

FSS 2022 - Module 13: Multilevel Analysis

Erik Fransen & Ella Roelant

StatUA
University of Antwerp, Belgium
STATUA  Universiteit
Antwerpen

September 12-16, 2022

Acknowledgements

- The slides are based on the slides of the course ‘Concepts of multilevel, longitudinal and mixed models’, by Prof. dr. Geert Verbeke, KULeuven and ‘FSS2016 - Module 2: Multilevel Analysis’ by dr. An Creemers.

References

Main References:

- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. New-York: Springer.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.
- Fitzmaurice, G.M., Laird, N.M., and Ware J.H. (2004). *Applied Longitudinal Analysis*. Hoboken, New Jersey: John Wiley & Sons.

Additional References:

- Demidenko, E. (2004). *Mixed Models: Theory and Applications*. New York: John Wiley & Sons.
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., and Schabenberger, O. (2005). *SAS for Mixed Models* (2nd ed.). Cary: SAS Press.
- Verbeke, G. and Molenberghs, G. (1997). *Linear Mixed Models In Practice: A SAS Oriented Approach*, Lecture Notes in Statistics 126. New-York: Springer.
- West, B.T., Welch, K.B., and Galecki, A.T. (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton: Chapman & Hall/CRC.

Software

Statistical software to fit multilevel models:

SAS

R: **package nlme**, package lme4, ...

Others: JMP Pro, (SPSS), ...

Part I

Introduction

Overview

① Data Structures

- Hierarchical Data
- Correlated Data

② Model Families

Hierarchical Data

Hierarchical data are obtained when the sample is taken at multiple, hierarchically ordered, levels.

Examples:

- Measurements taken on patients, at multiple visits after their treatment.
- Growth curves of children, animals, plants, . . .
- Survey in which all members from each of a sample of families are questioned.
- Survey in which 10 habitants from each of a sample of cities are questioned.
- Exam results from students from a sample of schools.
- ...

Hierarchical Data

- These are examples of **two-level** data structures, but extensions to multiple levels are possible:

10 cities

- In each: 5 schools
- In each: 5 students
- Each student given the test twice

- Terminologies:

- Repeated Measures
- Longitudinal Data
- Multilevel Data
- ...

Correlated Data

Example: Longitudinal Body Weight

- Consider a body weight experiment in which body weight is measured on a daily basis, for a sample of participants.
- It is natural to assume body weights from different subjects to be independent from each other.
- Body weights measured on the same subject are expected to be correlated.

**Should this correlation be accounted for in the analysis?
If yes, how?**

Correlated Data

Example: Comparing BMI between Males and Females

- Suppose interest is in comparing the average BMI between males and females, based on 100 observations from each population.
⇒ Natural analysis: **Two-sample, unpaired *t*-test**

- Suppose the 100 males and 100 females are married couples.
 - The BMI of spouses is likely to be correlated.
⇒ Natural analysis: **Paired *t*-test**

Correlated Data

Conclusion

- Hierarchical data structures often yield data which cannot be assumed independent.
- From a statistical perspective, the key issue in modelling hierarchical data is how to account for the association between observations.
- Alternative Terminology:
 - Repeated Measures
 - Longitudinal Data
 - Multilevel Data
 - **Correlated Data**
 - ...

Overview of Model Families

- Since hierarchical data are correlated, all traditional models in statistics need a counterpart for correlated data.
- Many different models have been proposed in the statistical literature.
- We focus on mixed models which explicitly model the various levels in the data structure.

| Cross-Sectional Data | ⇒ | Hierarchical Data |
|------------------------------|---|--|
| Linear Regression Models | ⇒ | Linear Mixed Models |
| Generalized Linear Models | ⇒ | Generalized Linear Mixed Models |
| Non-Linear Regression Models | ⇒ | Non-Linear Mixed Models |

Part II

Linear Mixed Models

Overview

3 Motivating Examples

- The Captopril Data
- The Lizard Data
- The Paired *t*-Test Revisited

4 The Linear Mixed Model

5 More Examples

6 Estimation of Random Effects

7 Concluding Remarks

The Captopril Data

Background

- 15 patients with hypertension
- The response of interest is the supine blood pressure, before and after treatment with CAPTOPRIL.
- Research Question:

How does treatment affect BP?

The Captopril Data

The Data

| Patient | Before | | After | |
|---------|--------|-----|-------|-----|
| | SBP | DBP | SBP | DBP |
| 1 | 210 | 130 | 201 | 125 |
| 2 | 169 | 122 | 165 | 121 |
| 3 | 187 | 124 | 166 | 121 |
| 4 | 160 | 104 | 157 | 106 |
| 5 | 167 | 112 | 147 | 101 |
| 6 | 176 | 101 | 145 | 85 |
| 7 | 185 | 121 | 168 | 98 |
| 8 | 206 | 124 | 180 | 105 |
| 9 | 173 | 115 | 147 | 103 |
| 10 | 146 | 102 | 136 | 98 |
| 11 | 174 | 98 | 151 | 90 |
| 12 | 201 | 119 | 168 | 98 |
| 13 | 198 | 106 | 179 | 110 |
| 14 | 148 | 107 | 129 | 103 |
| 15 | 154 | 100 | 131 | 82 |

Summary Statistics

| Average (mm Hg) |
|--------------------|
| Diastolic Before: |
| Diastolic After: |
| Systolic Before: |
| Systolic After: |

Diastolic Before: 112.333

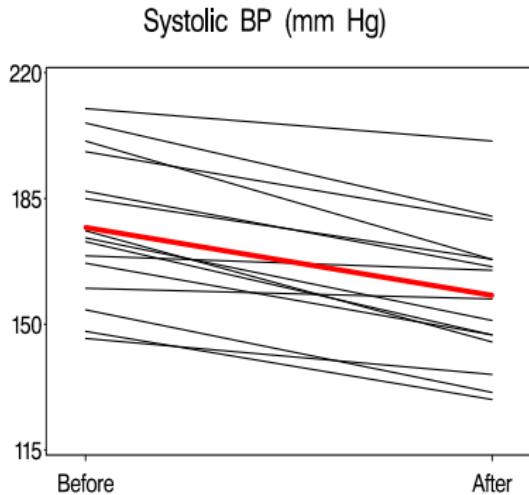
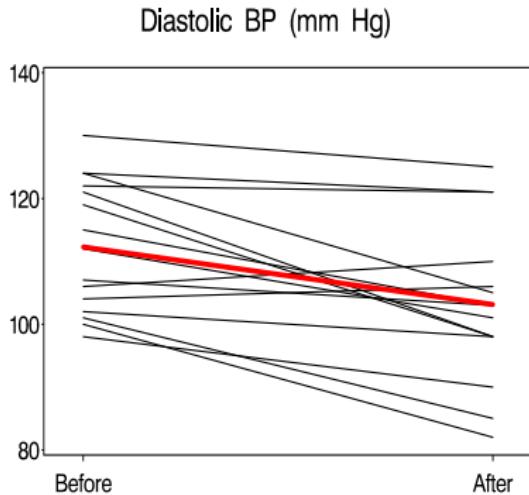
Diastolic After: 103.067

Systolic Before: 176.933

Systolic After: 158.000

The Captopril Data

Individual & Average Profiles



Paired *t*-Test

- Consider an analysis of diastolic BP:

| Average (mm Hg) | |
|-------------------|---------|
| Diastolic Before: | 112.333 |
| Diastolic After: | 103.067 |

- There is an average decrease of more than 9 mm HG.
- The classical analysis of paired data is based on comparisons within subjects:

$$\Delta_i = Y_{i1} - Y_{i2}, \quad i = 1, 2, \dots, 15$$

positive $\Delta_i \Rightarrow$ corresponds to a decrease of the BP
negative $\Delta_i \Rightarrow$ corresponds to an increase

Paired *t*-Test

- Testing for treatment effect is now equivalent to testing whether the average difference μ_Δ equals zero.
- R output

```
> round(summaryStats,digits=3)
      N      Mean Std Dev
BPbefore 15 112.333 10.472
BPafter   15 103.067 12.555
difference 15    9.267  8.614
> |
```

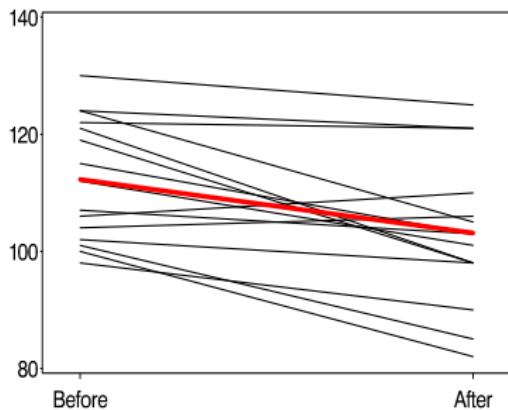
```
> round(resultPairedT,digits=3)
          DF t-value p-value
[1,] 14    4.166  0.001
> |
```

- Hence, the average change in BP is significantly different from zero ($p = 0.0010$).

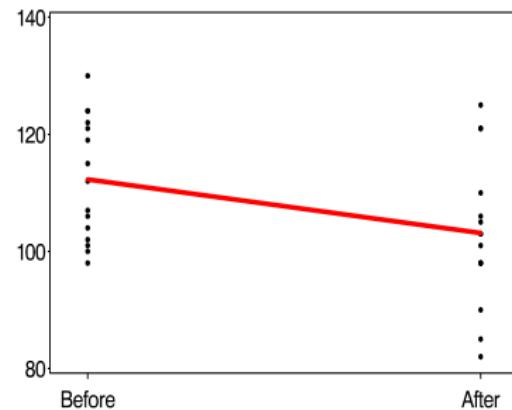
Paired vs. Unpaired *t*-Test

- What if the Captopril data were analyzed using an unpaired *t*-test?

Paired *t*-test



Unpaired *t*-test



Paired vs. Unpaired *t*-Test

- Results from unpaired and paired *t*-tests, respectively:

Unpaired

```
> round(resultVarTest,digits=3)
      Num DF Den DF t-value p-value
[1,]     14    14   1.437  0.506
>
> round(resultTTest,digits=3)
      DF t-value p-value
[1,] 28 -2.195  0.037
```

Paired

```
> round(resultPairedT,digits=3)
      DF t-value p-value
[1,] 14  4.166  0.001
> |
```

Paired vs. Unpaired *t*-Test

- Comparing results from unpaired and paired analysis:

| | | Unpaired | Paired |
|-----------------------------------|---------------|----------|--------|
| Estimated Difference | \bar{d} | 9.2667 | 9.2667 |
| Std. Err. of Estimated Difference | $se(\bar{d})$ | 4.2214 | 2.2243 |
| Test Statistic Value | t | 2.20 | 4.17 |
| <i>p</i> -value | p | 0.0366 | 0.0010 |

- Although both tests lead to a significant result, there is a serious difference in *p*-values.
- ⇒ Ignoring the paired nature of the data can lead to wrong conclusions!

Conclusion

15 x 2 measurements \neq 30 x 1 measurement

- The correlation *cannot* be ignored in the analysis.
- In the paired *t*-test, the correlation problem is circumvented by taking within-subject differences,

$$\Delta_i = Y_{i1} - Y_{i2}, \quad i = 1, 2, \dots, 15.$$

- How can this be extended to:
 - multiple measurements per subject?
 - include covariate information?
 - multiple levels in the data structure?

The Lizard Data

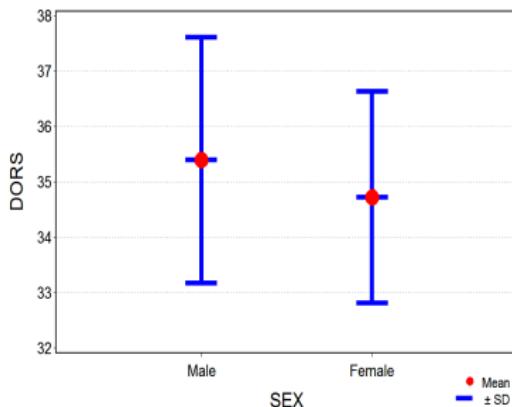
Background

- Data on 102 lizards
- Response of Interest: number of dorsal shells
- Research Question:

Is the number of dorsal shells gender-related?

The Lizard Data

Two-Sample *t*-Test



```
> round(summaryTable,digits=3)
      N   Mean Std Dev
Male  48 35.396 2.219
Female 54 34.722 1.907

> round(resultVarTest,digits=3)
      Num DF Den DF t-value p-value
[1,]    47    53   1.354  0.285

> round(resultTTest,digits=3)
      DF t-value p-value
[1,] 100   1.648  0.102
```

Hence, the small observed difference is not significant ($p = 0.1024$).

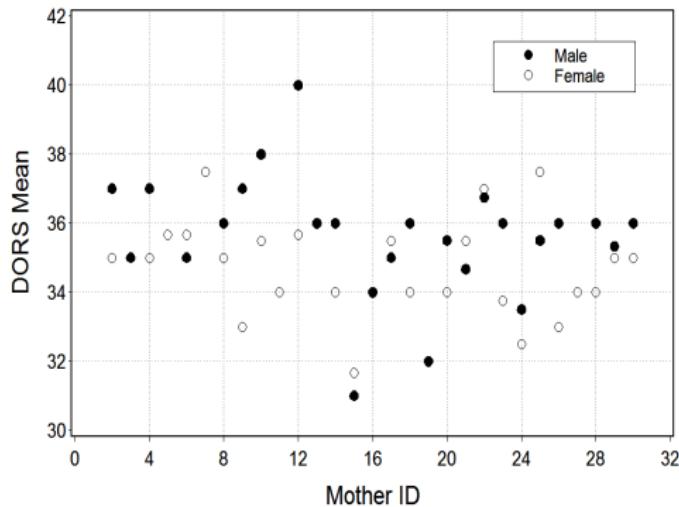
The Lizard Data

Data Features

- A typical aspect of the data is that some animals have the same mother.
 - We have 102 lizards from 29 mothers.
 - Mother effects might be present.
- ⇒ **Hence, a comparison between male and female animals should be based on within-mother comparisons.**

The Lizard Data

Within-Mother Comparisons



- much between-mother variability
- often, males (considerably) higher than females
- in cases where females higher than males, small differences

The Lizard Data

Remarks

- The non-significant *t*-test result may be due to the between-mother variability.
- This is an example of **clustered data**.
⇒ Observations are clustered within mothers.
- Measurements within mothers can be expected to be more alike than measurements from different mothers.
- We expect **correlated observations within mothers** and **independent observations between mothers**.
- How can we correct for differences between mothers?

Two-Way ANOVA

- An obvious first choice to test for a 'sex' effect, correcting for 'mother' effects, is a two-way ANOVA with factors 'sex' and 'mother'.
- The mother effect then represents the variability between mothers.
- Let Y_{ij} be the j^{th} measurement on the i^{th} mother, and let x_{ij} be 0 for males and 1 for females.
- The model then equals:

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}.$$

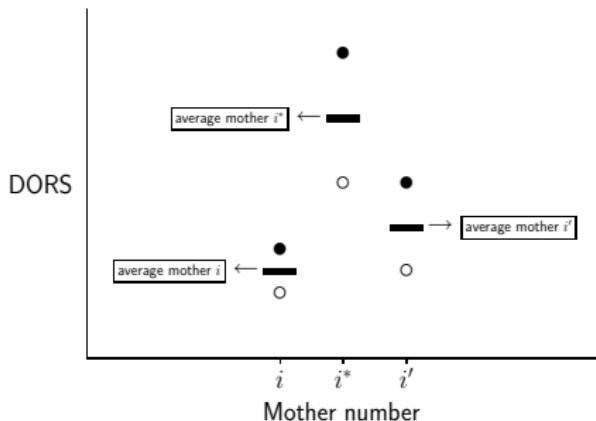
Two-Way ANOVA

- Parameter interpretation:
 - Overall mean: μ
 - Gender effect: β is the parameter of interest, the average difference between males and females.
 - Mother effect: α_i
- Since the model is overparameterized, restrictions are needed, e.g.,

$$\sum_i \alpha_i = 0.$$
- Residual distribution: $\varepsilon_{ij} \sim N(0, \sigma_{res}^2)$.

Two-Way ANOVA

Graphical Representation



R Program

```

library(car)
resultsAnova<-summary(lm(DORS ~ SEX +
    MOTHC, data=lizard))
resultsType3<-Anova(lm(DORS ~ SEX +
    MOTHC, data=lizard),type=3)

overallTest<-round(
  c(resultsAnova$fstatistic,
    pf(resultsAnova$fstatistic[1],
    resultsAnova$fstatistic[2],
    resultsAnova$fstatistic[3],
    lower.tail=FALSE)),digits=3)
overallTest

type3Test<-resultsType3[c(2,3),]
type3Test

```

Two-Way ANOVA

R Output

```
> overallTest
```

| value | numdf | dendf | value |
|-------|--------|--------|-------|
| 3.982 | 29.000 | 72.000 | 0.000 |

```
> type3Test
```

Anova Table (Type III tests)

Response: DORS

| | Sum Sq | Df | F value | Pr(>F) |
|-------|---------|----|---------|---------------|
| SEX | 16.725 | 1 | 7.1948 | 0.009062 ** |
| MOTHC | 256.938 | 28 | 3.9474 | 1.421e-06 *** |

- Note the highly significant mother effect.
- We now also obtain a significant gender effect.
- Many degrees of freedom are spent to the estimation of the mother effect, which is not even of interest.

Mixed Models

- Note the different nature of the two factors:
 - SEX: defines 2 groups of interest
 - MOTHER: defines 29 groups not of real interest. A new sample would imply other mothers.
- In practice, one considers the factor 'mother' as a **random factor**.
- The factor 'sex' is a **fixed effect**.
- Thus, the model is a **mixed model**.
- In general, models can contain multiple fixed and/or random factors.

Mixed Models

- The model is still of the form:

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}.$$

- But the fact that mothers can be assumed to be randomly selected from a population of mothers is reflected in the additional assumption

$$\alpha_i \sim N(0, \sigma_{\text{moth}}^2).$$

- Note that we still have that the α_i have mean zero. Before, we had the restriction $\sum_i \alpha_i = 0$.
- The normality assumption for the α_i is natural and mathematically convenient, **but not necessarily realistic**.
- Finally, all α_i and ε_{ij} are assumed independent.

Fitting Mixed Models in R

- Mixed models can be fitted using the *lme* function in the *nlme* package in R.
- Mixed model with 'sex' as fixed and 'mother' as random effect:

```
library(nlme)

# Parameter estimates:
resultMixed<-summary(lme(DORS~SEX , random = ~1|MOTHC, data = lizard))
fixEffSolution<-resultMixed$tTable
fixEffSolution

# Type 3 tests of fixed effects:
type3Results<-anova.lme(resultMixed,type="marginal")
type3Results

# Covariance parameter estimates
covParEst<-VarCorr(resultMixed)
rownames(covParEst)<-c("MOTHC", "Residual")
covParEst
```

Fitting Mixed Models in R

R Output

```
> fixEffSolution
      Value Std.Error DF   t-value     p-value
(Intercept) 35.4903631 0.3422179 72 103.706928 3.932505e-80
SEX2        -0.8289251 0.3220116 72  -2.574209 1.210397e-02

> type3Results
      numDF denDF   F-value p-value
(Intercept)     1    72 10755.127 <.0001
SEX            1    72     6.627  0.0121

> covParEst
MOTHC = pdLogChol(1)
      Variance StdDev
MOTHC     1.780446 1.334334
Residual  2.249915 1.499972
```

Fitting Mixed Models in R

- Estimation method is *iterative*.
- Note the significant difference between male and female animals ($p = 0.0121$).
- With the *t*-test, ignoring the mother effect, this was $p = 0.1024$.
- The average difference between males and females is estimated as $\hat{\beta} = 0.8289$.
- Covariance parameter estimates:
 - σ_{moth}^2 represents the **variability between** mothers: $\hat{\sigma}_{moth}^2 = 1.78$
 - σ_{res}^2 represents the **variability within** mothers: $\hat{\sigma}_{res}^2 = 2.25$

The Hierarchical vs. Marginal Model

- Our mixed model was given by:

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij},$$

$$\alpha_i \sim N(0, \sigma_{moth}^2), \quad \varepsilon_{ij} \sim N(0, \sigma_{res}^2), \quad \text{independent}$$

- The above model can be rewritten as:

$$Y_{ij} | \alpha_i \sim N(\mu + \alpha_i + \beta x_{ij}, \sigma_{res}^2), \quad \text{independent}$$

$$\alpha_i \sim N(0, \sigma_{moth}^2), \quad \text{independent}$$

- Each equation then corresponds to one level in the multilevel data structure.
- The model is therefore called the **hierarchical model**.

The Hierarchical vs. Marginal Model

- The hierarchical model implies a specific **marginal model**, i.e. the model which describes the marginal distribution of the outcomes:

- Normal distribution
- Mean:

$$E(Y_{ij}) = \mu + \beta x_{ij}$$

- Variance:

$$\begin{aligned}Var(Y_{ij}) &= Var(\mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}) \\&= Var(\alpha_i + \varepsilon_{ij}) \\&= Var(\alpha_i) + Var(\varepsilon_{ij}) \\Var(Y_{ij}) &= \sigma_{\text{moth}}^2 + \sigma_{\text{res}}^2\end{aligned}$$

The Hierarchical vs. Marginal Model

- The hierarchical model implies a specific **marginal model**, i.e. the model which describes the marginal distribution of the outcomes:

- Covariance between observations from different mothers i and i^* :

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{i^*k}) &= \text{Cov}(\mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}, \mu + \alpha_{i^*} + \beta x_{i^*k} + \varepsilon_{i^*k}) \\ &= \text{Cov}(\alpha_i, \alpha_{i^*}) + \text{Cov}(\alpha_i, \varepsilon_{i^*k}) + \text{Cov}(\varepsilon_{ij}, \alpha_{i^*}) + \text{Cov}(\varepsilon_{ij}, \varepsilon_{i^*k}) \\ &= 0 \end{aligned}$$

- Covariance between observations j and k , $j \neq k$, from the same mother i :

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{ik}) &= \text{Cov}(\mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}, \mu + \alpha_i + \beta x_{ik} + \varepsilon_{ik}) \\ &= \text{Cov}(\alpha_i, \alpha_i) + \text{Cov}(\alpha_i, \varepsilon_{ik}) + \text{Cov}(\varepsilon_{ij}, \alpha_i) + \text{Cov}(\varepsilon_{ij}, \varepsilon_{ik}) \\ &= \text{Var}(\alpha_i) \\ &= \sigma_{\text{moth}}^2 \end{aligned}$$

The Hierarchical vs. Marginal Model

- The total variability, correcting for gender differences is decomposed as **within-cluster variability** and **between-cluster variability**:

$$\begin{aligned}\sigma^2 &= \sigma_{moth}^2 + \sigma_{res}^2 \\ 4.03 &= 1.78 + 2.25\end{aligned}$$

- The ‘mother’ factor explains $1.78/4.03 = 44\%$ of the total variability, after correction for gender.

The Hierarchical vs. Marginal Model

- Observations from different mothers are assumed independent.
- Observations from the same mother are correlated with correlation coefficient:

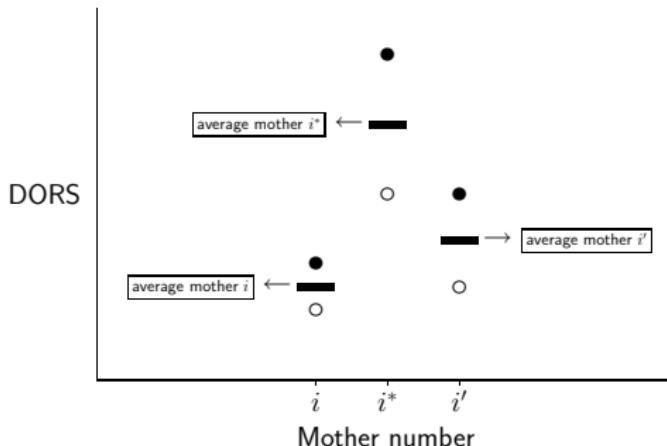
$$\rho_I = \frac{Cov(Y_{ij}, Y_{ik})}{\sqrt{Var(Y_{ij})}\sqrt{Var(Y_{ik})}} = \frac{\sigma_{moth}^2}{\sigma_{moth}^2 + \sigma_{res}^2} = \frac{1.78}{1.78 + 2.25} = 0.44$$

⇒ **intraclass correlation**

- Note how the mixed model accounts for the correlation in the data through the random effects α_i .
 - between-cluster variability >> within-cluster variability ⇒ ρ_I high
 - between-cluster variability << within-cluster variability ⇒ ρ_I low

The Hierarchical vs. Marginal Model

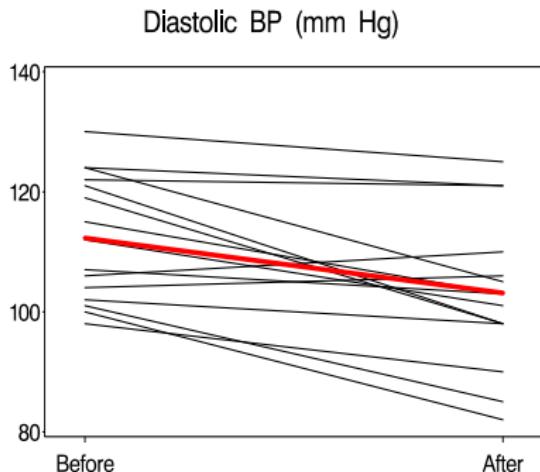
Graphical Representation



- Much between-cluster variability implies that observations from the same cluster are 'more alike' than observations from different clusters.

The Captopril Data

- A paired *t*-test analysis of the Captopril data yields:



```
> round(summaryStats,digits=3)
```

| | N | Mean | Std Dev |
|------------|----|---------|---------|
| BPbefore | 15 | 112.333 | 10.472 |
| BPafter | 15 | 103.067 | 12.555 |
| difference | 15 | 9.267 | 8.614 |

```
> |
```

```
> round(resultPairedT,digits=3)
```

| | DF | t-value | p-value |
|-------|----|---------|---------|
| [1,] | 14 | 4.166 | 0.001 |

```
> |
```

The Captopril Data

An alternative analysis could be based on a mixed model.

- Let Y_{ij} be the observation for the i^{th} subject, taken at time point $t_j = 0, 1$:

$$t_j = \begin{cases} 0, & \text{if before treatment} \\ 1, & \text{if after treatment} \end{cases}$$

- The mixed model is then of the form:

$$Y_{ij} = \mu + \alpha_i + \beta t_j + \varepsilon_{ij},$$

$$\alpha_i \sim N(0, \sigma_{\text{subj}}^2), \quad \varepsilon_{ij} \sim N(0, \sigma_{\text{res}}^2), \quad \text{independent}$$

The Captopril Data

$$Y_{ij} = \mu + \alpha_i + \beta t_j + \varepsilon_{ij},$$

$$\alpha_i \sim N(0, \sigma_{subj}^2), \quad \varepsilon_{ij} \sim N(0, \sigma_{res}^2), \quad \text{independent}$$

- The α_i are **subject-specific effects**, reflecting that some patients naturally have higher BPs than others, irrespective of the treatment.
- Assuming that subjects are randomly sampled from a population of patients, it is natural to assume the α_i to be random.
- The α_i reflect the variability between patients.

Analysis in R

R Program

```
captoprilDiaLong$TIME<-
  as.factor(ifelse
    (captoprilDiaLong$TIME=="before",0,1))

# Parameter estimates:
resultMixed<-summary(lme
  (BP~TIME , random = ~1|ID,
   data = captoprilDiaLong))
fixEffSolution<-resultMixed$tTable
fixEffSolution

# Type 3 tests of fixed effects:
type3Results<-anova.lme(resultMixed)
type3Results

# Covariance parameter estimates
covParEst<-VarCorr(resultMixed)
covParEst
```

R Output

```
> fixEffSolution
      Value Std.Error DF t-value p-value
(Intercept) 112.333333 2.984995 14 37.632676 1.811654e-15
TIME1        -9.266667  2.224247 14 -4.166204 9.510697e-04

> type3Results
      numDF denDF F-value p-value
(Intercept)     1     14 1511.6285 <.0001
TIME          1     14  17.3573  0.001

> covParEst
ID = pdLogChol(1)
      Variance StdDev
(Intercept) 96.54832 9.825900
Residual    37.10457 6.091352
```

Analysis in R

- The average difference in BP is estimated as $\hat{\beta} = -9.27$.
- We obtain the same result as with the paired *t*-test:

$$F = t^2 = \left(\frac{9.27}{\frac{8.61}{\sqrt{15}}} \right)^2 = 17.36, \quad 14 \text{ degrees of freedom}$$

- Covariance parameter estimates:
 - σ_{subj}^2 represents the **variability between** patients: $\hat{\sigma}_{subj}^2 = 96.55$
 - σ_{res}^2 represents the **variability within** patients: $\hat{\sigma}_{res}^2 = 37.10$

The Hierarchical vs. Marginal Model

- The mixed model can again be viewed as a hierarchical model:

$$Y_{ij} | \alpha_i \sim N(\mu + \alpha_i + \beta t_j, \sigma_{res}^2), \quad \text{independent}$$

$$\alpha_i \sim N(0, \sigma_{subj}^2), \quad \text{independent}$$

- The implied marginal model is again a normal one:

- Expectation: $E(Y_{ij}) = \mu + \beta t_j$

- Variance: $Var(Y_{ij}) \equiv \sigma^2 = \sigma_{subj}^2 + \sigma_{res}^2$

- Observations from different patients are independent.

- Observations from the same patient are correlated:

$$\rho_I = Corr(Y_{i1}, Y_{i2}) = \frac{\sigma_{subj}^2}{\sigma_{subj}^2 + \sigma_{res}^2}$$

The Hierarchical vs. Marginal Model

- In our example, the total variability, not explained by the systematic treatment effect, equals:

$$\sigma^2 = \sigma_{subj}^2 + \sigma_{res}^2 = 96.55 + 37.10 = 133.65.$$

- The between-subject variability accounts for $96.55/133.65 = 72.24\%$ of all variability.
- The within-subject correlation is:

$$\rho_I = \text{Corr}(Y_{i1}, Y_{i2}) = \frac{\sigma_{subj}^2}{\sigma_{subj}^2 + \sigma_{res}^2} = \frac{96.55}{96.55 + 37.10} = 0.7224.$$

⇒ large amount of between-subject variability, relative to the within-subject variability

The Hierarchical vs. Marginal Model

- The intraclass correlation *does not* equal the Pearson correlation between the BP before and after treatment, e.g., $r = 0.7343$.

```
> cor(captoprilDiaWide$BPbefore, captoprilDiaWide$BPafter)  
[1] 0.7343002  
> |
```

- The reason for this difference is that the Pearson correlation does not assume the variances of the BP before and after treatment to be equal.

The Hierarchical vs. Marginal Model

- The mixed model used assumes constant variance:

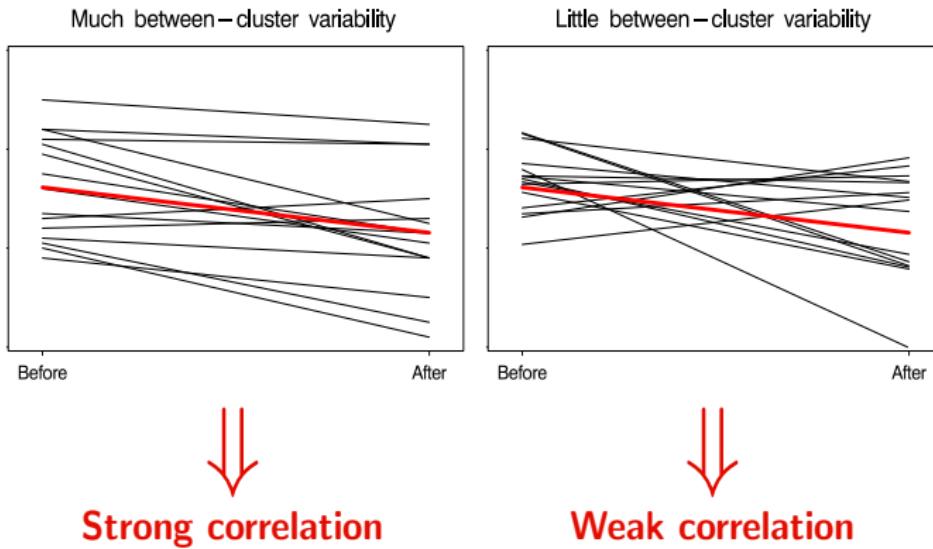
$$\sigma^2 = \text{Var}(Y_{i1}) = \text{Var}(Y_{i2}) = \sigma_{\text{subj}}^2 + \sigma_{\text{res}}^2 = 133.65 \Rightarrow \sigma = 11.56.$$

- Summary statistics for both measurements:

```
> round(summaryStats,digits=3)
      N   Mean Std Dev
BPbefore 15 112.333 10.472
BPafter  15 103.067 12.555
difference 15    9.267   8.614
> |
```

The Hierarchical vs. Marginal Model

Graphical Representation



Conclusion

- The simplest example of clustered data are paired observations, typically analyzed using a paired *t*-test.
- Traditionally, the within-pair correlation is circumvented by taking within-pair differences $\Delta_i = Y_{i1} - Y_{i2}$, which are then analyzed using a one-sample *t*-test.
- Hence, mixed models can be viewed as an extension of the paired *t*-test to:
 - more than 2 observations per cluster
 - unbalanced data: unequal number of measurements per cluster
 - models with covariates, e.g., 'sex', or others
 - models with multiple random effects (more later)

Overview

3 Motivating Examples

4 The Linear Mixed Model

- The Growth Curves Data
- The Linear Mixed Model
- Hierarchical versus Marginal Model
- Concluding Remarks

5 More Examples

6 Estimation of Random Effects

7 Concluding Remarks

The Growth Curves Data

Background

- Taken from Goldstein (1979)
- The height of 20 schoolgirls, with small, medium or tall mothers, was measured over a 4-year period:

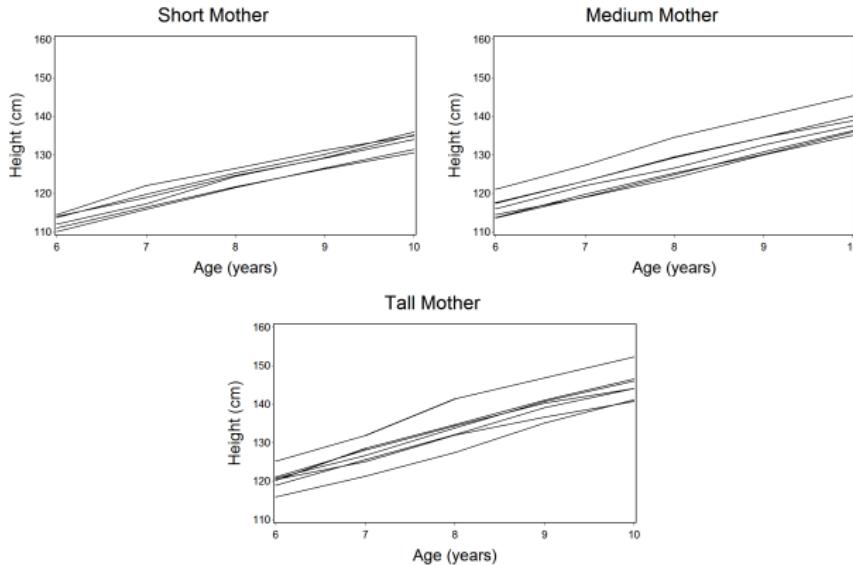
| | Mother's Height | Children Numbers |
|----------------|------------------|------------------|
| small mothers | < 155 cm | 1 → 6 |
| medium mothers | [155 cm, 164 cm] | 7 → 13 |
| tall mothers | > 164 cm | 14 → 20 |

- Research Question:

Is growth related to height of the mother?

The Growth Curves Data

Individual Profiles



The Growth Curves Data

Remarks

- almost perfect linear relation between Age and Height
- much variability between girls
- little variability within girls
- fixed number of measurements per subject
- measurements taken at fixed time points

The Model

- We will assume a linear relation between Age and Height, possibly different for the different groups.
- With **cross-sectional data**, the appropriate model would be an **ANCOVA model**:
 - Covariate: Age
 - Factor: Group
 - Interaction: Age*Group
- With **longitudinal data**, the observations are **clustered within children**, implying within-child correlation.
- Correction for the variability between children is done through a **random child effect**.

The Model

- As before, let Y_{ij} be the j^{th} height measurement for the i^{th} cluster (e.g., child), taken at time t_j (e.g., age).
- GROUP has 3 levels (1: short, 2: medium, 3: tall)

The Model

- Our model is of the form:

$$Y_{ij} = \beta_1 + \beta_2 t_j + \beta_3(\text{GROUP2}) + \beta_4(\text{GROUP3}) \\ + \beta_5(\text{GROUP2})t_j + \beta_6(\text{GROUP3})t_j + b_i + \varepsilon_{ij}$$

which leads to

$$Y_{ij} = \begin{cases} \beta_1 + \beta_2 t_j + b_i + \varepsilon_{ij}, & \text{if short mother} \\ \beta_1 + \beta_3 + (\beta_2 + \beta_5)t_j + b_i + \varepsilon_{ij}, & \text{if medium mother} \\ \beta_1 + \beta_4 + (\beta_2 + \beta_6)t_j + b_i + \varepsilon_{ij}, & \text{if tall mother} \end{cases}$$

- As before, it is assumed that random effects b_i are normal with mean zero and variance σ_{child}^2 .
- The errors ε_{ij} are normal with mean zero and variance σ_{res}^2 .

Analysis in R

R Program

```
# Parameter estimates:  
resultMixed<-summary(lme(HEIGHT~AGE + GROUP + AGE:GROUP ,  
    random = ~1|CHILD, data = growthcurves))  
fixEffSolution<-resultMixed$tTable  
fixEffSolution  
  
# Type 3 tests of fixed effects:  
type3Results<-anova.lme(resultMixed,type="marginal")  
type3Results  
  
# Covariance parameter estimates  
covParEst<-VarCorr(resultMixed)  
covParEst
```

Analysis in R

Relevant R Output

```
> covParEst
CHILD = pdLogChol(1)
      Variance StdDev
(Intercept) 8.9603304 2.9933811
Residual     0.7695751 0.8772543
```

- Covariance parameter estimates:

- σ_{child}^2 represents the **variability between** children: $\hat{\sigma}_{child}^2 = 8.96$
- σ_{res}^2 represents the **variability within** children: $\hat{\sigma}_{res}^2 = 0.77$

Analysis in R

Relevant R Output

```
> fixEffSolution
      Value Std.Error DF   t-value     p-value
(Intercept) 81.3000000 1.5296800 77 53.1483693 1.907493e-62
AGE          5.2700000 0.1132530 77 46.5329650 3.915043e-58
GROUP2       1.6742857 2.0846025 17  0.8031678 4.329645e-01
GROUP3       1.8228571 2.0846025 17  0.8744387 3.940655e-01
AGE:GROUP2   0.2971429 0.1543379 77  1.9252749 5.788775e-02
AGE:GROUP3   0.9785714 0.1543379 77  6.3404486 1.431054e-08

> type3Results
      numDF denDF   F-value p-value
(Intercept)    1     77 2824.7492 <.0001
AGE            1     77 2165.3168 <.0001
GROUP          2     17   0.4601  0.6389
AGE:GROUP      2     77  21.6615 <.0001
```

- The hypothesis of interest is $H_0 : \beta_5 = \beta_6 = 0$, which corresponds to testing the interaction AGE*GROUP.
 - highly significant difference between the slopes from the three groups ($p < 0.0001$)

The Hierarchical vs. Marginal Model

- The mixed model can again be viewed as a hierarchical model:

$$Y_{ij}|b_i = \begin{cases} N(\beta_1 + \beta_2 t_j + b_i, \sigma_{res}^2), & \text{if short mother} \\ N(\beta_1 + \beta_3 + (\beta_2 + \beta_5) t_j + b_i, \sigma_{res}^2), & \text{if medium mother} \\ N(\beta_1 + \beta_4 + (\beta_2 + \beta_6) t_j + b_i, \sigma_{res}^2), & \text{if tall mother} \end{cases}$$

The Hierarchical vs. Marginal Model

- The implied marginal model is again a normal one:
 - Expectation:

$$E(Y_{ij}) = \begin{cases} \beta_1 + \beta_2 t_j, & \text{if short mother} \\ \beta_1 + \beta_3 + (\beta_2 + \beta_5) t_j, & \text{if medium mother} \\ \beta_1 + \beta_4 + (\beta_2 + \beta_6) t_j, & \text{if tall mother} \end{cases}$$

- Variance: $Var(Y_{ij}) \equiv \sigma^2 = \sigma_{child}^2 + \sigma_{res}^2$
- Observations from different children are independent.
- Observations from the same child are correlated:

$$\rho_I = Corr(Y_{i1}, Y_{i2}) = \frac{\sigma_{child}^2}{\sigma_{child}^2 + \sigma_{res}^2}$$

The Hierarchical vs. Marginal Model

- In our example, the total variability, not explained by the systematic trends, equals:

$$\sigma^2 = \sigma_{child}^2 + \sigma_{res}^2 = 8.96 + 0.77 = 9.73.$$

- The between-child variability accounts for $8.96/9.73 = 92\%$ of all variability.
- The within-child correlation is:

$$\rho_I = \text{Corr}(Y_{i1}, Y_{i2}) = \frac{\sigma_{child}^2}{\sigma_{child}^2 + \sigma_{res}^2} = \frac{8.96}{8.96 + 0.77} = 0.9209.$$

⇒ large amount of between-child variability, relative to the within-child variability

General linear hypotheses

- To obtain inferences about specific contrasts involving the fixed effects: use *glht* in package *multcomp*
- Pairwise comparisons between the slopes are obtained using the following program:

```
library(multcomp)
groupComp<-rbind("medium - small" = c(0,0,0,0,1,0),
                  "tall - small" = c(0,0,0,0,0,1),
                  "tall - medium" = c(0,0,0,0,-1,1))
summary(glht(resultMixed,linfct=groupComp),test=adjusted("none"))
```

General linear hypotheses

- “medium - small”: $\beta_5 = 0$

$$[\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5 \quad \beta_6] \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \beta_1*0 + \beta_2*0 + \beta_3*0 + \beta_4*0 + \beta_5*1 + \beta_6*0 = 0$$

- “tall - small”: $\beta_6 = 0$

$$[\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5 \quad \beta_6] \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \beta_1*0 + \beta_2*0 + \beta_3*0 + \beta_4*0 + \beta_5*0 + \beta_6*1 = 0$$

General linear hypotheses

- “tall - medium”: $\beta_6 - \beta_5 = 0$

$$[\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5 \quad \beta_6] \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

$$\begin{aligned} &= \beta_1 * 0 + \beta_2 * 0 + \beta_3 * 0 + \beta_4 * 0 - \beta_5 * 1 + \beta_6 * 1 \\ &= -\beta_5 + \beta_6 = 0 \end{aligned}$$

General linear hypotheses

Relevant R Output

```
> summary(glht(resultMixed, linfct=groupComp), test=adjusted("none"))

Simultaneous Tests for General Linear Hypotheses

Fit: lme.formula(fixed = HEIGHT ~ AGE + GROUP + AGE:GROUP, data = growthcurves,
random = ~1 | CHILD)

Linear Hypotheses:
Estimate Std. Error z value Pr(>|z|)
medium - small == 0 0.2971 0.1543 1.925 0.0542 .
tall - small == 0 0.9786 0.1543 6.340 2.29e-10 ***
tall - medium == 0 0.6814 0.1483 4.595 4.32e-06 ***
---
Signif. codes: 0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1
(Adjusted p values reported -- none method)
```

- The difference between the slopes is mainly explained from the difference between the third group on one hand, and the other two groups on the other hand.

Random Intercepts Model

- The model, used to describe the growth curves, was:

$$Y_{ij} = \begin{cases} (\beta_{1S} + b_i) + \beta_{2S}t_j + \varepsilon_{ij}, & \text{if short mother} \\ (\beta_{1M} + b_i) + \beta_{2M}t_j + \varepsilon_{ij}, & \text{if medium mother} \\ (\beta_{1T} + b_i) + \beta_{2T}t_j + \varepsilon_{ij}, & \text{if tall mother} \end{cases}$$

- This can be interpreted as an ANCOVA model, but with child-specific intercepts b_i .
- Such a b_i represents the deviation of the intercept of a specific child from the average intercept in the group to which that child belongs.

Random Intercepts Model

That is why the mixed model can also be interpreted as a **subject-specific regression model**, i.e. a regression model with *subject-specific* regression *parameters*.

Remarks

- The growth-curve dataset is an example of a longitudinal dataset.
- In **longitudinal data**, there is a natural ordering of the measurements within clusters.
- The ordering is of primary interest.
- Our random-intercepts model implies very strong assumptions:
 - Parallel profiles within all 3 groups
 - Constant variance: $\sigma^2 = \sigma_{child}^2 + \sigma_{res}^2$
 - Constant correlation within children: $\sigma_{child}^2 / (\sigma_{child}^2 + \sigma_{res}^2)$

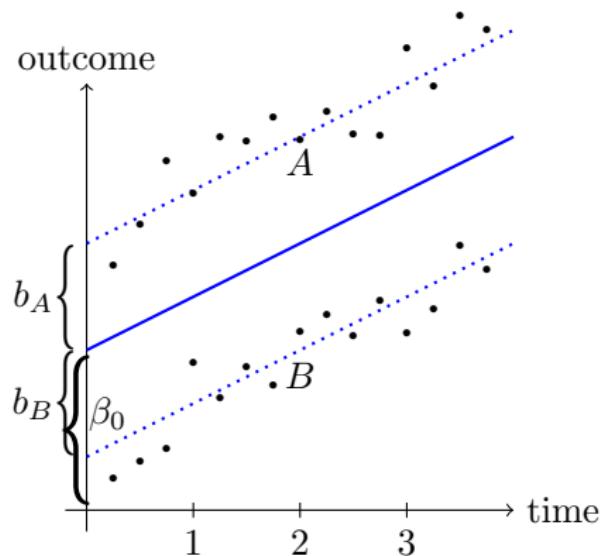
Remarks

- Hence, the marginal model implicitly assumes that the variance remains constant over time and that the correlation is the same between **any two** measurements from the same subject
- In the case of longitudinal data, this is often not realistic.
- For example, the covariance and correlation matrix of the residuals from the ANCOVA model respectively equal:

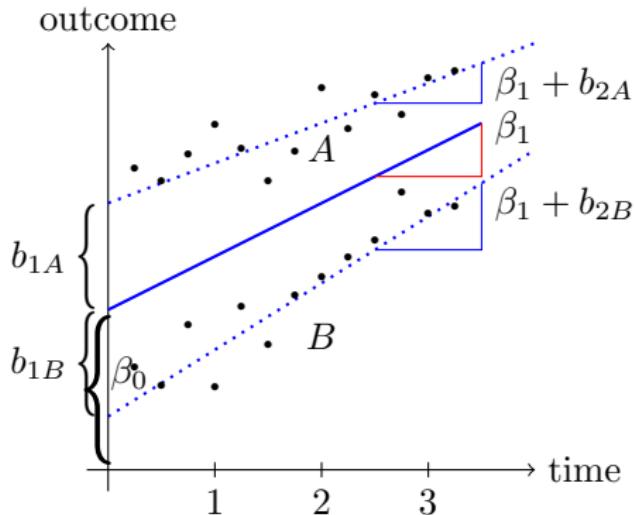
$$\begin{pmatrix} 8.7041 & 9.6119 & 11.4005 & 10.2351 & 8.5174 \\ 9.6119 & 11.3896 & 13.1437 & 11.9719 & 10.2474 \\ 11.4005 & 13.1437 & 15.8781 & 14.3981 & 12.6611 \\ 10.2351 & 11.9719 & 14.3981 & 13.4490 & 12.0644 \\ 8.5174 & 10.2474 & 12.6611 & 12.0644 & 12.0655 \end{pmatrix} \quad \begin{pmatrix} 1.0000 & 0.9654 & 0.9697 & 0.9460 & 0.8311 \\ 0.9654 & 1.0000 & 0.9774 & 0.9673 & 0.8742 \\ 0.9697 & 0.9774 & 1.0000 & 0.9853 & 0.9147 \\ 0.9460 & 0.9673 & 0.9853 & 1.0000 & 0.9471 \\ 0.8311 & 0.8742 & 0.9147 & 0.9471 & 1.0000 \end{pmatrix}$$

- This is the key motivation to further extend our mixed model.

Random Intercepts Model: Graphical



Random Slopes model: Graphical



Random Slopes

- One way to extend the random-intercepts model is to also allow the slopes to be subject-specific:

$$Y_{ij} = \begin{cases} (\beta_{1S} + b_{1i}) + (\beta_{2S} + b_{2i})t_j + \varepsilon_{ij}, & \text{if short mother} \\ (\beta_{1M} + b_{1i}) + (\beta_{2M} + b_{2i})t_j + \varepsilon_{ij}, & \text{if medium mother} \\ (\beta_{1T} + b_{1i}) + (\beta_{2T} + b_{2i})t_j + \varepsilon_{ij}, & \text{if tall mother} \end{cases}$$

- As before, the random effects are assumed to be normally distributed with mean zero:

$$\mathbf{b}_i = (b_{i1}, b_{i2})' \sim N(\mathbf{0}, \mathbf{D})$$

- The residuals ε_{ij} are still i.i.d. $N(0, \sigma^2)$ independent of the random effects \mathbf{b}_i .

Random Slopes

- D then equals the (2×2) covariance matrix of the random effects:

$$D = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$$

- Interpretation:

- d_{11} represents the variance of the intercepts b_{1i} : $Var(b_{1i}) = d_{11}$
- d_{22} represents the variance of the slopes b_{2i} : $Var(b_{2i}) = d_{22}$
- d_{12} represents the covariance between the intercepts b_{1i} and the slopes b_{2i} : $Cov(b_{1i}, b_{2i}) = d_{12}$
- The correlation between the intercepts and slopes then equals:

$$Corr(b_{1i}, b_{2i}) = \frac{d_{12}}{\sqrt{d_{11}}\sqrt{d_{22}}}$$

Analysis in R

R Program

```
resultMixedWithSlope<-summary(lme(HEIGHT~ AGE + GROUP + AGE:GROUP,  
random = ~1 + AGE|CHILD, data = growthcurves))  
  
# Covariance/correlation parameters  
VarCorr(resultMixedWithSlope)  
getVarCov(resultMixedWithSlope)  
  
# Type 3 tests of fixed effects:  
anova.lme(resultMixedWithSlope,type="marginal")
```

- As before, fixed effects are to be specified in the model form, while random effects are specified in the random statement.
- The VarCorr and getVarCov options requests the printout of the variances (d_{11}, d_{22}, σ^2), together with $\text{Corr}(b_{1i}, b_{2i})$, respectively the matrix D .

Analysis in R

Relevant R Output

```
> VarCorr(resultMixedWithSlope)
CHILD = pdLogChol(1 + AGE)
      Variance StdDev   Corr
(Intercept) 7.6028699 2.7573302 (Intr)
AGE          0.1330550 0.3647671 -0.441
Residual     0.4758166 0.6897946
```

- Covariance parameters:

- d_{11} represents the variability in subject-specific intercepts: $\hat{d}_{11} = 7.6029$
- d_{22} represents the variability in subject-specific slopes: $\hat{d}_{22} = 0.1331$
- d_{12} represents the covariance between subject-specific intercepts and subject-specific slopes: $\hat{d}_{12} = -0.4437$
- σ^2 represents the variability within children: $\hat{\sigma}^2 = 0.4758$

Analysis in R

Relevant R Output

```
> getVarCov(resultMixedWithSlope)
Random effects variance covariance matrix
      (Intercept)      AGE
(Intercept)    7.60290 -0.44371
AGE           -0.44371  0.13306
Standard Deviations: 2.7573 0.36477
```

The correlation between subject-specific intercepts and subject-specific slopes is estimated as:

$$\widehat{\text{Corr}}(b_{1i}, b_{2i}) = \frac{\widehat{d}_{12}}{\sqrt{\widehat{d}_{11}}\sqrt{\widehat{d}_{22}}} = -0.4412.$$

Analysis in R

Relevant R Output

```
> anova.lme(resultMixedWithSlope, type="marginal")
      numDF denDF  F-value p-value
(Intercept)     1     77 3691.444 <.0001
AGE             1     77  922.500 <.0001
GROUP          2     17    0.601  0.5594
AGE:GROUP       2     77    9.229  0.0003
```

- We still get a highly significant AGE*GROUP interaction term ($p = 0.0003$).

Analysis in R

Relevant R Output

```
> anova.lme(resultMixedWithSlope, type="marginal")
   numDF denDF F-value p-value
(Intercept)     1    77 3691.444 <.0001
AGE             1    77  922.500 <.0001
GROUP           2    17   0.601  0.5594
AGE:GROUP       2    77   9.229  0.0003
```

```
> type3Results
   numDF denDF F-value p-value
(Intercept)     1    77 2824.7492 <.0001
AGE             1    77 2165.3168 <.0001
GROUP           2    17   0.4601  0.6389
AGE:GROUP       2    77   21.6615 <.0001
```

NOW

Random Intercepts + Slopes

BEFORE

Random Intercepts Only

- Note the differences in test results for the fixed effects, when compared to those from the earlier random intercepts model.

Hierarchical Model

- The mixed model can again be viewed as a hierarchical model:

$$Y_{ij} | \boldsymbol{b}_i = \begin{cases} N [(\beta_{1S} + b_{1i}) + (\beta_{2S} + b_{2i})t_j, \sigma^2], & \text{if short mother} \\ N [(\beta_{1M} + b_{1i}) + (\beta_{2M} + b_{2i})t_j, \sigma^2], & \text{if medium mother} \\ N [(\beta_{1T} + b_{1i}) + (\beta_{2T} + b_{2i})t_j, \sigma^2], & \text{if tall mother} \end{cases}$$

Marginal Model

- The implied marginal model is again a normal one:
 - The expectation is the same as under the random intercepts model:

$$E(Y_{ij}) = \begin{cases} \beta_{1S} + \beta_{2St_j}, & \text{if short mother} \\ \beta_{1M} + \beta_{2Mt_j}, & \text{if medium mother} \\ \beta_{1T} + \beta_{2Tt_j}, & \text{if tall mother} \end{cases}$$

- Variance:

$$\begin{aligned} Var(Y_{ij}) &= Var(\beta_{1S} + b_{1i} + \beta_{2St_j} + b_{2it_j} + \varepsilon_{ij}) \\ &= Var(b_{1i} + b_{2it_j} + \varepsilon_{ij}) \\ &= Var(b_{1i}) + Var(b_{2it_j}) + 2Cov(b_{1i}, b_{2it_j}) + Var(\varepsilon_{ij}) \\ Var(Y_{ij}) &= d_{11} + d_{22}t_j^2 + 2d_{12}t_j + \sigma^2 \end{aligned}$$

Marginal Model

- The implied marginal model is again a normal one:
 - Covariance between observations from different children i and i^* :

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{i^*k}) &= \\ \text{Cov}(\beta_{1S} + b_{1i} + \beta_{2S}t_j + b_{2i}t_j + \varepsilon_{ij}, \beta_{1S} + b_{1i^*} + \beta_{2S}t_k + b_{2i^*}t_k + \varepsilon_{i^*k}) &= \\ \text{Cov}(b_{1i} + b_{2i}t_j + \varepsilon_{ij}, b_{1i^*} + b_{2i^*}t_k + \varepsilon_{i^*k}) & \\ \text{Cov}(Y_{ij}, Y_{i^*k}) &= 0 \end{aligned}$$

- Covariance between observations j and k , $j \neq k$, from the same child i :

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{ik}) &= \\ \text{Cov}(\beta_{1S} + b_{1i} + \beta_{2S}t_j + b_{2i}t_j + \varepsilon_{ij}, \beta_{1S} + b_{1i} + \beta_{2S}t_k + b_{2i}t_k + \varepsilon_{ik}) &= \\ \text{Cov}(b_{1i} + b_{2i}t_j + \varepsilon_{ij}, b_{1i} + b_{2i}t_k + \varepsilon_{ik}) & \\ = \text{Cov}(b_{1i}, b_{1i}) + \text{Cov}(b_{1i}, b_{2i}t_k) + \text{Cov}(b_{2i}t_j, b_{1i}) + \text{Cov}(b_{2i}t_j, b_{2i}t_k) & \\ = \text{Var}(b_{1i}) + \text{Cov}(b_{1i}, b_{2i})t_k + \text{Cov}(b_{2i}, b_{1i})t_j + \text{Cov}(b_{2i}, b_{2i})t_j t_k & \\ \text{Cov}(Y_{ij}, Y_{ik}) &= d_{11} + d_{12}(t_j + t_k) + d_{22}t_j t_k \end{aligned}$$

Marginal Model

- The implied marginal model is again a normal one:
 - Correlation between observations j and k , $j \neq k$, from the same child i :

$$\text{Corr}(Y_{ij}, Y_{ik}) = \frac{d_{11} + d_{12}(t_j + t_k) + d_{22}t_j t_k}{\sqrt{d_{11} + d_{22}t_j^2 + 2d_{12}t_j + \sigma^2} \sqrt{d_{11} + d_{22}t_k^2 + 2d_{12}t_k + \sigma^2}}$$

Variance & Association Structure

- Note how extending the random intercepts model with random slopes yields a more flexible covariance structure.

Random Intercepts Random Intercepts + Slopes

| | | |
|-----------------------|-------------------------------------|--|
| $Var(Y_{ij})$ | $\sigma_{child}^2 + \sigma_{res}^2$ | $d_{11} + d_{22}t_j^2 + 2d_{12}t_j + \sigma^2$ |
| $Cov(Y_{ij}, Y_{ik})$ | σ_{child}^2 | $d_{11} + d_{12}(t_j + t_k) + d_{22}t_j t_k$ |

- Further extension of the random effects structure would allow for even more flexible variance and correlations functions.
- Conclusion:

The role of random effects is to model the variance and association structure.

Conclusion & Terminology

- The linear mixed model is a linear regression model with two sets of regression parameters:
 - ⇒ **Fixed Effects:** β
 - ⇒ **Random Effects:** $b_i \sim N(\mathbf{0}, D)$
- The fixed effects are used to model the average outcome.
- The random effects are used to model the covariance structure.
- All parameters in D , jointly with the residual variance σ^2 , are called **variance components**.

Overview

3 Motivating Examples

4 The Linear Mixed Model

5 More Examples

- The Rat Data
- The Diabetes Project Leuven

6 Estimation of Random Effects

7 Concluding Remarks

The Rat Data

Background

- Taken from Dentistry, Leuven
- Data on 50 male Wistar rats
- Research Question:

**How does craniofacial growth
depend on testosterone production?**

The Rat Data

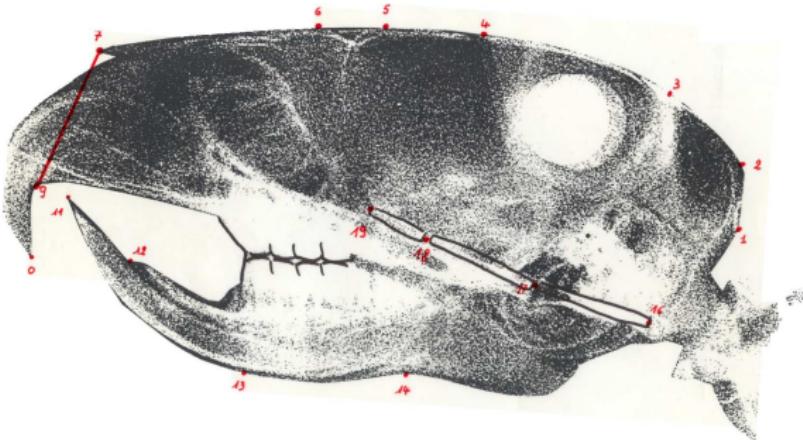
Study Design

- Randomized Experiment \Rightarrow 50 male Wistar rats are randomized to:
 - Control (15 rats)
 - Low dose of Decapeptyl (18 rats)
 - High dose of Decapeptyl (17 rats)
- treatment starts at the age of 45 days
- measurements taken every 10 days, from day 50 onwards

The Rat Data

The Response

- The responses are distances (pixels) between well defined points on x-ray pictures of the skull of each rat.



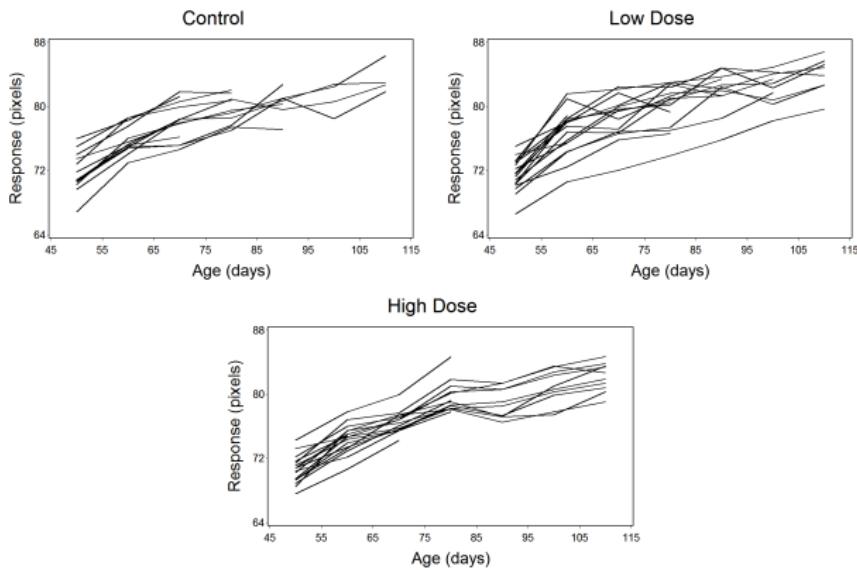
The Rat Data

The Response

- The responses are distances (pixels) between well defined points on x-ray pictures of the skull of each rat.
- Measurements with respect to the roof, base and height of the skull were taken.
- Here, we consider only one response, reflecting the height of the skull.

The Rat Data

Individual Profiles



The Rat Data

Complication

- Dropout due to anaesthesia: 56%

| Age (days) | Number of Observations | | | |
|------------|------------------------|-----|------|-------|
| | Control | Low | High | TOTAL |
| 50 | 15 | 18 | 17 | 50 |
| 60 | 13 | 17 | 16 | 46 |
| 70 | 13 | 15 | 15 | 43 |
| 80 | 10 | 15 | 13 | 38 |
| 90 | 7 | 12 | 10 | 29 |
| 100 | 4 | 10 | 10 | 24 |
| 110 | 4 | 8 | 10 | 22 |

The Rat Data

Remarks

- much variability between rats, much less variability within rats
- fixed number of measurements scheduled per subject, but not all measurements available due to dropout, for known reason
- measurements taken at fixed time points

A Linear Mixed Model

- Since linear mixed models assume a linear regression for each cluster separately, they can also be used for unbalanced data, i.e. data with unequal number of measurements per cluster.
- Note that this was also the case for the lizard data.
- Individual profiles show very similar evolutions for all rats (apart from measurement error).
- This suggests a random-intercepts model.
- Non-linearity can be accounted for by using a logarithmic transformation of the time scale:

$$\text{Age}_{ij} \longrightarrow t_{ij} = \ln[1 + (\text{Age}_{ij} - 45)/10]$$

A Linear Mixed Model

- We define the model as:

$$Y_{ij} = (\beta_0 + b_i) + (\beta_1 + \beta_2 L_i + \beta_3 H_i)t_{ij} + \varepsilon_{ij}$$

- L_i, H_i are indicator variables:

$$L_i = \begin{cases} 1, & \text{if low dose} \\ 0, & \text{otherwise} \end{cases} \quad H_i = \begin{cases} 1, & \text{if high dose} \\ 0, & \text{otherwise} \end{cases}$$

A Linear Mixed Model

- We then get the following model:

$$Y_{ij} = (\beta_0 + b_i) + (\beta_1 + \beta_2 L_i + \beta_3 H_i)t_{ij} + \varepsilon_{ij}$$

$$= \begin{cases} \beta_0 + b_i + (\beta_1 + \beta_2)t_{ij} + \varepsilon_{ij}, & \text{if low dose} \\ \beta_0 + b_i + (\beta_1 + \beta_3)t_{ij} + \varepsilon_{ij}, & \text{if high dose} \\ \beta_0 + b_i + \beta_1 t_{ij} + \varepsilon_{ij}, & \text{if control} \end{cases}$$

- Parameter Interpretation:

β_0 \Rightarrow average response at the start of the treatment
(independent of treatment)

β_1 \Rightarrow average time effect for control group

β_2, β_3 \Rightarrow deviation of average time effect in control group
for each treatment group

b_i \Rightarrow subject-specific intercepts

Fitting the Model in R

R Program

```
rat<-read.table("rat.txt",header=T,sep="\t")
rat$RAT<-as.factor(rat$RAT)
rat$logage<-log(1+(rat$AGE-45)/10)
str(rat)

resultMixed<-summary(lme(Y ~ logage + TREAT:logage, random = ~1|RAT, data = rat))

# Covariance parameters
VarCorr(resultMixed)
resultMixed$sigma^2

# Type 3 effects:

type3Results<-anova.lme(resultMixed,type="marginal")
type3Results

# Parameter estimates:
resultMixed$tTable
```

- Note the parameterization of the fixed effects.

Fitting the Model in R

Relevant R Output

```
> VarCorr(resultMixed)
RAT = pdLogChol(1)
      Variance StdDev
(Intercept) 3.564927 1.888101
Residual     1.444779 1.201990
```

- A lot of variability between rats, while little variability within rats:
 - σ_{rat}^2 represents the **variability between** rats: $\hat{\sigma}_{rat}^2 = 3.5649$
 - σ_{res}^2 represents the **variability within** rats: $\hat{\sigma}_{res}^2 = 1.4448$
- As before, the variance and correlation structure need to be explored to check model fit.

Fitting the Model in R

Relevant R Output

```
> resultMixed$Table
            Value Std.Error DF      t-value      p-value
(Intercept) 68.6073947 0.3312302 199 207.1290154 2.274693e-234
logage       7.3138409 0.2807696 199 26.0492607 4.862077e-66
logage:TREAThigh -0.4427191 0.3411844 199 -1.2975946 1.959288e-01
logage:TREATlow  0.1930477 0.3391107 199  0.5692765 5.698105e-01
> type3Results<-anova.lme(resultMixed,type="marginal")
> type3Results
    numDF denDF  F-value p-value
(Intercept)     1    199 42902.43 <.0001
logage          1    199   678.56 <.0001
logage:TREAT    2    199     2.32  0.1013
```

- The *p*-value for the TREAT:logage effect: it tests if logage effect in all three treatment groups is equal

The Diabetes Project Leuven

Background

- The **Diabetes Project Leuven**
- In Belgium, general practitioners (GPs) cannot rely on structured assistance of dieticians or diabetes nurse educators in their practice.
- The DPL intends to study the effect of implementing a structured model for chronic diabetes care on patients' clinical outcomes.
- GPs will be offered assistance and can redirect patients to the diabetes care team, consisting of a nurse educator, a dietician, an ophthalmologist and an internal medicine doctor.

The Diabetes Project Leuven

Study Design

- In the DPL, two programs were implemented and GPs were randomized to one of two groups:
 - **LIP:** Low Intervention Program (group A)
 - **HIP:** High Intervention Program (group R)
- Patients were measured twice:
 - when the program was initiated (time T0)
 - after one year (time T1)

The Diabetes Project Leuven

Study Design

- In the DPL, two programs were implemented and GPs were randomized to one of two groups:
 - **LIP:** Low Intervention Program (group A)
 - **HIP:** High Intervention Program (group R) \implies we focus on this
 - \Rightarrow 61 GPs
 - \Rightarrow 1577 patients
 - \Rightarrow number of patients per GP varies $\in [5,138]$, with a median of 47
- Patients were measured twice:
 - when the program was initiated (time T0)
 - after one year (time T1)

The Diabetes Project Leuven

The Response

- The outcome studied here is HbA1c, glycosylated hemoglobin:
 - HbA1c is a molecule in red blood cells that attaches to glucose (blood sugar).
 - High values of HbA1c reflect more glucose in blood.
 - In diabetes patients, HbA1c gives a good estimate of how well diabetes is being managed over the last 2 or 3 months.
 - Non-diabetics have values between 4% and 6%.
 - HbA1c above 7% means diabetes is poorly controlled, implying higher risk for long-term complications.

A Variety of Multilevel Models

- Let Y_{ijk} be the k^{th} measurement of HbA1c, for the j^{th} patient, of the i^{th} GP.
- We have 3-level data, hence random effects can enter the models at various levels.
- Several models for studying the longitudinal evolutions will be illustrated and compared:
 - no random effects
 - random GP effects
 - random patient effects
 - random effects for GP and patient

A Variety of Multilevel Models

Model 1: No Random Effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + \varepsilon_{ijk}, \\ \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

• R Program

```
resultNoRandom<-summary(lm(hba1c ~ time, data=diabetes))
resultNoRandom$coefficients
resultNoRandom$sigma^2
```

A Variety of Multilevel Models

Model 1: No Random Effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + \varepsilon_{ijk}, \\ \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

• Relevant R Output

```
> resultNoRandom$coefficients
            Estimate Std. Error    t value   Pr(>|t|)
(Intercept) 7.1357346 0.02822558 252.81095 0.0000000e+00
time        -0.3899089 0.04076358  -9.56513 2.271321e-21
>
> resultNoRandom$sigma^2
[1] 1.230875
```

A Variety of Multilevel Models

Model 2: Random GP Effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + \varepsilon_{ijk}, \\ a_i \sim N(0, \sigma_{GP}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

• R Program

```
resultRandomGP<-summary(lme(hba1c ~time, random = ~1|mdnr, data = diabetes))
VarCorr(resultRandomGP)
resultRandomGP$tTable
```

A Variety of Multilevel Models

Model 2: Random GP Effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + \varepsilon_{ijk}, \\ a_i \sim N(0, \sigma_{GP}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

• Relevant R Output

```
> VarCorr(resultRandomGP)
mdnr = pdLogChol(1)
      Variance   StdDev
(Intercept) 0.07097033 0.2664026
Residual    1.17090058 1.0820816
>
> resultRandomGP$tTable
      Value Std.Error DF   t-value   p-value
(Intercept) 7.1695211 0.04519497 2906 158.635377 0.000000e+00
time        -0.3872861 0.03978444 2906  -9.734613 4.64537e-22
>
```

A Variety of Multilevel Models

Model 3: Random Patient Effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + b_{j(i)} + \varepsilon_{ijk}, \\ b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

• R Program

```
resultRandomPatient<-summary(lme(hb1c ~time, random = ~1|md_patient, data = diabetes))  
VarCorr(resultRandomPatient)  
resultRandomPatient$tTable
```

A Variety of Multilevel Models

Model 3: Random Patient Effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + b_{j(i)} + \varepsilon_{ijk}, \\ b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

• Relevant R Output

```
> VarCorr(resultRandomPatient)
md_patient = pdLogChol(1)
      Variance StdDev
(Intercept) 0.6675274 0.8170235
Residual    0.5831360 0.7636334
>
> resultRandomPatient$tTable
        Value Std.Error   DF   t-value   p-value
(Intercept) 7.1392446 0.02838195 1571 251.54169 0.00000e+00
time        -0.3785085 0.02850557 1395 -13.27841 5.60818e-38
```

A Variety of Multilevel Models

Model 4: Random Effects for GP and Patient

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk},$$
$$a_i \sim N(0, \sigma_{GP}^2), \quad b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

• R Program

```
resultRandomPatientGP<-summary(lme(hba1c ~time, random = ~1| mdnr / md_patient, data = diabetes))
```

OR

```
resultRandomPatientGP<-summary(lme(hba1c ~time, random = ~1| mdnr / patientnr, data = diabetes))
```

```
VarCorr(resultRandomPatientGP)  
resultRandomPatientGP$table
```

A Variety of Multilevel Models

Model 4: Random Effects for GP and Patient

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk},$$
$$a_i \sim N(0, \sigma_{GP}^2), \quad b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

• Relevant R Output

```
> VarCorr(resultRandomPatientGP)
      Variance     StdDev
mdnr =      pdLogChol(1)
(Intercept) 0.05441162  0.2332630
patientnr = pdLogChol(1)
(Intercept) 0.61713060  0.7855766
Residual    0.58372742  0.7640206
>
> resultRandomPatientGP$tTable
    Value   Std.Error   DF   t-value   p-value
(Intercept) 7.1668523 0.04241276 1511 168.97866 0.0000000e+00
time        -0.3780113 0.02851348 1395 -13.25728 7.202624e-38
.
```

A Variety of Multilevel Models

Summary of Results

| Parameter | Model 1 Est (s.e.) | Model 2 Est (s.e.) | Model 3 Est (s.e.) | Model 4 Est (s.e.) |
|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Fixed Effects: | | | | |
| β_0 | 7.1357(0.0282) | 7.1695(0.0452) | 7.1392(0.0284) | 7.1669(0.0424) |
| β_1 | -0.3899(0.0408) | -0.3873(0.0398) | -0.3785(0.0285) | -0.3780(0.0285) |
| Variance Components: | | | | |
| σ_{GP}^2 | --- | 0.0710 | --- | 0.0544 |
| σ_{PAT}^2 | --- | --- | 0.6675 | 0.6171 |
| σ_{res}^2 | 1.2309 | 1.1709 | 0.5831 | 0.5837 |

A Variety of Multilevel Models

Variance Components

- No standard errors reported for variance components, since standard Z-tests do not produce correct tests (more later).
- The various models use different decompositions of the total variability:
 - Model 1: $\hat{\sigma}^2 = \hat{\sigma}_{res}^2 = 1.2309$
 - Model 2: $\hat{\sigma}^2 = \hat{\sigma}_{GP}^2 + \hat{\sigma}_{res}^2 = 0.0709 + 1.1710 = 1.2419$
 - Model 3: $\hat{\sigma}^2 = \hat{\sigma}_{PAT}^2 + \hat{\sigma}_{res}^2 = 0.6675 + 0.5831 = 1.2506$
 - Model 4: $\hat{\sigma}^2 = \hat{\sigma}_{GP}^2 + \hat{\sigma}_{PAT}^2 + \hat{\sigma}_{res}^2 = 0.0544 + 0.6171 + 0.5837 = 1.2552$

A Variety of Multilevel Models

Fixed Effects

- Inclusion of random effects has little effect on estimation of fixed effects but has severe impact on the standard errors:
 - larger standard errors for between-cluster effects (intercept β_0)
 - smaller standard errors for within-cluster effects (time β_1)
- There is a significant decrease in HbA1c, under all models.

Correlation Structure

- The models also imply specific correlation structures.

A Variety of Multilevel Models

Correlation Structure

- For instance, the marginal association structure implied by Model 4:
 - Observations from different GPs, i and i^* , $i \neq i^*$, are not correlated:

$$\widehat{\text{Corr}}(Y_{ijk}, Y_{i^*j^*k^*}) = 0$$

- Observations from same GP but different patients, j and j^* , $j \neq j^*$, are correlated:

$$\widehat{\text{Corr}}(Y_{ijk}, Y_{ij^*k^*}) = \frac{\widehat{\sigma}_{GP}^2}{\widehat{\sigma}_{GP}^2 + \widehat{\sigma}_{PAT}^2 + \widehat{\sigma}_{res}^2} = \frac{0.0544}{1.2552} = 0.0433$$

- Observations k and k^* , $k \neq k^*$, from same patient, are correlated:

$$\widehat{\text{Corr}}(Y_{ijk}, Y_{ijk^*}) = \frac{\widehat{\sigma}_{GP}^2 + \widehat{\sigma}_{PAT}^2}{\widehat{\sigma}_{GP}^2 + \widehat{\sigma}_{PAT}^2 + \widehat{\sigma}_{res}^2} = \frac{0.0544 + 0.6171}{1.2552} = 0.5350$$

Including Covariates at Various Levels

- Additional covariates can be added to explain variability at the different levels, or to study what patient and/or GP characteristics are related to time trends.
- We extend Model 4 with the following covariates:
 - at GP level: practice form ('one', 'two', 'more')
 - at patient level: BMI at baseline
 - if patient is a newly diagnosed diabetic (1: yes, 0: no)
- We will investigate the effect of each covariate separately.
- Obviously, models with multiple covariates are possible as well.

Including Covariates at Various Levels

Model 5: Correcting for Different Practice Forms

- R Program: Model with interaction

```
resultModel4Practice<-summary(lme(hba1c ~ practice + time + time:practice,  
      random = ~1| mdnr / patientnr, data = diabetes))  
VarCorr(resultModel4Practice)  
type3Results<-anova.lme(resultModel4Practice,type="marginal")  
type3Results  
resultModel4Practice$tTable
```

- The main effect of practice reflects differences at baseline between the various practice forms.

Including Covariates at Various Levels

Model 5: Correcting for Different Practice Forms

- Relevant R Output: Model with interaction

```
> resultModel4Practice$tTable
      Value Std.Error DF t-value p-value
(Intercept) 7.14075424 0.07505541 1511 95.1397649 0.000000e+00
practiceMor -0.11584856 0.11054521 58 -1.0479745 2.989979e-01
practiceOne 0.14067824 0.10015453 58 1.4046119 1.654697e-01
time -0.36592696 0.04975659 1393 -7.3543412 3.262007e-13
practiceMor:time 0.05051692 0.07467195 1393 0.6765179 4.988242e-01
practiceOne:time -0.06161613 0.06680061 1393 -0.9223888 3.564856e-01
```

- We notice lower average values of HbA1c the more GPs work together in group practices at baseline.
- The change over time is the largest for the practice with one GP.

Including Covariates at Various Levels

Model 5: Correcting for Different Practice Forms

- Relevant R Output: Model with interaction

```
> VarCorr(resultModel4Practice)          > type3Results
      Variance     StdDev
mdnr = pdLogChol(1)                   numDF  denDF   F-value p-value
(Intercept)  0.0543227  0.2330714    (Intercept)      1  1511 9051.575 <.0001
patientnr = pdLogChol(1)             practice       2     58   3.083  0.0534
(Intercept)  0.61524495  0.7843755   time         1 1393   54.086 <.0001
Residual     0.58381160  0.7640756  practice:time  2 1393    1.279  0.2786
```

- The practice form has no significant effect on the change over time ($p = 0.2786$).

Including Covariates at Various Levels

Model 5: Correcting for Different Practice Forms

- Relevant R Output: Model without interaction

```
> type3Results
      numDF denDF   F-value p-value
(Intercept)     1  1511  9705.303 <.0001
practice        2     58    2.165  0.1239
time            1 1395   175.673 <.0001
> resultModel4Practiceb$tTable
          Value Std.Error   DF      t-value      p-value
(Intercept) 7.14633474 0.07254021 1511  98.5154963 0.000000e+00
practiceMor -0.09170839 0.10470239    58 -0.8758959 3.846998e-01
practiceOne  0.11180451 0.09509153    58   1.1757568 2.444958e-01
time         -0.37789065 0.02851107 1395 -13.2541738 7.472853e-38
```

- The differences between types of practices are not significant $p = 0.1239$.

Including Covariates at Various Levels

Model 6: Correcting for Different BMI at Baseline

- R Program: Model with interaction

```
resultModel4BMI0<-summary(lme(hba1c ~ bmi0 + time + time:bmi0,  
    random = ~1| mdnr / patientnr, data = diabetes, na.action='na.omit'))  
VarCorr(resultModel4BMI0)  
type3Results<-anova.lme(resultModel4BMI0,type="marginal")  
type3Results  
resultModel4BMI0$tTable
```

Including Covariates at Various Levels

Model 6: Correcting for Different BMI at Baseline

- Relevant R Output: Model with interaction

```
> VarCorr(resultModel4BMI0)
    Variance     StdDev
mdnr =      pdLogChol(1)
(Intercept) 0.05847266  0.2418112
patientnr = pdLogChol(1)
(Intercept) 0.60477825  0.7776749
Residual    0.54682843  0.7394785
> type3Results<-anova.lme(resultModel4BMI0,type="marginal")
> type3Results
      numDF denDF   F-value p-value
(Intercept)     1 1391 1607.3015 <.0001
bmi0          1 1391   10.5529  0.0012
time           1 1323    0.9416  0.3321
bmi0:time     1 1323    1.6277  0.2023
> resultModel4BMI0$tblTable
      Value   Std.Error   DF   t-value      p-value
(Intercept) 6.632476791 0.165434873 1391 40.0911649 3.037833e-234
bmi0        0.017497573 0.005386309 1391   3.2485275 1.187736e-03
time       -0.157243969 0.162050691 1323  -0.9703382 3.320553e-01
bmi0:time  -0.006847709 0.005367405 1323  -1.2757951 2.022519e-01
```

- The average time trend is not significantly related to the initial BMI level ($p = 0.2023$).

Including Covariates at Various Levels

Model 6: Correcting for Different BMI at Baseline

- Relevant R Output: Model without interaction

```
> VarCorr(resultModel4BMI1)
      Variance StdDev
mdnr = pdLogChol(1)
(Intercept) 0.05831911 0.2414935
patientnr = pdLogChol(1)
(Intercept) 0.60391594 0.7771203
Residual    0.54746473 0.7399086
> type3Results<-anova.lme(resultModel4BMI1,type="marginal")
> type3Results
      numDF denDF   F-value p-value
(Intercept)     1 1391 2073.3443 <.0001
bmi0          1 1391     9.0037  0.0027
time          1 1324   160.7733 <.0001
> resultModel4BMI1$Table
      Value Std.Error DF t-value p-value
(Intercept) 6.72726232 0.147741548 1391 45.533991 6.592364e-278
bmi0        0.01429784 0.004764978 1391   3.000609 2.742357e-03
time       -0.36080145 0.028455182 1324 -12.679640 7.444150e-35
```

- We find a significantly higher baseline value for HbA1c as the initial BMI is larger ($p = 0.0027$).

Including Covariates at Various Levels

Model 7: Correcting for New Diagnosis

- R Program: Model with interaction

```
resultModel4NewDiag<-summary(lme(hba1c ~ new0 + time + time:new0,
    random = ~1| mdnr / patientnr, data = diabetes, na.action='na.omit'))
VarCorr(resultModel4NewDiag)
type3Results<-anova.lme(resultModel4NewDiag,type="marginal")
type3Results
resultModel4NewDiag$tTable
```

- The main effect of new0 reflects differences at baseline between newly diagnosed diabetics and others.

Including Covariates at Various Levels

Model 7: Correcting for New Diagnosis

- Relevant R Output: Model with interaction

```
> VarCorr(resultModel4NewDiag)          > type3Results
    Variance   StdDev
mdnr = pdLogChol(1)      0.2358830    numDF  denDF   F-value p-value
(Intercept)  0.0556408          1     1405 26553.653 <.0001
patientnr = pdLogChol(1)  0.6004118          1     1405   14.666  1e-04
(Intercept)  0.7748624          1     1321   115.153 <.0001
Residual     0.5385284          1     1321    52.824 <.0001
```

- The evolution of HbA1c differs significantly between newly diagnosed patients and others ($p < 0.0001$).

Including Covariates at Various Levels

Model 7: Correcting for New Diagnosis

- Relevant R Output: Model with interaction

```
> resultModel14NewDiag$tTable
    Value   Std.Error   DF   t-value   p-value
(Intercept) 7.1172232 0.04367656 1405 162.952917 0.0000000e+00
new0         0.4029861 0.10522766 1405   3.829660 1.339705e-04
time        -0.3148272 0.02933822 1321  -10.730958 8.143912e-26
new0:time   -0.7762397 0.10680215 1321   -7.268016 6.221721e-13
```

- Newly diagnosed patients have higher average HbA1c at baseline.
- Newly diagnosed patients have a steeper decrease in HbA1c compared to others.

Including Covariates at Various Levels

Model 7: Correcting for New Diagnosis

- We can test if, after one year, both groups are at the same level:

```
comp<-rbind("Equal at time 1" = c(0,1,0,1))  
  
summary(glht(resultModel4NewDiag,linfct=comp),test=adjusted("none"));
```

- Output:

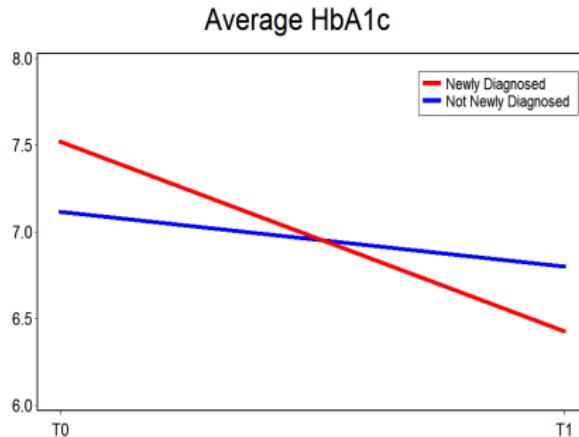
```
Linear Hypotheses:  
Estimate Std. Error z value Pr(>|z|)  
Equal at time 1 == 0 -0.3733 0.1106 -3.376 0.000736 ***  
---  
Signif. codes: 0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1  
(Adjusted p values reported -- none method)
```

- Hence, after one year, the newly diagnosed diabetics have lower average values for HbA1c ($p = 0.0007$).

Including Covariates at Various Levels

Model 8: Correcting for New Diagnosis

- Graphically:



- This can be explained by the fact that the disease gets worse over time, hence HbA1c is more difficult to keep under control.

Overview

3 Motivating Examples

4 The Linear Mixed Model

5 More Examples

6 **Estimation of Random Effects**

- Empirical Bayes Estimation
- Average versus Cluster-Specific Prediction

7 Concluding Remarks

Empirical Bayes Estimation

- Random effects reflect how specific clusters deviate from the population average.
- For example, for the Leuven diabetes project, Model 4 for the outcome Y_{ijk} being the k^{th} measurement of HbA1c, for the j^{th} patient, of the i^{th} GP, was given by:

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk},$$
$$a_i \sim N(0, \sigma_{GP}^2), \quad b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

- The parameter a_i expresses how the average HbA1c level of patients treated by GP i differs from the overall population average.
- The parameter $b_{j(i)}$ expresses how the average of patient j treated by GP i differs from the average of that specific GP.

Empirical Bayes Estimation

- Estimation of the random effects can be helpful for detecting outlying profiles or clusters.
- Since these parameters are assumed to be stochastic, Bayesian methods are applied.
- Posterior Means:

$$\hat{a}_i = E(a_i | Y_{ijk}, \forall j, k)$$

$$\hat{b}_{j(i)} = E(b_{j(i)} | Y_{ijk}, \forall k)$$

- The so-obtained estimates are called **Empirical Bayes (EB)** estimates.

Empirical Bayes Estimation

- The **Empirical Bayes (EB)** estimates are the expected random effects, conditionally on the observed data for that specific cluster.
- In practice, histograms and/or scatterplots of EB estimates are used to detect outlying clusters.

Example: Diabetes Project Leuven

- We reconsider Model 4 given by:

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk}.$$

- The parameters a_i and $b_{j(i)}$ represent GP and patient effects, respectively.
- Histograms and scatterplots will be used to study the EB estimates for a_i and $b_{j(i)}$.

Example: Diabetes Project Leuven

- R code for Calculation of the EB Estimates

```
ranef(resultRandomPatientGP)
```

- The result is a list with 2 objects (since random effects on 2 levels have to be estimated)

```
EBestGP<-ranef(resultRandomPatientGP)$mdnr  
EBestPatient<-ranef(resultRandomPatientGP)$patientnr
```

Example: Diabetes Project Leuven

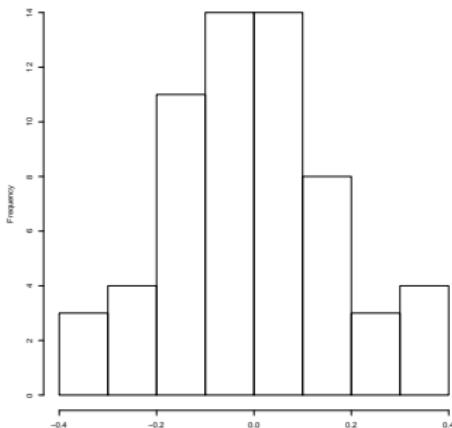
- EB estimates:

```
> EBestGP                                > EBestPatient
      (Intercept)                      (Intercept)
2    0.219797106                         2/1   0.0015997147
5    0.021124051                         2/2   -0.0323460754
6    0.183757540                         2/3   2.1062387005
7    0.014469716                         2/4   -0.2699666060
8    0.082367157                         2/5   -0.3039123961
10   0.155146799                         2/6   1.6988892193
11   -0.078765981                        2/7   -0.1002376555
13   -0.250076338                        2/8   -0.5075871366
...
142   0.084117192                         165/18 -0.4023543127
147   -0.224360577                        165/19 -0.1647337821
150   0.090871573                        165/20 -0.1647337821
151   0.364368779                        165/21 -0.4363001028
154   -0.045344065                        165/22 -0.4363001028
155   -0.016372486                        165/23 -1.1831074849
156   0.267672136                        165/24  0.1747241188
165   -0.235204182                        165/25  0.4462904396
> |                                > |
```

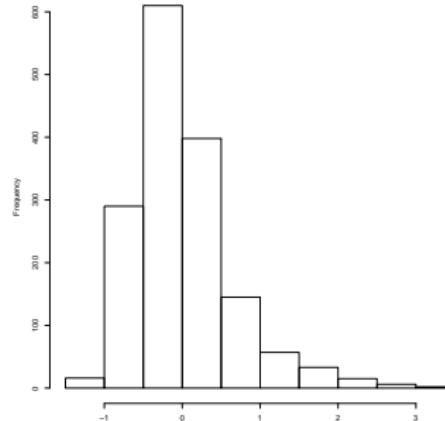
Example: Diabetes Project Leuven

- Histograms of both sets of EB estimates:

EB Estimates of GP Effects



EB Estimates of Patient Effects

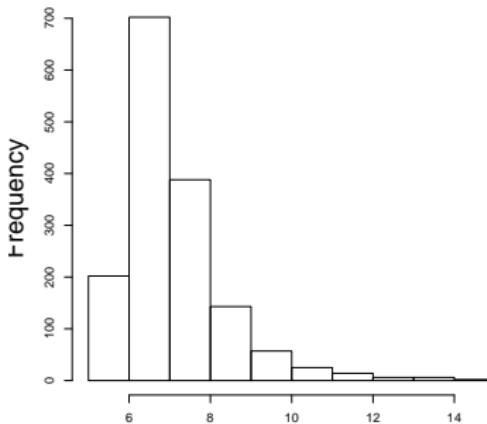


- We notice some patients with extremely large HbA1c values. The largest estimated $b_{j(i)}$ is $\hat{b}_{2(140)} = 3.46$.

Example: Diabetes Project Leuven

- This patient was newly diagnosed, with initial BMI equal to 26.40, initial HbA1c equal to 14.3%, and no follow-up measurement after one year.

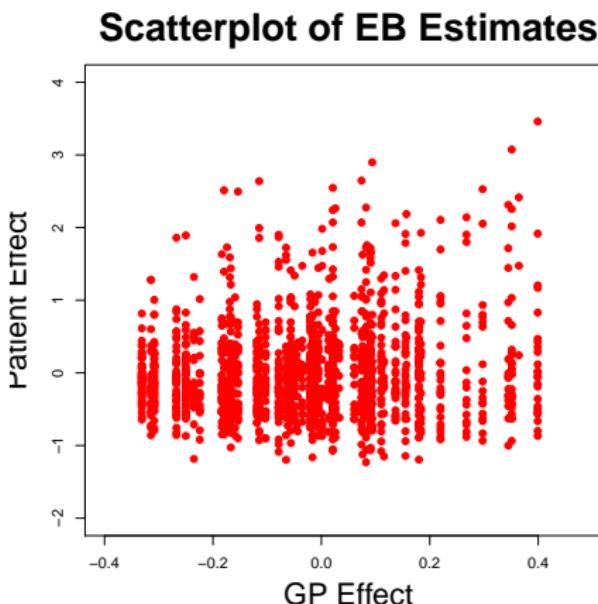
Initial HbA1c Levels



- An initial HbA1c level of 14.3% is extremely high.

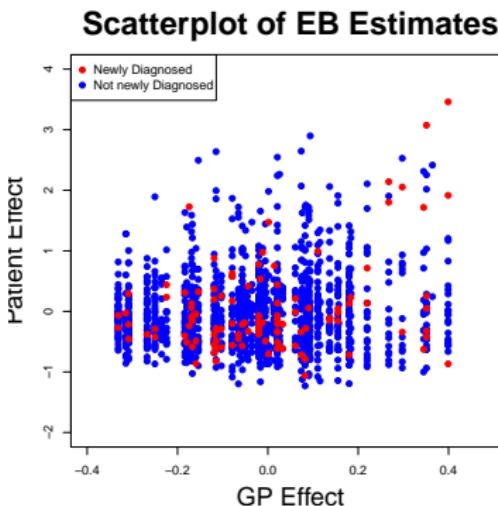
Example: Diabetes Project Leuven

- Scatterplot of EB estimates for patients versus GPs:



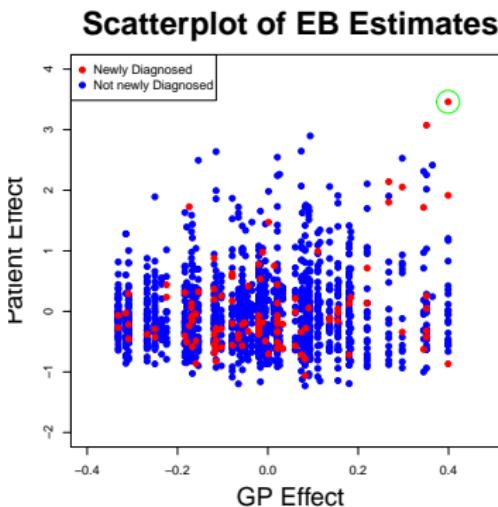
Example: Diabetes Project Leuven

- Plots can also be made in relation with patient or GP characteristics:



Example: Diabetes Project Leuven

- Note how the patient with the largest estimate for $b_{j(i)}$ was treated by the GP with the largest estimated a_i .



Example: Diabetes Project Leuven

- It is therefore worthwhile to repeat the analysis with this subject removed from the data.
- The estimate \hat{a}_{140} drops from 0.40 to 0.27 and four other GPs have now a higher estimate for their effect a_i .
- EB estimates can also be calculated based on other models which include patient or GP characteristics as covariates/factors.
- Extreme EB estimates then reflect that a specific GP or patient within GP is outlying, while this extreme behavior cannot be explained by the covariates in the model.

Average vs. Cluster-Specific Prediction

- Once the EB estimates have been calculated, predictions can be obtained both at the cluster level, as well as on the population average level.
- Reconsider Model 4 for the Leuven diabetes project:

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk}.$$

- Predictions:

- on population average level: $\widehat{E}(Y_{ijk}) = \widehat{\beta}_0 + \widehat{\beta}_1 t_k$

- on cluster level: $\widehat{Y}_{ijk} = \widehat{\beta}_0 + \widehat{\beta}_1 t_k + \widehat{a}_i + \widehat{b}_{j(i)}$

Example: Diabetes Project Leuven

- R Program for Predictions

```
predictedMeans<-predict(resultRandomPatientGP,level=0)
predictedGPlevel<-predict(resultRandomPatientGP,level=1)
predictedPatientLevel<-predict(resultRandomPatientGP,level=2)

predictedData<-data.frame(diabetes$hba1c,diabetes$mdnr,diabetes$patientnr,
    diabetes$time,predictedMeans,predictedGPlevel,predictedPatientLevel)
names(predictedData)<-c("hba1c","mdnr","patientnr","time","predictedMeans",
    "predictedGPlevel","predictedPatientLevel")
```

- The level= option in the predict-statement, gives the level on which the prediction should be calculated.

Example: Diabetes Project Leuven

- Data set with predictions:

```
> predictedData[1:8,]
   hb1alc mdnr patientnr time predictedMeans predictedGPlevel predictedPatientLevel
1    6.4     2          1    0      7.166852      7.386649      7.388249
2    8.0     2          1    1      6.788841      7.008638      7.010238
3    6.7     2          2    0      7.166852      7.386649      7.354303
4    7.6     2          2    1      6.788841      7.008638      6.976292
5   11.2     2          3    0      7.166852      7.386649      9.492888
6    9.4     2          3    1      6.788841      7.008638      9.114877
7    6.8     2          4    0      7.166852      7.386649      7.116683
8    6.8     2          4    1      6.788841      7.008638      6.738671

> predictedData[2960:2968,]
   hb1alc mdnr patientnr time predictedMeans predictedGPlevel predictedPatientLevel
2960    6.2   165         21    1      6.788841      6.553637      6.117337
2961    6.2   165         22    0      7.166852      6.931648      6.495348
2962    6.0   165         22    1      6.788841      6.553637      6.117337
2963    5.0   165         23    0      7.166852      6.931648      5.748541
2964    5.0   165         23    1      6.788841      6.553637      5.370529
2965    7.5   165         24    0      7.166852      6.931648      7.106372
2966    6.5   165         24    1      6.788841      6.553637      6.728361
2967    7.7   165         25    0      7.166852      6.931648      7.377939
2968    7.1   165         25    1      6.788841      6.553637      6.999927
```

Example: Diabetes Project Leuven

- Components needed to calculate predictions:

```
>
> resultRandomPatientGP$tTable
      Value   Std.Error   DF   t-value      p-value
(Intercept) 7.1668523 0.04241276 1511 168.97866 0.0000000e+00
time        -0.3780113 0.02851348 1395 -13.25728 7.202624e-38
.
```

| > EBestGP | | > EBestPatient | |
|-----------|--------------|----------------|---------------|
| | (Intercept) | | (Intercept) |
| 2 | 0.219797106 | 2/1 | 0.0015997147 |
| 5 | 0.021124051 | 2/2 | -0.0323460754 |
| 6 | 0.183757540 | 2/3 | 2.1062387005 |
| 7 | 0.014469716 | 2/4 | -0.2699666060 |
| 8 | 0.082367157 | 2/5 | -0.3039123961 |
| 10 | 0.155146799 | 2/6 | 1.6988892193 |
| 11 | -0.078765981 | 2/7 | -0.1002376555 |
| 13 | -0.250076338 | 2/8 | -0.5075871366 |

...

...

Example: Diabetes Project Leuven

- Population average HbA1c values at baseline and after one year:

$$\hat{E}(Y_{ij1}) = 7.1668 - 0.3780 \times 0 = 7.1668$$

$$\hat{E}(Y_{ij2}) = 7.1668 - 0.3780 \times 1 = 6.7888$$

- Subject-specific predicted HbA1c values for first three patients treated by GP 2:

$$\hat{Y}_{211} = 7.1668 + 0.2198 + 0.0016 = 7.3882$$

$$\hat{Y}_{212} = 6.7888 + 0.2198 + 0.0016 = 7.0102$$

$$\hat{Y}_{221} = 7.1668 + 0.2198 - 0.0323 = 7.3543$$

$$\hat{Y}_{222} = 6.7888 + 0.2198 - 0.0323 = 6.9763$$

$$\hat{Y}_{231} = 7.1668 + 0.2198 + 2.1062 = 9.4929$$

$$\hat{Y}_{232} = 6.7888 + 0.2198 + 2.1062 = 9.1148$$

Overview

3 Motivating Examples

4 The Linear Mixed Model

5 More Examples

6 Estimation of Random Effects

7 Concluding Remarks

- Tests for Variance Components
- Distributional Assumptions for Random Effects
- Missing Data Issues

Concluding Remarks

- Many examples of linear mixed models for longitudinal or clustered data have been discussed.
- Most emphasis was on model formulation, R implementation and interpretation of results.
- A number of issues have not been discussed:
 - Estimation Methods (ML, REML, ...)
 - Inference (F -test, t -test, LR test, Wald test, ...)
 - Model Checking
 - Influence Analysis
 - ...

Concluding Remarks

- These topics are much more difficult and technical than in classical linear models for cross-sectional data, and are therefore outside the scope of this course.
- Three illustrations are given:
 - Tests for Variance Components
 - Distributional Assumptions for Random Effects
 - Missing Data Issues

Tests for Variance Components

- In a number of situations, a researcher might be interested in studying the covariance structure in the data:
 - to give insight into the random variation in the data
 - to test the need for random effects in the model
 - ...
- Thus, it might sometimes be of interest to test whether variance components equal zero.
- For example, for the Leuven diabetes project, it may be of interest to know whether there is any variability between GPs.

Example: Diabetes Project Leuven

- Let's reconsider the Leuven diabetes project.
- As before, let Y_{ijk} being the k^{th} measurement of HbA1c, for the j^{th} patient, of the i^{th} GP. Model 4 was given by:

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk},$$
$$a_i \sim N(0, \sigma_{GP}^2), \quad b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

- Absence of any heterogeneity between GPs would be reflected in:

$$\sigma_{GP}^2 = 0.$$

- It is therefore of interest to test:

$$H_0 : \sigma_{GP}^2 = 0 \quad \text{versus} \quad H_A : \sigma_{GP}^2 > 0.$$

Example: Diabetes Project Leuven

- The default output from R is:

```
> VarCorr(resultRandomPatientGP)
      Variance     StdDev
mdnr = pdLogChol(1)
(Intercept) 0.05441162  0.2332630
patientnr = pdLogChol(1)
(Intercept) 0.61713060  0.7855766
Residual    0.58372742  0.7640206
.
```

- In contrast to, e.g., fixed effects, R does *not* report test-statistics, nor *p*-values.

Remarks

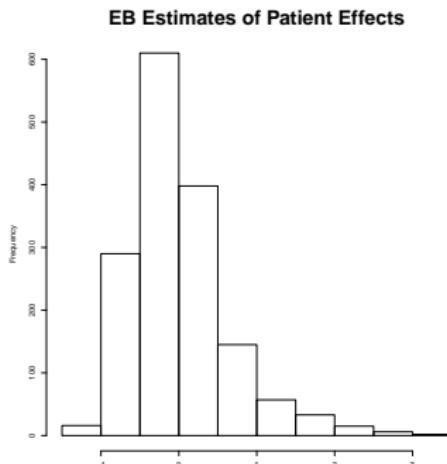
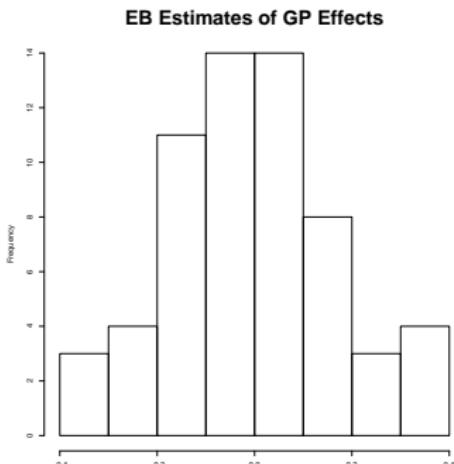
- The well-known p -values based on a $N(0, 1)$ approximation to the Z -statistic *cannot reflect the correct sampling variability* in the estimation of the variance components as these are estimated under the restriction of being positive.
- This so-called **boundary** problem requires correction of the classical p -values.
- The correction depends on the model, and sometimes requires simulation methods.
- Therefore, R does not report any test-statistics, nor p -values.

Distributional Assumptions for Random Effects

- In all the linear mixed models discussed here, we assume that the random effects are normally distributed.
- This distributional assumption needs to be somehow checked.
- In previous sections, we saw how the EB estimates were used for diagnostic purposes (e.g., to detect outlying clusters).
- However, the EB estimates may not necessarily reflect the normality (or lack thereof) of the random effects.
- We continue the analysis of the Leuven diabetes project (Model 4).

Example: Diabetes Project Leuven

- Histograms of the EB estimates of GP and patient effects were:



- The histograms seem to suggest that the normality assumption for the random effects a_i and $b_{j(i)}$ is questionable.

Remarks

- However, one should realize that the precision with which a_i and $b_{j(i)}$ are estimated depends on many aspects, and can vary from patient to patient and from GP to GP.
- So, the above histograms do not necessarily reflect non-normality of the random effects a_i and $b_{j(i)}$.
- Outlying EB estimates can be the reflection of a random effect estimated with very little precision.
 - The differences in precision can be corrected for by standardizing the EB estimates.
 - However, standardized EB estimates still do not necessarily reflect the correct random effects distribution.

Example: Simulated Data

- To illustrate this, consider a small simulation example:

- 1000 profiles with 5 measurements, balanced

- 1000 random intercepts sampled from:

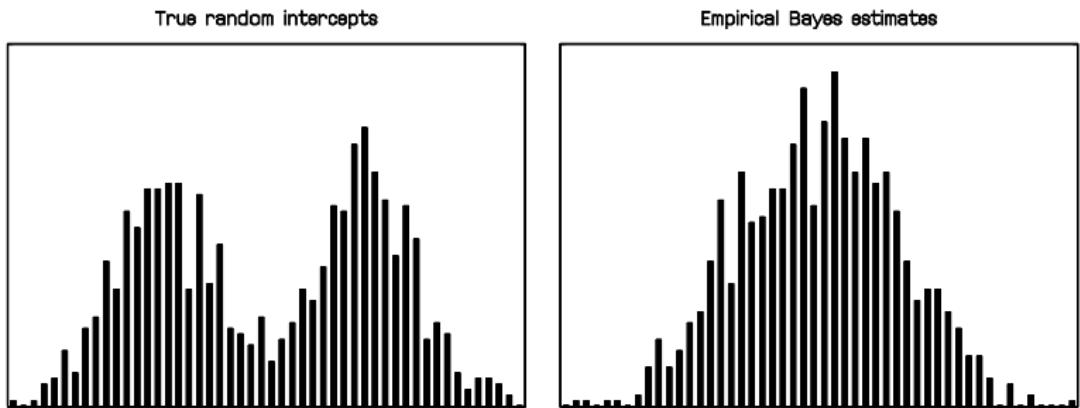
$$\frac{1}{2}N(-2, 1) + \frac{1}{2}N(2, 1)$$

- error components ε_{ij} with variance $\sigma^2 = 30$

- data analyzed assuming normality for the intercepts

Example: Simulated Data

- Histograms of sample intercepts and EB estimates:



- Apparently, the model assumption sometimes forces the estimates to satisfy the assumption.

Conclusion

The normality assumption for random effects cannot be tested within the context of the linear mixed model.

Model extensions are needed.

- Fortunately, inferences about the fixed effects are very robust with respect to model deviations, provided the data set contains sufficient **independent** clusters:
 - Lizard Data: sufficient mothers
 - Rat Data: sufficient rats
 - Growth Curves Data: sufficient children
 - Diabetes Project Leuven: sufficient GPs

Missing Data Issues

- A key feature of mixed models is that they can be used to model unbalanced data.
- In the context of longitudinal data, this includes situations where not all subjects have the same number of repeated measurements, or where subjects are measured at different time points.
- Mixed models are therefore often used in contexts with **missing data**, e.g., subjects left the study prematurely.
- However although mixed models can technically handle such unbalanced data sets, the obtained results can be severely biased in cases where missingness is related to the outcome studied.

Missing Data Issues

- General Principle:

Dropout related to the outcome can imply biased results.

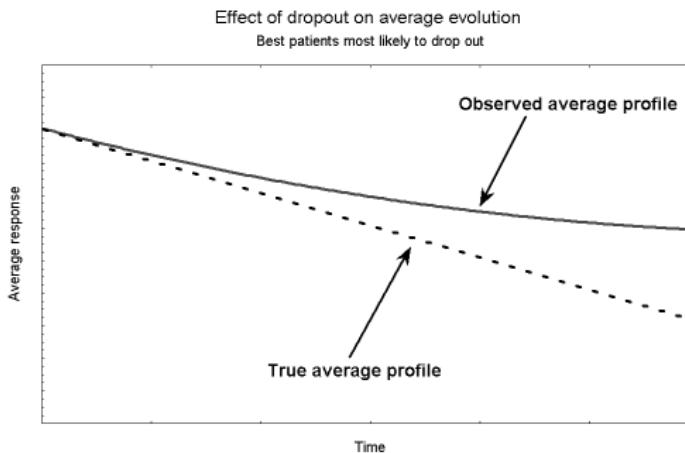
- Unrelated Dropout:

- subjects moving
- subjects dying of other causes
- lost blood samples
- ...

- If dropout is unrelated to the outcome, the obtained sample can be considered as a random sub-sample, which is still a random sample from the original population.

Related Dropout

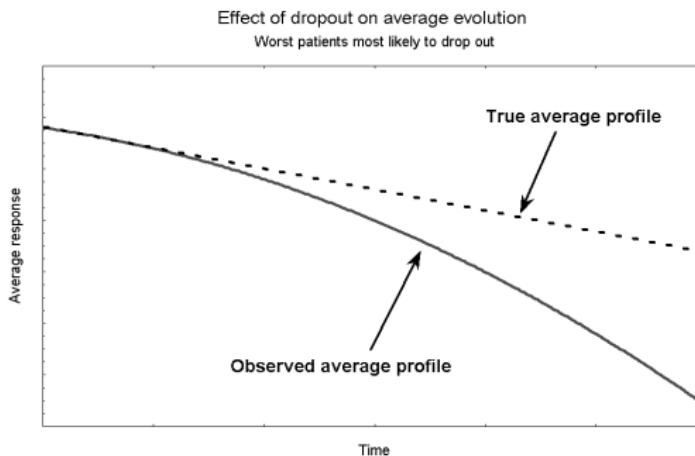
- ‘Best’ patients most likely to drop out:



⇒ Over-Pessimistic

Related Dropout

- ‘Worst’ patients most likely to drop out:

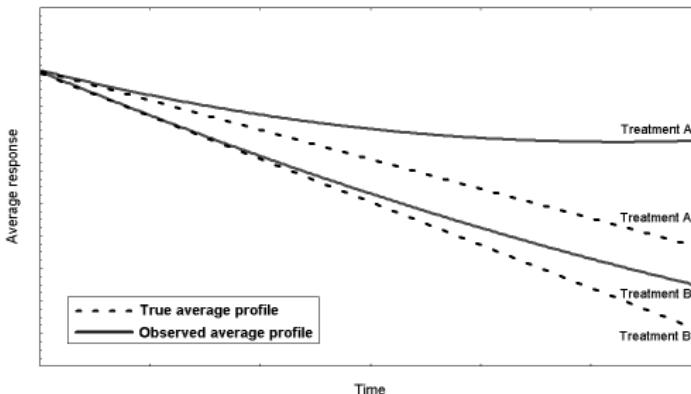


⇒ Over-Optimistic

Related Dropout

- ‘Best’ patients most likely to drop out, but dropout rate dependent on treatment:

Effect of dropout on average evolutions
Best patients most likely to drop out, dropout rate treatment dependent



⇒ Biased estimation of treatment effect