

Models for longitudinal 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.
 - ⇒ Can we describe the time evolution for each individual with a linear function?
- **Level 2:** Changes between individuals.
 - ⇒ Are the individuals different at the beginning of the study?
 - ⇒ Do they evolve differently over time?

Distinguishing quality

Longitudinal Studies: Repeated measurements over time (Waves)

- ⇒ Metric: Time, age, weeks since treated...
- ⇒ Spacing: Equal time intervals?
- ⇒ Time structure: All individual measured at the same time points?
- ⇒ 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. Randomized to
 - 58 early intervention program.
 - 45 control group.
- Evaluated on ages 12, 18, and 24 months.
- Research question: Effect of the early intervention on the evolution of cognitive performance?

Effect of early dietary intervention on children IQ

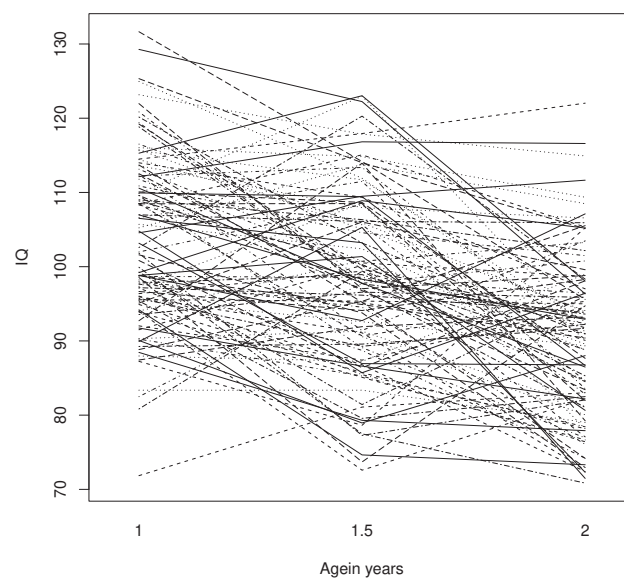
id	program	age	cog
1	1	1.00	106.98
1	1	1.50	98.31
1	1	2.00	92.91
2	1	1.00	108.86
2	1	1.50	100.29
2	1	2.00	85.30
3	1	1.00	112.52
3	1	1.50	96.77
3	1	2.00	83.43
4	1	1.00	90.24
4	1	1.50	85.27
4	1	2.00	76.41
5	1	1.00	105.71
5	1	1.50	102.40
5	1	2.00	88.79
6	1	1.00	93.89
6	1	1.50	85.10
6	1	2.00	76.66
7	1	1.00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

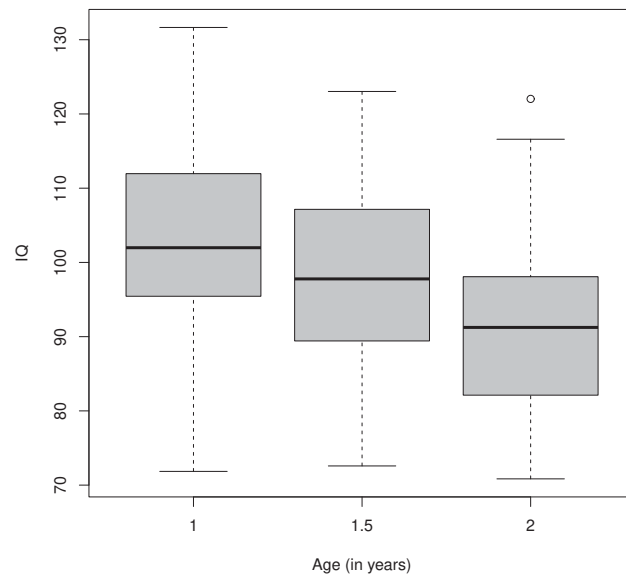
```
> ## Reading in the early.int data
>
> early.int1 <- read.table("earlyint.txt", header=T, sep=",")
>
> ## Attach data to the search path
>
> attach(early.int1)
>
> ## Spaghettiplot
>
> n=length(unique(id))
> interaction.plot(age,id,cog, xlab="Agein years", ylab="IQ",
+   legend=F)
>
```

Means per time point and group

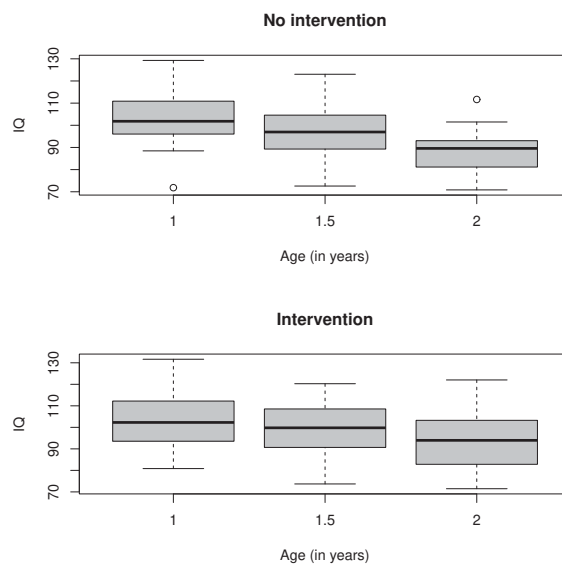
			Program	
			0	1
Age	Resp	Statistics		
1	IQ	n	45	58
		Mean	103.93	102.93
		Sd	11.01	11.78
1.5	IQ	n	45	58
		Mean	96.91	99.18
		Sd	11.93	12.02
2	IQ	n	45	58
		Mean	87.68	92.99
		Sd	9.05	12.13

```
> ## 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



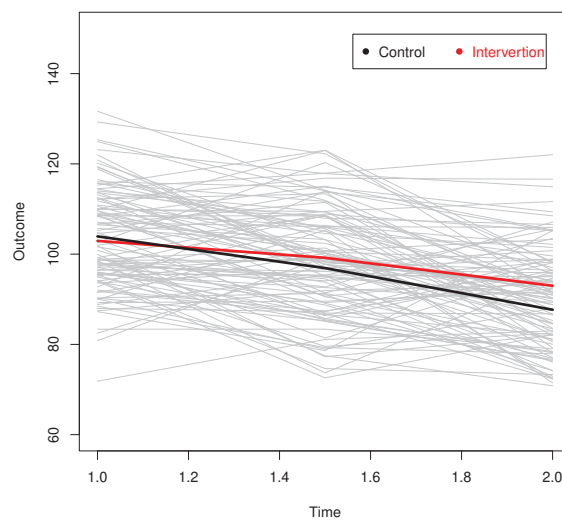
Boxplot per program



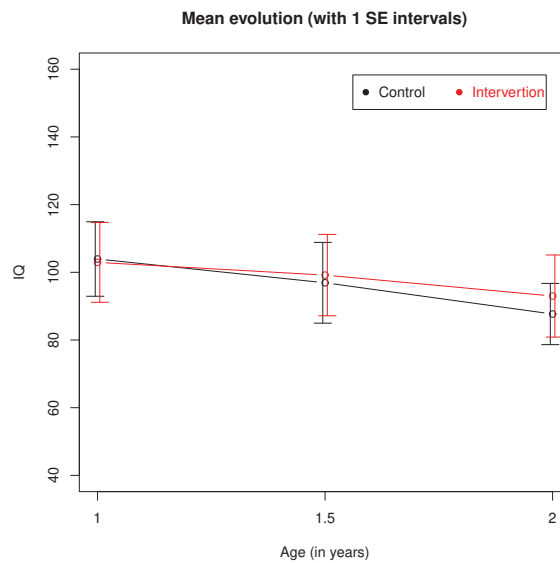
R code

```
> ## 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           #
#####

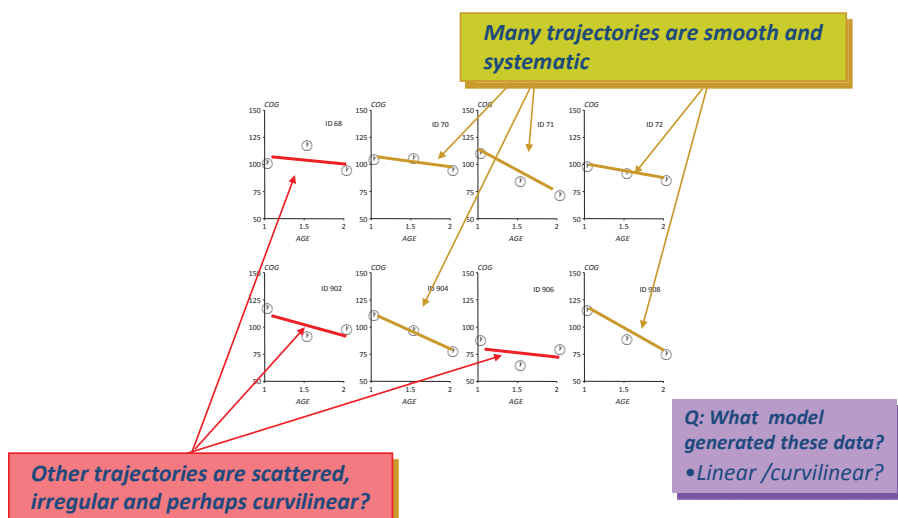
errbar=function(x,y,height,width,lty=1,col="black")
arrows(x,y,x,y+height,angle=90,length=width,lty=lty,
col=col)
arrows(x,y,x,y-height,angle=90,length=width,lty=lty,
col=col)

> ## 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

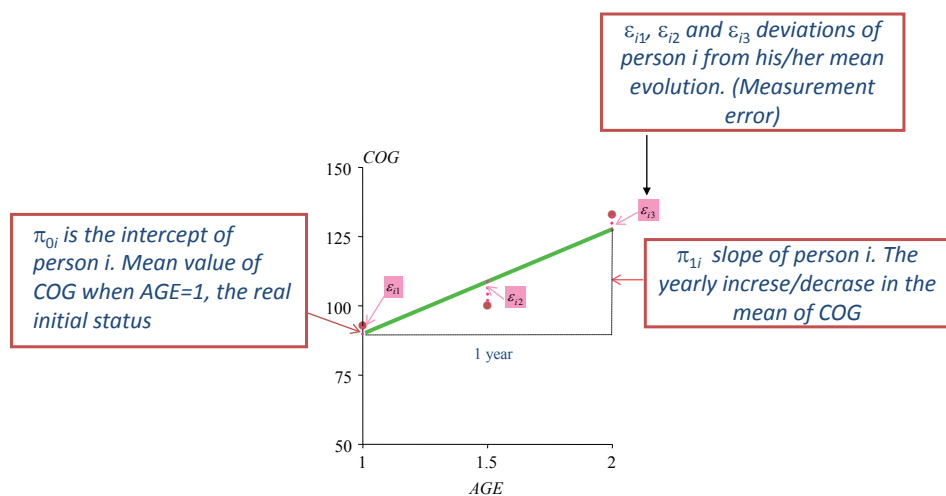
Model

Model for subject i

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

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

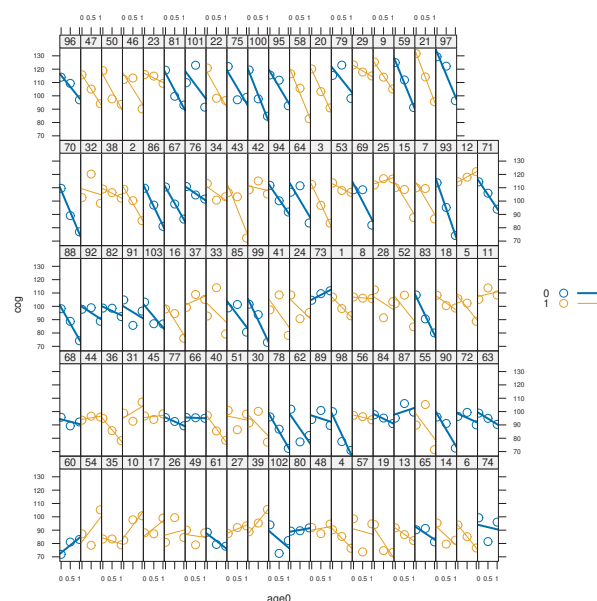
Interpretation of the model



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 $\pm 45^\circ$, 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



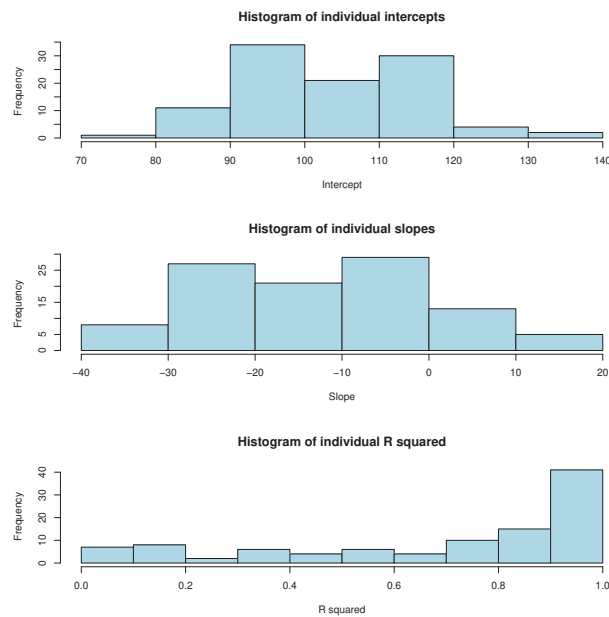
Linear regression per person: R code

```
>## 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



Ariel Alonso

Multilevel Models

27 / 98

Multilevel models

Level 1

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

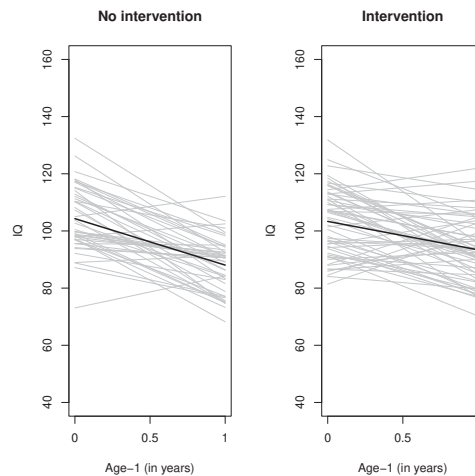
- ⇒ Structural part of the level 1. How individuals evolve.
- ⇒ Random part of Level 1. How individuals deviate from their own evolution.
- ⇒ Why do π_{0i} and π_{1i} vary?
- ⇒ Is due to the effect of the intervention program?

Ariel Alonso

Multilevel Models

28 / 98

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

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

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

Level 2

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

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

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

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

$$\begin{cases} Y_{ij} = \pi_{0i} + \pi_{1i}(\text{Age}_{ij} - 1) + \varepsilon_{ij} \\ \pi_{0i} = \gamma_{00} + \gamma_{01} \text{PROG}_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{11} \text{PROG}_i + b_{1i} \end{cases}$$

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

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

Distributional Assumptions

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

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

One single model

Model

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

→ Random effects

$\varepsilon_{ij} \rightarrow$ Error

Distributional Assumptions

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

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

Expected evolution: Control

Control group $PROG_i = 0$

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

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

Intervention group $PROG_i = 1$

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

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

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

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

⇒ Parameters are estimated via

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

⇒ R: lmer (packages: nlme, lme4 or arm)

A 2-stage Model Formulation: A bit of theory

Stage 1

- Response Y_{ij} for i th subject, measured at time t_{ij} , $i = 1, \dots, N$, $j = 1, \dots, n_i$
- Response vector \mathbf{Y}_i for i th subject: $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})'$
- **Stage 1 model:**

$$\mathbf{Y}_i = \mathbf{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$$

- \mathbf{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
- $\boldsymbol{\varepsilon}_i \sim N(0, \Sigma_i)$, often $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$
- Note that the above model describes the observed variability within subjects

Dietary intervention example

The 1-stage model

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

can be rewritten in matrix form as

$$\mathbf{Y}_i = \mathbf{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$$

where

$$\underbrace{\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix}}_{\mathbf{Y}_i} = \underbrace{\begin{pmatrix} 1 & \text{Age}_{i1} - 1 \\ 1 & \text{Age}_{i2} - 1 \\ 1 & \text{Age}_{i3} - 1 \end{pmatrix}}_{\mathbf{Z}_i} \underbrace{\begin{pmatrix} \pi_{0i} \\ \pi_{1i} \end{pmatrix}}_{\boldsymbol{\beta}_i} + \underbrace{\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \end{pmatrix}}_{\boldsymbol{\varepsilon}_i}$$

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:**

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

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

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

$$\beta_i = K_i\beta + b_i$$

where

$$\underbrace{\begin{pmatrix} \beta_i \\ \pi_{0i} \\ \pi_{1i} \end{pmatrix}}_{\beta_i} = \underbrace{\begin{pmatrix} 1 & PROG_i & 0 & 0 \\ 0 & 0 & 1 & PROG_i \end{pmatrix}}_{K_i} \underbrace{\begin{pmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}}_{b_i}$$

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:

$$\begin{cases} Y_i = Z_i\beta_i + \varepsilon_i \\ \beta_i = K_i\beta + b_i \end{cases}$$

- and plugging β_i into the expression for Y_i

$$\Rightarrow Y_i = \underbrace{Z_i K_i}_{X_i} \beta + Z_i b_i + \varepsilon_i = X_i \beta + Z_i b_i + \varepsilon_i$$

The general linear mixed-effects model

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

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

Hierarchical versus marginal model

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

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

- It can be rewritten as:

$$f(\mathbf{Y}_i | \mathbf{b}_i) = N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \Sigma_i),$$

$$f(\mathbf{b}_i) = N(0, \mathbf{D})$$

Hierarchical versus marginal model

- It is therefore also called a hierarchical model:

- A model for \mathbf{Y}_i given \mathbf{b}_i : $f(\mathbf{Y}_i|\mathbf{b}_i)$
- A model for \mathbf{b}_i : $f(\mathbf{b}_i)$

- Marginally, we have that \mathbf{Y}_i is distributed as:

$$f(\mathbf{Y}_i) = \int f(\mathbf{Y}_i|\mathbf{b}_i)f(\mathbf{b}_i) d\mathbf{b}_i = N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \boldsymbol{\Sigma}_i)$$

- Hence, very specific assumptions are made about the dependence of the mean and covariance on the covariates \mathbf{X}_i and \mathbf{Z}_i :

- **Implied mean** : $\mathbf{X}_i\boldsymbol{\beta}$
- **Implied covariance** : $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \boldsymbol{\Sigma}_i$

- The hierarchical model always implies a marginal one, **NOT** vice versa

Estimation of the Marginal Model

- Recall that the general linear mixed model equals

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

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

- The implied marginal model equals

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

- 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:
 - β : vector of fixed effects (as before)
 - α : vector of all variance components in D and Σ_i
 - $\theta = (\beta', \alpha')'$: vector of all parameters in marginal model
- Marginal likelihood function:

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

- If α were known, MLE of β equals

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

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

Estimation of the Marginal Model

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

Maximum Likelihood Estimation (ML)

- $\hat{\alpha}_{ML}$ obtained from maximizing

$$L_{ML}(\alpha, \hat{\beta}(\alpha))$$

with respect to α

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

Restricted Maximum Likelihood Estimation (REML)

- We first combine all models

$$Y_i \sim N(X_i\beta, V_i)$$

into one model

$$Y \sim N(X\beta, V)$$

in which

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}, \quad V(\alpha) = \begin{pmatrix} V_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_N \end{pmatrix}$$

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

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

Restricted Maximum Likelihood Estimation (REML)

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

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

with respect to θ , i.e., with respect to α and β simultaneously.

- $L_{REML}(\alpha, \hat{\beta}(\alpha))$ is the likelihood of the error contrasts \mathbf{U} , and is often called the REML likelihood function. It is **NOT** the likelihood for the original data \mathbf{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 (LRT) 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 LRT, requires ML

Fitting the model: R code

```
> ## Installing the packages
>
> install.packages("lme4")
> install.packages("arm")
> install.packages("nlme")
>
> ## Loading the packages
>
> library(lme4)
> library(lattice)
> library(nlme)
> library(arm)
> library(car)
>
> ## 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

- $(1 + \text{age0}|id)$ subject specific part: $b_{0i} + b_{1i}(\text{Age}_{ij} - 1)$

- Intercept is default: $(\text{age0}|id)$

- $\text{age0} * \text{program}$: Fixed effects

$$\gamma_{00} + \gamma_{01} \text{PROG}_i + \gamma_{10}(\text{Age}_{ij} - 1) + \gamma_{11} \text{PROG}_i(\text{Age}_{ij} - 1)$$

- Default estimation procedure is REML.

- $\text{REML} = \text{FALSE}$ 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
>
```

Inference for the Fixed Effects

- Estimate for β :

$$\hat{\beta}(\alpha) = \left(\sum_{i=1}^N \mathbf{x}_i' \mathbf{W}_i \mathbf{x}_i \right)^{-1} \sum_{i=1}^N \mathbf{x}_i' \mathbf{W}_i \mathbf{y}_i,$$

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

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

$$\text{Var}(\hat{\beta}(\alpha)) = \left(\sum_{i=1}^N \mathbf{x}_i' \mathbf{W}_i \mathbf{x}_i \right)^{-1}$$

- In practice one again replaces α by its ML or REML estimate

Approximate Wald Test

- For any known matrix L , consider testing

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

- Wald test statistic:

$$G = \hat{\beta}' L' \left[L \underbrace{\left(\sum_{i=1}^N \mathbf{x}_i' \mathbf{V}_i^{-1}(\alpha) \mathbf{x}_i \right)^{-1}}_{\text{Var}(\hat{\beta})^{-1}} L' \right]^{-1} L \hat{\beta}$$

- Conditional on α the asymptotic null distribution of G is χ^2 with $\text{rank}(L)$ degrees of freedom

Approximate t -test and F -test

- Wald test based on

$$\text{Var}(\hat{\beta}(\alpha)) = \left(\sum_{i=1}^N \mathbf{x}_i' \mathbf{W}_i \mathbf{x}_i \right)^{-1}$$

- In practice α is replaced by an estimate but...
- The variability introduced from replacing α by some estimate is not taken into account in Wald tests
- Therefore, Wald tests will only provide valid inferences in sufficiently large samples
- 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 \mathbf{L} , consider testing

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

- F test statistic:

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

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

Approximate t -test and F -test

- Approximate null-distribution of F is F with numerator degrees of freedom equal to $\text{rank}(\mathbf{L})$
- 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 ($\text{rank}(\mathbf{L}) = 1$) 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] Imer, 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

```
> ## 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 -1      1
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
>
```

Getting p-values

```
> ## Get the KR-approximated degrees of freedom
>
> require(pbkrtest)
> early.lmer1.df.KR <- get_Lb_ddf(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,df=early.lmer1.df.KR, 2 * (1 - pt(abs(early.lmer1.coef[,3]),
  early.lmer1.df.KR)))
> early.lmer1.p.KR
      Estimate Std. Error   t value      df
(Intercept) 104.3007437   1.727403 60.3800996 105.2575 0.000000e+00
age0        -16.2554565   1.885981 -8.6190984 105.2575 7.416290e-14
program      -0.9646326   2.301962 -0.4190479 105.2575 6.760350e-01
age0:program   6.3187112   2.513286  2.5141232 105.2575 1.344683e-02
>
```

Getting p-values with lmerTest

```
## Another way to get the p-values require(lmerTest) and refit the model
>
> require(lmerTest)
> early.lmer1<-lmer(cog~1+age0*program+(1 + age0|id), REML = FALSE, data=early.int1)
> summary(early.lmer1)

Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: cog ~ 1 + age0 * program + (1 + age0 | id)
Data: early.int1

Random effects:
Groups   Name      Variance Std.Dev. Corr
id       (Intercept) 84.02    9.166
age0     age0       39.44    6.281   -0.55
Residual                60.31    7.766
Number of obs: 309, groups: id, 103

Fixed effects:
      Estimate Std. Error    df t value Pr(>|t|)
(Intercept) 104.3007   1.7274 102.9999  60.380 < 2e-16 ***
age0        -16.2555   1.8860 103.0001  -8.619 8.55e-14 ***
program      -0.9646   2.3020 102.9999  -0.419  0.6761
age0:program   6.3187   2.5133 103.0001   2.514  0.0135 *
---
>
```

lmerTest: anova function

```
## Another way to get the p-values require(lmerTest) and refit the model
>
> require(lmerTest)
>
> ## Type III analysis the same as the one obtained with the summary function
>
> anova(early.lmer1)
>
Type III Analysis of Variance Table with Satterthwaite's method
      Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
age0      4480.2   4480.2     1    103  74.2889 8.547e-14 ***
program      10.6     10.6     1    103   0.1756  0.67605
age0:program   381.2    381.2     1    103   6.3208  0.01348 *
---
>
> ## Type I sequential model building
>
> anova(early.lmer1, type=1)
>
Type I Analysis of Variance Table with Satterthwaite's method
      Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
age0      6256.8   6256.8     1    103 103.7471 < 2e-16 ***
program      89.7     89.7     1    103   1.4869  0.22549
age0:program  381.2    381.2     1    103   6.3208  0.01348 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Getting p-values with lme

```
> ## Fitting the model with lme
>
> require(nlme)
> early.lme1<-lme(cog~1+age0*program, random=~1+age0|id, method = "ML", data=early.int1)
> summary(early.lme1)

Linear mixed-effects model fit by maximum likelihood
Data: early.int1
      AIC      BIC    logLik
2332.532 2362.398 -1158.266

Random effects:
Formula: ~1 + age0 | id
Structure: General positive-definite, Log-Cholesky parametrization
      StdDev  Corr
(Intercept) 9.166182 (Intr)
age0        6.280528 -0.554
Residual    7.765843

Fixed effects: cog ~ 1 + age0 * program
              Value Std.Error DF t-value p-value
(Intercept) 104.30074  1.738689 204 59.98815  0.0000
age0       -16.25546  1.898308 204 -8.56313  0.0000
program     -0.96463  2.317003 101 -0.41633  0.6781
age0:program  6.31871  2.529714 204  2.49780  0.0133

Number of Observations: 309
Number of Groups: 103
>
```

Testing the effects in R with lme

```
> ## CI lme
>
> intervals(early.lme1)

Approximate 95% confidence intervals

Fixed effects:
      lower      est.      upper
(Intercept) 100.894899 104.3007437 107.706589
age0         -19.973972 -16.2554565 -12.536941
program      -5.531097  -0.9646326   3.601831
age0:program   1.363362   6.3187112  11.274060

Random Effects:
Level: id
      lower      est.      upper
sd((Intercept))  7.2611304  9.1661817  11.571048
sd(age0)         3.1375541  6.2805285  12.571907
cor((Intercept),age0) -0.7944973 -0.5540066 -0.163302

Within-group standard error:
      lower      est.      upper
6.774495  7.765843  8.902259
>
```

Likelihood ratio test

$$H_0 : \beta \in \Theta_0 \quad H_1 : \beta \in \Theta_0^C$$

- Notation:

- L_{ML} : Likelihood function
- $\hat{\beta}_{ML0}$: Point in Θ_0 that maximizes L_{ML}
- $\hat{\beta}_{ML}$: Point in $\Theta = \Theta_0 \cup \Theta_0^C$ that maximizes L_{ML}

- Test statistic:

$$-2 \ln \lambda_N = -2 \ln \left[\frac{L_{ML}(\hat{\beta}_{ML0})}{L_{ML}(\hat{\beta}_{ML})} \right] \xrightarrow{H_0} \chi^2(df)$$

- $df = \dim(\Theta) - \dim(\Theta_0)$.

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.noprogram)
- Program has an effect only on the intercept $\gamma_{11} = 0$ (early.lmer1.intprog)
- Program has an effect on both intercept and slope (early.lmer1)

Likelihood ratio tests in R

```
> ## Likelihood ratio tests
> early.lmer1.noprogram<-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)
> anova(early.lmer1.noprogram,early.lmer1.intprog,early.lmer1)
>
Data: early.int1
Models:
early.lmer1.noprogram: cog ~ 1 + age0 + (1 + age0 | id)
early.lmer1.intprog: cog ~ 1 + age0 + program + (1 + age0 | id)
early.lmer1: cog ~ 1 + age0 * program + (1 + age0 | id)

```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
early.lmer1.noprogram	6	2336.8	2359.2	-1162.4	2324.8				
early.lmer1.intprog	7	2336.7	2362.8	-1161.3	2322.7	2.0840	1	0.14885	
early.lmer1	8	2332.5	2362.4	-1158.3	2316.5	6.1345	1	0.01326	*

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Effect of early dietary intervention on children IQ

Conclusions level 2 model

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

- ⇒ Children in the intervention and control group have the same average initial scores. Expected?
- ⇒ 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

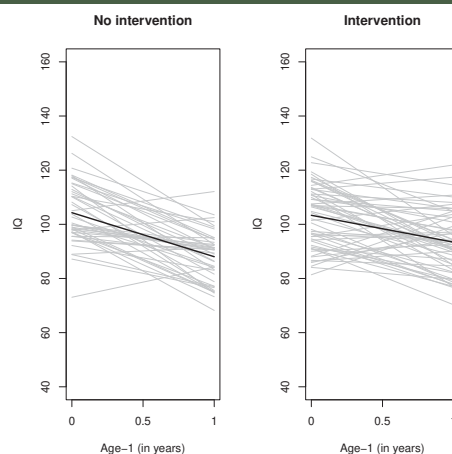
$$\mathbf{Y}_i | \mathbf{b}_i \sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \boldsymbol{\Sigma}_i), \quad \mathbf{b}_i \sim N(0, \mathbf{D})$$

- Marginally,

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

- Thus, random effects \mathbf{b}_i reflect how the evolution of the i th subject deviates from the expected evolution $\mathbf{X}_i \boldsymbol{\beta}$, i.e., how the evolution of the i th subject deviates from the average evolution in the population
- Estimation of \mathbf{b}_i helpful for detecting outlying profiles or predicting individual trajectories

Assessing the random effects



- ⇒ \mathbf{b}_i reflect how the evolution for the i th subject deviates from the average
- ⇒ Some subjects are above/below the average at the beginning of the study
- ⇒ 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
- $f(\mathbf{b}_i) = N(0, \mathbf{D})$ can be interpreted as the prior distribution of \mathbf{b}_i , i.e., its distribution before the data are collected
- Hence, it is natural to base the prediction of \mathbf{b}_i on the posterior distribution $f(\mathbf{b}_i | \mathbf{Y}_i)$ using Bayesian methods

Assessing the random effects

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

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

$$\propto \exp \left\{ -\frac{1}{2} (\mathbf{b}_i - \mathbf{DZ}_i' \mathbf{W}_i (\mathbf{Y}_i - \mathbf{X}_i \beta))' \Lambda_i^{-1} (\mathbf{b}_i - \mathbf{DZ}_i' \mathbf{W}_i (\mathbf{Y}_i - \mathbf{X}_i \beta)) \right\}$$

for some positive definite matrix Λ_i

- Posterior distribution:

$$\mathbf{b}_i | \mathbf{Y}_i \sim N(\mathbf{DZ}_i' \mathbf{W}_i (\mathbf{Y}_i - \mathbf{X}_i \beta), \Lambda_i)$$

Assessing the random effects

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

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

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

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

- Inferences for \mathbf{b}_i should account for the variability in $\hat{\mathbf{b}}_i$

Assessing the random effects

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

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

R Output

```
> ## We are going to work with the full model for illustration
> summary(early.lmer1)

Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: cog ~ 1 + age0 * program + (1 + age0 | id)
Data: early.int1

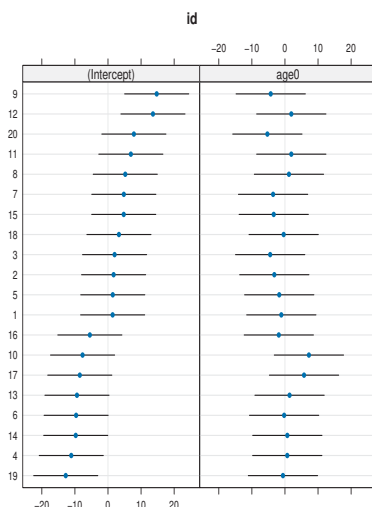
            AIC      BIC    logLik deviance df.resid
      2332.5    2362.4   -1158.3   2316.5      301

Scaled residuals:
      Min       1Q   Median       3Q      Max
-2.25362 -0.59088  0.02131  0.56850  2.29366

Random effects:
Groups   Name              Variance Std.Dev. Corr
id       (Intercept)  84.02      9.166
         age0         39.44      6.281   -0.55
Residual             60.31      7.766
Number of obs: 309, groups: id, 103

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  104.3007    1.7274 102.9999  60.380 < 2e-16 ***
age0        -16.2555    1.8860 103.0001  -8.619 8.55e-14 ***
program      -0.9646    2.3020 102.9999  -0.419 0.6761
age0:program   6.3187    2.5133 103.0001   2.514 0.0135 *
---
```

Effect of early dietary intervention on children IQ



```
## Plotting the random intercept and slope for a subset of
## the data. In this case the first 20 individuals given
## in s

r.int=raneef(early.lmer1, condVar=TRUE)
s=1:20

r.int=lapply(r.int, function(x) {
  s2=which(rownames(x) %in% s)
  x=x[s2, ]

  attributes(x)$postVar=attributes(x)$postVar[, , s2]

  return(x)
})

class(r.int)="ranef.mer"
dotplot(r.int)
```

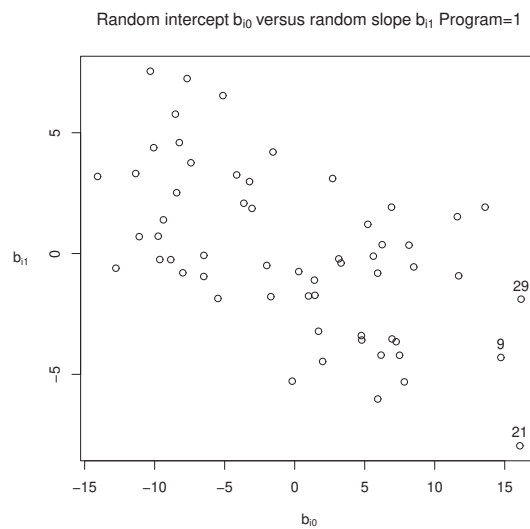
R Output

```
> ## 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=raneef(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[1:58,],
+ main="Random intercept (b0i) versus random slope (b1i).Program=1")
>
```

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



Best Linear Unbiased Prediction (BLUP)

- Often, parameters of interest are linear combinations of fixed effects in β and random effects in \mathbf{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

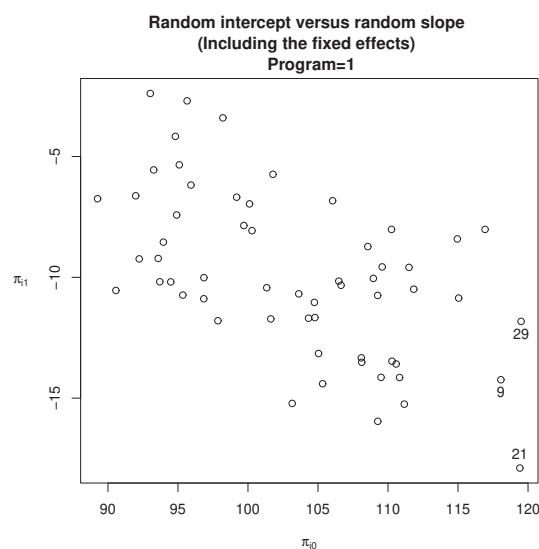
$$\pi_{1i} = \gamma_{10} + b_{1i}$$

- In general, suppose $\mathbf{u} = \mathbf{X}'_{\beta}\beta + \mathbf{X}'_b\mathbf{b}_i$ is of interest
- Conditionally on α , $\hat{\mathbf{u}} = \mathbf{X}'_{\beta}\hat{\beta} + \mathbf{X}'_b\hat{\mathbf{b}}_i$ is BLUP:
 - Linear in the observations \mathbf{Y}_i
 - Unbiased for \mathbf{u}
 - Minimum variance among all unbiased linear estimators

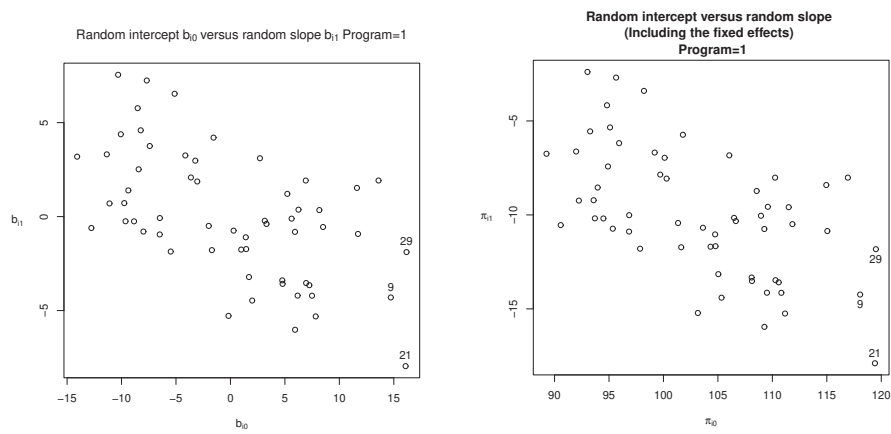
Intercept and slope: OLS versus LMM estimates

```
> ## Creating the subject specific intercepts and slopes
> ## Here we have to use the model fitted with lme
>
> ind.coef=coef(early.lme1)
> head(ind.coef)
      (Intercept)      age0      program age0:program
1    105.70683   -17.35531   -0.9646326    6.318711
2    106.00156   -19.47223   -0.9646326    6.318711
3    106.29715   -20.72297   -0.9646326    6.318711
4     93.19743   -15.55611   -0.9646326    6.318711
5    105.74548   -17.98354   -0.9646326    6.318711
6     94.66773   -16.50345   -0.9646326    6.318711
>
> prog=early.int1[early.int1$age0==0,]$program
> int.subject=ind.coef[,1]+ind.coef[,3]*prog
> slope.subject=ind.coef[,2]+ind.coef[,4]*prog
>
> plot(int.subject[1:58],slope.subject[1:58], xlab=expression(pi[i0]), ylab="",
+ main="Random intercept versus random slope (Including the fixed effects)
+ Program=1")
> mtext(expression(pi[i1]), side = 2, line = 3, las = 1)
>
```

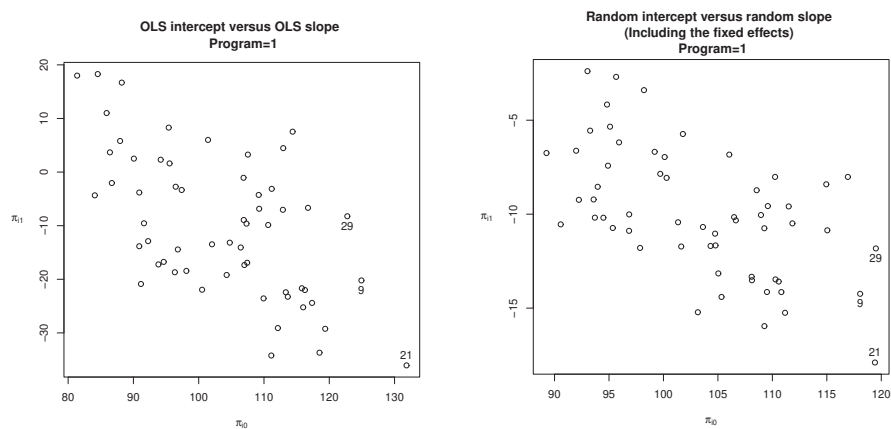
Random intercept ($\hat{\pi}_{0i}$) versus random slope ($\hat{\pi}_{1i}$)



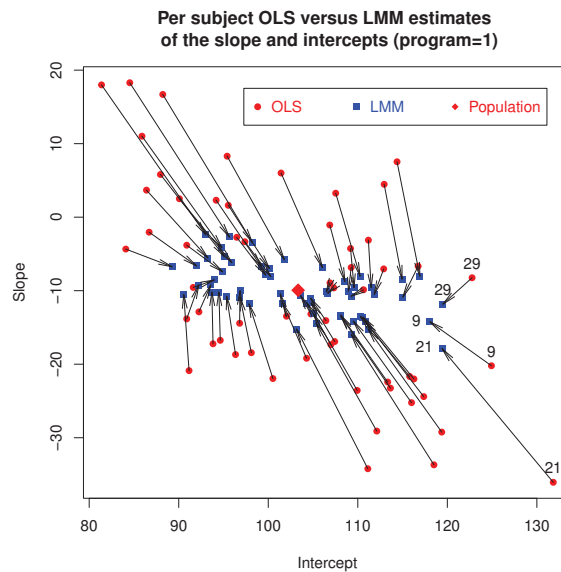
$(\hat{b}_{0i}, \hat{b}_{1i})$ versus $(\hat{\pi}_{0i}, \hat{\pi}_{1i})$



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 \hat{b}_i

- Consider the prediction of the evolution of the i th subject:

$$\begin{aligned}\hat{Y}_i &\equiv X_i \hat{\beta} + Z_i \hat{b}_i \\ &= X_i \hat{\beta} + Z_i D Z_i' V_i^{-1} (Y_i - X_i \hat{\beta}) \\ &= (I_{n_i} - Z_i D Z_i' V_i^{-1}) X_i \hat{\beta} + Z_i D Z_i' V_i^{-1} Y_i \\ &= \Sigma_i V_i^{-1} X_i \hat{\beta} + (I_{n_i} - \Sigma_i V_i^{-1}) Y_i\end{aligned}$$

- Let us look more closely at this expression

Shrinkage Estimators \hat{b}_i

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

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

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

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

Shrinkage Estimators $\hat{\mathbf{b}}_i$

- Note that the population average $\mathbf{X}_i\hat{\boldsymbol{\beta}}$ gets much weight if the residual variability Σ_i is 'large' in comparison to the total variability \mathbf{V}_i .
- This phenomenon is usually called shrinkage

The observed data are shrunk towards the prior average profile $\mathbf{X}_i\boldsymbol{\beta}$.

- This is also reflected in the fact that for any linear combination $\boldsymbol{\lambda}'\mathbf{b}_i$ of random effects,

$$\text{var}(\boldsymbol{\lambda}'\hat{\mathbf{b}}_i) \leq \text{var}(\boldsymbol{\lambda}'\mathbf{b}_i).$$

Models for cluster data

Ariel Alonso Abad

Catholic University of Leuven