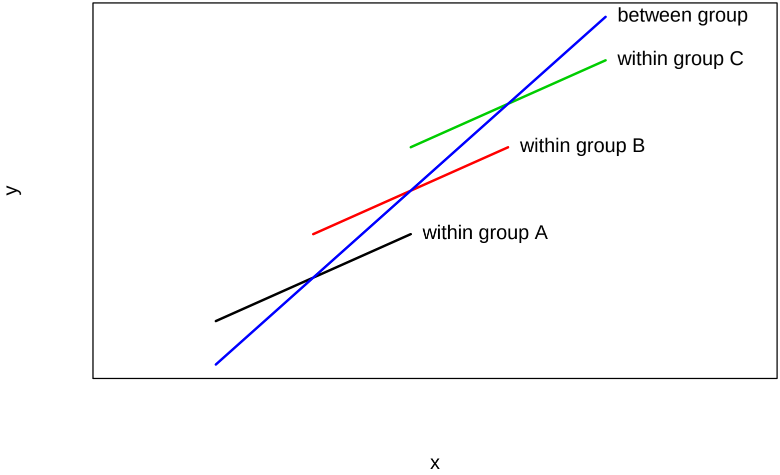
- p level-one explanatory variables $x_{1,ij}, \ldots, x_{p,ij}$ associated with individual i in group j.
- q level-two explanatory variables $z_{1j}, \ldots, x_{qj}$ associated with group j.

Difference between **within-group** and **between-group** regression:

- The within-group regression coefficient expresses the effect of the explanatory variable within a given group.
- The between-group regression coefficient expresses the effect of the group mean of the explanatory variable on the group mean of the response variable.

Random Intercept Model

Difference between **within-group** and **between-group** regression:



Random Intercept Model

Example: pocket money for children in families.

- This will depend on the child's age as also on the average age of the children in the family.
- The within-group regression coefficient measures the effect of age differences within a given family
- The between-group regression coefficient measures the effect of average age on the average pocket money received by the children in the family.

Random Intercept Model

Example: pocket money for children in families.
Denote age of child i in family j by x_{ij}, and the average age of all children in family j by $z_j = \overline{x}_{\bullet j}$. In the model

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + \epsilon_{ij}$$

the within-group and between-group coefficient are forced to be equal. If we add z_j as additional explanatory variable, we obtain

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} \overline{x}_{\bullet j} + u_{0j} + \epsilon_{ij} \\ &= (\gamma_{00} + \gamma_{01} \overline{x}_{\bullet j} + u_{0j}) + \gamma_{10} x_{ij} + \epsilon_{ij} \end{aligned}$$

resulting in the **within-group** j regression line

$$\mathrm{E}(y_{ij}) = \gamma_{00} + \gamma_{01} \overline{x}_{\bullet j} + \gamma_{10} x_{ij}$$

Random Intercept Model

Example: pocket money for children in families.

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} \overline{x}_{\bullet j} + u_{0j} + \epsilon_{ij}$$

Averaging this model over all elements in group j gives

$$\begin{aligned}\overline{y}_{\bullet j} &= \gamma_{00} + \gamma_{10} \overline{x}_{\bullet j} + \gamma_{01} \overline{x}_{\bullet j} + u_{0j} + \overline{\epsilon}_{\bullet j} \\ &= \gamma_{00} + (\gamma_{10} + \gamma_{01}) \overline{x}_{\bullet j} + u_{0j} + \overline{\epsilon}_{\bullet j}\end{aligned}$$

resulting in the **between-group** regression line

$$\mathrm{E}(\overline{y}_{\bullet j}) = \gamma_{00} + (\gamma_{10} + \gamma_{01}) \overline{x}_{\bullet j}$$

with regression coefficient $\gamma_{10} + \gamma_{01}$.

Random Intercept Model

The average IQ of all pupils in school j is already contained in the variable `sch_iqv`.

```
> mlmod <- lmer(langPOST ~ IQ_verb + sch_iqv + (1|schoolnr),
+               data = mlbook_red, REML = FALSE)
> summary(mlmod)

Random effects:
 Groups   Name        Variance Std.Dev.
 schoolnr (Intercept)  8.68    2.946
 Residual             40.43    6.358
Number of obs: 3758, groups:  schoolnr, 211

Fixed effects:
            Estimate Std. Error t value
(Intercept) 41.11378    0.23181  177.36
IQ_verb      2.45361    0.05549   44.22
sch_iqv      1.31242    0.26160    5.02
```

Random Intercept Model

The parameters of the random part of the model and the estimated intercept variance are in

```
> VarCorr(mlmod)
 Groups   Name        Std.Dev.
 schoolnr (Intercept) 2.9461
 Residual             6.3584

> VarCorr(mlmod)$schoolnr[1,1]
[1] 8.679716
```

Random Intercept Model

For other methods for the objects produced by `lmer`, see

```
> methods(class="merMod")
 [1] anova          as.function    coef           confint
 [5] deviance       df.residual    drop1          extractAIC
 [9] family         fitted         fixef          formula
[13] fortify        getL           getME          hatvalues
[17] isGLMM         isLMM          isNLMM         isREML
[21] logLik         model.frame    model.matrix ngrps
[25] nobs           plot           predict        print
[29] profile        qqmath         ranef          refit
[33] refitML        residuals      show           sigma
[37] simulate       summary        terms          update
[41] VarCorr        vcov           weights
```

Random Intercept Model

Denote now the average IQ of pupils in school j by $\overline{x}_{\bullet j}$, then the model states

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} \overline{x}_{\bullet j} + u_{0j} + \epsilon_{ij}$$

with

- within-group coefficient γ_{10} estimated by 2.45,
- between-group coefficient $\gamma_{10} + \gamma_{01}$ estimated by $2.45 + 1.31 = 3.77$, (a pupil with a given IQ is predicted to obtain a higher language test score if (s)he is in a class with higher average IQ score),
- difference between within-group and between-group coefficient is tested by the respected t-value of 5.02 (highly significant).

Random Intercept Model

What can we say about the **latent** (unobservable) random effects u_{0j}?

Consider the empty model

$$\begin{aligned} y_{ij} &= \gamma_{00} + u_{0j} + \epsilon_{ij} \\ &= \beta_{0j} + \epsilon_{ij} \end{aligned}$$

Since these are no parameters we cannot estimate them.

However, we are able to **predict** these quantities by using the **Empirical Bayes** method.

Random Intercept Model

$$y_{ij} = \gamma_{00} + u_{0j} + \epsilon_{ij} = \beta_{0j} + \epsilon_{ij}$$

We started with the prior model $u_{0j} \overset{iid}{\sim} \text{Normal}(0, \tau_0^2)$
Then we took a sample $y_{1j}, \ldots, y_{n_j j}$ from the jth group assuming that the conditional model $y_{ij}|u_{0j} \overset{ind}{\sim} \text{Normal}(\gamma_{00} + u_{0j}, \sigma^2)$ holds.
If we only use group j then β_{0j} would be estimated by

$$\hat{\beta}_{0j} = \overline{y}_{\bullet j}$$

Using the entire sample we would estimate the population mean γ_{00} by the overall mean, i.e.

$$\hat{\gamma}_{00} = \overline{y}_{\bullet\bullet} = \frac{1}{\sum_j n_j} \sum_{j=1}^{N} \sum_{i=1}^{n_j} y_{ij}$$

Random Intercept Model

$$y_{ij} = \gamma_{00} + u_{0j} + \epsilon_{ij} = \beta_{0j} + \epsilon_{ij}$$

Now combine these two sources of information using a weighted average. This results in the **empirical Bayes** (posterior mean) estimator

$$\hat{\beta}_{0j}^{EB} = \lambda_j \hat{\beta}_{0j} + (1 - \lambda_j)\hat{\gamma}_{00}$$

with optimal weights

$$\lambda_j = \frac{\tau_0^2}{\tau_0^2 + \sigma^2 / n_j}$$

The weight λ_j somehow evaluates the **reliability** of the jth group mean $\hat{\beta}_{0j} = \overline{y}_{\bullet j}$ as an estimator of the true mean $\gamma_{00} + u_{0j}$.
If explanatory variables are in the model, the same principle can be applied.

Random Intercept Model

The ratio of the 2 weights

$$\frac{\lambda_j}{1-\lambda_j} = \frac{\frac{\tau_0^2}{\tau_0^2+\sigma^2/n_j}}{\frac{\sigma^2/n_j}{\tau_0^2+\sigma^2/n_j}} = \frac{\tau_0^2}{\sigma^2/n_j}$$

is the ratio of the true variance τ_0^2 to the error variance σ^2/n_j.

Since these parameters are usually unknown, we substitute their estimates in order to calculate $\hat{\beta}_{0j}^{EB}$.

These posterior means can be used to detect groups with unexpected high/low values of their response (given their predictors).

Random Intercept Model

Model: Denote the average IQ of pupils in school j by $\overline{x}_{\bullet j}$, then

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} \overline{x}_{\bullet j} + u_{0j} + \epsilon_{ij}$$

Q: How should parents choose a school for their kids?
A: Good schools are those where the students on average achieve more than expected on the basis of their IQ.

The level-two residual u_{j0} contains this information and has to be estimated from the data. Comparison is sometimes based on associated confidence intervals based on comparative (posterior) standard errors

$$se^c(\hat{u}_{0j}^{EB}) = se(\hat{u}_{0j}^{EB} - u_{0j})$$

or on diagnostic standard errors

$$se^d(\hat{u}_{0j}^{EB}) = se(\hat{u}_{0j}^{EB})$$

Random Intercept Model

Again: comparative standard errors

$$se^c(\hat{u}_{0j}^{EB}) = se(\hat{u}_{0j}^{EB} - u_{0j})$$

and diagnostic standard errors

$$se^d(\hat{u}_{0j}^{EB}) = se(\hat{u}_{0j}^{EB})$$

An interesting property is that the sum of both variances equals the random intercept variance, i.e.

$$\mathrm{var}(\hat{u}_{0j}^{EB} - u_{0j}) + \mathrm{var}(\hat{u}_{0j}^{EB}) = \tau_0^2.$$

Thus,

$$\mathrm{var}(\hat{u}_{0j}^{EB}) = \tau_0^2 - \mathrm{var}(\hat{u}_{0j}^{EB} - u_{0j}).$$

Random Intercept Model

Conditional means (and variances) of the random effects are obtained as follows (`ranef` stands for random effects)

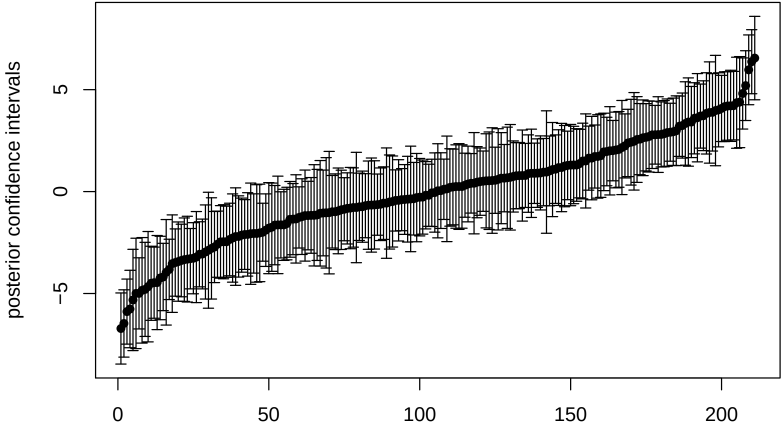
```
> pmu <- ranef(mlmod, condVar=TRUE)
> # posterior means
> postmean <- pmu$schoolnr[,1]
> # comparative (posterior) variances
> postvar <-  attr(pmu$schoolnr,'postVar')[1,1,]
> # comparative standard deviations
> postsd <- sqrt(postvar)
> # diagnostic variances
> diagvar <- VarCorr(mlmod)$schoolnr[1,1] - postvar

> # bounds of 95% comparative intervals
> # (testing equality of level-two residuals)
> lower <- postmean - 1.39*postsd
> upper <- postmean + 1.39*postsd
```

Random Intercept Model

Caterpillar plot (comparative 95 % confidence intervals for the random effects)

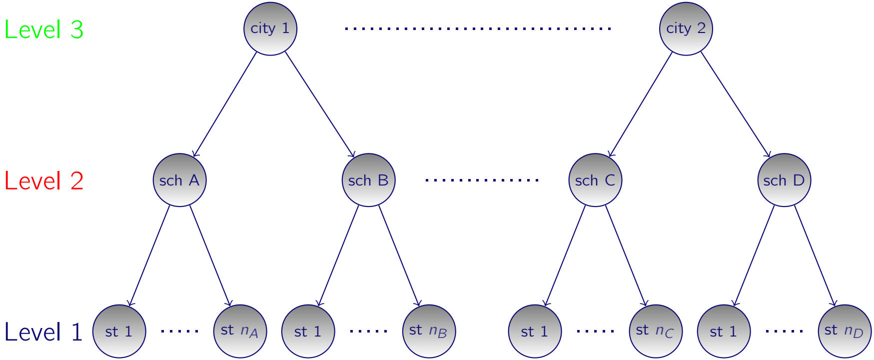
```
> perm <- order(postmean, lower, upper)
> pm_sort <- postmean[perm]
> upper_sort <- upper[perm]
> lower_sort <- lower[perm]

> library(Hmisc)
> errbar(1:211, pm_sort, upper_sort, lower_sort)
```

Random Intercept Model



Hierarchical Linear Model

Multilevel or Hierarchical Models:



Hierarchical Linear Model

In addition to the intercept, also the effect of x could **randomly depend** on the group, i.e. in the model

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

also the slope β_{1j} could have a random part. Thus, we have

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Substitution in the model results in

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + u_{1j} x_{ij} + \epsilon_{ij}$$

Hierarchical Linear Model

Random intercept and random slope model:

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + u_{1j} x_{ij} + \epsilon_{ij}$$

Assume that the random effects (u_{0j}, u_{1j}) are independent pairs across j from a bivariate normal with zero means $(0, 0)$ and

$$\begin{aligned} \text{var}(u_{0j}) &= \tau_{00} = \tau_0^2 \\ \text{var}(u_{1j}) &= \tau_{11} = \tau_1^2 \\ \text{cov}(u_{0j}, u_{1j}) &= \tau_{01} \end{aligned}$$

Again, the (u_{0j}, u_{1j}) are not individual parameters, but their variances and covariance are of interest.
This is again a linear model for the mean, and a parameterized covariance within groups with independence between groups.

Hierarchical Linear Model

Random slope model for the language scores: denote the average IQ of all pupils in school j by $\overline{x}_{\bullet j}$, then the model now states

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}x_{\bullet j} + u_{0j} + u_{1j}x_{ij} + \epsilon_{ij}$$

```
> ransl <- lmer(langPOST ~ IQ_verb + sch_iqv
+               + (IQ_verb|schoolnr), data = mlbook_red,
+               REML = FALSE)
> summary(ransl)

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 schoolnr (Intercept)  8.877   2.9795
          IQ_verb      0.195   0.4416   -0.63
 Residual             39.685   6.2996
Number of obs: 3758, groups:  schoolnr, 211
```

Thus, $\widehat{\text{var}}(u_{0j}) = \hat{\tau}_0^2 = 8.88$, $\widehat{\text{var}}(u_{1j}) = \hat{\tau}_1^2 = 0.19$, and $\widehat{\text{var}}(\epsilon_{ij}) = \hat{\sigma}^2 = 39.68$,

Hierarchical Linear Model

Second part of the R output:

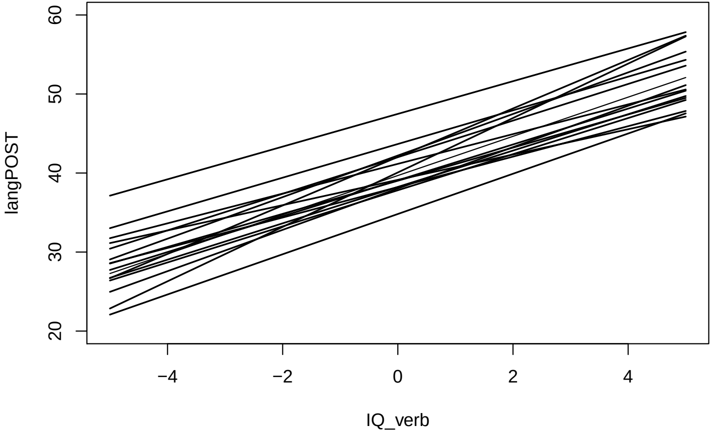
```
Fixed effects:
            Estimate Std. Error t value
(Intercept)  41.1275     0.2336  176.04
IQ_verb       2.4797     0.0643   38.57
sch_iqv       1.0285     0.2622    3.92

Correlation of Fixed Effects:
        (Intr) IQ_vrb
IQ_verb -0.279
sch_iqv -0.003 -0.188
```

Estimated model:

$$\hat{\mathrm{E}}(y_{ij}|u_{0j}, u_{1j}) = 41.13 + 2.48x_{ij} + 1.03x_{\bullet j} + u_{0j} + u_{1j}x_{ij}$$

Hierarchical Linear Model

15 randomly chosen models with $u_{0j} \overset{iid}{\sim} \text{Normal}(0, 8.9)$ and $u_{1j} \overset{iid}{\sim} \text{Normal}(0.0.2)$ for school $j = 1$ with $\overline{IQ}_j = -1.4$:



Hierarchical Linear Model

General formulation of a two-level model:

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\gamma} + \mathbf{Z}_j \mathbf{u}_j + \boldsymbol{\epsilon}_j$$

with

$$\begin{bmatrix} \boldsymbol{\epsilon}_j \\ \mathbf{u}_j \end{bmatrix} \overset{ind}{\sim} \text{Normal}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_j & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix} \right)$$

Often we simplify and consider a model with $\boldsymbol{\Sigma}_j = \sigma^2 \mathbf{I}$ but also other structures are possible (e.g. time series).
The above model is equivalent to

$$\mathbf{y}_j \sim \text{Normal}\left(\mathbf{X}_j \boldsymbol{\gamma}, \mathbf{Z}_j \boldsymbol{\Omega} \mathbf{Z}_j^\top + \boldsymbol{\Sigma}_j \right)$$

a special case of a **linear mixed model**.

MLE versus REML

- Because of the unknown mean parameter μ, even for a simple random sample, the MLE $\hat{\sigma}^2$ of the variance is biased, where
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$$
- Similar problems occur when estimating the variance components in linear mixed models.
- The **RE**stricted/**RE**sidual **M**aximum **L**ikelihood (**REML**) estimator tries to solve this problem.

We have just shown that marginally

$$\mathbf{y}_j \sim \text{Normal}(\mathbf{X}_j\boldsymbol{\beta}, \mathbf{V}_j = \mathbf{Z}_j\boldsymbol{\Omega}\mathbf{Z}_j^\top + \boldsymbol{\Sigma}_j)$$

Idea: get rid of all unknown parameters $\boldsymbol{\beta}$ in the mean model.

MLE versus REML

$$\mathbf{y}_j \sim \text{Normal}(\mathbf{X}_j\boldsymbol{\beta}, \mathbf{V}_j = \mathbf{Z}_j\boldsymbol{\Omega}\mathbf{Z}_j^\top + \boldsymbol{\Sigma}_j)$$

Q: How to get rid of all unknown parameters in the mean model?

Consider a linear combination of the response $\mathbf{k}^\top\mathbf{y}_j$, such that $\mathbf{k}^\top\mathbf{X}_j = \mathbf{0}$. Then

$$\mathbf{k}^\top\mathbf{y}_j \sim \text{Normal}(\mathbf{0}, \mathbf{k}^\top\mathbf{V}_j\mathbf{k})$$

Maximize the likelihood of $\mathbf{k}^\top\mathbf{y}_j$ with no fixed effects included.

This gives estimates of the random effects parameter.

Now it is easy to get estimators of fixed effects parameters in $\mathbf{V}$.

Generally, REML estimates have smaller bias

Generalized Linear Mixed Models

Extend the model to the linear exponential family, e.g. student i in university j takes an exam and the result can be modeled as

$$\Pr(y_{ij} = \text{"success"}) = \text{logit}^{-1}(\mathbf{x}_{ij}^\top \boldsymbol{\gamma} + \mathbf{z}_j^\top \mathbf{u}_j)$$

again with $\mathbf{u}_j \overset{ind}{\sim} \text{Normal}(\mathbf{0}, \boldsymbol{\Omega})$.
Thus, assume that conditional on the random effects, the response distribution is a linear exponential family with pdf

$$f(\mathbf{y}|\mathbf{u}; \boldsymbol{\gamma})$$

and the random effects are from a zero mean normal distribution with pdf

$$f(\mathbf{u}; \boldsymbol{\Omega})$$

The likelihood function corresponds to the marginal pdf which is

$$f(\mathbf{y}; \boldsymbol{\gamma}, \boldsymbol{\Omega}) = \int f(\mathbf{y}|\mathbf{u}; \boldsymbol{\gamma}) f(\mathbf{u}; \boldsymbol{\Omega}) d\mathbf{u}$$

Generalized Linear Mixed Models

The MLEs $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\Omega}}$ jointly maximize this integral, which is

$$\begin{aligned} f(\mathbf{y}; \boldsymbol{\gamma}, \boldsymbol{\Omega}) &= \int f(\mathbf{y}|\mathbf{u}; \boldsymbol{\gamma}) f(\mathbf{u}; \boldsymbol{\Omega}) d\mathbf{u} \\ &= \prod_{j=1}^{N} \int \prod_{i=1}^{n_j} f(y_{ij}|\mathbf{u}_j; \boldsymbol{\gamma}) f(\mathbf{u}_j; \boldsymbol{\Omega}) d\mathbf{u}_j \end{aligned}$$

but very often there does not even exist an explicit form of it.

The normal–normal model discussed before is an exception because this is a **conjugate** pair of distributions.

Laplace or **Gauss-Hermite** approximations can be utilized to approximate the likelihood function above.

Multilevel Logistic Model

Gelman and Hill (2007) consider a **multilevel logistic model** for survey responses y_{ij} that equal 1 for supporters of the Republican candidate (G.W. Bush) and 0 for the Democrat (M. Dukakis) in the presidential election 1988.

Their model is based on the predictors sex and ethnicity (African American or other) as also on the State of the respondent.

$$\Pr(y_{ij} = 1|u_{0j}) = \text{logit}^{-1}(\gamma_{00} + u_{0j} + \gamma_{10}\text{female}_{ij} + \gamma_{20}\text{black}_{ij})$$

with state-specific random intercepts $u_{0j} \overset{iid}{\sim} \text{Normal}(0, \tau_0^2)$.

```
> mean(female)
[1] 0.5886913
> mean(black)
[1] 0.07615139
```

Multilevel Logistic Model

This model is fitted in R now using the function glmer

```
> M1 <- glmer (y ~ black + female + (1|state),
+                   family=binomial(link="logit"))
> display(M1)

            coef.est coef.se
(Intercept)  0.45     0.10
black       -1.74     0.21
female      -0.10     0.10

Error terms:
 Groups   Name        Std.Dev.
 state    (Intercept) 0.41
 Residual             1.00
---
number of obs: 2015, groups: state, 49
AIC = 2666.7, DIC = 2531.5
deviance = 2595.1
```

Multilevel Logistic Model

The average intercept is 0.45 with standard error 0.10, the coefficients for black and female are $-1.74(0.21)$ and $-0.10(0.10)$. Furthermore, $\hat{\tau}_0^2 = 0.41$.

Empirical Bayes estimates of all state-specific intercepts are available by

```
> coef(M1)
$state
    (Intercept)     black      female
1   0.990578098 -1.741612 -0.09704731
3   0.686196961 -1.741612 -0.09704731
4   0.314917122 -1.741612 -0.09704731
5   0.306467230 -1.741612 -0.09704731
:
```

Connecting to Social Network Analysis

Variance components (individual variance within groups and variance between groups) in multilevel models are especially interesting in the social network context (from P.P. Pare):

- interpretation as a measure of sociability of behaviors
- the larger the between group variance the more social is the behavior
- if 100% variance is within group and 0% between groups, the behavior is purely individual
- if 0% variance is within group and 100% between groups, the behavior is purely social (individuals behave in perfect conformity with their own group and all the variation is between groups)
- in reality, there is often a division of the variance within and between groups, but different behaviors can be compared in regard to their level of sociability