Multilevel Models

Longitudinal and Cluster Data

Ariel Alonso Abad

Catholic University of Leuven

Studying changes over time

- Longitudinal data: Studying changes over time.
- Exploratory data analysis.
- Introduction to multilevel models.
- Formulation and interpretation of the models.
- Implementation in R.
- Statistical inference
  - Fixed effects
  - Random effects

Studying changes over time

- Changes over time play pivotal role in science.
- Original ideas
  - ⇒ British astronomer George Biddel Airy 1861.
  - ⇒ Laird and Ware (1982): Life sciences.
  - ⇒ Goldstein (1979): Humanities.
- Computing power and software available in the 1990s.
- Synthesis: Intra and inter individual changes need to be modeled.

Different names similar models

- Individual growth models.
- Random coefficient models.
- Multilevel models.
- Mixed models.
- Hierarchical (linear) models.
- Growth curve models.

Why multilevel?

- Level 1: Changes within individuals.

  $\Rightarrow$ Can we describe the time evolution for each individual with a linear function?

- Level 2: Changes between individuals.

  $\Rightarrow$ Are the individuals different at the beginning of the study?

  $\Rightarrow$ Do they evolve differently over time?

Distinguishing quality

Longitudinal Studies: Repeated measurements over time (Waves)

$\Rightarrow$ Metric: Time, age, weeks since treated...

$\Rightarrow$ Spacing: Equal time intervals?

$\Rightarrow$ Time structure: All individual measured at the same time points?

$\Rightarrow$ Balanced: Same number of measurements for all individuals?

Cross-Sectional Studies: Only one measurement per subject. Nothing can be concluded about time changes.

Effect of early dietary intervention on children IQ

- 103 African American, low income families.
  - 58 early intervention program.
  - 45 control group.
- Evaluated on ages 12, 18, and 24 months.
- Research question: Effect of the early intervention on cognitive performance?

Effect of early dietary intervention on children IQ

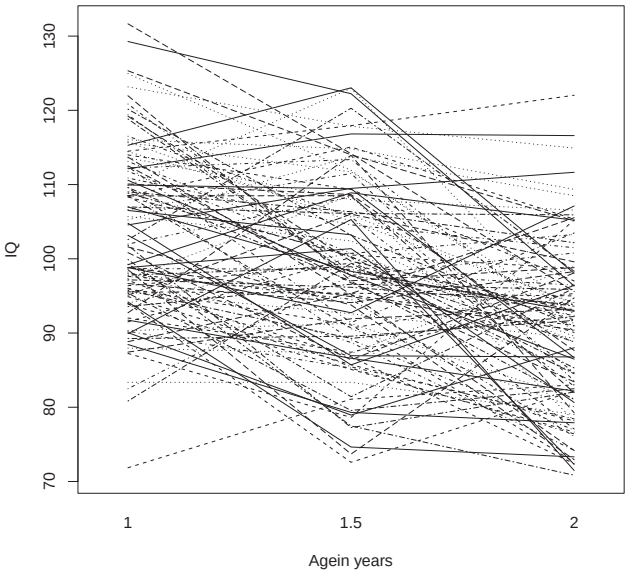
id	program	age	cog
1	1	1.00	106.98
1	1	1.50	98.31
1	1	2.00	92.91
2	1	1.00	108.86
2	1	1.50	100.29
2	1	2.00	85.30
3	1	1.00	112.52
3	1	1.50	96.77
3	1	2.00	83.43
4	1	1.00	90.24
4	1	1.50	85.27
4	1	2.00	76.41
5	1	1.00	105.71
5	1	1.50	102.40
5	1	2.00	88.79
6	1	1.00	93.89
6	1	1.50	85.10
6	1	2.00	76.66
7	1	1.00	109.94
⋮	⋮	⋮	⋮

- Fully balanced: Age=1.0, 1.5 and 2.0 years.
- PROGRAM: 1-intervention, 0-control.
- COG is a nationally normed scale.

Exploratory analysis

- Spaghettiplot: Individual profiles. Points are joined with lines.
- Descriptive tables.
- Box plots.
- Mean plots.
- Individual regressions.

Effect of early intervention: Spaghetti-plot



Spaghettiplot: R code

Let us get started with R:

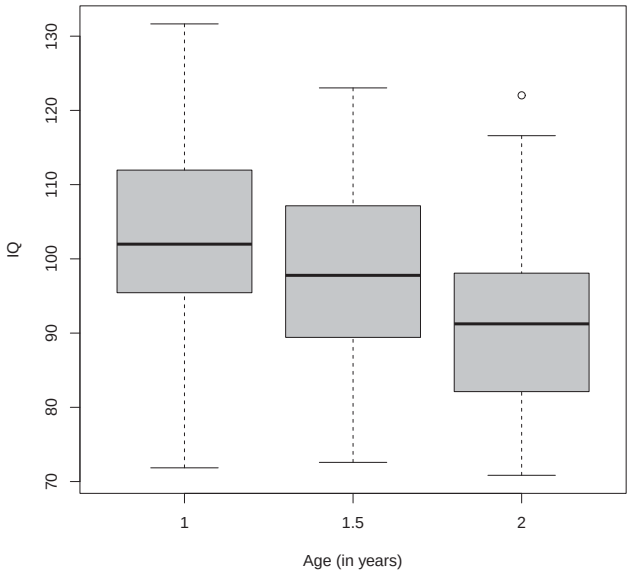
```
> ## Reading in the early.int data
>
> early.int1 <- read.table("earlyint.txt", header=T, sep=",")
> early.int1.table <- xtable(early.int1[1:24,])
> print(early.int1.table)
>
> ## Attach data to the search path
>
> attach(early.int1)
>
> ## Spaghettiplot
>
> n=length(unique(id))
> interaction.plot(age,id,cog, xlab="Agein years", ylab="IQ",
+   legend=F)
>
```


Means per time point and group

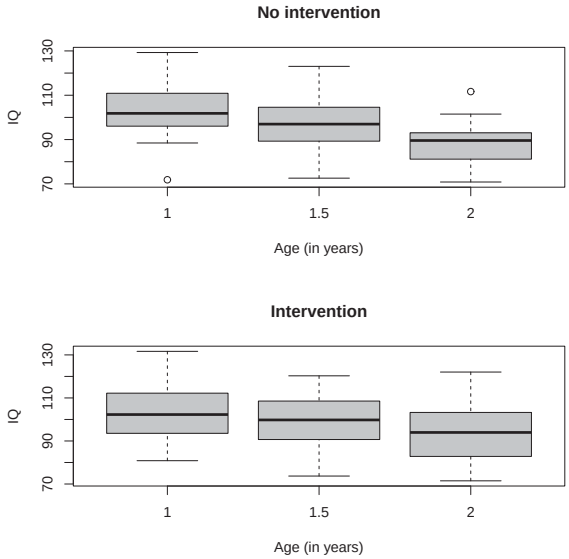
			Program	
			0	1
Age	Resp	Statistics		
1	IQ	n	45	58
		Mean	103.93	102.93
		Sd	11.01	11.78
1.5	IQ	n	45	58
		Mean	96.91	99.18
		Sd	11.93	12.02
2	IQ	n	45	58
		Mean	87.68	92.99
		Sd	9.05	12.13

```
> ## Descriptives
>
> ## Mean:
> early.mean=tapply(cog,list(age,program),mean)
>
> ## Standard deviation:
> early.sd=tapply(cog,list(age,program),sd)
>
> ## Variance:
> early.var=tapply(cog,list(age,program),var)
>
> ## Frequency:
> early.n=table(age,program)
>
```

Boxplot



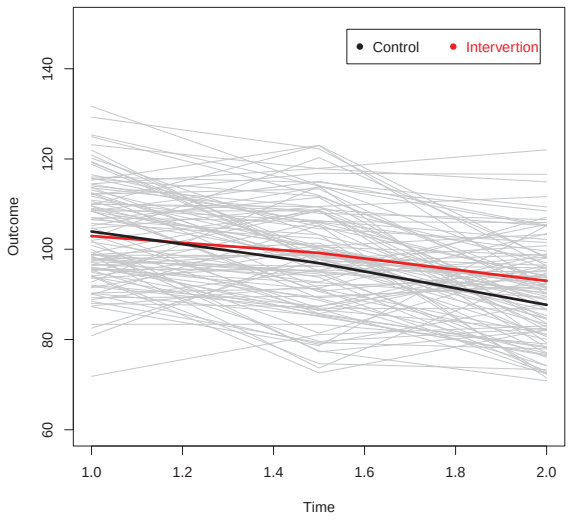
Boxplot per program



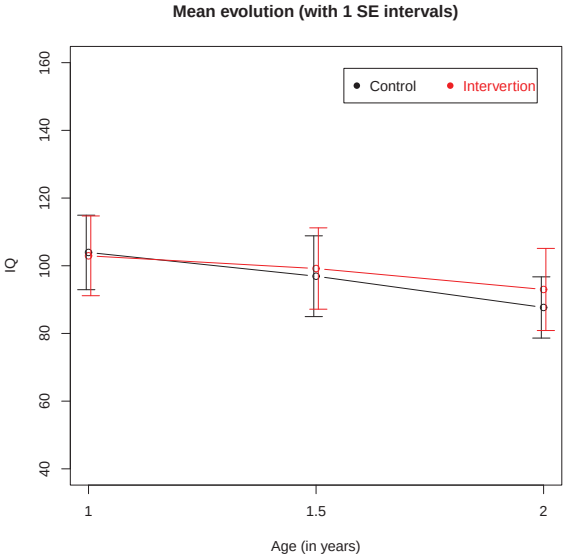
R code

```
> ## Boxplots:
>
> boxplot(cog~age,xlab="Age (in years)",ylab="IQ")
>
> ## Boxplots per program
>
> par(mfrow=c(2,1))
> boxplot(cog[program==0]~age[program==0],main="No intervention",
+ main="No intervention",xlab="Age (in years)",ylab="IQ")
>
> boxplot(cog[program==1]~age[program==1],main="Intervention",
+ main="No intervention",xlab="Age (in years)",ylab="IQ")
>
```


Mean evolution



Mean evolution



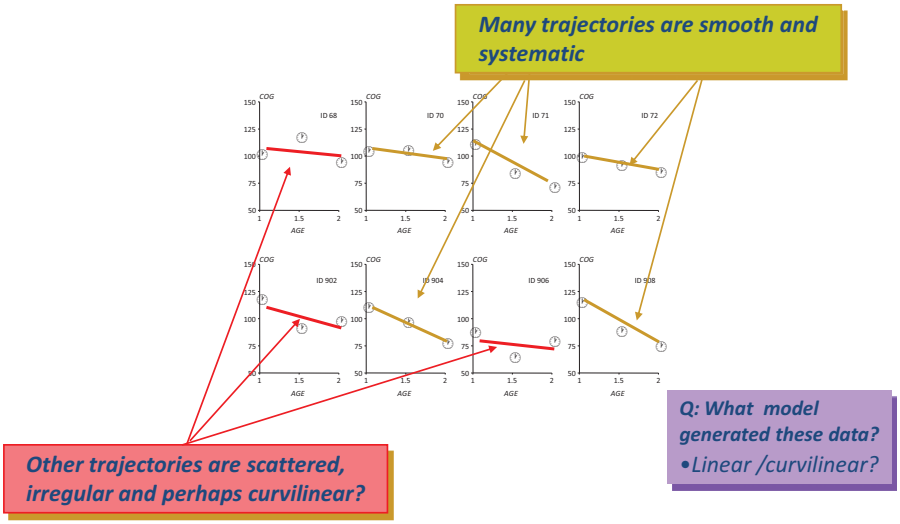
R code

```
####################################################################
#           General function to plot error bars                    #
####################################################################

errbar=function(x,y,height,width,lty=1,col="black")
{arrows(x,y,x,y+height,angle=90,length=width,lty=lty,
col=col)
arrows(x,y,x,y-height,angle=90,length=width,lty=lty,
col=col)}

> ## Plotting mean evolutions
>
> plot(age[id==1],early.mean[,1],type="b",xlim=c(1,2),
+ ylim=c(40,160),xlab="Age (in years)",ylab="IQ",axes=F,
+ main="Mean evolution (with 1 SE intervals)")
> axis(side=1,at=c(1,1.5,2),labels=c(1,1.5,2))
> axis(side=2,at=seq(40,160,20))
>
> box()
> points(age[id==1],early.mean[,2],type="b",col="red")
> errbar(age[id==1]-.005,early.mean[,1],early.sd[,1],.1)
> errbar(age[id==1]+.005,early.mean[,2],early.sd[,2],.1,col="red")
>
```

Correlations: R code

```
> ## Reshaping the data into a wide form
> early.int2 <- reshape(early.int1,
+  timevar = "age", idvar = c("id", "program"), direction = "wide")
> early.int2
>
     id program     cog.1   cog.1.5     cog.2
1     1       1 106.98289  98.31060  92.91342
2     2       1 108.86019 100.29307  85.29502
3     3       1 112.52438  96.76684  83.42649
4     4       1  90.24428  85.27380  76.41052
5     5       1 105.70738 102.39839  88.78872
6     6       1  93.88987  85.09601  76.66209
.........................................
>
> ## Correlation between the IQ scores  at  different ages
> cor(early.int2[,3:5])
>
            cog.1   cog.1.5     cog.2
cog.1   1.0000000 0.5816070 0.3263912
cog.1.5 0.5816070 1.0000000 0.4371109
cog.2   0.3263912 0.4371109 1.0000000
>
```


Linear regression per person



Linear regression per person

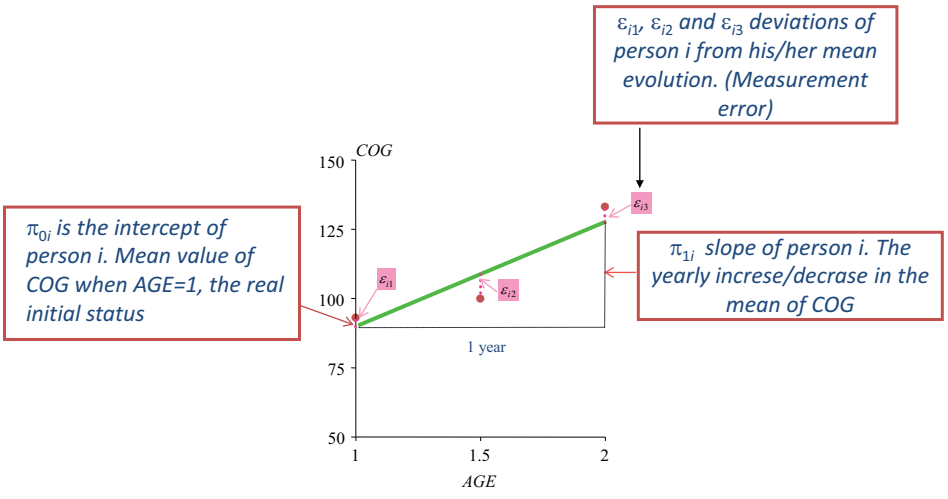
Model

Model for subject i

$$Y_{ij} = \pi_{0i} + \pi_{1i}(Age_{ij} - 1) + \varepsilon_{ij}$$

- Y_{ij} denotes COG for subject i at Age_{ij}.
- π_{0i} intercept for subject i at $Age_{ij} = 1$.
- π_{1i} slope for subject i.
- ε_{ij} error term $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$.

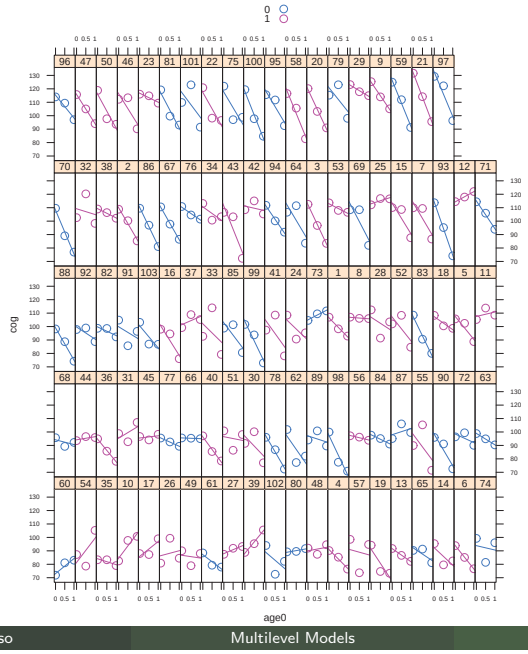
Interpretation of the model



Linear regression per person: Trellis graph

- The aspect ratio of the panels (ratio of the height to the width) chosen according to an algorithm described in Cleveland (1993) to facilitate comparison of slopes
- The effect is to have the slopes of the lines on the page distributed around ± 45, thereby making it easier to detect systematic changes in slopes
- The panels have been ordered (from left to right starting at the bottom row) by increasing intercept
- If there were a correlation between initial status (intercept) and rate of change (slope) then slopes would show an increasing trend (or a decreasing trend) in the left to right, bottom to top ordering.

Linear regression per person



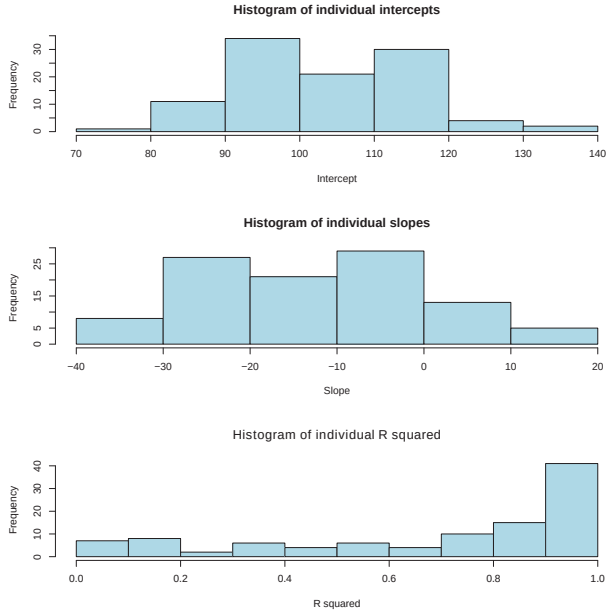
Linear regression per person: R code

```
>## Creating the time variable
>
early.int1$age0<-early.int1$age-1
>
> ## Displaying the linear regression per person
>
> cf<-sapply(early.int1$id, function(x)
+    coef(lm(cog~age0, data=subset(early.int1, id==x))))
>
> Sx<-reorder(early.int1$id, cf[1,])
>
> xyplot(cog ~ age0|Sx,groups=program,data=early.int1,
+ type=c('p','r'),auto.key=T,aspect="xy",
+ par.settings=list(axis.text=list(cex=0.6),
+ fontsize=list(text=8, points=10)),
+ scales=list(
+ x=list(
+ at=c(0,0.5,1),
+ labels=c("0","0.5","1")))
)
>
```


Linear regression per person: R code

```
> ## Linear regression per participant of cog on age
>
> ## Coefficients
> lin.reg.coef <- by(early.int1, early.int1$id,
+          function(data) coef(lm(cog ~ age0, data=data)))
> lin.reg.coef1 <- unlist(lin.reg.coef)
> names(lin.reg.coef1) <- NULL
> lin.reg.coef2=matrix(lin.reg.coef1,length(lin.reg.coef1)/2,2,byrow = TRUE)
>
>## R squared
> lin.reg.r.squared <- by(early.int1, early.int1$id,
+          function(data) summary(lm(cog ~ age, data=data))$r.squared )
lin.reg.r.squared1<- as.vector(unlist(lin.reg.r.squared))
>
>## Histograms
> par(mfrow=c(3,1))
> hist(lin.reg.coef2[,1],xlab="Intercept",col="lightblue",main="Histogram of individual intercepts")
> hist(lin.reg.coef2[,2],xlab="Slope",col="lightblue",main="Histogram of individual slopes")
> hist(lin.reg.r.squared1,xlab="R squared",col="lightblue",main="Histogram of individual R squared")
>
```

Between subject variability



Multilevel models

Level 1

$$Y_{ij} = \boxed{\pi_{0i} + \pi_{1i}(Age_{ij} - 1)} + \boxed{\varepsilon_{ij}}$$

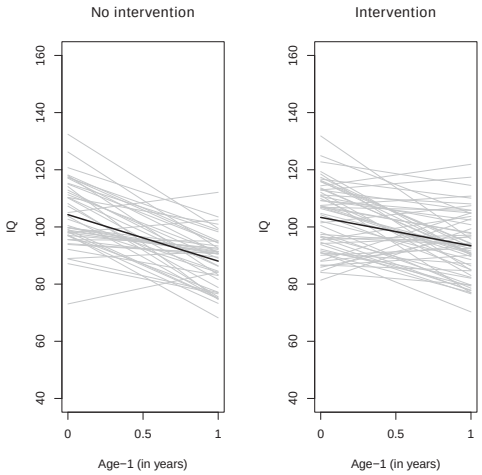
$\Rightarrow$ Structural part of the level 1. How individuals evolve.

$\Rightarrow$ Random part of Level 1. How individuals deviate from their own evolution.

$\Rightarrow$ Why do π_{0i} and π_{1i} vary?

$\Rightarrow$ Is due to the effect of the intervention program?

Multilevel models



⇒ Program is not the entire story.

⇒ How can we handle the unexplained variability?

Linear regression per person and group: R code

```
> ## Plotting individual regression lines per group
>
> reg.coef=cbind(lin.reg.coef2, early.int1[early.int1$age==1,]$program)
>
> mean.int<-tapply(reg.coef[,1],reg.coef[,3],mean)
> mean.slope<-tapply(reg.coef[,2],reg.coef[,3],mean)
>
> par(mfrow=c(1,2))
> plot(age,cog,type="n",xlim=c(1,2),ylim=c(40,160),main="No intervention",
+      xlab="Age-1 (in years)",ylab="IQ",axes=F)
> axis(side=1,at=c(1,1.5,2),labels=c(1,1.5,2))
> axis(side=2,at=seq(40,160,20))
> box()
> for (i in 1:103)
+ {if (reg.coef[i,3]==0)
+ {curve(cbind(1,x)%*%reg.coef[i,1:2],add=T,col="gray")}}
> curve(cbind(1,x)%*%c(mean.int[1],mean.slope[1]),add=T,lwd=2)
>
> plot(age,cog,type="n",xlim=c(1,2),ylim=c(40,160),main="Intervention",
+      xlab="Age-1 (in years)",ylab="IQ",axes=F)
> axis(side=1,at=c(1,1.5,2),labels=c(1,1.5,2))
> axis(side=2,at=seq(40,160,20))
> box()
> for (i in 1:103)
+ {if (reg.coef[i,3]==1)
+ {curve(cbind(1,x)%*%reg.coef[i,1:2],add=T,col="gray")}}
> curve(cbind(1,x)%*%c(mean.int[2],mean.slope[2]),add=T,lwd=2)
```

Multilevel models

Level 1

$$Y_{ij} = \pi_{0i} + \pi_{1i}(Age_{ij} - 1) + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2_{\varepsilon})$$

Level 2

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} PROG_i + b_{0i} & \text{explaining the intercept} \\ \pi_{1i} = \gamma_{10} + \gamma_{11} PROG_i + b_{1i} & \text{explaining the slope} \end{cases}$$

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_0 & \sigma_{01} \\ \sigma_{01} & \sigma^2_1 \end{pmatrix} \right)$$

Parameters interpretation

Symbol	Definition
σ^2_0	Level 2 residual variance in true intercept π_{0i} across all individuals in the population, after controlling for program participation
σ^2_1	Level 2 residual variance in true slope π_{1i} across all individuals in the population, after controlling for program participation
σ_{01}	Level 2 residual covariance between true intercept π_{0i} and slope π_{1i} across all individuals in the population, after controlling for program participation

Explaining variation:

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} PROG_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{11} PROG_i + b_{1i} \end{cases}$$

Control Group $PROG_i = 0$

$$\begin{cases} \pi_{0i} = \gamma_{00} + b_{0i} \\ \pi_{1i} = \gamma_{10} + b_{1i} \end{cases}$$

Intervention Group $PROG_i = 1$

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{11} + b_{1i} \end{cases}$$

Parameters interpretation

Symbol	Definition
σ_0^2	Level 2 residual variance in true intercept π_{0i} across all individuals in the population, after controlling for program participation
σ_1^2	Level 2 residual variance in true slope π_{1i} across all individuals in the population, after controlling for program participation
σ_{01}	Level 2 residual covariance between true intercept π_{0i} and slope π_{1i} across all individuals in the population, after controlling for program participation

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix} \right)$$

Final model

Hierarchical model

$$\begin{cases} Y_{ij} = & \pi_{0i} + \pi_{1i}(Age_{ij} - 1) + \varepsilon_{ij} \\ \pi_{0i} = & \gamma_{00} + \gamma_{01} PROG_i + b_{0i} \\ \pi_{1i} = & \gamma_{10} + \gamma_{11} PROG_i + b_{1i} \end{cases}$$

Distributional assumptions

$$\begin{cases} \varepsilon_{ij} \sim & N(0, \sigma_\varepsilon^2) \\ \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim & N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix} \right) \end{cases}$$

One single model

Model

$$Y_{ij} = \overbrace{\gamma_{00} + \gamma_{01} PROG_i + \gamma_{10}(Age_{ij} - 1) + \gamma_{11} PROG_i (Age_{ij} - 1)}^{\text{Fixed effects}} +$$
$$b_{0i} + b_{1i}(Age_{ij} - 1) + \rightarrow \text{Random effects}$$
$$\varepsilon_{ij} \rightarrow \text{Error}$$

Distributional Assumptions

$$\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$$
$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_0 & \sigma_{01} \\ \sigma_{01} & \sigma^2_1 \end{pmatrix} \right)$$

Expected evolution: Control

Control group $PROG_i = 0$

$$Y_{ij} = \gamma_{00} + \gamma_{10}(Age_{ij} - 1) + b_{0i} + b_{1i}(Age_{ij} - 1) + \varepsilon_{ij}$$

$$\mathrm{E}\left(Y_{ij} | PROG_i = 0\right) = \gamma_{00} + \gamma_{10}(Age_{ij} - 1)$$

Intervention group $PROG_i = 1$

$$Y_{ij} = (\gamma_{00} + \gamma_{01}) + (\gamma_{10} + \gamma_{11})(Age_{ij} - 1) + b_{0i} + b_{1i}(Age_{ij} - 1) + \varepsilon_{ij}$$

$$\mathrm{E}\left(Y_{ij} | PROG_i = 1\right) = (\gamma_{00} + \gamma_{01}) + (\gamma_{10} + \gamma_{11})(Age_{ij} - 1)$$

Hypotheses of interest

Hierarchical model

$$\begin{cases} Y_{ij} = & \pi_{0i} + \pi_{1i}(Age_{ij} - 1) + \varepsilon_{ij} \\ \pi_{0i} = & \gamma_{00} + \gamma_{01} PROG_i + b_{0i} \\ \pi_{1i} = & \gamma_{10} + \gamma_{11} PROG_i + b_{1i} \end{cases}$$

Hypotheses of interest

$$H_0 : \gamma_{01} = 0 \quad H_1 : \gamma_{01} \neq 0$$

$$H_0 : \gamma_{11} = 0 \quad H_1 : \gamma_{11} \neq 0$$

Fitting the model

Model

$$\begin{aligned} Y_{ij} = \; & \gamma_{00} + \gamma_{01} PROG_i + \gamma_{10}(Age_{ij} - 1) + \gamma_{11} PROG_i (Age_{ij} - 1) + \\ & b_{0i} + b_{1i}(Age_{ij} - 1) + \varepsilon_{ij} \end{aligned}$$

$\Rightarrow$ Parameters are estimated via

- **Maximum likelihood (ML).**
- **Restricted maximum likelihood (REML).**
- **What is that?**

$\Rightarrow$ R: lmer (packages lme4 or arm)

A 2-stage Model Formulation: A bit of theory

Stage 1

- Response Y_{ij} for ith subject, measured at time t_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$
- Response vector $\boldsymbol{Y_i}$ for ith subject: $\boldsymbol{Y_i} = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})'$
- **Stage 1 model**:

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon_i}$$

- $\boldsymbol{Z}_i$ is a $(n_i \times q)$ matrix of known covariates
- $\boldsymbol{\beta}_i$ is a q-dimensional vector of subject-specific regression coefficients
- $\boldsymbol{\varepsilon_i} \sim N(\mathbf{0}, \Sigma_i)$, often $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}_{n_i}$
- Note that the above model describes the observed variability within subjects

Dietary intervention example

The 1-stage model

$$Y_{ij} = \pi_{0i} + \pi_{1i}(Age_{ij} - 1) + \varepsilon_{ij}$$

can be rewritten in matrix form as

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon_i}$$

where

$$\overbrace{\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix}}^{\boldsymbol{Y}_i} = \overbrace{\begin{pmatrix} 1 & Age_{i1} - 1 \\ 1 & Age_{i2} - 1 \\ 1 & Age_{i3} - 1 \end{pmatrix}}^{\boldsymbol{Z}_i} \overbrace{\begin{pmatrix} \pi_{0i} \\ \pi_{1i} \end{pmatrix}}^{\boldsymbol{\beta}_i} + \overbrace{\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \end{pmatrix}}^{\boldsymbol{\varepsilon_i}}$$

A 2-stage Model Formulation: A bit of theory

Stage 2

- Between-subject variability can now be studied from relating the $\boldsymbol{\beta}_i$ to known covariates
- **Stage 2 model**:

$$\boldsymbol{\beta}_i = \boldsymbol{K}_i\boldsymbol{\beta} + \boldsymbol{b}_i$$

- $\boldsymbol{K}_i$ is a $(q \times p)$ matrix of known covariates
- $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression parameters
- $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$

Dietary intervention example

The 2-stage model

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} PROG_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{11} PROG_i + b_{1i} \end{cases}$$

can be rewritten in matrix form as

$$\boldsymbol{\beta}_i = \boldsymbol{K}_i\boldsymbol{\beta} + \boldsymbol{b}_i$$

where

$$\overbrace{\begin{pmatrix} \pi_{0i} \\ \pi_{1i} \end{pmatrix}}^{\boldsymbol{\beta}_i} = \overbrace{\begin{pmatrix} 1 & PROG_i & 0 & 0 \\ 0 & 0 & 1 & PROG_i \end{pmatrix}}^{\boldsymbol{K}_i} \overbrace{\begin{pmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{pmatrix}}^{\boldsymbol{\beta}} + \overbrace{\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}}^{\boldsymbol{b}_i}$$

The general linear mixed-effects model

- A 2-stage approach can be performed explicitly in the analysis
- Combining the two stages into one model leads to:

$$\begin{cases} \boldsymbol{Y}_i &= \boldsymbol{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon_i} \\ \boldsymbol{\beta}_i &= \boldsymbol{K}_i\boldsymbol{\beta} + \boldsymbol{b}_i \end{cases}$$

- and plugging $\boldsymbol{\beta}_i$ into the expression for $\boldsymbol{Y}_i$

$$\Rightarrow \boldsymbol{Y}_i = \underbrace{\boldsymbol{Z}_i\boldsymbol{K}_i}_{\boldsymbol{X}_i}\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon_i} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon_i}$$

The general linear mixed-effects model

$$\begin{cases} \boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon_i} \\ \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}), \quad \boldsymbol{\varepsilon_i} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i), \\ \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \boldsymbol{\varepsilon_1}, \ldots, \boldsymbol{\varepsilon_N} \text{ independent} \end{cases}$$

- Terminology:
  - Fixed effects: $\boldsymbol{\beta}$
  - Random effects: $\boldsymbol{b}_i$
  - Variance components: elements in $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_i$

Hierarchical versus marginal model

- The general linear mixed model (LMM) is given by:

$$\begin{cases} \boldsymbol{Y}_i = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i + \varepsilon_{\boldsymbol{i}} \\ \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}), \quad \varepsilon_{\boldsymbol{i}} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i), \\ \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \varepsilon_{\boldsymbol{1}}, \ldots, \varepsilon_{\boldsymbol{N}} \text{ independent} \end{cases}$$

- It can be rewritten as:

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \sim N(\boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i, \boldsymbol{\Sigma}_i),$$

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$$

Hierarchical versus marginal model

- It is therefore also called a hierarchical model:
  - A model for $\boldsymbol{Y}_i$ given $\boldsymbol{b}_i$
  - A model for $\boldsymbol{b}_i$
- Marginally, we have that $\boldsymbol{Y}_i$ is distributed as:

$$\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i\beta, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_i)$$

- Hence, very specific assumptions are made about the dependence of the mean and covariance on the covariates $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$:
  - **Implied mean :** $\boldsymbol{X}_i\beta$
  - **Implied covariance :** $\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_i$
- The hierarchical model always implies a marginal one, **NOT** vice versa

Estimation of the Marginal Model

- Recall that the general linear mixed model equals

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon_i}$$

$$\left.\begin{array}{l} \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}) \\ \boldsymbol{\varepsilon_i} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i) \end{array}\right\} \text{ independent}$$

- The implied marginal model equals

$$\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_i)$$

- Note that inferences based on the marginal model do not explicitly assume the presence of random effects representing the natural heterogeneity between subjects

Estimation of the Marginal Model

- Notation:
  - $\boldsymbol{\beta}$: vector of fixed effects (as before)
  - $\boldsymbol{\alpha}$: vector of all variance components in $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_i$
  - $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$: vector of all parameters in marginal model

$$Y_{ij} = \overbrace{\gamma_{00} + \gamma_{01}PROG_i + \gamma_{10}(Age_{ij} - 1) + \gamma_{11}PROG_i(Age_{ij} - 1)}^{\text{Fixed effects}} +$$

$$b_{0i} + b_{1i}(Age_{ij} - 1) + \rightarrow \text{Random effects}$$

$$\varepsilon_{ij} \rightarrow \text{Error}$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}\right)$$

$$\boldsymbol{\beta} = (\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11})' \text{ and } \boldsymbol{\alpha} = (\sigma_\varepsilon^2, \sigma_0^2, \sigma_{01}, \sigma_1^2)'$$

Estimation of the Marginal Model

- Notation:
  - $\boldsymbol{\beta}$: vector of fixed effects (as before)
  - $\boldsymbol{\alpha}$: vector of all variance components in $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_i$
  - $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$: vector of all parameters in marginal model
- Marginal likelihood function:

$$L_{ML}(\boldsymbol{\theta}) = \prod_{i=1}^{N} \left\{ (2\pi)^{-n_i/2} \left| \boldsymbol{V}_i(\boldsymbol{\alpha}) \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})' \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha}) (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \right) \right\}$$

- If $\boldsymbol{\alpha}$ were known, MLE of $\boldsymbol{\beta}$ equals

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{W}_i \boldsymbol{X}_i \right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{W}_i \boldsymbol{y}_i,$$

  where $\boldsymbol{W}_i$ equals $\boldsymbol{V}_i^{-1}$.

Estimation of the Marginal Model

- In most cases, $\boldsymbol{\alpha}$ is not known, and needs to be replaced by an estimate $\widehat{\boldsymbol{\alpha}}$
- Two frequently used estimation methods for $\boldsymbol{\alpha}$:
  - Maximum likelihood
  - Restricted maximum likelihood

Maximum Likelihood Estimation (ML)

- $\widehat{\boldsymbol{\alpha}}_{ML}$ obtained from maximizing

$$L_{ML}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))$$

with respect to $\boldsymbol{\alpha}$

- The resulting estimate $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}_{ML})$ for $\boldsymbol{\beta}$ will be denoted by $\widehat{\boldsymbol{\beta}}_{ML}$
- $\widehat{\boldsymbol{\alpha}}_{ML}$ and $\widehat{\boldsymbol{\beta}}_{ML}$ can also be obtained from maximizing $L_{ML}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, i.e., with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ simultaneously.

Restricted Maximum Likelihood Estimation (REML)

- We first combine all models

$$\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{V}_i)$$

into one model

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

in which

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_N \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{pmatrix}, \quad \boldsymbol{V}(\boldsymbol{\alpha}) = \begin{pmatrix} \boldsymbol{V}_1 & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{V}_N \end{pmatrix}$$

- The data are transformed orthogonal to $\boldsymbol{X}$ ($\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{0}$):

$$\boldsymbol{U} = \boldsymbol{A}'\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{A}'\boldsymbol{V}(\alpha)\boldsymbol{A})$$

Restricted Maximum Likelihood Estimation (REML)

- The MLE of $\boldsymbol{\alpha}$, based on $\boldsymbol{U}$, is called the REML estimate and is denoted by $\widehat{\boldsymbol{\alpha}}_{REML}$
- The resulting estimate $\widehat{\boldsymbol{\beta}}(\widehat{\alpha}_{REML})$ for $\boldsymbol{\beta}$ will be denoted by $\widehat{\boldsymbol{\beta}}_{REML}$
- $\widehat{\boldsymbol{\alpha}}_{REML}$ and $\widehat{\boldsymbol{\beta}}_{REML}$ can also be obtained from maximizing

$$L_{REML}(\boldsymbol{\theta}) = \left| \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{W}_i(\boldsymbol{\alpha}) \boldsymbol{X}_i \right|^{-\frac{1}{2}} L_{ML}(\boldsymbol{\theta})$$

  with respect to $\boldsymbol{\theta}$, i.e., with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ simultaneously.

- $L_{REML}\left(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\right)$ is the likelihood of the error contrasts $\boldsymbol{U}$, and is often called the REML likelihood function. It is **NOT** the likelihood for the original data $\boldsymbol{Y}$

Restricted versus Maximum Likelihood Estimation

- The **MLE** is **negatively biased** for the estimation of **variance components**, but the bias gets smaller for larger sample sizes (asymptotically unbiased)
- **REML** is **unbiased** for the estimation of **variance components** and, therefore, it may be a better option for small sample sizes
- Likelihood ratio tests based on REML require **exactly the same fixed effects specification** in both models (Why?). So, comparing models with different fixed effects (a common scenario) using an LR test, requires ML

Fitting the model: R code

```
> ## Installing the packages
>
> install.packages("lme4")
> install.packages("arm")
> install.packages("pbkrtest")
>
> ## Loading the packages
>
> library(lme4)
> library(lattice)
> library(arm)
> library(car)
> library(pbkrtest)
>
> ## Creating the time variable
>
> early.int1$age0<-early.int1$age-1
>
> ## Fitting the model with ML
>
> early.lmer1<-lmer(cog~1+age0*program+(1 + age0|id), REML = FALSE,
+                   data=early.int1)
>
```


R code: Remarks

- $(1 + age0|id)$ subject specific part: $b_{0i} + b_{1i}(Age_{ij} - 1)$
- Intercept is default: $(age0|id)$
- $age0 * program$: Fixed effects

$$\gamma_{00} + \gamma_{01} PROG_i + \gamma_{10}(Age_{ij} - 1) + \gamma_{11} PROG_i(Age_{ij} - 1)$$

- Default estimation procedure is REML. $REML = FALSE$ calculates MLE!

R Output

```
> summary(early.lmer1)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: cog ~ 1 + age0 * program + (1 + age0 | id)
   Data: early.int1

     AIC      BIC   logLik deviance df.resid
  2332.5   2362.4  -1158.3   2316.5      301

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.25361 -0.59088  0.02132  0.56849  2.29366

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 84.02    9.166
          age0        39.44    6.280    -0.55
 Residual             60.31    7.766
Number of obs: 309, groups:  id, 103

Fixed effects:
             Estimate Std. Error t value
(Intercept)  104.3007     1.7274   60.38
age0         -16.2555     1.8860   -8.62
program       -0.9646     2.3020   -0.42
age0:program   6.3187     2.5133    2.51

Correlation of Fixed Effects:
            (Intr) age0   progrm
age0        -0.629
program     -0.750  0.472
age0:progrm  0.472 -0.750 -0.629
>
```


R Output

```
> display(early.lmer1)
>
lmer(formula = cog ~ 1 + age0 * program + (1 + age0 | id), data = early.int1,
    REML = FALSE)
             coef.est coef.se
(Intercept)  104.30     1.73
age0         -16.26     1.89
program       -0.96     2.30
age0:program   6.32     2.51

Error terms:
 Groups   Name        Std.Dev. Corr
 id       (Intercept) 9.17
          age0        6.28     -0.55
 Residual             7.77
---
number of obs: 309, groups: id, 103
AIC = 2332.5, DIC = 2316.5
deviance = 2316.5
>
> anova(early.lmer1)
>
Analysis of Variance Table
             Df Sum Sq Mean Sq  F value
age0          1 6256.8  6256.8 103.7473
program       1  134.8   134.8   2.2344
age0:program  1  381.2   381.2   6.3208
>
```

No p-values!!!

Inference for the Fixed Effects

- Estimate for $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{W}_i \boldsymbol{X}_i \right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{W}_i \boldsymbol{y}_i,$$

  where $\boldsymbol{W}_i = \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})$ and $\boldsymbol{\alpha}$ replaced by its ML or REML estimate

- Conditional on $\boldsymbol{\alpha}$, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is asymptotically multivariate normal with mean $\boldsymbol{\beta}$ and covariance

$$\text{Var}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})) = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{W}_i \boldsymbol{X}_i \right)^{-1}$$

- In practice one again replaces $\boldsymbol{\alpha}$ by its ML or REML estimate

Approximate Wald Test

- For any known matrix L, consider testing

$$H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}, \quad \text{versus} \quad H_A : \boldsymbol{L}\boldsymbol{\beta} \neq \boldsymbol{0}$$

- Wald test statistic:

$$G = \widehat{\boldsymbol{\beta}}' \boldsymbol{L}' \left[\boldsymbol{L} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1}(\widehat{\boldsymbol{\alpha}}) \boldsymbol{X}_i \right)^{-1} \boldsymbol{L}' \right]^{-1} \boldsymbol{L}\widehat{\boldsymbol{\beta}}$$

- Asymptotic null distribution of G is χ^2 with rank(L) degrees of freedom

Approximate t-test and F-test

- Wald test based on

$$\text{Var}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})) = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{W}_i \boldsymbol{X}_i\right)^{-1}$$

- Variability introduced from replacing $\boldsymbol{\alpha}$ by some estimate is not taken into account in Wald tests
- Therefore, Wald tests will only provide valid inferences in sufficiently large samples
- In practice, this is often solved by replacing the χ^2 distribution by an appropriate F-distribution (and the normal by a t).

Approximate t-test and F-test

- For any known matrix $\boldsymbol{L}$, consider testing

$$H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}, \quad \text{versus} \quad H_A : \boldsymbol{L}\boldsymbol{\beta} \neq \boldsymbol{0}$$

- F test statistic:

$$F = \frac{\widehat{\boldsymbol{\beta}}' \boldsymbol{L}' \left[\boldsymbol{L}\left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1}(\widehat{\boldsymbol{\alpha}}) \boldsymbol{X}_i\right)^{-1} \boldsymbol{L}'\right]^{-1} \boldsymbol{L}\widehat{\boldsymbol{\beta}}}{\text{rank}(\boldsymbol{L})}.$$

- Approximate null-distribution of F is F with numerator degrees of freedom equal to rank($\boldsymbol{L}$)

Approximate t-test and F-test

- Approximate null-distribution of F is F with numerator degrees of freedom equal to rank($\boldsymbol{L}$)
- Denominator degrees of freedom to be estimated from the data:
  - Satterthwaite approximation
  - Kenward and Roger approximation
  - ...
- In the context of longitudinal data, all methods typically lead to large numbers of degrees of freedom, and therefore also to very similar p-values.
- For univariate hypotheses (rank($\boldsymbol{L}$) $= 1$) the F-test reduces to a t-test

Testing fixed effects in LMM

> Perhaps I can try again to explain why I don't quote p-values or, more to the point, why I do not take the "obviously correct" approach of attempting to reproduce the results provided by SAS. Let me just say that, although there are those who feel that the purpose of the R Project - indeed the purpose of any statistical computing whatsoever - is to reproduce the p-values provided by SAS, I am not a member of that group.
>
> Douglas Bates at [R] lmer, p-values and all that
>
> https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

Testing fixed effects in LMM

> Most of the research on tests for the fixed-effects specification in a mixed model begin with the assumption that these statistics will have an F distribution with a known numerator degrees of freedom and the only purpose of the research is to decide how to obtain an approximate denominator degrees of freedom. I don't agree.
>
> Douglas Bates at [R] lmer, p-values and all that
>
> https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

Testing fixed effects in LMM

- In general it is **not** clear that the null distribution of the computed ratio of sums of squares is really an F distribution, for any choice of denominator degrees of freedom.
- When the responses are normally distributed and the design is balanced, nested etc. (i.e. the classical LMM situation), the scaled deviances and differences in deviances are exactly F-distributed and looking at the experimental design (i.e., which treatments vary/are replicated at which levels) tells us what the relevant degrees of freedom are.

Testing fixed effects in LMM

- When the data are not classical (crossed, unbalanced), we might still **assume** that the deviances are approximately F-distributed but that we don't know the real degrees of freedom this is what the Satterthwaite, Kenward-Roger, Fai-Cornelius, among other approximations are supposed to do

- Situation worsens when dealing with discrete responses (binary, Poisson, etc)

Testing the effects in R

```
> ## Estimating the fixed effects via bootstrap
> fixed.boot=bootMer(early.lmer1,  fixef, use.u = TRUE, nsim = 250)
> fixed.boot
>
Call:
bootMer(x = early.lmer1, FUN = fixef, nsim = 250, use.u = TRUE)

Bootstrap Statistics :
       original     bias    std. error
t1* 104.3007437  0.1185616    1.018903
t2* -16.2554565 -0.1540370    1.697468
t3*  -0.9646326 -0.2478112    1.346218
t4*   6.3187112  0.2910760    2.206772
>
> summary(fixed.boot)
>
               R  original bootBias bootSE  bootMed
(Intercept)  250 104.30074  0.11856 1.0189 104.3584
age0         250 -16.25546 -0.15404 1.6975 -16.3408
program      250  -0.96463 -0.24781 1.3462  -1.1463
age0:program 250   6.31871  0.29108 2.2068   6.3985
>
```

Testing the effects in R

```
> ## Calculating confidence intervals for the fixed effects via Wald, bootstrap and profile likelihood
> confint(early.lmer1,par=5:8,method="Wald",oldNames = FALSE) # Only for fixed effects vc will return NA
>
                             2.5 %     97.5 %
(Intercept)              100.915099 107.686389
age0                     -19.951908 -12.559005  ## Significant
program                   -5.476393   3.547128  ## Not significant
age0:program               1.392766  11.244657  ## Significant
> confint(early.lmer1,method="boot",boot.type ="perc",oldNames = FALSE,nsim=500)
>
                               2.5 %       97.5 %
sd_(Intercept)|id          6.9406669  11.17900578
cor_age0.(Intercept)|id   -1.0000000  -0.06533987
sd_age0|id                 0.7563938   9.45484576
sigma                      6.7327885   8.78499590
(Intercept)              100.5754354 108.09445268
age0                     -20.3249215 -12.16307504  ## Significant
program                   -5.4742982   4.03860091  ## Not significant
age0:program               1.8325498  11.24816569  ## Significant
> confint(early.lmer1, level = 0.95,method="profile",oldNames = FALSE)
>
                             2.5 %     97.5 %
sd_(Intercept)|id          7.009249  11.406182
cor_age0.(Intercept)|id          NA         NA
sd_age0|id                 0.000000   9.975352
sigma                      6.814978   8.953279
(Intercept)              100.883287 107.718200
age0                     -19.986640 -12.524273  ## Significant
program                   -5.518786   3.589521  ## Not significant
age0:program               1.346481  11.290942  ## Significant
There were 50 or more warnings (use warnings() to see the first 50)
>
```


Testing the effects in R

```
> ## Get the KR-approximated degrees of freedom
>
> early.lmer1.df.KR <- get_ddf_Lb(early.lmer1, fixef(early.lmer1))
>
> ## Get p-values from the t-distribution using the t-values and approximated
> ## degrees of freedom
>
> early.lmer1.coef=coef(summary(early.lmer1))
> early.lmer1.p.KR <- cbind(early.lmer1.coef,2 * (1 - pt(abs(early.lmer1.coef[,3]), early.lmer1.df.KR)))
> early.lmer1.p.KR
>
               Estimate Std. Error    t value
(Intercept)  104.3007437   1.727402 60.3801366 0.000000e+00
age0         -16.2554565   1.885979 -8.6191067 7.416290e-14  ## Significant
program       -0.9646326   2.301961 -0.4190482 6.760348e-01  ## Not significant
age0:program   6.3187112   2.513284  2.5141257 1.344676e-02  ## Significant
>
```

Likelihood ratio test

- Comparison of nested models with different mean structures, but equal covariance structure
- Null hypothesis of interest equals $H_0 : \boldsymbol{\beta} \in \Theta_{\boldsymbol{\beta},0}$, for some subspace $\Theta_{\boldsymbol{\beta},0}$ of the parameter space $\Theta_{\boldsymbol{\beta}}$ of the fixed effects $\boldsymbol{\beta}$.
- Notation:
  - L_{ML}: ML likelihood function
  - $\widehat{\boldsymbol{\beta}}_{ML,0}$: MLE under H_0
  - $\widehat{\boldsymbol{\beta}}_{ML}$: MLE under general model

Likelihood ratio test

- Test statistic:

$$-2 \ln \lambda_N = -2 \ln \left[\frac{L_{ML}(\widehat{\boldsymbol{\beta}}_{ML,0})}{L_{ML}(\widehat{\boldsymbol{\beta}}_{ML})} \right]$$

- Asymptotic null distribution: χ^2 with d.f. equal to the difference in dimension of $\Theta_{\boldsymbol{\beta}}$ and $\Theta_{\boldsymbol{\beta},0}$.

Early dietary intervention study

Hierarchical model

$$\begin{cases} Y_{ij} = & \pi_{0i} + \pi_{1i}(Age_{ij} - 1) + \varepsilon_{ij} \\ \pi_{0i} = & \gamma_{00} + \gamma_{01} PROG_i + b_{0i} \\ \pi_{1i} = & \gamma_{10} + \gamma_{11} PROG_i + b_{1i} \end{cases}$$

Three models considered for the second level

- No effect of program $\gamma_{01} = \gamma_{11} = 0$ (early.lmer1.noprog)
- Program has an effect only on the intercept $\gamma_{11} = 0$ (early.lmer1.intprog)
- Program has an effect only on the evolution $\gamma_{01} = 0$ (early.lmer1.slopeprog)

Likelihood ratio tests in R

```
> ## Likelihood ratio tests
> early.lmer1.noprog<-lmer(cog~1+age0+(1 + age0|id), REML = FALSE, data=early.int1)
> early.lmer1.intprog<-lmer(cog~1+age0+program+(1 + age0|id), REML = FALSE, data=early.int1)
> early.lmer1.slopeprog<-lmer(cog~1+age0+age0:program+(1 + age0|id), REML = FALSE, data=early.int1)
> anova(early.lmer1.noprog,early.lmer1.intprog,early.lmer1.slopeprog)
>
Data: early.int1
Models:
early.lmer1.noprog: cog ~ 1 + age0 + (1 + age0 | id)
early.lmer1.intprog: cog ~ 1 + age0 + program + (1 + age0 | id)
early.lmer1.slopeprog: cog ~ 1 + age0 + age0:program + (1 + age0 | id)
                     Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
early.lmer1.noprog    6 2336.8 2359.2 -1162.4   2324.8
early.lmer1.intprog   7 2336.7 2362.8 -1161.3   2322.7 2.084      1     0.1489  ## \gamma_{01}=0
early.lmer1.slopeprog  7 2330.7 2356.8 -1158.3   2316.7 5.959      0     <2e-16  ## \gamma_{11} is not 0
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
>
```

Effect of early dietary intervention on children IQ

Conclusions level 2 model

$$\begin{cases} \hat{\pi}_{0i} = & 103.758 + b_{0i} \\ \hat{\pi}_{1i} = & -15.882 + 5.656 PROG_i + b_{1i} \end{cases}$$

$\Rightarrow$ Children in the intervention and control group have the same average initial scores. Expected?

$\Rightarrow$ The average cognitive performance decreased in both groups but less in the intervention group.

Assessing the random effects

- Empirical Bayes inference
- Best linear unbiased prediction
- Example: Early dietary intervention
- Shrinkage
- Example: Early dietary intervention
- A theoretical illustration

Assessing the random effects

- Recall that the general linear mixed model equals

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \sim N(\boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i, \boldsymbol{\Sigma}_i), \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$$

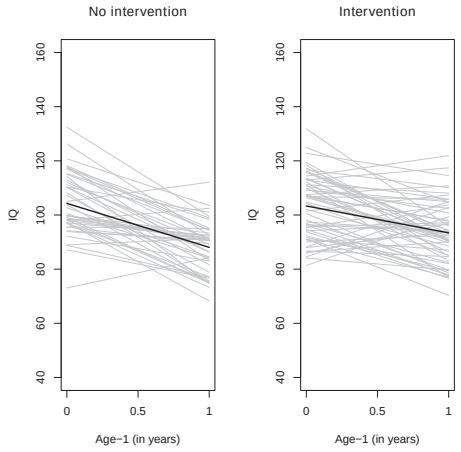
- Marginally,

$$\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i\beta, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_i)$$

- Thus, random effects $\boldsymbol{b}_i$ reflect how the evolution of the ith subject deviates from the expected evolution $\boldsymbol{X}_i\beta$, i.e., how the evolution of the ith subject deviates from the average evolution in the population

- Estimation of $\boldsymbol{b}_i$ helpful for detecting outlying profiles or predicting individual trajectories

Assessing the random effects



$\Rightarrow$ $\boldsymbol{b}_i$ reflect how the evolution for the ith subject deviates from the average

$\Rightarrow$ Some subjects are above/below the average at the beginning of the study

$\Rightarrow$ The evolution of individual subjects differs from the average evolution

Assessing the random effects

- The term "estimates" of the random effects is some times used in the literature
- Random effects are not, strictly speaking, parameters but unobserved random variables
- One does not estimate the random effects in the same sense that one estimates parameters
- $f(\boldsymbol{b}_i) = N(\boldsymbol{0}, \boldsymbol{D})$ can be interpreted as the prior distribution of $\boldsymbol{b}_i$, i.e., its distribution before the data are collected
- Hence, it is natural to base the prediction of $\boldsymbol{b}_i$ on the posterior distribution $f(\boldsymbol{b}_i|\boldsymbol{Y}_i)$ using Bayesian methods

Assessing the random effects

- Applying Bayes theorem the posterior density of $\boldsymbol{b}_i$ is

$$f(\boldsymbol{b}_i|\boldsymbol{Y}_i) = \frac{f(\boldsymbol{Y}_i|\boldsymbol{b}_i)\ f(\boldsymbol{b}_i)}{\int f(\boldsymbol{Y}_i|\boldsymbol{b}_i)\ f(\boldsymbol{b}_i)\, d\boldsymbol{b}_i} \propto f(\boldsymbol{Y}_i|\boldsymbol{b}_i)\ f(\boldsymbol{b}_i)$$

$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{b}_i - \boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{W}_i(\boldsymbol{Y}_i - \boldsymbol{X}_i\beta)\right)' \boldsymbol{\Lambda}_i^{-1}\left(\boldsymbol{b}_i - \boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{W}_i(\boldsymbol{Y}_i - X_i\beta)\right)\right\}$$

  for some positive definite matrix $\boldsymbol{\Lambda}_i$

- Posterior distribution:

$$\boldsymbol{b}_i|\boldsymbol{Y}_i \sim N\left(\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{W}_i(\boldsymbol{Y}_i - \boldsymbol{X}_i\beta), \boldsymbol{\Lambda}_i\right)$$

Assessing the random effects

- Posterior mean $E[\boldsymbol{b}_i|\boldsymbol{Y}_i]$ used to predict $\boldsymbol{b}_i$

$$\widehat{\boldsymbol{b}}_i(\boldsymbol{\theta}) = E[\boldsymbol{b}_i|\boldsymbol{Y}_i] = \int \boldsymbol{b}_i f(\boldsymbol{b}_i|\boldsymbol{Y}_i)\, d\boldsymbol{b}_i = \boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{W}_i(\boldsymbol{\alpha})(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$$

- $\widehat{\boldsymbol{b}}_i(\boldsymbol{\theta})$ is normally distributed with covariance matrix

$$\text{var}(\widehat{\boldsymbol{b}}_i(\boldsymbol{\theta})) = \boldsymbol{D}\boldsymbol{Z}_i'\left\{\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{X}_i\left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{W}_i\boldsymbol{X}_i\right)^{-1}\boldsymbol{X}_i'\boldsymbol{W}_i\right\}\boldsymbol{Z}_i\boldsymbol{D}$$

- Inferences for $\boldsymbol{b}_i$ should account for the variability in $\widehat{\boldsymbol{b}}_i$

Assessing the random effects

- Parameters in $\boldsymbol{\theta}$ are replaced by their ML or REML estimates, obtained from fitting the marginal model

- $\widehat{\boldsymbol{b}}_i = \widehat{\boldsymbol{b}}_i(\widehat{\boldsymbol{\theta}})$ is called the Empirical Bayes estimate/prediction of $\boldsymbol{b}_i$

R Output

```
> summary(early.lmer1) ## We are goint to work with  the full model for illustration
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: cog ~ 1 + age0 * program + (1 + age0 | id)
   Data: early.int1

     AIC      BIC   logLik deviance df.resid
  2332.5   2362.4  -1158.3   2316.5      301

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.25361 -0.59088  0.02132  0.56849  2.29366

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 84.02    9.166
          age0        39.44    6.280    -0.55
 Residual             60.31    7.766
Number of obs: 309, groups:  id, 103

Fixed effects:
             Estimate Std. Error t value
(Intercept)  104.3007     1.7274   60.38
age0         -16.2555     1.8860   -8.62
program       -0.9646     2.3020   -0.42
age0:program   6.3187     2.5133    2.51

Correlation of Fixed Effects:
            (Intr) age0   progrm
age0        -0.629
program     -0.750  0.472
age0:progrm  0.472 -0.750 -0.629
>
```

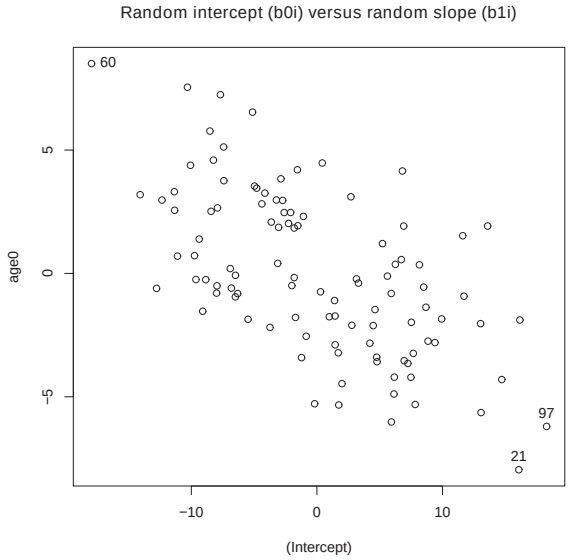

R Output

```
> ## Random effects covariance matrix
>
> D.early=unclass(VarCorr(early.lmer1))$id
> D.early
>
            (Intercept)      age0
(Intercept)    84.01915 -31.89344
age0          -31.89344  39.44432
attr(,"stddev")
(Intercept)         age0
   9.166196    6.280471
>
attr(,"correlation")
            (Intercept)       age0
(Intercept)   1.0000000 -0.5540131
age0         -0.5540131  1.0000000
>
```

Predicting the random effects in R

```
> ## Predicted random effects
> early.lmer1.re=ranef(early.lmer1)$id
>
> head(early.lmer1.re,10)
>
   (Intercept)       age0
1     1.406077 -1.0998359
2     1.700796 -3.2167090
3     1.996373 -4.4674210
4   -11.103346  0.6994383
5     1.444727 -1.7280502
6    -9.633042 -0.2478932
7     4.787545 -3.5797912
8     5.221731  1.2096573
9    14.723746 -4.3027959
10   -7.682804  7.2427588
>
> plot(early.lmer1.re,
+ main="Random intercept (b0i) versus random slope (b1i)")
>
```


Random intercept ($\widehat{b}_{0i}$) versus random slope ($\widehat{b}_{1i}$)



Best Linear Unbiased Prediction (BLUP)

- Often, parameters of interest are linear combinations of fixed effects in $\boldsymbol{\beta}$ and random effects in $\boldsymbol{b}_i$
- For example, a subject-specific slope is the sum of the average slope and the subject-specific random slope. In the case study a child that did not receive the intervention has slope

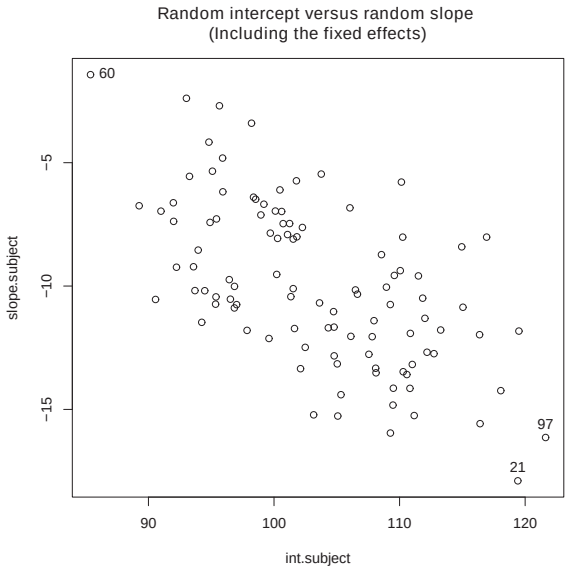
$$\pi_{1i} = \gamma_{10} + b_{1i}$$

- In general, suppose $\boldsymbol{u} = \boldsymbol{\lambda}'_{\beta}\boldsymbol{\beta} + \boldsymbol{\lambda}'_{b}\boldsymbol{b}_i$ is of interest
- Conditionally on $\boldsymbol{\alpha}$, $\widehat{\boldsymbol{u}} = \boldsymbol{\lambda}'_{\beta}\widehat{\boldsymbol{\beta}} + \boldsymbol{\lambda}'_{b}\widehat{\boldsymbol{b}_i}$ is BLUP:
  - linear in the observations $\boldsymbol{Y}_i$
  - unbiased for $\boldsymbol{u}$
  - minimum variance among all unbiased linear estimators

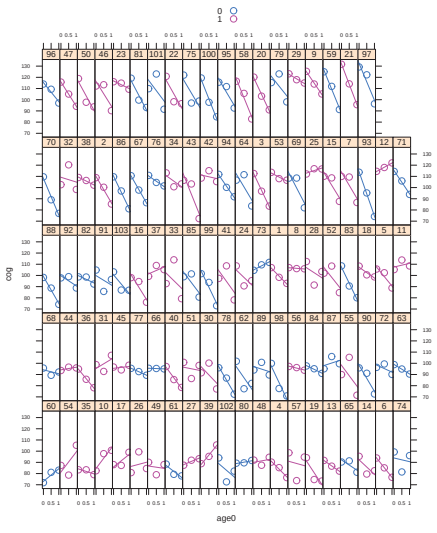
Intercept and slope: OLS versus LMM estimates

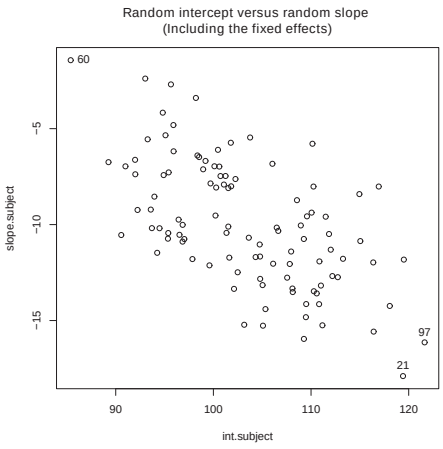
```
> ## Creating the subject specific intercepts and slopes
>
> ind.coef=coef(early.lmer1)$id
> head(ind.coef)
>
  (Intercept)      age0    program age0:program
1    105.7068 -17.35529 -0.9646326     6.318711
2    106.0015 -19.47217 -0.9646326     6.318711
3    106.2971 -20.72288 -0.9646326     6.318711
4     93.1974 -15.55602 -0.9646326     6.318711
5    105.7455 -17.98351 -0.9646326     6.318711
6     94.6677 -16.50335 -0.9646326     6.318711
>
> int.subject=ind.coef[,1]+ind.coef[,3]
> slope.subject=ind.coef[,2]+ind.coef[,4]
> plot(int.subject,slope.subject, main="Random intercept versus random
+ slope \n(Including the fixed effects)")
>
```

Random intercept ($\widehat{\pi}_{0i}$) versus random slope ($\widehat{\pi}_{1i}$)

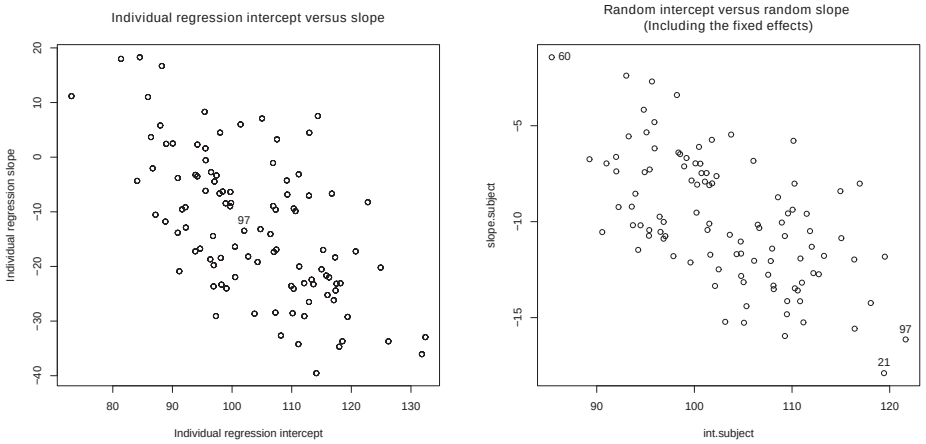


OLS versus LMM estimates

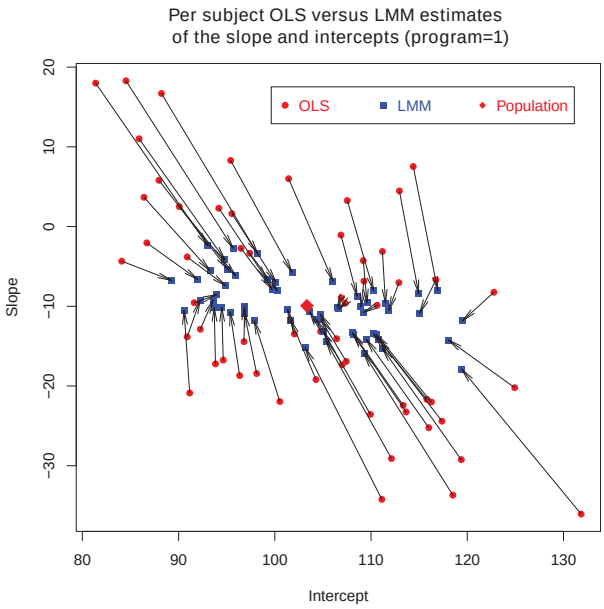




OLS versus LMM estimates



OLS versus LMM estimates



OLS versus LMM estimates

- In general, the per-subject slopes and intercepts from the mixed-effects model (LMM) are closer to the population estimates than are the within-subject OLS estimates
- This pattern is sometimes described as a shrinkage of coefficients toward the population values
- John Tukey chose to characterize this process in terms of the estimates for individual subjects "borrowing strength" from each other
- In a mixed-effects model we assume that the levels of a grouping factor are a selection from a population and, as a result, can be expected to share characteristics to some degree

Shrinkage Estimators $\widehat{b}_i$

- Consider the prediction of the evolution of the ith subject:

$$\begin{aligned}\widehat{\boldsymbol{Y}}_i &\equiv \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i \\ &= \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}) \\ &= \left(\boldsymbol{I}_{n_i} - \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\right)\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{Y}_i \\ &= \boldsymbol{\Sigma}_i\boldsymbol{V}_i^{-1}\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \left(\boldsymbol{I}_{n_i} - \Sigma_i\boldsymbol{V}_i^{-1}\right)\boldsymbol{Y}_i\end{aligned}$$

- Let us look more closely at this expression

Shrinkage Estimators $\widehat{b}_i$

$\widehat{Y}_i$ is a weighted mean of two factors

$$\widehat{Y}_i \equiv \boxed{\Sigma_i V_i^{-1} X_i \widehat{\beta}} + \boxed{\left(I_{n_i} - \Sigma_i V_i^{-1}\right) Y_i}$$

$\Rightarrow$ Factor 1: Population-averaged profile $X_i\widehat{\beta}$ with weight $\Sigma_i V_i^{-1}$

$\Rightarrow$ Factor 2: Individual data Y_i with weight $\left(I_{n_i} - \Sigma_i V_i^{-1}\right)$

Shrinkage Estimators $\widehat{b}_i$

- Note that the population average $X_i\widehat{\beta}$ gets much weight if the residual variability Σ_i is 'large' in comparison to the total variability V_i.
- This phenomenon is usually called shrinkage

> The observed data are shrunk towards the prior average profile $X_i\beta$.

- This is also reflected in the fact that for any linear combination $\lambda' b_i$ of random effects,

$$\text{var}(\lambda'\widehat{b}_i) \leq \text{var}(\lambda' b_i).$$

Models for Clustered Data

The Rat Pup Example

The data come from a study in which 30 female rats were randomly assigned to receive one of three doses of an experimental compound (variable **treat** with levels: high, low or control). Although 10 female rats were initially assigned to receive each treatment dose, three of the female rats in the high-dose group died, so there are no data for their litters. In addition, litter sizes (variable **lts**) varied widely, ranging from 2 to 18 pups. The sex of the pups was also recorded (variable **sex**)

Objective of the study: To compare the birth weights (variable **w**) of pups from litters born to female rats that received the high- and low-dose treatments to the birth weights of pups from litters that received the control treatment.

Jose Pinheiro and Doug Bates, (2000) Mixed-Effects Models in S and S-PLUS.

The Rat Pup Example

- Two-level clustered data from a cluster randomized trial
- Each litter (cluster) was randomly assigned to a specific level of treatment
- Rat pups (units of analysis) nested within litters
- Birth weights of rat pups within the same litter are likely to be correlated because the pups shared the same maternal environment

Exploring the data in R

```
> ## Reading the data
> ratpup <- read.table("rat_pup.dat", h = T)
> ratpup$sex1[ratpup$sex == "Female"] <- 1
> ratpup$sex1[ratpup$sex == "Male"] <- 0
> attach(ratpup)
>
> ## Table describing the  data
> g <- function(x)c(N=length(x),Mean=mean(x,na.rm=TRUE),
+ SD=sd(x,na.rm=TRUE), Min=min(x,na.rm=TRUE),Max=max(x,na.rm=TRUE))
> summarize(weight,by=llist(treatment,sex),g)
>
  treatment    sex weight     Mean        SD  Min  Max
1   Control Female     54 6.116111 0.6851179 3.68 7.57
2   Control   Male     77 6.471039 0.7537880 4.57 8.33
3      High Female     32 5.851562 0.6001887 4.48 7.68
4      High   Male     33 5.918485 0.6909058 5.01 7.70
5       Low Female     65 5.837538 0.4504964 4.75 7.73
6       Low   Male     61 6.025082 0.3803403 5.25 7.13
>
```


Exploring the data in R

Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females

Exploring the data in R

Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females

Exploring the data in R

Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females
- Sample mean birth weight of males are consistently higher than those of females within all levels of treatment

Exploring the data in R

Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females
- Sample mean birth weight of males are consistently higher than those of females within all levels of treatment

Exploring the data in R

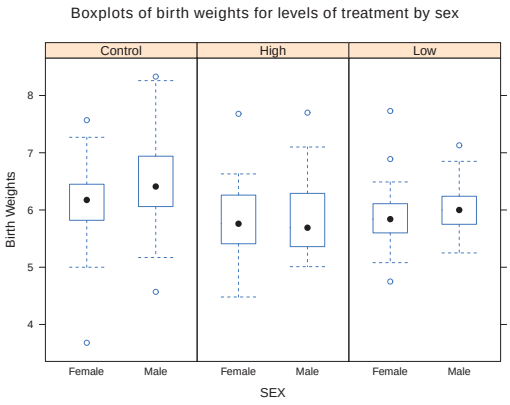
Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females
- Sample mean birth weight of males are consistently higher than those of females within all levels of treatment

Exploring the data in R

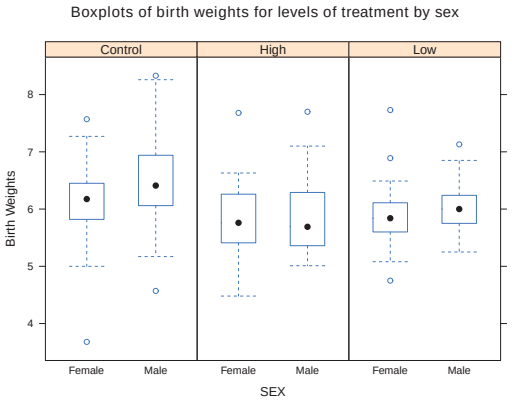
```
> ## Comparing the distributions of birth weights
> ## for each treatment by sex combination
>
> library(lattice)  # trellis graphics
> library(grid)
>
> bwplot(weight ~ sex|treatment, data=ratpup,aspect = 2,
+  ylab="Birth Weights", xlab="SEX",
+  main = "Boxplots of birth weights for levels of treatment by sex")
>
```


Birth weights for levels of treatment by sex



- Males appear to have a higher median birth weight than females in the low and control groups, but not in the high group
- The distribution of birth weight appears to be roughly symmetric at each level of treatment and sex

Birth weights for levels of treatment by sex

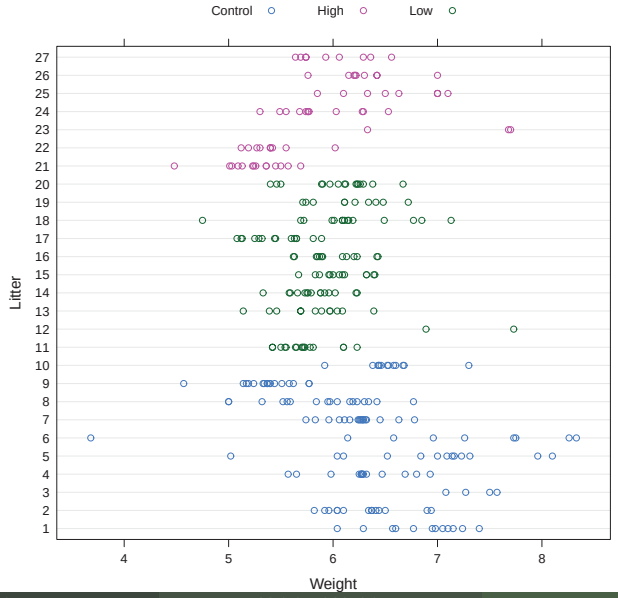


- Lower birth weight for the high- and low-dose treatments compared to the control group
- Variance of the birth weight is similar for males and females within each treatment but appears to differ across treatments

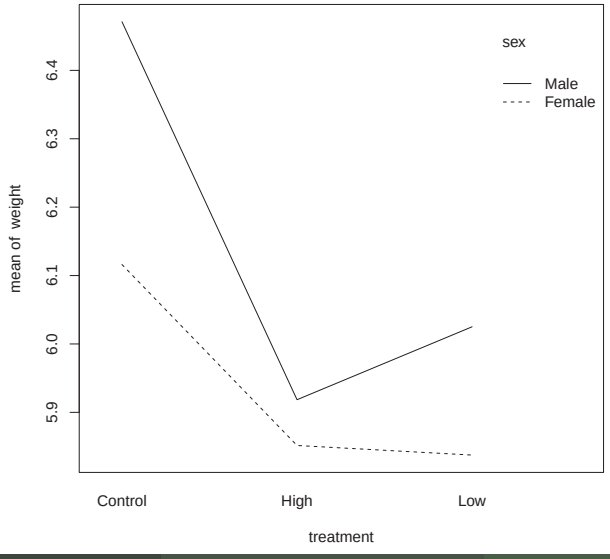
Exploring the data in R

```
> ## Comparing the distributions of birth weights for each treatment
>
> dotplot(litterid ~ weight,group=treatment, data =ratpup,
+ xlab="Weight", ylab="Litter",
+ auto.key=list(space="top", column=3, cex=.8, title="",
+               cex.title=1, lines=FALSE, points=TRUE) )
> with(ratpup, interaction.plot(treatment,sex,weight))
>
```

Exploring the data in R



Exploring the data in R



Hierarchical model

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} sex_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$$

and sex_{ij} a level 1 indicator variable for female.

Level 2:

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} treat_{1i} + \gamma_{02} treat_{2i} + \gamma_{03} ls_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{20} treat_{1i} + \gamma_{30} treat_{2i} \end{cases}$$

where $treat_{1i}$ and $treat_{2i}$ are level 2 indicator variables for high and low treatment levels, ls_i is the litter size and $b_{0i} \sim N\left(0, \sigma^2_b\right)$

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} sex_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$$

Level 2:

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} treat_{1i} + \gamma_{02} treat_{2i} + \gamma_{03} ls_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{20} treat_{1i} + \gamma_{30} treat_{2i} \end{cases}$$

$\Rightarrow$ Birth weights of pups vary **within** litter due to differences in gender and in other unaccounted factors (ε_{ij})

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} sex_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$$

Level 2:

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} treat_{1i} + \gamma_{02} treat_{2i} + \gamma_{03} ls_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{20} treat_{1i} + \gamma_{30} treat_{2i} \end{cases}$$

$\Rightarrow$ Birth weights vary **between** litters due to differences in treatment, litter size and other litter-specific characteristics unaccounted for by the model (b_{0i})

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} sex_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$$

Level 2:

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} treat_{1i} + \gamma_{02} treat_{2i} + \gamma_{03} ls_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{20} treat_{1i} + \gamma_{30} treat_{2i} \end{cases}$$

$\Rightarrow$ Birth weights vary **between** litters due to differences in treatment, litter size and other litter-specific characteristics unaccounted for by the model (b_{0i})

$\Rightarrow$ Notice that treatment may affect males and females pups differently

One single model

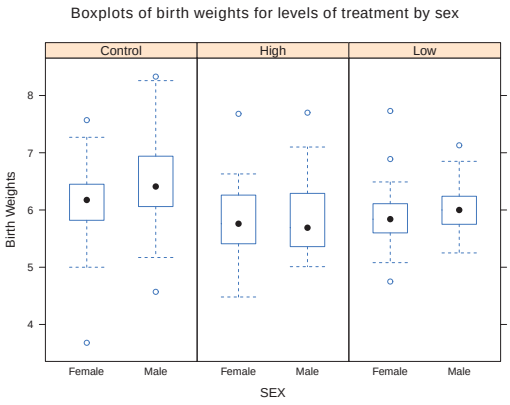
Model

$$\begin{aligned} w_{ij} = \quad & \gamma_{00} + \gamma_{01} treat_{1i} + \gamma_{02} treat_{2i} + \gamma_{03} ls_i + \\ & \gamma_{10} sex_{ij} + \gamma_{20} treat_{1i} sex_{ij} + \gamma_{30} treat_{2i} sex_{ij} + \\ & b_{0i} + \varepsilon_{ij} \end{aligned}$$

Distributional Assumptions

$$b_{0i} \sim N\left(0, \sigma_b^2\right) \text{ and } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

Modeling the covariance structure



- Previous model assumes that the within litter variability σ_ε^2 is constant across treatment
- The variances of the birth weights are similar for males and females within each treatment but appear to differ across treatments

Covariance structure: Testing homoscedasticity

Hence, one wants to test if the variance of the residuals (σ^2_ε) is the same (homogeneous) for the three treatment groups (high, low, and control)

$$H_0 : \sigma^2_{high} = \sigma^2_{low} = \sigma^2_{control} = \sigma^2_\varepsilon$$

- REML-based likelihood ratio test to compare two models (**mean structure stays the same**):

  Model 1: All three variances equal (meanfull.hom)

  Model 2: All three variances different (meanfull.het)

- The asymptotic null distribution of this test statistic is a χ^2 with 2 degrees of freedom

Covariance structure: Testing homoscedasticity

- At this moment the lmer() function does not allow users to fit models with heterogeneous error variance structures
- Therefore, we will work with the function lme() from the package nlme
- lme() and lmer() are similar but there are some differences in syntax and output that will be explained in the following

Fitting the homocedastic model in R

Model 1

```
> ## Fitting a homocedastic model
>
> library(nlme)
>
> meanfull.hom <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                     random = ~1 | litterid, ratpup, method = "REML")
>
```

- The factor() function is not necessary for treatment, because the original treatment variable has string values High, Low, and Control, and will therefore be considered as a factor automatically
- We also do not need to declare sex1 as a factor, because it is an indicator variable having only values of 0 and 1

Fitting the homocedastic model in R

Model 1

```
> ## Fitting a homocedastic model
>
> library(nlme)
>
> meanfull.hom <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                     random = ~1 | litterid, ratpup, method = "REML")
>
```

- lme() treats the lowest level (alphabetically or numerically) of a factor as the reference category. This means that "Control" will be the reference category of treatment. The reference level can be changed using

$$treatment = relevel(treatment, ref = "High")$$

Fitting the homocedastic model in R

Model 1

```
> ## Fitting a homocedastic model
>
> library(nlme)
>
> meanfull.hom <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                     random = ~1 | litterid, ratpup, method = "REML")
>
```

- *random* $= \tilde{}1|$*litterid*, includes a random effect (intercept) for each level of litter in the model
- *method* $=$ "REML", specifies that the default REML estimation method is to be used

Fitting the homocedastic model in R

```
> summary(meanfull.hom)
>
Linear mixed-effects model fit by REML
 Data: ratpup
       AIC      BIC    logLik
  419.1043 452.8775 -200.5522

Random effects:
 Formula: ~1 | litterid
        (Intercept) Residual
StdDev:   0.3106722 0.404337

Fixed effects: weight ~ treatment + sex1 + litsize + treatment:sex1
                      Value  Std.Error  DF   t-value p-value
(Intercept)         8.323340 0.27333009 292 30.451605  0.0000
treatmentHigh      -0.906057 0.19154238  23 -4.730320  0.0001
treatmentLow       -0.467040 0.15818328  23 -2.952521  0.0071
sex1               -0.411688 0.07315410 292 -5.627679  0.0000
litsize            -0.128382 0.01875336  23 -6.845819  0.0000
treatmentHigh:sex1  0.107023 0.13176318 292  0.812239  0.4173
treatmentLow:sex1   0.083866 0.10568189 292  0.793568  0.4281

.............................................................

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-7.47250744 -0.50014749  0.02911668  0.57348178  3.00962055

Number of Observations: 322
Number of Groups: 27
>
```

Fitting the homocedastic model in R

```
> anova(meanfull.hom)
>
                numDF denDF  F-value p-value
(Intercept)         1   292 9093.772  <.0001
treatment           2    23    5.082  0.0149
sex1                1   292   52.602  <.0001
litsize             1    23   47.374  <.0001
treatment:sex1      2   292    0.466  0.6282
>
```

- The anova() function performs a series of Type I (or sequential) F-tests for the fixed effects in the model, each of which are conditional on the preceding terms in the model specification
- For example, the F-test for sex1 is conditional on the treatment effects, but the F-test for treatment is not conditional on the sex1 effect

Fitting the homocedastic model in R


```
> anova(meanfull.hom)
>
                numDF denDF  F-value p-value
(Intercept)         1   292 9093.772  <.0001
treatment           2    23    5.082  0.0149
sex1                1   292   52.602  <.0001
litsize             1    23   47.374  <.0001
treatment:sex1      2   292    0.466  0.6282
>
```


Model fitted using REML

The model was fitted using REML and, therefore, different mean structures cannot be compare!

Homocedastic model: Predicting the random effects

```
> ## Display the random effects (EBLUPs) from the model.
>
> random.effects(meanfull.hom)
   (Intercept)
1   0.17480024
2  -0.07362296
3  -0.17490203
4  -0.05376249
5   0.34446954
6  -0.05480208
7   0.39153638
8  -0.02616704
9  -0.61772106
10  0.09017150
11  0.04136696
12  0.02072931
13 -0.35981737
14  0.01847368
15  0.03549783
16  0.06416884
17 -0.39636862
18  0.42802095
19 -0.18110865
20  0.32903707
21 -0.27813901
22 -0.49096620
23  0.26053476
24  0.17537803
25  0.23748827
26  0.22966911
27 -0.13396497
>
```


Fitting the heterocedastic model in R

Model 2

```
> ## Fitting a heterocedastic model
>
> meanfull.het <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                     random = ~1 | litterid, ratpup, method = "REML",
+                     weights = varIdent(form = ~1 | treatment))
>
```

- The arguments of the lme() function are the same as those used to fit Model 1, with the addition of the weights argument
- The argument

$$weights = varIdent(form = \tilde{} 1|treatment)$$

sets up a heterogeneous residual variance structure, with observations at different levels of treatment having different residual variance parameters

Fitting the heterocedastic model in R

```
> summary(meanfull.het)
>
Linear mixed-effects model fit by REML
 Data: ratpup
       AIC     BIC    logLik
  381.8847 423.163 -179.9423

Random effects:
 Formula: ~1 | litterid
        (Intercept)  Residual
StdDev:   0.3134846 0.5147948

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | treatment
 Parameter estimates:
  Control       Low      High
1.0000000 0.5649830 0.6394383

Fixed effects: weight ~ treatment + sex1 + litsize + treatment:sex1
                     Value  Std.Error  DF   t-value p-value
(Intercept)         8.345294 0.27464753 292 30.385468  0.0000
treatmentHigh      -0.903277 0.19215903  23 -4.700672  0.0001
treatmentLow       -0.466292 0.15908908  23 -2.931013  0.0075
sex1               -0.408131 0.09303486 292 -4.386865  0.0000
litsize            -0.130007 0.01848708  23 -7.032332  0.0000
treatmentHigh:sex1  0.094666 0.12919527 292  0.732737  0.4643
treatmentLow:sex1   0.076013 0.10811858 292  0.703053  0.4826

.............................................................

>
```


Fitting the heterocedastic model in R

```
Random effects:
 Formula: ~1 | litterid
        (Intercept)  Residual
StdDev:   0.3134846 0.5147948

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | treatment
 Parameter estimates:
  Control       Low      High
1.0000000 0.5649830 0.6394383
```

- Random effects portion of the output: Estimated residual standard deviation equal to 0.5147948
- Parameter estimates: Values by which the residual standard deviation should be multiplied to obtain the estimated standard deviation of the residuals in each treatment group
- This multiplier is 1.0 for the control group (the reference). Multipliers for the low and high treatment groups are very similar

Heterocedastic versus homocedastic model

The variance of the residuals (σ^2_ε) is the same (homogeneous) for the three treatment groups

$$H_0 : \sigma^2_{high} = \sigma^2_{low} = \sigma^2_{control} = \sigma^2_\varepsilon$$

```
> ## Heterocedastic versus homocedastic model
>
> anova(meanfull.hom,meanfull.het)
>
             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
meanfull.hom     1  9 419.1043 452.8775 -200.5522
meanfull.het     2 11 381.8847 423.1630 -179.9423 1 vs 2 41.21964  <.0001
>
```


Heterocedastic model

```
Random effects:
 Formula: ~1 | litterid
        (Intercept)  Residual
StdDev:   0.3134846 0.5147948

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | treatment
 Parameter estimates:
  Control       Low      High
1.0000000 0.5649830 0.6394383
```

- $\sigma_{high} = 0.329179$, $\sigma_{low} = 0.290850$ and $\sigma_{control} = 0.5147948$
- Is $\sigma^2_{high} = \sigma^2_{low}$?

High-low dose: Equal residual variance

Hence, one wants to test if the variance of the residuals in the high and low dose groups are the same

$$H_0 : \sigma^2_{high} = \sigma^2_{low}$$

- REML-based likelihood ratio test to compare two models (**mean structure stays the same**):

  Model 2: All three variances different (meanfull.het)

  Model 3: $\sigma^2_{high} = \sigma^2_{low}$ (meanfull.hilo)

- The asymptotic null distribution of this test statistic is a χ^2 with 1 degrees of freedom

High-low dose: Equal residual variance

```
> ## High-low dose: Equal residual variance
>
> ratpup$trtgrp[treatment=="Control"] <- 1
> ratpup$trtgrp[treatment == "Low" | treatment == "High"] <- 2
>
> meanfull.hilo <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                      random = ~1 | litterid, ratpup, method = "REML",
+                      weights = varIdent(form = ~1 | trtgrp))
>
> summary(meanfull.hilo)
> anova(meanfull.hilo)
>
```

Fitting the heterocedastic model in R

```
> summary(meanfull.hilo)
>
Linear mixed-effects model fit by REML
 Data: ratpup
       AIC      BIC    logLik
  381.0807 418.6065 -180.5404

Random effects:
 Formula: ~1 | litterid
        (Intercept)  Residual
StdDev:   0.3145679 0.5147878

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | trtgrp
 Parameter estimates:
        1         2
1.0000000 0.5905487

Fixed effects: weight ~ treatment + sex1 + litsize + treatment:sex1
                      Value  Std.Error  DF   t-value p-value
(Intercept)         8.350351 0.27567833 292 30.290196  0.0000
treatmentHigh      -0.901844 0.19140146  23 -4.711793  0.0001
treatmentLow       -0.466596 0.15999337  23 -2.916347  0.0078
sex1               -0.408195 0.09303540 292 -4.387529  0.0000
litsize            -0.130383 0.01856367  23 -7.023574  0.0000
treatmentHigh:sex1  0.092026 0.12461723 292  0.738473  0.4608
treatmentLow:sex1   0.076397 0.10939797 292  0.698337  0.4855
.............................................................
>
```


High-low dose: Equal residual variance

Hence, one wants to test if the variance of the residuals in the high and low dose groups are the same

$$H_0 : \sigma^2_{high} = \sigma^2_{low}$$

```
> ## High-low dose: Equal residual variance
>
> anova(meanfull.het,meanfull.hilo)
>
              Model df      AIC      BIC    logLik   Test  L.Ratio p-value
meanfull.het      1 11 381.8847 423.1630 -179.9423
meanfull.hilo     2 10 381.0807 418.6065 -180.5404 1 vs 2 1.196053  0.2741
>
```

Is there a litter effect?

- Can the random effects (b_{0i}) associated with the litter–specific intercepts be omitted from Model 3?
- One do not directly test the significance of the random litter–specific intercepts, but rather tests a hypothesis related to the variance of the random litter effects.
- The null and alternative hypotheses can be written as follows:

$$H_0 : \sigma_b^2 = 0 \text{ versus } H_1 : \sigma_b^2 > 0$$

Is there a litter effect?

- Although hypothesis tests are often phrased in terms of parameter restrictions, they basically compare the quality of the fit obtained from two nested models
- Likelihood ratio tests (LRTs) are a valuable tool to compare nested models
- An approximate reference distribution for a LRT is the χ^2_γ where γ, the degrees of freedom, is determined by the difference in the number of parameters for the models H_1 and H_0
- Hence, the LRT for testing $H_0 : \sigma_b^2 = 0$ versus $H_1 : \sigma_b^2 > 0$ has an approximate reference distribution χ^2_1

Is there a litter effect?

- However, the argument for using a χ^2_1 distribution **does not apply** when the parameter value being tested is on the boundary of the parametric space
- The asymptotic null distribution of the test statistic is a mixture of χ^2 distributions, with 0 and 1 degrees of freedom, and equal weights of 0.5
- As shown in Pinheiro and Bates (2000) Section 2.5, the p-value from the χ^2_1 distribution will be "conservative"in the sense that it is larger than a simulation-based p-value would be
- In the worst-case scenario the χ^2_1-based p-value will be twice as large as it should be

Is there a litter effect?

```
> ## Is there a litter effect?
>
> meanfull.hilo.nolitter <- gls(weight ~ treatment + sex1 + litsize +
+         treatment:sex1, data = ratpup, weights = varIdent(form = ~1 | trtgrp))
>
> summary(meanfull.hilo.nolitter)
>
```

Is there a litter effect?

```
Generalized least squares fit by REML
  Model: weight ~ treatment + sex1 + litsize + treatment:sex1
  Data: ratpup
       AIC      BIC   logLik
  489.6521 523.4252 -235.826

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | trtgrp
 Parameter estimates:
         1         2
1.0000000 0.7060188

Coefficients:
                      Value  Std.Error   t-value p-value
(Intercept)         8.201712 0.15902776  51.57409  0.0000
treatmentHigh      -0.976414 0.10624042  -9.19060  0.0000
treatmentLow       -0.456018 0.08700180  -5.24147  0.0000
sex1               -0.339911 0.10616682  -3.20167  0.0015
litsize            -0.121478 0.01008518 -12.04524  0.0000
treatmentHigh:sex1  0.180960 0.14941228   1.21114  0.2267
treatmentLow:sex1   0.076386 0.13035758   0.58597  0.5583

.....................................................

Residual standard error: 0.5980885
Degrees of freedom: 322 total; 315 residual
```


Is there a litter effect?

Is there a litter effect?

$$H_0 : \sigma_b^2 = 0 \text{ versus } H_1 : \sigma_b^2 > 0$$

```
> ## Is there a litter effect?
>
> anova(meanfull.hilo.nolitter,meanfull.hilo)
>
                       Model df      AIC      BIC    logLik   Test  L.Ratio p-value
meanfull.hilo.nolitter     1  9 489.6521 523.4252 -235.8260
meanfull.hilo              2 10 381.0807 418.6065 -180.5404 1 vs 2 110.5713  <.0001
>
```

Modeling the mean structure

```
> ## Fitting the final model using ML
>
> meanfull.hilo.ml <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                         random = ~1 | litterid, ratpup, method = "ML",
+                         weights = varIdent(form = ~1 | trtgrp))
>
> summary(meanfull.hilo.ml)
```


Modeling the mean structure

```
> summary(meanfull.hilo.ml)
>
Linear mixed-effects model fit by maximum likelihood
 Data: ratpup
       AIC      BIC    logLik
  357.1317 394.8773 -168.5659

Random effects:
 Formula: ~1 | litterid
        (Intercept)  Residual
StdDev:   0.2882595 0.5123784

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | trtgrp
 Parameter estimates:
        1         2
1.0000000 0.5897706
Fixed effects: weight ~ treatment + sex1 + litsize + treatment:sex1
                     Value  Std.Error  DF  t-value p-value
(Intercept)          8.350608 0.26150064 292 31.93341  0.0000
treatmentHigh       -0.904757 0.18092616  23 -5.00070  0.0000
treatmentLow        -0.466869 0.15105108  23 -3.09080  0.0052
sex1                -0.406590 0.09357754 292 -4.34495  0.0000
litsize             -0.130402 0.01755814  23 -7.42689  0.0000
treatmentHigh:sex1   0.093026 0.12521954 292  0.74290  0.4581
treatmentLow:sex1    0.075602 0.10998665 292  0.68737  0.4924
 ...........................................................

Number of Observations: 322
Number of Groups: 27
```

Modeling the mean structure

```
> anova(meanfull.hilo.ml)
>
               numDF denDF   F-value p-value
(Intercept)        1   292 10274.678  <.0001
treatment          2    23     4.810  0.0180
sex1               1   292    59.906  <.0001
litsize            1    23    55.438  <.0001
treatment:sex1     2   292     0.315  0.7303
>
```


Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} sex_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim \begin{cases} N(0, 0.51^2), \text{Control} \\ N(0, 0.30^2), \text{Low/High dose} \end{cases}$$

Level 2:

$$\begin{cases} \pi_{0i} = 8.35 - 0.90 treat_{1i} - 0.47 treat_{2i} - 0.13 ls_i + b_{0i} \\ \pi_{1i} = -0.41 \\ b_{0i} \sim N(0, 0.29^2) \end{cases}$$

$\Rightarrow$ Birth weights of pups vary **within** litter due to differences in gender and in other unaccounted factors (ε_{ij})

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} sex_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim \begin{cases} N(0, 0.51^2), \text{Control} \\ N(0, 0.30^2), \text{Low/High dose} \end{cases}$$

Level 2:

$$\begin{cases} \pi_{0i} = 8.35 - 0.90 treat_{1i} - 0.47 treat_{2i} - 0.13 ls_i + b_{0i} \\ \pi_{1i} = -0.41 \\ b_{0i} \sim N(0, 0.29^2) \end{cases}$$

$\Rightarrow$ Litters in the high/low dose have pups with smaller average birth weights. In addition, litter size has a negative impact on the average birth weight and there is extra variability from other unknown factors

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} sex_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim \begin{cases} N(0, 0.51^2), \text{Control} \\ N(0, 0.30^2), \text{Low/High dose} \end{cases}$$

Level 2:

$$\begin{cases} \pi_{0i} = 8.35 - 0.90 treat_{1i} - 0.47 treat_{2i} - 0.13 ls_i + b_{0i} \\ \pi_{1i} = -0.41 \\ b_{0i} \sim N(0, 0.29^2) \end{cases}$$

$\Rightarrow$ Litters in the high/low dose have pups with smaller average birth weights. In addition, litter size has a negative impact on the average birth weight and there is extra variability from other unknown factors

$\Rightarrow$ Treatment affects males and females pups equally