Essentials of Mixed and Longitudinal Modeling

Roula Tsonaka

Department of Biomedical Data Sciences, Section Medical Statistics

Leiden University Medical Center, The Netherlands

s.tsonaka@lumc.nl

Master Statistics and Data Science

Chapter 1: Introduction to Repeated Measurements

Master Statistics and Data Science

Outline

Outline

Repeated Measurements studies

Repeated measures designs arise when an outcome of interest is repeatedly measured on the same unit.

- Examples:
  - **Longitudinal data:** Patients are followed up in time.
  - **Clustered data:** Measurements collected from members of the same family or from patients in the same hospital.
  - **Multilevel data:** Measurements collected from students within the same classroom in the same school.
- *Equivalent terms:* Repeated measurements, clustered observations, multilevel data, correlated data or dependent data.
- In this course we will study more extensively longitudinal designs.

Outline

Example: ARMD Trial

- Reference: Interferon alpha-IIA is ineffective for patients with choroidal neovascularization secondary to age-related macular degeneration. Results of a prospective randomized placebo-controlled clinical trial. Archives of Ophthalmology, 115, 865-872.
- Randomized multi-center clinical trial comparing an experimental treatment (interferon-a) versus placebo in 240 patients diagnosed with Age-Related Macular Degeneration Trial (ARMD).
- Response: Visual acuity score based on patient's ability to read lines of letters on standardized vision charts.
- Covariates:
  - Group: Placebo versus the highest dose (6 million units daily) of interferon-α.
  - Follow-up: Baseline and after randomization at 4, 12, 24, and 52 weeks.

	ID	Treat	VASbase	Weeks	VAS
1	1	Active	59	0	59
2	1	Active	59	4	55
3	1	Active	59	12	45
4	2	Active	65	0	65
5	2	Active	65	4	70
6	2	Active	65	12	65

Example: ARMD Trial

- Research question: Assess the impact of treatment on progression of visual acuity score (VAS).

- Number of measurements recorded at each visit:

```
  0   4  12  24  52
240 231 227 214 195
```

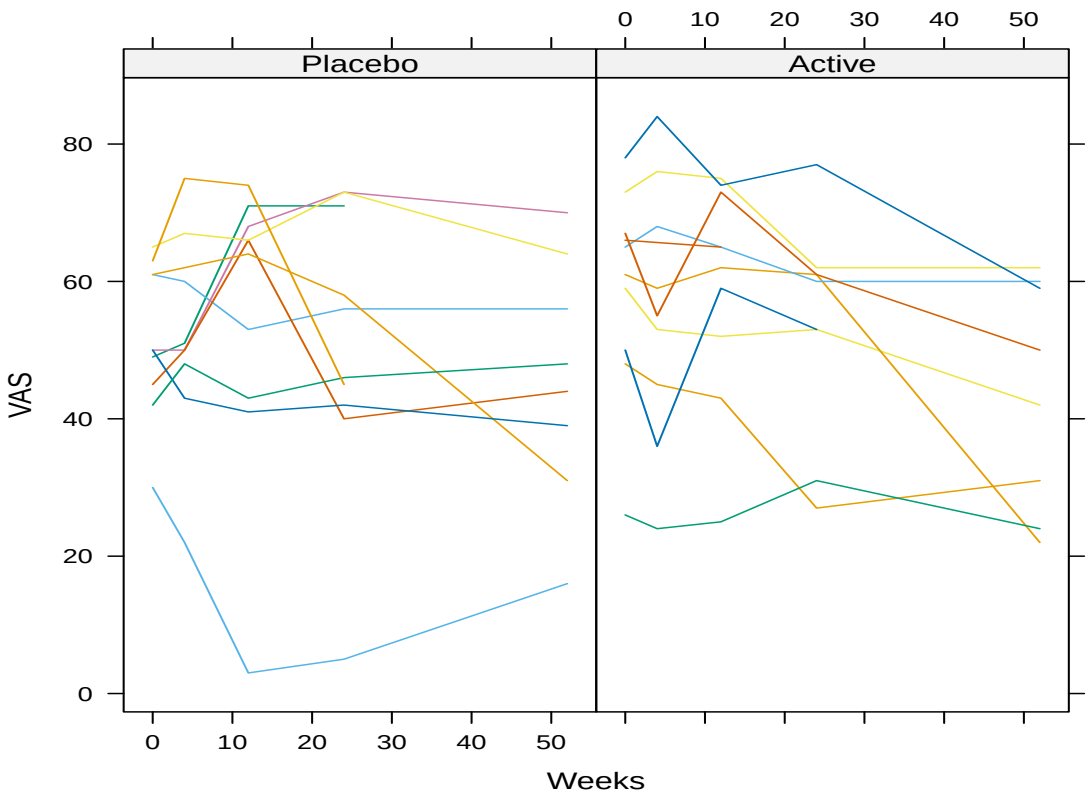
  note that not all measurements are recorded: 0,4,5,11,19 % missingness at each visit.

- Number of measurements recorded at each visit and treatment group:

```
            0   4  12  24  52
Placebo   119 117 117 112 105
Active    121 114 110 102  90
```

Example: ARMD Trial

- Sample individual VAS profiles in time per treatment group



- Substantial variability from patient to patient within each treatment group.
- VAS scores of some patients tend to decrease in time, whereas for others they remain relatively constant.

Outline

Repeated Measurements studies - Features

- Measurements taken on the same unit are expected to be **correlated**.
- Methods which ignore this correlation can lead to invalid inferences.
- Example: **Paired vs Unpaired t-test**
  - We will use VAS from ARMD trial at weeks 0 and 4.

```
       Min. 1st Qu. Median     Mean 3rd Qu. Max.
Week 0   20      45     56 54.84848      66   85
Week 4   12      42     53 52.45887      64   84
```

  - We are interested if VAS changes within first 4 weeks.
  - This is equivalent to testing:

$$H_0 : \mu_0 = \mu_4 \text{ versus } H_A : \mu_0 \neq \mu_4$$
$$\text{or}$$
$$H_0 : \Delta = 0 \text{ versus } H_A : \Delta \neq 0$$

    with $\Delta = \mu_0 - \mu_4$.
  - We use Paired t-test which is based on within subject mean differences Δ.

Repeated Measurements studies - Features

- **Paired t-test:** Statistically significant mean VAS change within first 4 weeks (p-value = 2e-05).

```
        Paired t-test

data:  data.$VASbase and data.$VAS
t = 4.3459, df = 230, p-value = 2.082e-05
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 1.306228 3.472993
sample estimates:
mean difference
        2.38961
```

- If we ignored the within subject correlations, i.e. use **Two-sample t-test**: Mean VAS change within first 4 weeks is not statistically significant at 5% (p-value = 0.0956).

```
        Welch Two Sample t-test

data:  data.$VASbase and data.$VAS
t = 1.67, df = 457.81, p-value = 0.0956
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4223452  5.2015660
sample estimates:
mean of x mean of y
 54.84848  52.45887
```

Repeated Measurements studies - Features

- The two-sample t-test does not take into account that the measurements are not independent.
- This affects the estimation of the standard error of $\hat{\Delta}$, i.e. $\hat{\sigma}_{\hat{\Delta}}$:
  - The paired t-test uses: $\sigma_{\hat{\Delta}} = \sqrt{\sigma^2_{\hat{\mu}_0} + \sigma^2_{\hat{\mu}_4} - 2 * \text{cov}(\hat{\mu}_0, \hat{\mu}_4)}$
  - While in the two-sample t-test $\sigma_{\hat{\Delta}} = \sqrt{\sigma^2_{\hat{\mu}_0} + \sigma^2_{\hat{\mu}_4}}$
- Ignoring dependence leads to higher standard errors $\Rightarrow$ higher p-values $\Rightarrow$ More difficult to obtain significance.
- The estimation of the mean change is often not affected $\hat{\Delta} = \hat{\mu}_0 - \hat{\mu}_4$, unless we have missing data (see later in slides).

Repeated Measurements studies - Features

Statistical methods which assume independent observations will not be valid for the analysis of repeated measures data.

Repeated Measurements studies - Features

- Longitudinal studies are a special case of repeated measures designs:
  Subjects are followed up in time.
- Observations have an ordering in time, in contrast to clustered data, e.g. patients in hospitals or siblings in families.
- Proper treatment of this ordering is needed.
- In clustered data measurements within units can be exchangeable (order is not important) $\Rightarrow$ important for modelling the type of correlation e.g. within hospitals or families.

Repeated Measurements studies - Features

- The number of repeated measurements can be different across units ⇒ **Unbalanced or incomplete data**.
- In longitudinal studies individuals may:
  - be lost to follow-up: Understanding reasons of missingness data crucial (see later in slides).
  - provide measurements at irregular points in time.
- Timing of visits important for modelling the type of correlation e.g. measurements further apart less correlated.

Repeated Measurements studies - Features

- At each timepoint one outcome and p covariates are observed.
- The covariates can be either
  - fixed (or time independent) e.g. gender, baseline patient characteristics or
  - time dependent (different between times within the same subject) e.g. visit times, BMI, etc.
- The covariates can be both continuous or categorical: We do not follow the (historical) distinction between ANOVA and linear regression models.

Repeated Measurements studies - Features

- The outcome variable can be: continuous, binary or count.
- Different analyses methods can be used:
  - Marginal Regression Models.
  - Linear Mixed Effects Models.
  - Generalized Estimating Equations.
  - Generalized Mixed Effects Models.

Outline

Repeated Measurements studies - Strengths

- Longitudinal studies allow to
  - study progression (e.g. aging, growth): changes over time.
  - study the effect of treatments or subject characteristics or environmental exposures on those changes (longitudinal effect).
  - separate longitudinal from cross-sectional effects.
    e.g. study for treatment effects at each time point (cross-sectional effect) and if treatment effect changes over time (longitudinal effect).

Repeated Measurements studies - Strengths

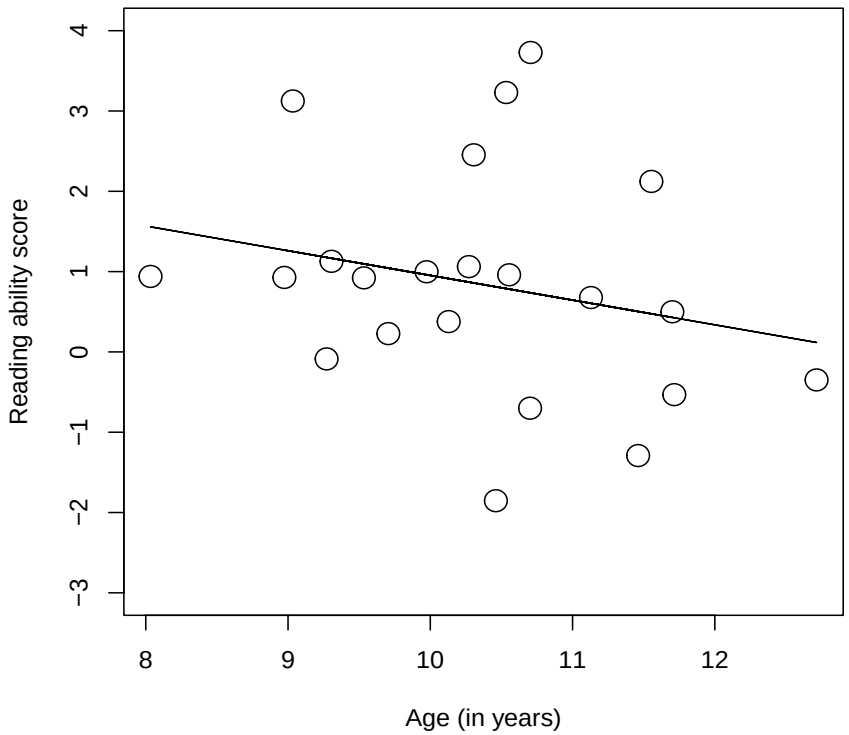
- **Longitudinal studies allow us to separate longitudinal from cross-sectional effects.**

- Example: Say we are interested on the relationship between reading ability and age (Diggle *et al.*2002).

- We can study this question by:
  - Cross-sectional study: one measurement on children of different age.
  - Longitudinal study: two measurements on chidren of different age.

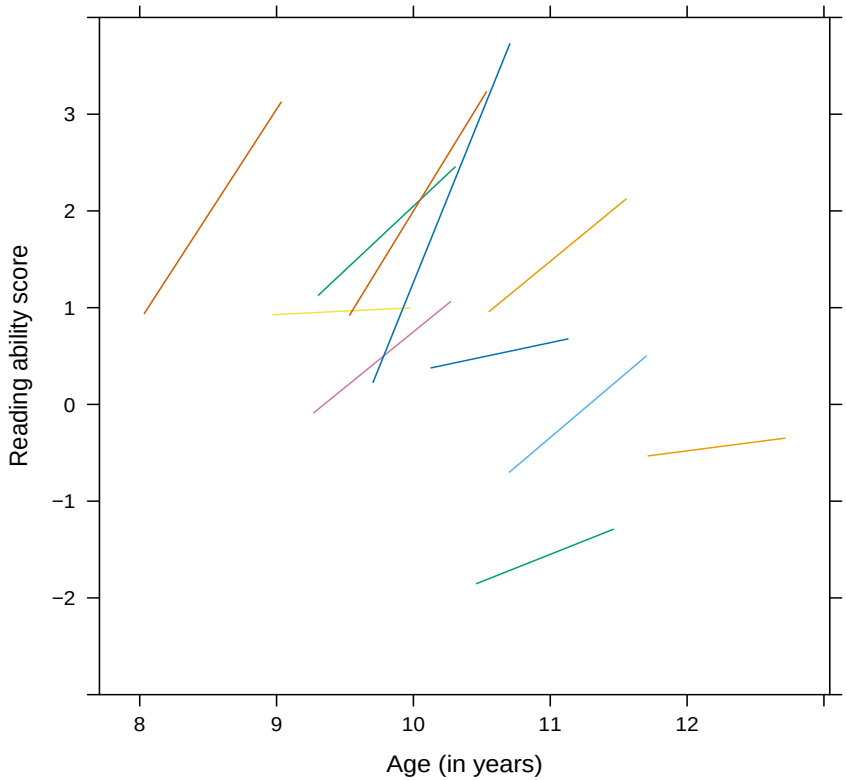
Repeated Measurements studies - Strengths

- Based on a cross-sectional study we conclude that there is negative relation between reading ability and age.



Repeated Measurements studies - Strengths

- The same measurements can arise when we collect 2 measurements per child.



- We can now conclude that there is a negative cross-sectional relation but a positive longitudinal trend.

Repeated Measurements studies - Chapter Summary

- Repeated measures designs arise when an outcome of interest is repeatedly measured on the same unit.
- Measurements taken on the same unit are expected to be correlated.
- Correlations should not be ignored $\Rightarrow$ wrong inference can arise: proper method are needed.
- Longitudinal studies are a special case of repeated measures designs:
  Subjects are followed up in time.
- Number of measurements on the same unit not necessarily equal.
- Longitudinal studies allow us to separate longitudinal from cross-sectional effects.

Chapter 2: Simple methods

Master Statistics and Data Science

Outline

Outline

Simple methods - Introduction

- Several challenges in analysing longitudinal data.
- Often simplistic methods are adopted which can lead to erroneous inference:
  - Ignore correlation.
  - Analysis at each time point separately.
  - Analysis of Area Under the Curve (AUC).
  - Analysis of endpoints.
  - Analysis of increments.
  - Analysis of covariance.
- We will review these methods and discuss their pros and cons.

Outline

Ignore correlation

- Idea:
  - Treat the responses from the same subject as independent observations.
- Pros:
  - Uses all available data.
- What can go wrong?
  - If data are missing, estimation of effects is valid only when data are missing completely at random.
  - Estimation of standard errors will be severely affected:
    - **overestimated for within-subject effects.**
    - **underestimated for between-subjects effects.**

Ignore correlation

- Example I:
- Let y_{i1}, y_{i2} measurements for the 2 eyes of subject i with $(i = 1, \ldots, n)$ and let a treatment is given only on the right eye.
- We estimate the treatment effect as $\hat{\Delta} = \bar{Y}_1 - \bar{Y}_2$, i.e. difference of sample mean of 2 treatments $\Rightarrow$ within subject effect.
- For the standard errors, it holds: $\text{Var}(\hat{\Delta}) = Var(\bar{Y}_1) + Var(\bar{Y}_2) - 2cov(\bar{Y}_1, \bar{Y}_2)$.
- If we ignore correlation between y_{i1}, y_{i2} then we imply $\text{Var}(\hat{\Delta}) = Var(\bar{Y}_1) + Var(\bar{Y}_2)$.
- Ignoring correlations leads to overestimation of standard errors for within subject effects $\Rightarrow$ higher p-values, i.e. difficult to reach significance.

Ignore correlation

- Example II:
- Let y_{i1}, y_{i2} measurements for the 2 eyes of subject i and let a treatment is given on both eyes for half of the subjects.
- Take the mean of the 2 measurements.
- We estimate the treatment effect as $\hat{\Delta} = \bar{Y}_A - \bar{Y}_B$, i.e. difference of sample means of 2 treatments $\Rightarrow$ between subjects effect.
- The two treatment groups A and B are independent: $\text{Var}(\hat{\Delta}) = Var(\bar{Y}_A) + Var(\bar{Y}_B)$.
- Here $Var(\bar{Y}_A) = Var(\bar{Y}_1) + Var(\bar{Y}_2) + 2cov(\bar{Y}_1, \bar{Y}_2)$
- Ignoring correlations leads to underestimation of standard errors for between subject effects $\Rightarrow$ smaller p-values, i.e. risk of false positives.

Analysis at each time point

- Idea: Let measurements collected at the same occasions for all subjects ⇒ analyse each occasion separately.
- Pros:
  - Uses all available data.
- What can go wrong?
  - Multiple testing problems.
  - Possible problems with missing data: assumes data missing completely at random.
  - Does not take advantage of existing within subject correlations ⇒ Power loss.
  - Difficult to interpret if interest on progression.

Analysis of Area Under the Curve (AUC)

- **Idea:**
  - Let $y_{i1}, y_{i2}, \ldots$ the measurements for subject i at $t_{i1}, t_{i2}, \ldots$ occasions.
  - The area under the profile of each subject i is computed as:

$$AUC_i = (t_{i2} - t_{i1}) \times \frac{(y_{i1} + y_{i2})}{2} + (t_{i3} - t_{i2}) \times \frac{(y_{i2} + y_{i3})}{2} + \ldots$$

  - The AUCs are further analyzed with cross-sectional analysis methods.
- **Pros:**
  - No problems of multiple testing.
  - Uses all available data.
  - Measurements can be collected at different time points per subject.
- **What can go wrong?**
  - Subjects need to have the same follow-up period to be comparable $\Rightarrow$ problems with missing data.
  - Cannot separate subjects with increasing or decreasing profiles $\Rightarrow$ the same AUC value.

Analysis of endpoints

- Idea: Use the measurements y_{in} of the last time point for each subject i in any cross-sectional analysis method.
- Pros:
  - No problems of multiple testing.
  - Measurements can be collected at different time points per subject.
- What can go wrong?
  - Applicable only in randomized trials where no differences at baseline.
  - Uses partial information, i.e. y_{in_i}.
  - The last time point must be the same for all subjects $\Rightarrow$ possible problems with missing data.
  - Does not consider differences over whole follow-up.

Analysis of covariance

- Idea: To perform analysis of endpoints, while correcting for differences at baseline, we can use analysis of covariance, where the first measurement is included as covariate in the model.
- Pros:
  - No problems of multiple testing.
  - Measurements can be collected at different time points per subject.
- What can go wrong?
  - Uses partial information, i.e. y_{i1}, y_{in_i}.
  - Does not take into account the variability of y_{i1}.

Analysis of increments

- **Idea:**
  - Let $y_{i1}, y_{i2}, \ldots, y_{in_i}$ the measurements for subject i.
  - Compute the subject-specific changes $y_{in_i} - y_{i1}$.
- **Pros:**
  - No problems of multiple testing.
  - Measurements can be collected at different time points per subject.
- **What can go wrong?**
  - Uses partial information, i.e. y_{i1}, y_{in_i}.
  - the last time point must be the same for all subjects.
  - possible problems with missing data.

Simple methods - Chapter summary

- Simple methods summarize the information for each subject in a single value.
- Then common statistical procedures for independent data can be applied.
- However, they have several disadvantages and their use can lead to erroneous inference:
  - not all available information is used.
  - severe bias in the presence of missing data.
- They are discussed for historical reasons and because they are unfortunately still applied in practice by non-experts.

Chapter 3: Linear models with correlated errors

Master Statistics and Data Science

Outline

Outline

Linear Regression Model - Introduction

- Also known as **multiple linear regression model**.
- Cross-sectional design: single measurement per subject.
- Response variable Y: Continuous (e.g. diastolic blood pressure).
- Goal: relate mean response to covariates.
- Example: Baseline measurements in the ARMD trial:
  - test for group mean VAS differences while adjusting for age, sex, etc.
  - identify factors associated with mean VAS at baseline.

Linear Regression - Model specification

- The model for each subject i is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

with

  - y_i dependent or response variable for subject i, $i = 1, \ldots, n$.
  - $x_{i1}, x_{i2}, \ldots, x_{ip}$ are the p independent variables (aka covariates)
  - $\beta_0, \beta_1, \ldots, \beta_p$ regression coefficients
  - ε_i the error term for subject i

Linear Regression - Model specification - Cont'd

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i, \ \varepsilon_i \sim N(0, \sigma^2)$$

- **Model assumptions:**
  - The error terms are normally distributed.
  - Y_i is assumed to have a normal distribution with mean $\mu_i = E(Y_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$ and variance σ^2:

$$Y_i \sim N(\mu_i, \sigma^2)$$

  - Linear relationship between expected response and covariates.
  - The measurements y_i are **independent** of each other.

- **Parameter interpretation:** change in mean response for a unit change in covariate e.g. x_1 given that all other covariates are held fixed.

Linear Regression - Example

- In the ARMD Trial, we consider only the measurements at baseline.
- Research question: Are there differences in mean VAS at baseline?

$$y_i = \beta_0 + \beta_1 \text{treat}_i + \varepsilon_i, \ \varepsilon_i \sim N(0, \sigma^2)$$

where

  - y_i baseline VAS for patient i
  - treat_i treatment allocation for patient i
  - β_0, β_1 regression coefficients
  - ε_i error terms for patient i

Linear Regression - Example

- Using the R function `lm(.)` we get:

```
Call:
lm(formula = VASbase ~ Treat, data = armd.)

Residuals:
    Min      1Q  Median      3Q     Max
-34.531  -9.531   1.469  11.469  29.725

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  55.2754     0.6205  89.077   <2e-16 ***
TreatActive  -0.7447     0.8910  -0.836    0.403
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.82 on 1105 degrees of freedom
Multiple R-squared:  0.0006319,        Adjusted R-squared:  -0.0002725
F-statistic: 0.6987 on 1 and 1105 DF,  p-value: 0.4034
```

- There is no statistically significant difference in mean VAS at baseline at 5% between the 2 groups (p-value $= 0.403$).

Linear Regression with Matrix Notation - Reminder

- We use matrix notation to describe the data collected.
- A matrix is a rectangular array of numbers arranged in rows and columns:
- Example: Let $\mathbf{A}$ a $k \times m$ matrix with k rows and m columns

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \dots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{km} \end{bmatrix}$$

  where a_{ij} is the jth element in row i.
- By convention we denote matrices with capital letters: $\mathbf{Y}, \mathbf{W}, \mathbf{Z}$, etc.
- A matrix with one column is called vector and denoted by lower case letters: $\mathbf{y}, \mathbf{w}, \mathbf{z}$, etc.

Linear Regression with Matrix Notation

- For each subject i

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i$$

- For all n subjects

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{y} = \mathbf{X} \times \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

- $\mathbf{Y}$ response vector
- $\mathbf{X}$ design matrix
- $\boldsymbol{\beta}$ parameter vector
- $\boldsymbol{\varepsilon}$ error vector

Linear Regression with Matrix Notation

$$\mathbf{y} = \mathbf{X} \times \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

- $\mathbf{y}$ response vector
- $\mathbf{X}$ design matrix
- $\boldsymbol{\beta}$ parameter vector
- $\boldsymbol{\varepsilon}$ measurement error vector

- For the error term ε_i we assume $\varepsilon_i \sim N(0, \sigma^2)$ and in matrix notation we write

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \sim N_n \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} \right)$$

where $\boldsymbol{\Sigma}$ is the variance-covariance matrix for the error terms.

Linear Regression with Matrix Notation - Reminder

- **Covariance** is a measure of the degree of association between two random variables: linear association!
- If X and Y are jointly distributed random variables with expectations μ_X and μ_Y, respectively, the covariance of X and Y is $\text{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$.
- Note that $\text{Cov}(X,X) = \text{Var}(X)$.
- X and Y are correlated if $\text{Cov}(X,Y) \neq 0$.
- $\text{Cov}(X,Y)$ can take any positive or negative value and its value depends on the scale of X and Y.
- $\text{Cov}(X,Y)$ relates to the correlation coefficient ρ

$$\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

where σ_X and σ_Y is the standard deviation of X and Y and $-1 \leq \rho \leq 1$.

Linear Regression with Matrix Notation - Reminder

- For any pair y_i and y_j with $i, j \in \{1, \ldots, n\}$ we compute $\text{Cov}(y_i, y_j)$.
- The dependence structure among all observations $\mathbf{y}$ is described by the variance-covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \text{Var}(y_1) & \text{Cov}(y_1, y_2) & \ldots & \text{Cov}(y_1, y_n) \\ \text{Cov}(y_2, y_1) & \text{Var}(y_2) & \ldots & \text{Cov}(y_2, y_n) \\ \vdots & \vdots & \vdots & \vdots \\ \text{Cov}(y_n, y_1) & \ldots & \ldots & \text{Var}(y_n) \end{bmatrix}.$$

- The correlation matrix $\mathbf{R} = \text{corr}(\mathbf{y})$ is then symmetric around the main diagonal:

$$\mathbf{R} = \begin{bmatrix} 1 & \text{corr}(y_1, y_2) & \ldots & \text{corr}(y_1, y_n) \\ \text{corr}(y_2, y_1) & 1 & \ldots & \text{corr}(y_2, y_n) \\ \vdots & \vdots & \vdots & \vdots \\ \text{corr}(y_n, y_1) & \ldots & \ldots & 1 \end{bmatrix}.$$

Linear Regression with Matrix Notation

- In the linear regression model the dependence structure among all observations $\mathbf{y}$ is described by the variance-covariance matrix:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}.$$

- $\text{Cov}(y_i, y_j) = 0$ for $i \neq j$ because the observations are independent $\Rightarrow$ scaled identity matrix.

- $\text{Cov}(y_i, y_i) = \sigma^2$.

- The correlation matrix is the identity matrix in this case.

Linear Regression - Estimation

- The parameters $\boldsymbol{\beta}$ and σ^2 can be estimated using the maximum likelihood method.
- The likelihood function is:

$$L(\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mathbf{X}_i\boldsymbol{\beta})^2\}$$

- The maximum likelihood estimates (MLEs) are:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\ \hat{\sigma}^2 &= \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \end{aligned}$$

Linear Regression - Estimation

- The MLE $\hat{\boldsymbol{\beta}}$ is equivalent to the ordinary least squares (OLS) estimate.
- In OLS the unbiased estimator of σ^2 is given by:

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$$

- The ML estimate of σ^2 is biased downwards $\Rightarrow$ the std errors of $\hat{\boldsymbol{\beta}}$ too small.

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}$$

- Restricted ML deals with the bias of ML (we will get back to it later).

Outline

Linear models with correlated errors - Introduction

- Multiple measurements per unit.
- Also known as **Marginal models** or **Multivariate models**.
- Measurements within subjects are correlated:
  - longitudinal studies e.g. follow-up subjects in time
  - cluster-randomized clinical trials, e.g. at hospital level
  - family studies....

Repeated measurements designs

- Contrary to linear regression here multiple measurements per subject - Typical data structure:

Unit	Response	Covariates			
1	y_{11}	$x_{11}^{(1)}$	$x_{11}^{(2)}$	$\dots$	$x_{11}^{(p)}$
1	y_{12}	$x_{12}^{(1)}$	$x_{12}^{(2)}$	$\dots$	$x_{12}^{(p)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
1	y_{1n_1}	$x_{1n_1}^{(1)}$	$x_{1n_1}^{(2)}$	$\dots$	$x_{1n_1}^{(p)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	y_{i1}	$x_{i1}^{(1)}$	$x_{i1}^{(2)}$	$\dots$	$x_{i1}^{(p)}$
i	y_{i2}	$x_{i2}^{(1)}$	$x_{i2}^{(2)}$	$\dots$	$x_{i2}^{(p)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
i	y_{in_i}	$x_{in_i}^{(1)}$	$x_{in_i}^{(2)}$	$\dots$	$x_{in_i}^{(p)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{n1}	$x_{n1}^{(1)}$	$x_{n1}^{(2)}$	$\dots$	$x_{n1}^{(p)}$
n	y_{n2}	$x_{n2}^{(1)}$	$x_{n2}^{(2)}$	$\dots$	$x_{n2}^{(p)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
n	y_{nn_n}	$x_{nn_n}^{(1)}$	$x_{nn_n}^{(2)}$	$\dots$	$x_{nn_n}^{(p)}$

- Here the repeated measurements per subject are correlated.

Repeated Measurements studies - Notation

- $i = 1, \ldots, n$: n independent units of analysis (e.g. patients, families).
- $j = 1, \ldots, n_i$: n_i visits for each subject i.
- Visit times for each subject i: $t_{i1}, t_{i2}, \ldots, t_{in_i}$.
- Measurements for each subject i: $y_{i1}, y_{i2}, \ldots, y_{in_i}$.
- Covariates for each subject i at visit j: $x_{ij}^{(1)}, x_{ij}^{(2)}, \ldots, x_{ij}^{(p)}$.

Linear models with correlated errors - Model specification

- For each sampling unit i (e.g. patient, school, family) with n_i correlated repeated measurements:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

- **$\mathbf{X}_i\boldsymbol{\beta}$ models the mean of the dependent variable Y.**

- For the error terms $\boldsymbol{\varepsilon}_i$ we assume

$$\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i),$$

  with $\boldsymbol{\Sigma}_i$ the variance-covariance matrix $(n_i \times n_i)$.

- **The covariance matrix $\boldsymbol{\Sigma}_i$ explicitly accounts for the correlations.**

- Note: we assume that sampling units are independent but measurements within unit correlated.

Linear models with correlated errors - Model specification

- For each subject i the response vector and error terms vector are:

$$\mathbf{y}_i \equiv \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i_{n_i}} \end{bmatrix} \qquad \boldsymbol{\varepsilon}_i \equiv \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{i_{n_i}} \end{bmatrix}$$

- β is the regression parameters vector

$$\boldsymbol{\beta} \equiv \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}$$

- $\mathbf{X}_i$ the design matrix with covariate information

$$\mathbf{X}_i \equiv \begin{bmatrix} 1 & x_{i1}^{(1)} & x_{i1}^{(2)} & \dots & x_{i1}^{(p)} \\ 1 & x_{i2}^{(1)} & x_{i2}^{(2)} & \dots & x_{i2}^{(p)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{in_i}^{(1)} & x_{in_i}^{(2)} & \dots & x_{in_i}^{(p)} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{1} & \mathbf{x}_i^{(1)} & \mathbf{x}_i^{(2)} & \dots & \mathbf{x}_i^{(p)} \end{bmatrix}.$$

Linear models with correlated errors - Model the mean I

- Mean model as in ANOVA/Linear Regression but we need to carefully study the research questions.
- Consider the ARMD longitudinal trial where VAS is measured at randomization and at 4, 12, 24, and 52 weeks afterwards in each treatment group.
- Potential research questions:
  1. Does the mean VAS progression differ between the 2 groups?
  2. If the mean VAS progression does not differ between the 2 groups, is it at the same level?
  3. Is the mean VAS progression constant over time?

Linear models with correlated errors - Model the mean I

- Consider the ARMD longitudinal trial where VAS is measured at randomization and at 4, 12, 24, and 52 weeks afterwards in each treatment group.
- Research question: Does the mean VAS progression differ between the 2 groups?
- We will study 2 ways to model the mean response:
  1. Unstructured model.
  2. (Semi-) Parametric model.

Linear models with correlated errors - Model the mean I

- We can assume a model with **unstructured mean**:

$$\boxed{y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_{2j} \text{TimeID}_{ij}^{(j)} + \beta_{3j} \text{TimeID}_{ij}^{(j)} \times \text{group}_i + \varepsilon_{ij}}$$

  with $i = 1, \ldots, 240$, $j = 0, \ldots, 4$, $\beta_{20} = \beta_{30} = 0$, TimeID_{ij} is the visit indicator and $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$.

- We do not assume a specific trend $\Rightarrow$ Time is used as a categorical variable.

- Also known as Analysis of Response Profiles.

- For $\boldsymbol{\Sigma}_i$ we assume unstructured covariance matrix (we will discuss this later).

Linear models with correlated errors - Model the mean I

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_{2j} \text{TimeID}_{ij}^{(j)} + \beta_{3j} \text{TimeID}_{ij}^{(j)} \times \text{group}_i + \varepsilon_{ij}$$

	Value	Std.Error	t-value	p-value
(Intercept)	55.304	1.368	40.442	0.000
WeeksF4	-1.289	0.762	-1.692	0.091
WeeksF12	-2.351	1.087	-2.162	0.031
WeeksF24	-5.992	1.312	-4.567	0.000
WeeksF52	-11.236	1.600	-7.021	0.000
TreatActive	-0.600	1.925	-0.312	0.755
WeeksF4:TreatActive	-2.157	1.082	-1.992	0.047
WeeksF12:TreatActive	-3.476	1.551	-2.241	0.025
WeeksF24:TreatActive	-3.245	1.883	-1.724	0.085
WeeksF52:TreatActive	-5.301	2.307	-2.298	0.022

- Coefficients' interpretation:
  - β_0: mean VAS in placebo group at baseline.
  - β_1: change in the mean VAS between the two groups at baseline.
  - β_{2j}: change in the mean VAS from baseline to week j in the placebo group.
  - $\beta_{2j} + \beta_{3j}$: change in the mean VAS from baseline to week j in the treatment group.
  - β_{3j}: change in the mean VAS from baseline to week j between treatment and placebo.

Linear models with correlated errors - Model the mean I

- Design matrix $\mathbf{X}_i$:
  `model.matrix(VAS ~ WeeksF * Treat, data = armd.)`

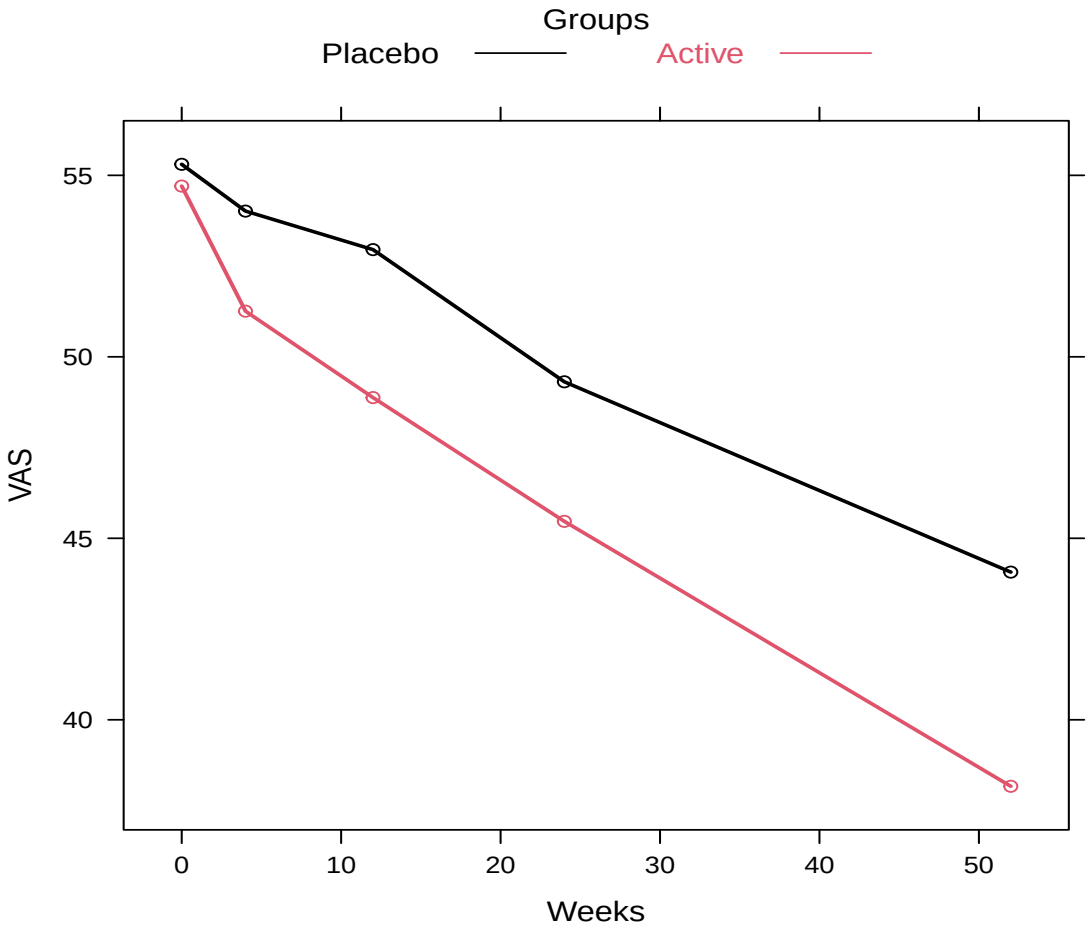
```
  (Intercept) WeeksF4 WeeksF12 WeeksF24 WeeksF52 TreatActive
4           1       0        0        0        0           1
5           1       1        0        0        0           1
6           1       0        1        0        0           1
7           1       0        0        1        0           1
8           1       0        0        0        1           1
  WeeksF4:TreatActive WeeksF12:TreatActive WeeksF24:TreatActive
4                   0                    0                    0
5                   1                    0                    0
6                   0                    1                    0
7                   0                    0                    1
8                   0                    0                    0
  WeeksF52:TreatActive
4                    0
5                    0
6                    0
7                    0
8                    1
```

- or without the column names

```
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
4    1    0    0    0    0    1    0    0    0     0
5    1    1    0    0    0    1    1    0    0     0
6    1    0    1    0    0    1    0    1    0     0
7    1    0    0    1    0    1    0    0    1     0
8    1    0    0    0    1    1    0    0    0     1
```

Linear models with correlated errors - Model the mean I

- The fitted profiles per group are:



Linear models with correlated errors - Model the mean I

- The estimated covariance matrix $\boldsymbol{\Sigma}_i$ is:

```
Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 220.93 202.26 189.12 181.82 141.87
[2,] 202.26 251.27 229.09 216.99 174.74
[3,] 189.12 229.09 295.81 259.23 224.42
[4,] 181.82 216.99 259.23 340.11 290.40
[5,] 141.87 174.74 224.42 290.40 351.05
  Standard Deviations: 14.864 15.851 17.199 18.442 18.736
```

- The estimated correlation matrix is:

```
      [,1]  [,2]  [,3]  [,4]  [,5]
[1,] 1.000 0.858 0.740 0.663 0.509
[2,] 0.858 1.000 0.840 0.742 0.588
[3,] 0.740 0.840 1.000 0.817 0.696
[4,] 0.663 0.742 0.817 1.000 0.840
[5,] 0.509 0.588 0.696 0.840 1.000
```

Linear models with correlated errors - Model the mean I

- The unstructured/saturated mean model perfectly describes the data.
- Minimal risk of model misspecification.
- However, too many parameters are used which may lead to loss of efficiency.
- More parsimonious structures can be considered.
- The unstructured mean model is most appealing when subject measured at the same time point and for short follow-up.
- Using a more parsimonious mean structure can increase the power of tests $\Rightarrow$ use a parametric or semi-parametric curve.
- When data are not collected at the same points in time, postulating a saturated mean model is not possible.

Linear models with correlated errors - Model the mean II

- We will use a parametric model for the mean.
- Consider again the ARMD longitudinal trial where VAS is measured at randomization and at 4, 12, 24, and 52 weeks afterwards in each treatment group.
- Research question: Does the mean VAS progression differ between the 2 groups?
- The simplest mean structure is a **linear trend** over time:

$$\boxed{y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + \varepsilon_{ij}}$$

with $i = 1, \ldots, 240$, $j = 0, \ldots, 4$, $\text{time}_{ij} \in \{0, 4, 12, 24, 52\}$, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$.

- For $\boldsymbol{\Sigma}_i$ we assume unstructured covariance matrix (we will discuss this later).

Linear models with correlated errors - Model the mean II

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + \varepsilon_{ij}$$

	Value	Std.Error	t-value	p-value
(Intercept)	55.312	1.355	40.834	0.000
TreatActive	-1.022	1.907	-0.536	0.592
Weeks	-0.209	0.029	-7.238	0.000
TreatActive:Weeks	-0.085	0.042	-2.028	0.043

- Coefficients' interpretation:
  - β_0: mean VAS in placebo group at baseline.
  - β_1: change in the mean VAS between the two groups at baseline.
  - β_2: change in the mean VAS for every week that passes by in the placebo group.
  - $\beta_2 + \beta_3$: change in the mean VAS for every week that passes by in the treatment group.
  - The change in the mean VAS between treatment and placebo for every week that passes by is β_3 (= -0.085) (p-value = 0.046).

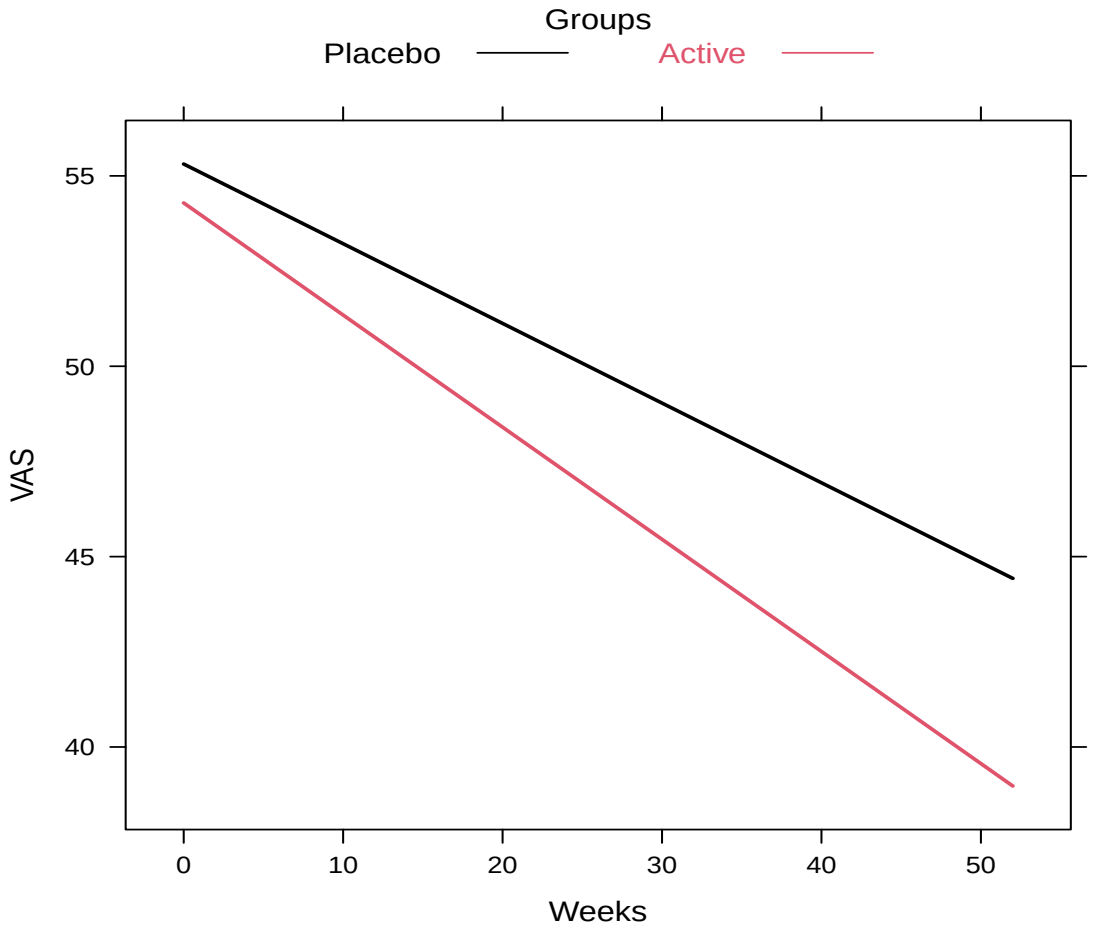
Linear models with correlated errors - Model the mean II

- Design matrix $\mathbf{X}_i$:
  `model.matrix(VAS ~ Treat + Weeks + Weeks:Treat, data = armd.)`

	(Intercept)	TreatActive	Weeks	TreatActive:Weeks
4	1	1	0	0
5	1	1	4	4
6	1	1	12	12
7	1	1	24	24
8	1	1	52	52

Linear models with correlated errors - Model the mean II

- The fitted profiles per group are:



- Constant rate of change for each group: derivative.

Linear models with correlated errors - Model the mean II

- The estimated covariance matrix $\boldsymbol{\Sigma}_i$ is:

```
Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 221.10 202.25 188.78 181.64 142.09
[2,] 202.25 253.73 231.29 219.30 175.52
[3,] 188.78 231.29 297.82 261.06 225.15
[4,] 181.64 219.30 261.06 342.09 291.10
[5,] 142.09 175.52 225.15 291.10 351.50
  Standard Deviations: 14.869 15.929 17.257 18.496 18.748
```

- The estimated correlation matrix is:

```
      [,1]  [,2]  [,3]  [,4]  [,5]
[1,] 1.000 0.854 0.736 0.660 0.510
[2,] 0.854 1.000 0.841 0.744 0.588
[3,] 0.736 0.841 1.000 0.818 0.696
[4,] 0.660 0.744 0.818 1.000 0.839
[5,] 0.510 0.588 0.696 0.839 1.000
```

Linear models with correlated errors - Model the mean III

- Consider the ARMD longitudinal trial where VAS is measured at randomization and at 4, 12, 24, and 52 weeks afterwards in each treatment group.
- Research question: Does the mean VAS progression differ between the 2 groups?
- When changes in the mean response are not linear over time, consider polynomials:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij}^2 + \beta_4 \text{time}_{ij} \times \text{group}_i + \beta_5 \text{time}_{ij}^2 \times \text{group}_i + \varepsilon_{ij}$$

with $i = 1, \ldots, 240$, $j = 0, \ldots, 4$, $\text{time}_{ij} \in \{0, 4, 12, 24, 52\}$, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$.

- We assume **quadratic trend** over time.
- For $\boldsymbol{\Sigma}_i$ we assume unstructured covariance matrix (we will discuss this later).

Linear models with correlated errors - Model the mean III

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij}^2 + \beta_4 \text{time}_{ij} \times \text{group}_i + \beta_5 \text{time}_{ij}^2 \times \text{group}_i + \varepsilon_{ij}$$

	Value	Std.Error	t-value	p-value
(Intercept)	55.337	1.356	40.798	0.000
TreatActive	-0.974	1.910	-0.510	0.610
I(Weeks)	-0.256	0.081	-3.167	0.002
I(Weeks^2)	0.001	0.001	0.621	0.535
TreatActive:I(Weeks)	-0.181	0.116	-1.555	0.120
TreatActive:I(Weeks^2)	0.002	0.002	0.899	0.369

- Coefficients' interpretation:
  - β_0: mean VAS in placebo group at baseline.
  - β_1: change in the mean VAS between the two groups at baseline.
  - $\beta_2 + 2\beta_3 \times \text{time}_{ij}$: rate of change of the mean VAS at time j in the placebo group.
  - $\beta_2 + 2\beta_3 \times \text{time}_{ij} + \beta_4 + 2\beta_5 \times \text{time}_{ij}$: rate of change of the mean VAS at time j in the treatment group.

Linear models with correlated errors - Model the mean III

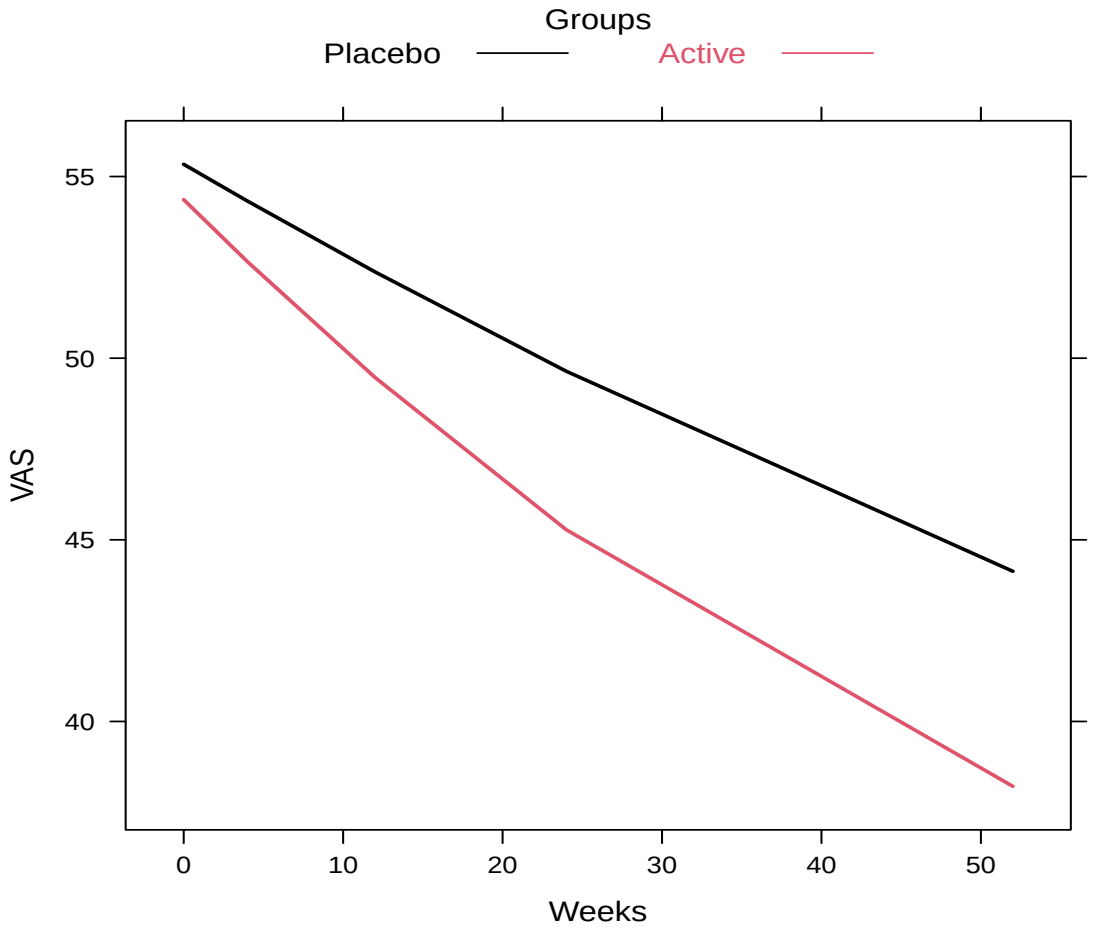
- Design matrix $\mathbf{X}_i$:

	(Intercept)	TreatActive	I(Weeks)	I(Weeks^2)	TreatActive:I(Weeks)	TreatActive:I(Weeks^2)
4	1	1	0	0	0	0
5	1	1	4	16	4	16
6	1	1	12	144	12	144
7	1	1	24	576	24	576
8	1	1	52	2704	52	2704

- Note the function: `I(.)`.

Linear models with correlated errors - Model the mean III

- The fitted profiles per group are:



Linear models with correlated errors - Model the mean III

- Note: it is often advisable to center time (i.e. subsract mean) to avoid collinearity. Alternatively, orthogonal polynomials.
- Interpretation of the coefficients with polynomials difficult.
- Easier to communicate results via fitted mean profiles.

Linear models with correlated errors - Model the mean IV

- Consider the ARMD longitudinal trial where VAS is measured at randomization and at 4, 12, 24, and 52 weeks afterwards in each treatment group.
- Research question: Does the mean VAS progression differ between the 2 groups?
- For non-linear changes in time polynomials can be difficult to interpret.
- Linear splines offer a flexible alternative.
- They assume piecewise linear changes over time:

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + \varepsilon_{ij}, \quad \text{when} \quad \text{time}_{ij} \leq t^* \\ y_{ij} &= \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + \\ &\quad \beta_4 (\text{time}_{ij} - t^*) + \beta_5 (\text{time}_{ij} - t^*) \times \text{group}_i + \varepsilon_{ij}, \quad \text{when} \quad \text{time}_{ij} > t^* \end{aligned}$$

with $i = 1, \ldots, 240$, $j = 0, \ldots, 4$, $\text{time}_{ij} \in \{0, 4, 12, 24, 52\}$, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$.

- For $\boldsymbol{\Sigma}_i$ we assume unstructured covariance matrix (we will discuss this later).

Linear models with correlated errors - Model the mean IV

- Or in a more compact way:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_{21} \text{time}_{ij} + \beta_{22}(\text{time}_{ij} - t^*)_+ + \beta_{31} \text{time}_{ij} \times \text{group}_i + \beta_{32}(\text{time}_{ij} - t^*)_+ \times \text{group}_i + \varepsilon_{ij}$$

where $u_+ = u$ when $u > 0$ and 0 when $u \leq 0$.

	Value	Std.Error	t-value	p-value
(Intercept)	55.318	1.356	40.798	0.000
TreatActive	-1.016	1.909	-0.532	0.595
Weeks	-0.244	0.054	-4.555	0.000
I(pmax(Weeks - 25, 0))	0.053	0.070	0.768	0.443
TreatActive:Weeks	-0.116	0.077	-1.504	0.133
TreatActive:I(pmax(Weeks - 25, 0))	0.050	0.101	0.499	0.618

- The breakpoints are called knots.
- k knots define $k+1$ intervals on which a linear progression is assumed.
- Choice of location and number of knots can be based on subject matter input or hypothesis testing (to discuss later).

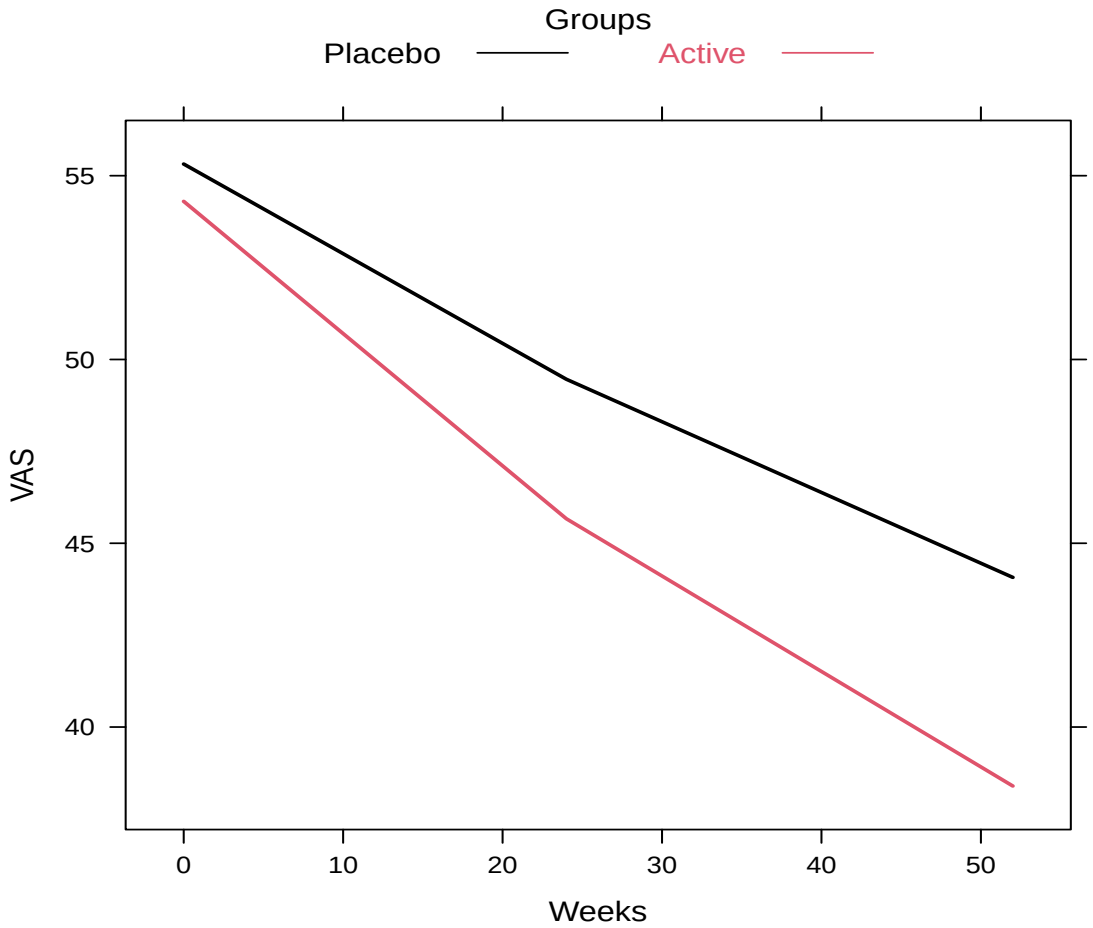
Linear models with correlated errors - Model the mean IV

- Design matrix $\mathbf{X}_i$:

```
  (Intercept) TreatActive Weeks I(pmax(Weeks - 25, 0)) TreatActive:Weeks
4           1           1     0                      0                 0
5           1           1     4                      0                 4
6           1           1    12                      0                12
7           1           1    24                      0                24
8           1           1    52                     27                52
  TreatActive:I(pmax(Weeks - 25, 0))
4                                  0
5                                  0
6                                  0
7                                  0
8                                 27
```

Linear models with correlated errors - Model the mean IV

- The fitted profiles per group are:



- Interpretation of model parameters.

Linear models with correlated errors - Model the mean V

- Consider the ARMD longitudinal trial where VAS is measured at randomization and at 4, 12, 24, and 52 weeks afterwards in each treatment group.
- Research question: Does the mean VAS progression differ between the 2 groups?
- For non-linear changes in time polynomials can be difficult to interpret and they are known to be sensitive to outliers.
- Linear splines offer a flexible alternative but are less smooth.
- Natural cubic or B-splines offer more flexible or smooth alternatives and are less sensitive to outliers:

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij}^2 + \\ &\quad \beta_4 (\text{time}_{ij} - t^*)_+^3 + \beta_5 \text{time}_{ij} \times \text{group}_i + \\ &\quad \beta_6 \text{time}_{ij}^2 \times \text{group}_i + \beta_7 (\text{time}_{ij} - t^*)_+^3 \times \text{group}_i + \varepsilon_{ij} \end{aligned}$$

where $u_+ = u$ when $u > 0$ and 0 when $u \leq 0$. with $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$.

- For $\boldsymbol{\Sigma}_i$ we assume unstructured covariance matrix (we will discuss this later).

Linear models with correlated errors - Model the mean V

	Value	Std.Error	t-value	p-value
(Intercept)	55.330	1.356	40.801	0.000
TreatActive	-0.992	1.909	-0.520	0.603
ns(Weeks, knots = 25)1	-13.822	2.565	-5.389	0.000
ns(Weeks, knots = 25)2	-8.393	1.190	-7.053	0.000
TreatActive:ns(Weeks, knots = 25)1	-7.061	3.678	-1.920	0.055
TreatActive:ns(Weeks, knots = 25)2	-3.157	1.735	-1.819	0.069

- The breakpoints are called knots.
- k knots define $k+1$ intervals on which a cubic polynomials over time is assumed.
- Cubic splines can be made to be smooth at the join points (knots) by forcing the first and second derivatives of the function to agree at the knots.
- Cubic splines can behave poorly outside bountary knots $\Rightarrow$ Natural cubic splines are linear in the tails.
- Choice of location and number of knots can be based on subject matter input. Often 3-4 knots at sample quantiles is sufficient.

Linear models with correlated errors - Model the mean V

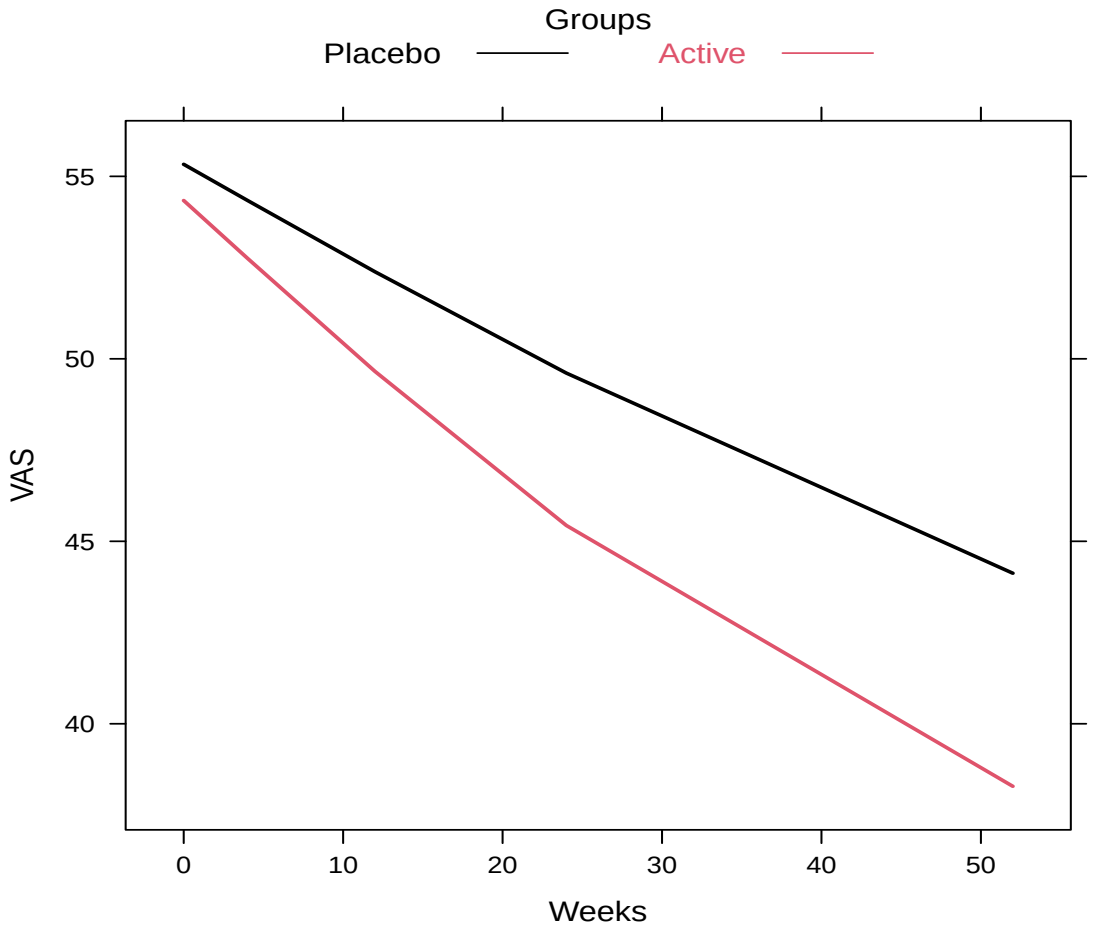
- Design matrix $\mathbf{X}_i$:

	(Intercept)	TreatActive	ns(Weeks, knots = 25)1	ns(Weeks, knots = 25)2
4	1	1	0.0000000	0.0000000
5	1	1	0.1187295	-0.0775279
6	1	1	0.3380256	-0.2056722
7	1	1	0.5534511	-0.2296913
8	1	1	0.3389098	0.7769202

	TreatActive:ns(Weeks, knots = 25)1	TreatActive:ns(Weeks, knots = 25)2
4	0.0000000	0.0000000
5	0.1187295	-0.0775279
6	0.3380256	-0.2056722
7	0.5534511	-0.2296913
8	0.3389098	0.7769202

Linear models with correlated errors - Model the mean V

- The fitted profiles per group are:



Linear models with correlated errors - Model the mean

- In all the models we have seen so far the mean is linear in the parameters.
- We will not consider models where the mean is non-linear in the parameters, e.g.:

$$\begin{aligned} E(Y) &= \beta_0 + \exp\{\beta_1 X\} \text{ or} \\ E(Y) &= \beta_0 (1 + \beta_1 \exp\{-\beta_2 X\})^{-1}. \end{aligned}$$

  these models are known as **non-linear** regression models.

- This does not mean that we do not allow for a non-linear relationship between the mean and covariates.
- In such cases, as we show, we use polynomials, splines, transformation of covariates, etc.

Linear models with correlated errors - Summary: Model the mean

- The unstructured mean model is most appealing when subject measured at the same time point and for short follow-up.
- A parametric curve for the mean (linear trend, polynomials or splines) is more parsimonious and thus can increase the power of tests.
- Interpretation of the coefficients with higher order polynomials difficult.
- Easier to communicate results via fitted mean profiles.

Linear models with correlated errors - Modelling the covariance

- Proper modelling of the covariance between the repeated measurements leads to correct standard errors and thus valid inferences.
- Even though mean model and covariance matrix are chosen separately they are interdependent.
- The covariance matrix describes the correlation of the pairwise residuals which depend on mean model.
- Let a repeated measures dataset:

Subject	Y_1	Y_2	$\ldots$	Y_N
1	y_{11}	y_{12}	$\cdots$	y_{1N}
2	y_{21}	y_{22}	$\cdots$	y_{2N}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{n1}	y_{n2}	$\cdots$	y_{2N}

- For any pair Y_i and Y_j with $i, j \in \{1, \ldots, N\}$ we compute $\text{Cov}(Y_i, Y_j)$.

Linear models with correlated errors - Modelling the covariance

- The dependence structure among all observations $\mathbf{y}$ is described by the variance-covariance matrix:

$$\mathbf{\Sigma} = \begin{bmatrix} \text{Var}(y_1) & \text{Cov}(y_1, y_2) & \dots & \text{Cov}(y_1, y_N) \\ \text{Cov}(y_2, y_1) & \text{Var}(y_2) & \dots & \text{Cov}(y_2, y_N) \\ \vdots & \vdots & \vdots & \vdots \\ \text{Cov}(y_N, y_1) & \dots & \dots & \text{Var}(y_N) \end{bmatrix}.$$

- Variances appear on the main diagonal.
- Covariances on the off-diagonal elements with $\Rightarrow Cov(y_1, y_2) = Cov(y_2, y_1)$.

Linear models with correlated errors - Modelling the covariance

- For each sampling unit i (e.g. patient, school, family) with n_i correlated repeated measurements:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,\ \boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$$

  with $\boldsymbol{\Sigma}_i$ the variance-covariance matrix.

- We need an appropriate choice for $\boldsymbol{\Sigma}_i$ in order to properly describe the correlations between the repeated measurements:
  - Unstructured
  - Compound Symmetry
  - M-dependence or Toeplitz
  - Autoregressive of order 1 (AR-1), etc.

Linear models with correlated errors - Modelling the covariance

- **Unstructured**
  - For 4 repeated measurements:

$$\Sigma_i = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

  - Number of parameters increases with follow-up time i.e. $\frac{N \times (N+1)}{2} \Rightarrow$ it is more suitable for short follow-up provided sufficient sample size.
  - Not suitable when data are collected at different time points for each subject.
  - It can be used in case of unbalanced data (i.e missing data) provided planned measurement at the same time points for all subjects.

Linear models with correlated errors - Modelling the covariance

- **Compound symmetry**
  - For 4 repeated measurements:

$$\Sigma_i = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

  - For each subject all pairwise correlations are the same ρ and variance constant across occations σ^2.
  - Not very realistic for longitudinal data $\Rightarrow$ we expect correlations to decrease in time and non-constant variance.
  - Perhaps more realistic in multilevel designs e.g. data from students on the same classroom or from members of the same family.
  - It can be extended to unequal variances.

Linear models with correlated errors - Modelling the covariance

- **Autoregressive of order 1 (AR-1)**
  - For each subject the correlation depends on the distance in time between two measurements $|t_j - t_k|$.
  - $\text{Corr}(\varepsilon_k, \varepsilon_j) = \rho^{|t_j - t_k|}$, $-1 \leq \rho \leq 1$

    $\Rightarrow$ Repeated measurements have a first-order autoregressive relationship.
  - For 4 repeated measurements collected at equal time points i.e. at $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$

$$\Sigma_i = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \sigma^2\rho^3 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ \sigma^2\rho^3 & \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

  - Because $-1 \leq \rho \leq 1$ the pairwise correlations decrease with distance in time between the repeated measurements.

Linear models with correlated errors - Modelling the covariance

- **Autoregressive of order 1 (AR-1)**
  - It can be extended for heterogeneous variances.
  - For each subject the correlation depends on the distance in time between two measurements $|t_j - t_k|$.
  - $\text{Corr}(\varepsilon_k, \varepsilon_j) = \rho^{|t_j - t_k|}$, $-1 \leq \rho \leq 1$

    $\Rightarrow$ Repeated measurements have a first-order autoregressive relationship.
  - For 4 repeated measurements collected at equal time points i.e. at $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$

$$\Sigma_i = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

  - Because $-1 \leq \rho \leq 1$ the pairwise correlations decrease with distance in time between the repeated measurements.

Linear models with correlated errors - Modelling the covariance

- M-dependence or Toeplitz
  - Correlations k occasions apart are the same.
  - Correlations $> M$ occasions apart are zero.
- For 4 repeated measurements: 1-dependence

$$\Sigma_i = \begin{bmatrix} \sigma^2 & \sigma_{12} & 0 & 0 \\ \sigma_{12} & \sigma^2 & \sigma_{12} & 0 \\ 0 & \sigma_{12} & \sigma^2 & \sigma_{12} \\ 0 & 0 & \sigma_{12} & \sigma^2 \end{bmatrix}$$

Linear models with correlated errors - Modelling the covariance

- M-dependence or Toeplitz
  - For 4 repeated measurements: 2-dependence

$$\Sigma_i = \begin{bmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & 0 \\ \sigma_{12} & \sigma^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{13} & \sigma_{12} & \sigma^2 & \sigma_{12} \\ 0 & \sigma_{13} & \sigma_{12} & \sigma^2 \end{bmatrix}$$

- Or the generalization for heterogeneous variances

$$\Sigma_i = \begin{bmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \rho_2\sigma_1\sigma_3 & 0 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 & \rho_1\sigma_1\sigma_2 & \rho_2\sigma_1\sigma_3 \\ \rho_2\sigma_1\sigma_3 & \rho_1\sigma_1\sigma_2 & \sigma_3^2 & \rho_1\sigma_1\sigma_2 \\ 0 & \rho_2\sigma_1\sigma_3 & \rho_1\sigma_1\sigma_2 & \sigma_4^2 \end{bmatrix}$$

- Appropriate when equally spaced measurements.
- Special case: Autoregressive of order 1.

Linear models with correlated errors - Modelling the covariance

- The covariance matrices we have seen so far are appropriate when data are collected at the same occassions for all subjects.

- For highly unbalanced data collected at different occassions per subject:
  - Continuous AR1.
  - Exponential serial correlation.
  - Gaussian serial correlation.

Linear models with correlated errors - Modelling the covariance

- In longitudinal studies it is expected that the correlation is a decreasing function of the time span between the repeated measurements.

- The correlation between any time points t_{ij} and t_{ik} is

$$Corr(\varepsilon_{ij}, \varepsilon_{ik}) = h[d(t_{ij}, t_{ik}), \phi]$$

where

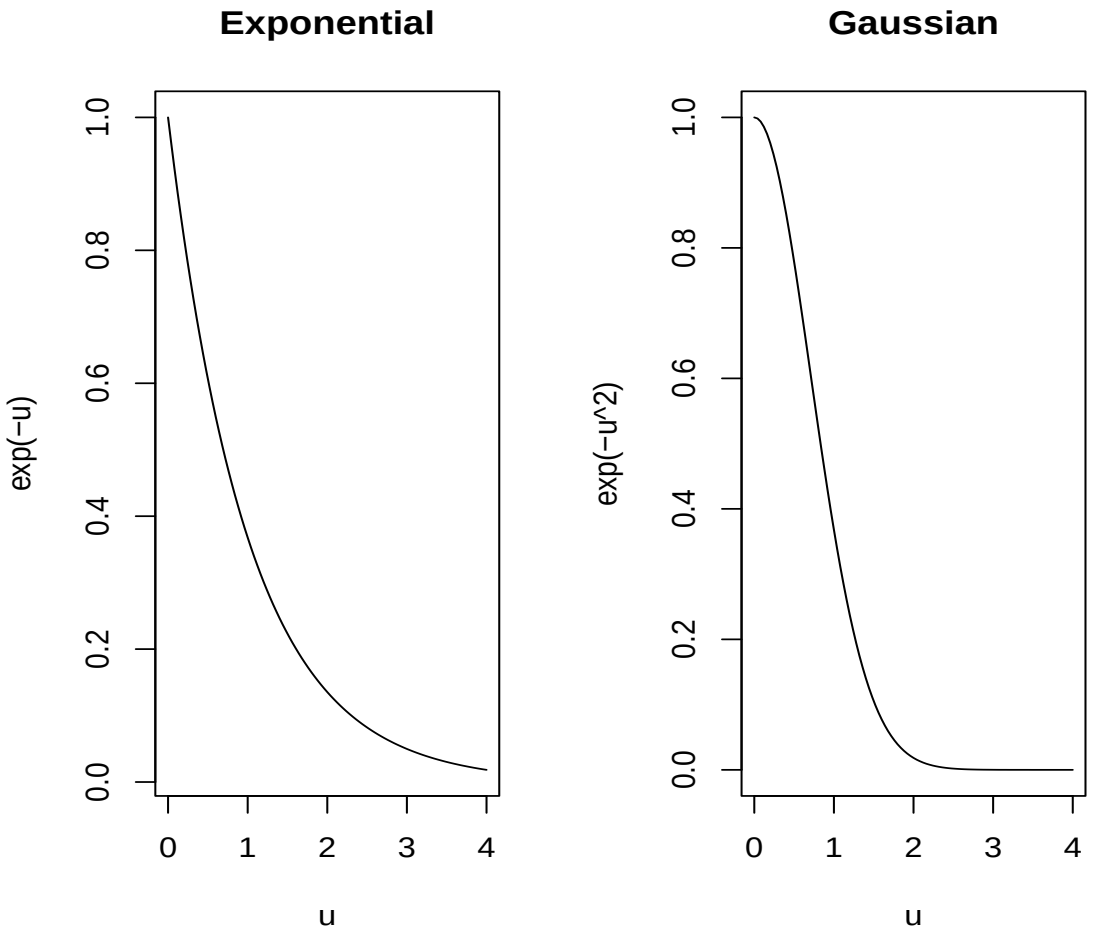
  - ϕ is a correlation parameter $\phi > 0$.
  - $d(t_{ij}, t_{ik})$ is a distance function.
  - $h(,)$ is a continuous function, monotonically nonincreasing with respect to $d(t_{ij}, t_{ik})$ at $d(t_{ij}, t_{ik}) = 0$.
  - $h(0, \phi) = 1$: this may not be realistic for human due to measurement error.
  - $h(0, \phi) = 1$ can be relaxed by using nugget effect ρ_0, i.e. allow discontinuity at $d(t_{ij}, t_{ik}) = 0$.

$$h_{\rho_0}[d(t_{ij}, t_{ik}),\ \phi] = \begin{cases} (1-\rho_0) h_{\rho_0}[d(t_{ij}, t_{ik}),\ \phi], & d(t_{ij}, t_{ik}) > 0 \\ 1-\rho_0, & d(t_{ij}, t_{ik}) = 0 \end{cases}$$

Linear models with correlated errors - Modelling the covariance

- Different distance metrics and correlation functions imply different correlation structures:
  - Continuous AR1: $Corr(\varepsilon_{ij}, \varepsilon_{ik}) = \phi^{d(t_{ij}, t_{ik})}$, $\phi > 0$.
  - Exponential serial correlation: $h(u) = exp(u/\phi)$ $(u = \mid t_{ij} t_{ik} \mid)$.
    - $h(0) = 1$ if multiple measurements are taken at the same occasion $\Rightarrow$ implies no measurement error which is not realistic for humans.
    - $h(u) = 0$ for measurements taken far apart $\Rightarrow$ it is rarely observed in longitudinal studies.
  - Gaussian serial correlation: $h(u) = exp((u/\phi)^2)$.

Linear models with correlated errors - Modelling the covariance



Linear models with correlated errors - Modelling the covariance

- Each statistical software specifies differently the covariance matrix, e.g., in `nlme` the marginal covariance matrix Σ_i is specified as:

$$\Sigma_i = \sigma^2 R_i$$

  where $R_i = \Lambda_i C_i \Lambda_i$, with Λ_i a diagonal matrix with nonnegative diagonal elements and C_i is a correlation matrix.

- Via Λ_i we allow for heteroscedasticity of variance within subject i.

- Via C_i we model the correlation between measurements within subject i.

Linear models with correlated errors - Summary: Modelling the covariance

- Reducing the parameters in the covariance matrix leads to more efficient inferences for the mean parameters.
- This is particularly useful when many repeated measurements are taken per subject.
- Covariance matrices allow the variance and correlation to change over time. They may also depend on covariates e.g. a different covariance matrix between males and females.
- All standard statistical software offer many covariance structures.

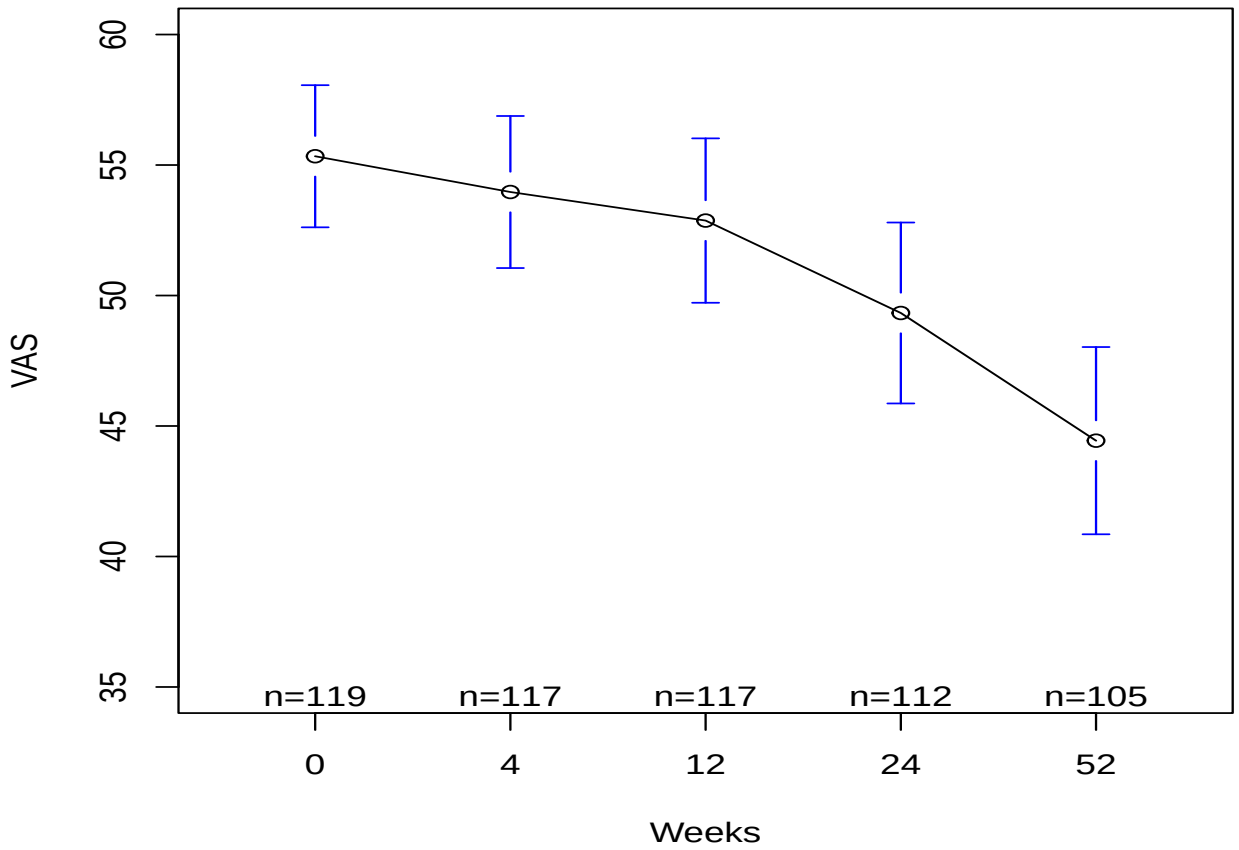
Outline

Linear models with correlated errors - Exploratory Data Analysis

- Linear models with correlated errors make assumptions about:
  - Mean structure: linear or nonlinear progression, covariates, etc.
  - Variance function: contant, heteskedasticity, etc.
  - Correlation struction: constant, serial, spatial.
- Data exploration is very useful as an additional tool in selecting the best fitting model.

Linear models with correlated errors - Exploring Mean Structure

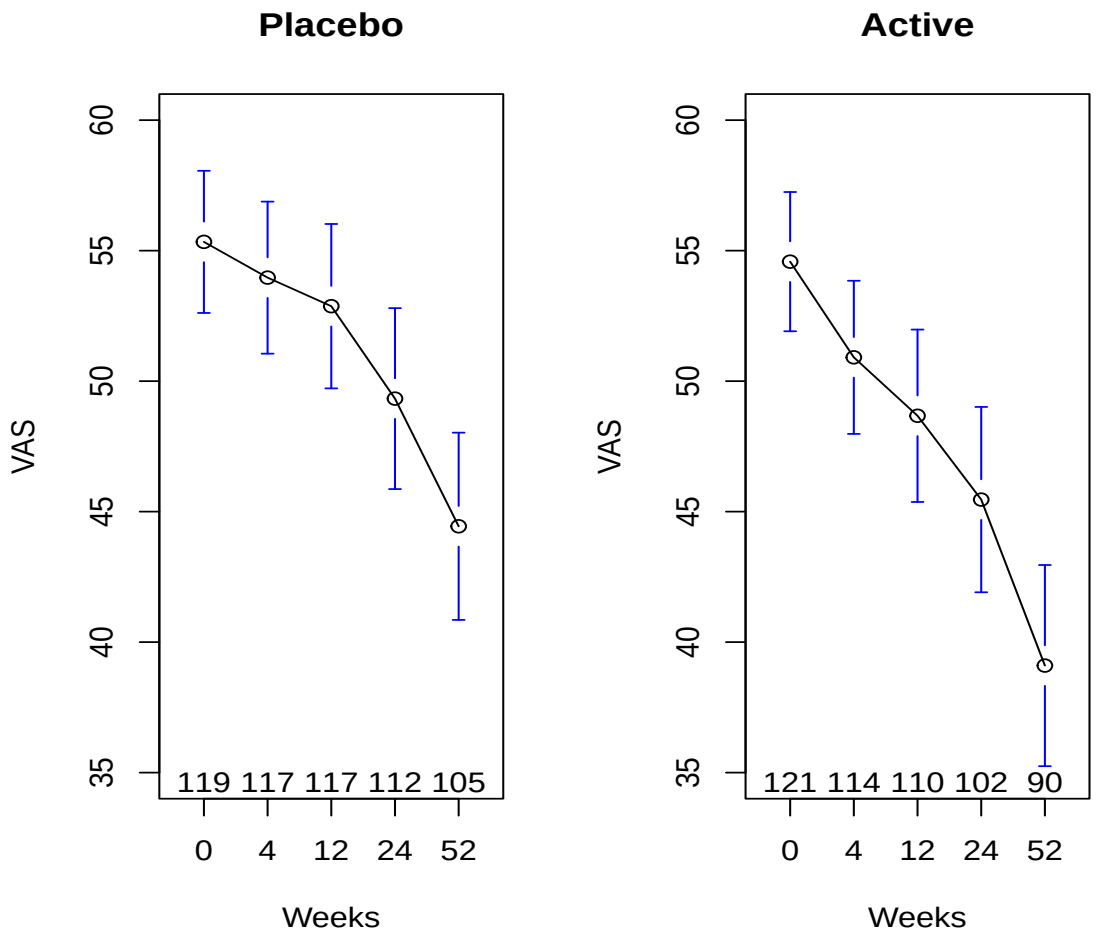
- Compute means and std errors for the means at each time point (balanced data).



- We observe non-linear mean evolution and increasing standard error due to dropout.

Linear models with correlated errors - Exploring Mean Structure

- Same plots for levels of important covariates.

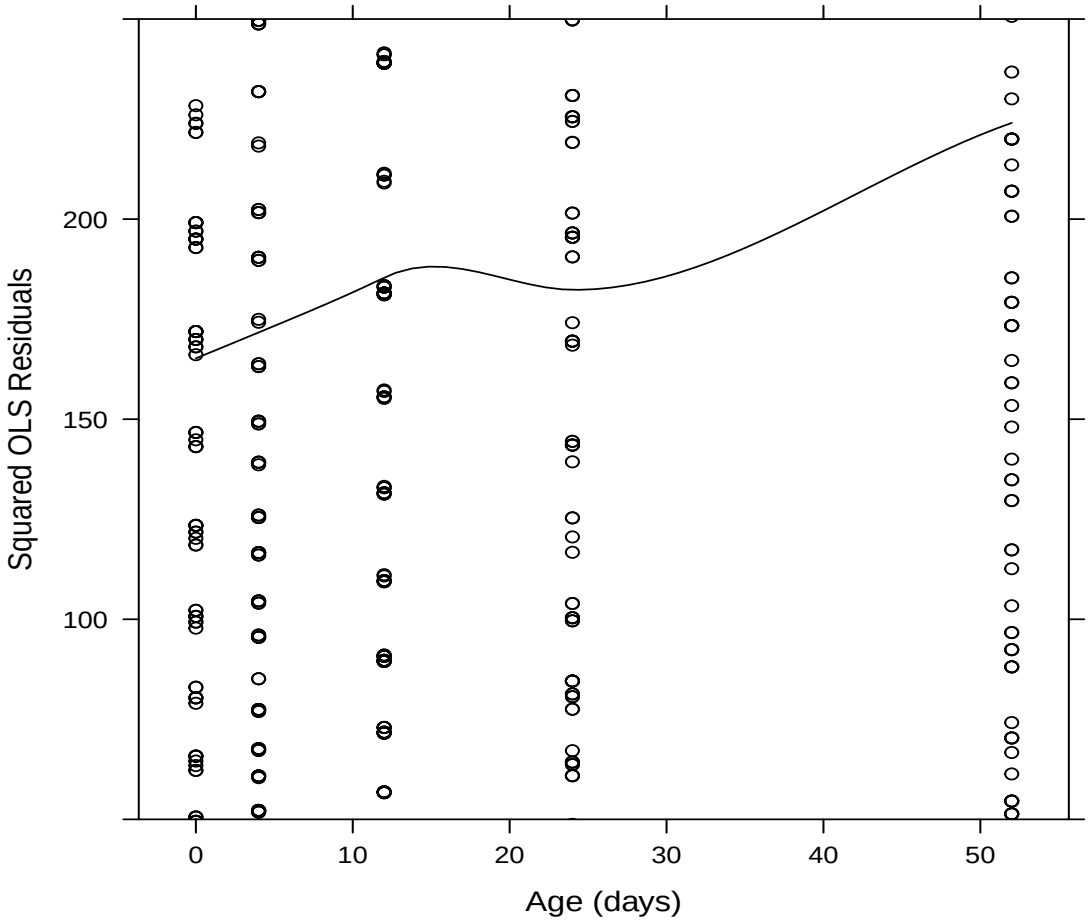


Linear models with correlated errors - Exploring Variance function

- The variance function is given by

$$\sigma^2(t) = E\left[y(t) - \mu(t)\right]^2$$

- Thus the variance function can be explored by studying the OLS residuals:



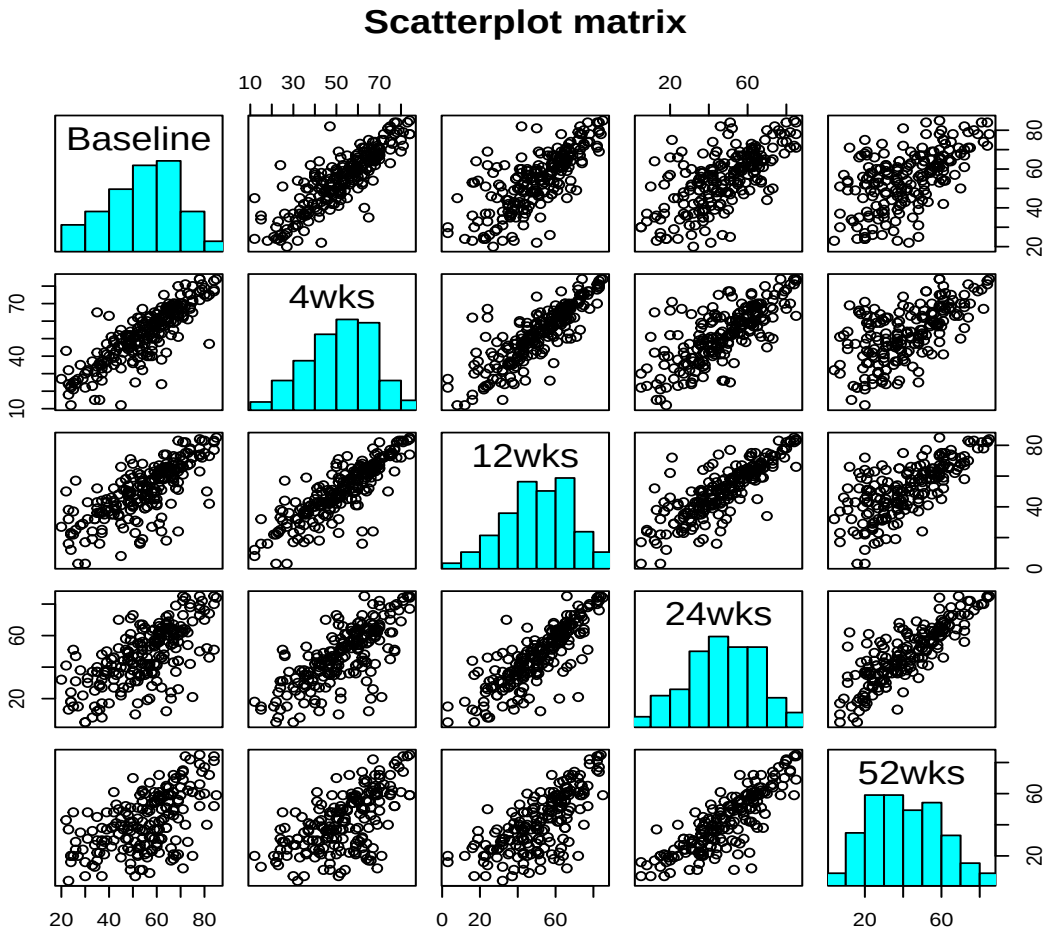
Linear models with correlated errors - Exploring correlation structure

- For balanced designs correlation matrix describes pairwise correlations:

	visual0	visual4	visual12	visual24	visual52
visual0	1.00	0.85	0.74	0.65	0.56
visual4	0.85	1.00	0.85	0.73	0.61
visual12	0.74	0.85	1.00	0.80	0.69
visual24	0.65	0.73	0.80	1.00	0.83
visual52	0.56	0.61	0.69	0.83	1.00

Linear models with correlated errors - Exploring correlation structure

- We can visualize correlations via pairwise scatterplots:



Outline

Linear models with correlated errors - Estimation

- Estimation of the model parameters θ:
  - mean model parameters $\boldsymbol{\beta}$
  - all variance-covariance parameters α in $\boldsymbol{\Sigma}$

  is done by maximizing the marginal likelihood of the marginal model.

- Marginal likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} \left\{ (2\pi)^{-n_i/2} \mid \boldsymbol{\Sigma}_i \mid^{-1/2} \exp \left[(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \right] \right\}$$

Linear models with correlated errors - Estimation

- For known $\boldsymbol{\Sigma}$, $\boldsymbol{\beta}$ is estimated by:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i$$

  where $\boldsymbol{\Sigma}_i$ is replaced by its Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) estimate and thus we get $\hat{\boldsymbol{\beta}}_{ML}$ or $\hat{\boldsymbol{\beta}}_{REML}$, respectively.

- ML estimation of the model parameters α.
  - $\hat{\alpha}_{ML}$ is obtained by maximizing $L_{ML}(\alpha, \hat{\boldsymbol{\beta}})$.

- REML estimation of the model parameters α.
  - $\hat{\alpha}_{REML}$ is obtained by maximizing $L_{REML}(\alpha, \hat{\boldsymbol{\beta}}) = \mid \sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \mid^{-1/2} L_{ML}(\alpha, \hat{\boldsymbol{\beta}})$.

Linear models with correlated errors - Estimation

- Why we need the REML?
- Variance Estimation in Normal Populations
  - Let a random sample $y_1, \ldots, y_n$ from $N(\mu, \sigma^2)$.
  - For known μ the ML estimate of σ^2 is $\hat{\sigma}^2 = \sum_{i=1}^{n} (y_i - \mu)/n$
  - $\hat{\sigma}^2$ is an unbiased estimate of σ^2
  - When μ is unkown $\hat{\sigma}^2 = \sum_{i=1}^{n} (y_i - \bar{y})/n$
- Having to estimate μ introduces bias in $\hat{\sigma}^2$.
- REML offers a solution to estimate σ^2, without estimating μ first by data transformation.

Linear models with correlated errors - Estimation

- Note that $L_{REML}(\theta)$ is NOT the likelihood for our original data $\mathbf{Y}$.
- REML applies a transformation in the repeated measurements Y based on the chosen covariates in the model:

$$L_{REML}(\alpha, \hat{\boldsymbol{\beta}}) = \mid \sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \mid^{-1/2} L_{ML}(\alpha, \hat{\boldsymbol{\beta}}).$$

- **Thus we should not compare the likelihoods of models fitted with REML that have different $\mathbf{X}\beta$ part.**

Linear models with correlated errors - Estimation - Example

- Consider a marginal model with unstructured covariance matrix for the ARMD data with linear progression in time that differs between the two treatment groups.
- Marginal covariance matrix using ML

```
         [,1]     [,2]     [,3]
[1,] 221.1011 202.2456 188.7807
[2,] 202.2456 253.7274 231.2886
[3,] 188.7807 231.2886 297.8161
```

- Marginal covariance matrix using REML

```
         [,1]     [,2]     [,3]
[1,] 222.9269 204.0261 190.4624
[2,] 204.0261 255.4767 232.9685
[3,] 190.4624 232.9685 299.4907
```

- We observe small differences in this dataset due to the relatively big sample size. In studies with less subjects the differences are more pronounced.
- Also the estimates for fixed effects and their standard errors are similar in this case.

Outline

Statistical Software to Fit Linear models with correlated errors

- Most popular programs in medical applications:
  - R function `gls(.)` in package `nlme`
  - SAS proc MIXED
- STATA
- SPSS
- MLwin (popular in social sciences)
- WINBUGS (Bayesian methodology)
- GENSTAT
- We will use mainly R, material is available on blackboard for SAS.

Outline

Fitting Linear models with correlated errors in R

- Marginal models can be fitted using function `gls()` from R package `nlme`.
- `gls()` has five main arguments:
  - `model:` a formula specifying the response vector and the covariates to be used.
  - `correlation:` a formula specifying the correlation structure.
  - `weights:` a formula specifying the within-group variance structure.
  - `data:` a data frame containing all the variables.
  - `method:` estimation approach used 'ML' or 'REML'.
- The data should be in long format!

Fitting Linear models with correlated errors in R

- R code for the various correlation structures (Unstructured mean model is assumed):
  - **Unstructured/General** correlation structure:

```
> model.unstr.homo <- gls(VAS ~ Timef * Treat, correlation = corSymm(form = ~ 1 | ID),
+                 data = armd., method = "REML")
```

  - **Compound symmetry** correlation structure:

```
> model.compsymm <- gls(VAS ~ Timef * Treat, correlation = corCompSymm(form = ~ 1 | ID),
+                       data = armd., method = "REML")
```

  - **AR1** correlation structure:

```
> model.ar1 <- gls(VAS ~ Timef * Treat, correlation = corAR1(form = ~ 1 | ID),
+                       data = armd., method = "REML")
```

  - **ARMA** correlation structure:

```
> model.arma <- gls(VAS ~ Timef * Treat, correlation = corARMA(form = ~ 1 | ID, p = 1, q = 2),
+                       data = armd., method = "REML")
```

Fitting Linear models with correlated errors in R

- R code for the various correlation structures (Unstructured mean model is assumed):
  - **Continuous AR1** correlation structure:

```
> model.car1 <- gls(VAS ~ Timef * Treat, correlation = corCAR1(form = ~ 1 | ID),
+                       data = armd., method = "REML")
```

  - **Exponential** spatial correlation structure:

```
> model.exp <- gls(VAS ~ Timef * Treat, correlation = corExp(form = ~ 1 | ID),
+                       data = armd., method = "REML")
```

  - **Gaussian** spatial correlation structure:

```
> model.gaus <- gls(VAS ~ Timef * Treat, correlation = corGaus(form = ~ 1 | ID),
+                       data = armd., method = "REML")
```

  - **Linear** spatial correlation structure:

```
> model.lin <- gls(VAS ~ Timef * Treat, correlation = corLin(form = ~ 1 | ID),
+                       data = armd., method = "REML")
```

  - **Rational quadratic** spatial correlation structure:

```
> model.ratio <- gls(VAS ~ Timef * Treat, correlation = corRatio(form = ~ 1 | ID),
+                       data = armd., method = "REML")
```

  - **Spherical** spatial correlation structure:

```
> model.spher <- gls(VAS ~ Timef * Treat, correlation = corSpher(form = ~ 1 | ID),
+                       data = armd., method = "REML")
```

Fitting Linear models with correlated errors in R

- we use the formula: `form = ~ 1 | ID`
  - Correlation applies to data of the same individual in `ID`.
  - Data from different individuals are uncorrelated.
  - The ordering of the data is used as a position variable when `~ 1` is used; a time variable denoting the ordering within each subject may be used as well. $\Rightarrow$ Important when visits are missing or unbalanced data.
- The default distance metric is `euclidean`.
- Nugget effect is allowed via `nugget = TRUE` argument.
- Specification of the `correlation` argument corresponds to specifying matrix C_i in the marginal covariance matrix $\Sigma_i = \Lambda_i C_i \Lambda_i$.
- More on the correlation matrices supported in R can be found by: `?corClasses`.

Fitting Linear models with correlated errors in R

- So far costant within group variances are assumed.
- To allow for within group heterogeneity we use the `weights` argument:
  - **Unstructured** covariance structure:

```
> model.unstr <- gls(VAS ~ Timef * Treat, correlation = corSymm(form = ~ 1 | ID),
+                    weights = varIdent(form = ~ 1|Timef), data = armd., method = "REML")
```

- Specification of the `weights` argument corresponds to specifying matrix Λ_i in the marginal covariance matrix $\Sigma_i = \Lambda_i C_i \Lambda_i$.
- In R several variance functions are supported such as: `varExp(.)`, `varPower(.)`, etc. $\Rightarrow$ use `?varClasses` for the complete list.

Fitting Linear models with correlated errors in R

- `summary()`:

```
> summary(model.unstr)
Generalized least squares fit by REML
  Model: VAS ~ Weeks * Treat
  Data: armd.
       AIC      BIC   logLik
  8375.699 8470.809 -4168.85

Correlation Structure: General
 Formula: ~1 | ID
 Parameter estimate(s):
 Correlation:
  1     2     3     4
2 0.855
3 0.737 0.842
4 0.662 0.745 0.819
5 0.509 0.588 0.696 0.840
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | Timef
 Parameter estimates:
Baseline     4wks    12wks    24wks    52wks
1.000000 1.070519 1.159072 1.242012 1.261392

Coefficients:
                    Value Std.Error  t-value p-value
(Intercept)       55.31142 1.3577460 40.73768  0.0000
Weeks             -0.20932 0.0290116 -7.21519  0.0000
TreatActive       -1.01977 1.9117089 -0.53343  0.5938
Weeks:TreatActive -0.08531 0.0421369 -2.02457  0.0432

 Correlation:
                  (Intr) Weeks  TrtAct
Weeks             -0.296
TreatActive       -0.710  0.210
Weeks:TreatActive  0.204 -0.689 -0.284

Standardized residuals:
```

Fitting Linear models with correlated errors in R

- `getVarCov()`:

```
> # patient 2 has all measurements
> getVarCov(model.unstr, individual = 2)
Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 222.93 204.03 190.46 183.16 143.26
[2,] 204.03 255.48 232.97 220.88 176.88
[3,] 190.46 232.97 299.49 262.72 226.83
[4,] 183.16 220.88 262.72 343.89 293.25
[5,] 143.26 176.88 226.83 293.25 354.70
  Standard Deviations: 14.931 15.984 17.306 18.544 18.834
> # corresponding correlation matrix
> cov2cor(getVarCov(model.unstr, individual = 2))
Marginal variance covariance matrix
        [,1]    [,2]    [,3]    [,4]    [,5]
[1,] 1.00000 0.85493 0.73712 0.66153 0.50946
[2,] 0.85493 1.00000 0.84223 0.74520 0.58759
[3,] 0.73712 0.84223 1.00000 0.81866 0.69596
[4,] 0.66153 0.74520 0.81866 1.00000 0.83967
[5,] 0.50946 0.58759 0.69596 0.83967 1.00000
  Standard Deviations: 1 1 1 1 1
```

Outline

Linear models with correlated errors - Model Building

- Linear models with correlated errors consist of two parts:
  - Mean model $\mathbf{X}_i\boldsymbol{\beta}$: describes how the covariates in the model explain the average of the repeated measurements.
  - Errors terms ε_{ij}: assumed covariance structure between the repeated measurements.
- Scientific interest is often on the mean.
- However, to obtain valid and efficient inferences for the mean, the covariance part needs to be correctly specified.

Linear models with correlated errors - Model Building

- The recommended strategy is as follows:
  - Consider an elaborate model for the mean response, including possible nonlinear and interaction terms.
  - Then select an appropriate covariance matrix $\mathbf{\Sigma}$ that adequately describes the correlations in the repeated measurements.
  - Finally, simplify the mean model by excluding non-significant terms.

Outline

Inference - Hypothesis testing for covariances

- To find the "optimal" covariance structure between models with the **same mean structure** we can use:
  1. Likelihood ratio test for nested models.
  2. Information criteria for non-nested models:
     - Akaike's Information Criterion (AIC).
     - Schwarz's Bayesian Information Criterion (BIC).

- Note:

  A Model A is nested in Model B if Model A is a "special case" of Model B.

  Loosely speaking, the nested Model A arises by imposing constraints on the parameters in the more general Model B (e.g., certain parameters are set to zero or equal to each other).

Hypothesis testing for covariances - Nested models

- The Likelihood Ratio Test (LRT) can be used to choose the error structure.
- The LRT is derived as:

$$LRT = \{2\log(L_A)\} - \{2\log(L_0)\} \sim \chi^2_{df}$$

where

  - L_0 is the value of the likelihood function under the null model.
  - L_A is the value of the likelihood function under the alternative model (i.e. more general model).
  - df the difference in the number of parameters between the two models.

- Note: We use REML instead of ML because REML reduces bias in ML estimates of covariance parameters.

Hypothesis testing for covariances - ARMD study

- Consider the ARMD study:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_{2j} \text{TimeID}_{ij} + \beta_{3j} \text{TimeID}_{ij} \times \text{group}_i + \varepsilon_{ij}.$$

- We want to test if modelling serial correlation is needed: we compare the unstructured with diagonal covariance matrix.

```
> model.unstr <- gls(VAS ~ Weeks * Treat, correlation = corSymm(form = ~ 1 | ID),
+                  weights = varIdent(form = ~ 1|Timef), data = armd., method = "REML")
> model.ind <- gls(VAS ~ Weeks * Treat,
+                   weights = varIdent(form = ~ 1|Timef), data = armd., method = "REML")
> anova(model.unstr, model.ind)
            Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.unstr     1 19 8375.699 8470.809 -4168.850
model.ind       2  9 9409.841 9454.893 -4695.921 1 vs 2 1054.142  <.0001
```

modelling serial correlation is required!

Hypothesis testing for covariances - Non-Nested models

- When we have non-nested models we cannot use the LRT.
- Alternatively we can use information criteria: AIC or BIC.

$$AIC = -2\log L(\hat{\theta}) + 2n_\theta$$

$$BIC = -2\log L(\hat{\theta}) + n_\theta \log(n)$$

where

  - $\log L(\hat{\theta})$ is the value of the log-likelihood function
  - n_θ the number of parameters in the model
  - n the number of subjects (independent units)

- When we compare two non-nested models we choose the model that has the lowest AIC/BIC value.

Hypothesis testing for covariances - ARMD study

- Consider the ARMD study:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_{2j} \text{TimeID}_{ij} + \beta_{3j} \text{TimeID}_{ij} \times \text{group}_i + \varepsilon_{ij}.$$

- We want to compare different correlation matrices: AR1 and Gaussian.
- Use R function `anova()`:

```
> anova(model.ar1, model.car1, model.exp, model.gaus)
           Model df      AIC      BIC    logLik
model.ar1      1 12 8378.591 8438.595 -4177.295
model.car1     2 12 8378.591 8438.595 -4177.295
model.exp      3 12 8378.591 8438.595 -4177.295
model.gaus     4 12 8686.818 8746.822 -4331.409
```

- Note that in this case the first 3 coincide.

Hypothesis testing for covariances - Summary

- Reducing the parameters in the covariance matrix leads to more efficient inferences for the mean parameters.
- This is particularly useful when many repeated measurements are taken per subject.
- Nested models should be compared using the LRT.
- The asymptotic distribution of the LRT, i.e. χ^2 may not always be valid $\Rightarrow$ more on this in random effects models.
- For non-nested models we use Information Criteria.

Inference - Hypothesis testing for β

- In practice we test hypotheses of the form:

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_A : \beta_3 \neq 0$$

or

$$H_0 : \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_A : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

  namely, the hypotheses involve only one or > 1 parameters.

- Two tests are often used:
  1. t-test (test for 1 parameter) or F-test (test for > 1 parameters).
  2. Likelihood Ratio Test.

Inference - Hypothesis testing for $\boldsymbol{\beta}$

- The ML estimate of $\boldsymbol{\beta}$ is given by:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i$$

  where $\boldsymbol{\Sigma}_i$ is replaced by its ML or REML estimate.

- The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1}$$

  where again $\boldsymbol{\Sigma}_i$ is replaced by its ML or REML estimate.

- Given $\hat{\boldsymbol{\Sigma}}_i$ it holds $\hat{\boldsymbol{\beta}} \sim N_p \left(\boldsymbol{\beta}, \text{Var}(\hat{\boldsymbol{\beta}}) \right)$.

Hypothesis testing for $\boldsymbol{\beta}$: t-test

- To test hypotheses of the form:

$$H_0 : \beta = 0 \qquad \text{vs} \qquad H_A : \beta \neq 0$$

  i.e. for a single parameter, we use the t-test statistic

$$W = \frac{\hat{\beta}}{s.e.(\hat{\beta})} \sim t_{df}$$

  where $\hat{\beta}$ is the ML estimate of β, $s.e.(\hat{\beta})$ is the standard error of $\hat{\beta}$, and df are the degrees of freedom.

- Note:
  - W does not follow an exact t distribution as in linear regression where $df = n - p$ with p # of regression parameters.
  - Instead we use approximate methods to estimate the degrees of freedom to account for the variability introduced by replacing $\boldsymbol{\Sigma}_i^{-1}$ by its estimate (details later).

Example: ARMD Study

- In R the t-test for each regression parameter in the model is given in the table:

```
> summary(model.unstr)$tTable
                       Value  Std.Error    t-value       p-value
(Intercept)       55.31142224 1.35774597 40.7376810 3.873036e-222
Weeks             -0.20932410 0.02901158 -7.2151903  9.996510e-13
TreatActive       -1.01976780 1.91170894 -0.5334326  5.938416e-01
Weeks:TreatActive -0.08530908 0.04213686 -2.0245714  4.315197e-02
```

Hypothesis testing for $\boldsymbol{\beta}$: F-test

- For a null hypothesis with more parameters, e.g.:

$$H_0 : \beta_2 = \beta_3 = 0 \text{ vs } H_A : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

  we use the F-test statistic.

- In the general case, for a known $k \times p$ contrasts matrix $\mathbf{L}$ and $\boldsymbol{\beta}$ the $p \times 1$ vector of fixed effects parameters the set of hypotheses is written as:

$$H_0 : \mathbf{L}\beta = 0 \text{ vs } H_A : \mathbf{L}\beta \neq 0$$

  and the F-test statistic is:

$$F = \frac{\hat{\beta}^T \mathbf{L}^T \left[\mathbf{L} \sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \mathbf{L}^T \right]^{-1} \mathbf{L}\hat{\beta}}{\text{rank}(\mathbf{L})} \sim F_{df_1, df_2}$$

  where df_1 are the numerator degrees of freedom equal to the rank of matrix $\mathbf{L}$ and df_2 are the denominator degrees of freedom which are estimated from the data.

Hypothesis testing for $\boldsymbol{\beta}$: F-test

- Several methods are available to estimate the denominator degrees of freedom:
  - Containment method
  - Satterthwaite approximation
  - Kenward and Roger approximation, etc.
- In longitudinal studies, all methods typically give similar p-values because of the high number of degrees of freedom.
- More attention should be paid when working with small samples:
  Kenward-Roger method is recommended in these situations.

Example: ARMD Study

- In `nlme` the function `anova(.)` gives the F-test (or t-test for 1 parameter) results.

```
> anova(model.unstr)
Denom. DF: 1103
            numDF  F-value p-value
(Intercept)     1 3168.881  <.0001
Weeks           1  140.402  <.0001
Treat           1    1.336  0.2479
Weeks:Treat     1    4.099  0.0432
```

- `anova(.)` by default gives the sequential F-tests (type I tests; depends on order of terms) and with argument `type = "marginal"` we get the marginal (type III tests) ones (not interpretable in the presence of interaction; independent of order of terms).

- `anova(., Terms = ...)` or `anova(., L = ...)` should be used.

- Note: `nlme` does not use the approximate F or t-test. The (denominator) degrees of freedom are $n - p$. This is used to compute also 95% CI.

- This is crucial for small studies i.e. up to 30-40 subjects. In such cases, bootstrap p-values should be prefered.

Example: ARMD Study

- Often hypotheses of interest not automatically computed.
- Consider the model for the ARMD study:

$$y_{ij} = \beta_0 + \beta_1 \text{time}_{ij} + \beta_2 \text{group}_i + \beta_3 \text{time}_{ij} \times \text{group}_i + \varepsilon_{ij}$$

- To test for different mean trajectories between the two groups, we need to test:

$$H_0 : \beta_2 = \beta_3 = 0 \text{ vs } H_A : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

- In R to test such hypothesis using the F-test we need to use the function `anova(.)`.
- But first we need to understand how to construct the contrasts matrix $\mathbf{L}$, i.e. write $H_0 : \beta_2 = \beta_3 = 0$ in the form $H_0 : \mathbf{L}\beta = 0$.

Hypothesis testing for $\boldsymbol{\beta}$: F-test

- Example: $H_0 : \beta_2 = \beta_3 = 0$
- It can be re-written as $H_0 : \mathbf{L}\beta = 0$ with:

$$\mathbf{L}\beta = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

Hypothesis testing for $\boldsymbol{\beta}$: F-test

- Example: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
- It can be re-written as $H_0 : \mathbf{L}\beta = 0$ with:

$$\mathbf{L}\beta = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

Hypothesis testing for $\boldsymbol{\beta}$: F-test

- Example: $H_0 : \beta_1 = \beta_2$
- It implies $H_0 : \beta_1 - \beta_2 = 0$
- It can be re-written as $H_0 : \mathbf{L}\beta = 0$ with:

$$\mathbf{L}\beta = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

Example: ARMD Study

- In the model for the ARMD study:

$$y_{ij} = \beta_0 + \beta_1 \text{time}_{ij} + \beta_2 \text{group}_i + \beta_3 \text{time}_{ij} \times \text{group}_i + \varepsilon_{ij}$$

we can test

$$H_0 : \beta_2 = 0 \text{ vs } H_A : \beta_2 \neq 0$$

using

```
> anova(model.unstr, Terms = c("Treat"))
Denom. DF: 1103
 F-test for: Treat
  numDF   F-value p-value
1     1 0.2845503  0.5938
```

or

```
> Ltreat <- rbind(c(0,0,1,0))
> anova(model.unstr, L = Ltreat)
Denom. DF: 1103
 F-test for linear combination(s)
[1] 1
  numDF   F-value p-value
1     1 0.2845503  0.5938
```

Example: ARMD Study

- To test for different mean trajectories between the two groups, we need to test:

$$H_0 : \beta_2 = \beta_3 = 0 \text{ vs } H_A : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

- This can be done using

```
> anova(model.unstr, Terms = c("Treat", "Weeks:Treat"))
Denom. DF: 1103
 F-test for: Treat, Weeks:Treat
  numDF  F-value p-value
1     2 2.717564  0.0665
```

or

```
> Ltreat <- rbind(c(0,0,1,0),
+                 c(0,0,0,1))
> anova(model.unstr, L = Ltreat)
Denom. DF: 1103
 F-test for linear combination(s)
  TreatActive Weeks:TreatActive
1           1                 0
2           0                 1
  numDF  F-value p-value
1     2 2.717564  0.0665
```

Hypothesis testing for $\boldsymbol{\beta}$: Likelihood Ratio Test

- We can use the likelihood ratio test (LRT) to test hypothesis:

$$H_0 : \beta_2 = \beta_3 = 0 \text{ vs } H_A : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

- LRT can be used to compare nested models different mean structures, but equal covariance structure.

- The implied models are:

$$\begin{aligned} H_0 : &\quad y_{ij} = \beta_0 + \beta_1 \text{time}_{ij} + \varepsilon_{ij} \\ H_A : &\quad y_{ij} = \beta_0 + \beta_1 \text{time}_{ij} + \beta_2 \text{group}_i + \beta_3 \text{time}_{ij} \times \text{group}_i + \varepsilon_{ij} \end{aligned}$$

  for both models we assume unstructured $\boldsymbol{\Sigma}_i$.

- The LRT is given by:

$$LRT = \{-2\log(L_0)\} - \{-2\log(L_A)\} \sim \chi^2_{df}$$

  with L_A and L_0 the values of ML under the alternative and null model, df the difference in the number of parameters between the two models.

Hypothesis testing for $\boldsymbol{\beta}$: Likelihood Ratio Test

- For the ARMD study, for the models under H_0 and H_A, respectively we get

```
> model.unstr0 <- gls(VAS ~ Weeks, correlation = corSymm(form = ~ 1 | ID),
+                 weights = varIdent(form = ~ 1|Timef), data = armd., method = "ML")
> model.unstr1 <- gls(VAS ~ Weeks * Treat, correlation = corSymm(form = ~ 1 | ID),
+                 weights = varIdent(form = ~ 1|Timef), data = armd., method = "ML")
>
> anova(model.unstr0, model.unstr1)
             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.unstr0     1 17 8371.332 8456.492 -4168.666
model.unstr1     2 19 8370.013 8465.191 -4166.006 1 vs 2 5.319739    0.07
```

Summary of Hypothesis testing for $\boldsymbol{\beta}$

- We can test hypotheses for $\boldsymbol{\beta}$ using the t-test, $F-$test or LRT.
- LRT should be applied on nested models.
- Important: The likelihood ratio test for comparing models with different mean parts is only valid when the models have been fitted using the ML method and not REML.
- REML likelihoods are based on different observations, which makes them no longer comparable with ML.
- The LRT may be 'liberal' in small samples, i.e., it gives smaller p-values than expected $\Rightarrow$ t- or F-test should be preferred.

Outline

Model diagnostics

- **All statistical models make assumptions.**
- For valid statistical inference we need to check if these assumptions hold.
- Assumptions for the multivariate model:

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \varepsilon_i$$

  - the error terms follow the normal distribution $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$
  - the error terms are independent of the covariates $\mathbf{X}_i$.
- It is also advisable to check if validity of model assumptions sensitive to outliers.

Model diagnostics

- Model diagnostics include:
  - Residual analysis.
- Recommendations can be extracted:
  - Do we need to transform response variable and/or covariates?
  - Do we need to change mean model? e.g. add interactions or polynomials?
  - Should we model covariance structure differently?
  - Is the model severely influnced by certain subjects?
- Diagnostic methods for standard linear regression models are well established in the statistics literature.
- For complex models they are more difficult to perform and interpret $\Rightarrow$ Often they should be treated with caution.
- Caution in interpreting plots is also needed with missing data.

Model diagnostics - Residuals

- A **residual** is a measure of how far an observation y_{ij} is from its predicted value $\hat{y}_{ij} = x_{ij}^T \hat{\boldsymbol{\beta}}$.
- For the multivariate model

$$\mathbf{Y}_i = \mathbf{X}_i \beta + \varepsilon_i, \text{ with } \varepsilon_i \sim (\mathbf{0}, \boldsymbol{\Sigma}_i)$$

  the residuals $\mathbf{e}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}$ is an estimate of ε_i with $\hat{\boldsymbol{\beta}}$ the (RE)ML estimate.
- In a correctly specified model $E(\mathbf{e}_i) = \mathbf{0}$ and $\text{cov}(\varepsilon_i) \approx \text{cov}(\mathbf{e}_i) = \hat{\boldsymbol{\Sigma}}$.
- We call $\mathbf{e}_i$ **raw** residuals vs **Pearson** and **Normalized** residuals.

Model diagnostics - Residuals

- In repeated measures designs $\mathbf{e}_i$ are correlated and do not have constant variance as in simple linear regression $\Rightarrow$ interpreting raw residual plots not meaningful.
- **Pearson residuals:**

$$e_{ij}^P = \frac{e_{ij}}{\sqrt{\widehat{\text{var}}(y_{ij})}}$$

  deal with the heteroscedasticity of raw residuals but they are still correlated.
- **Normalized residuals:** Using the Cholesky decomposition of $\hat{\mathbf{\Sigma}}_i = \mathbf{L}_i \mathbf{L}_i^T$ with $\mathbf{L}_i$ a lower triangular matrix

$$\mathbf{e}_i^N = \mathbf{L}_i^{-1} \mathbf{e}_i$$

  they should be approximately distributed as $N_{n_i}(\mathbf{0}, \mathbf{I})$.
- Studying the normalized residuals is advisable if available in software used.

Model diagnostics - Residuals

- **Normality assumption:** Normal probability plot (or Q-Q plot)
  - If data are normally distributed the residuals form a straight line around the diagonal.
  - If residuals deviate markedly from a straight line $\Rightarrow$ data are probably not normally distributed (e.g. a bow shape indicates a skewed distribution and a sigmoid shape indicates a symmetric but non-normal distribution) $\Rightarrow$ transformation of the data can be considered.
- **Zero mean:** Scatterplot of residuals versus fitted values and/or covariates.
  - e_{ij} should have mean zero at each time point.
  - If there is a trend over time or a non-zero mean at some time points $\Rightarrow$ the population mean has not been correctly specified as a function of time.
  - Similarly determine if a covariate could be added to the current model.

Model diagnostics - R

- For the ARMD trial we consider the model:

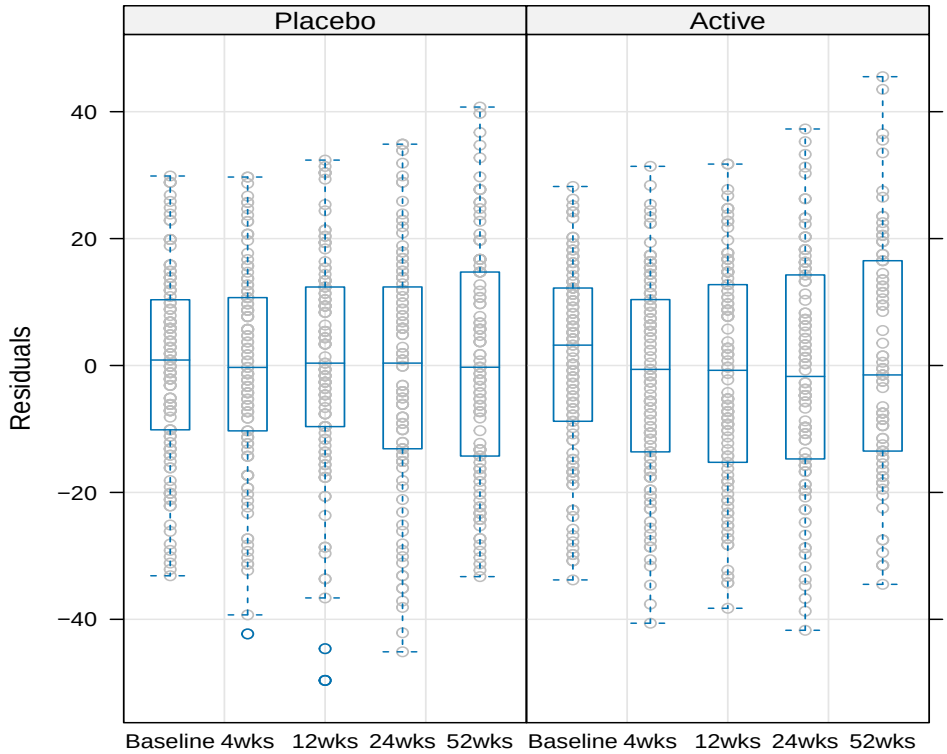
$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + \varepsilon_{ij}$$

  where $\varepsilon_i \sim (\mathbf{0}, \mathbf{\Sigma}_i)$ with $\mathbf{\Sigma}_i$ an unstructured covariance matrix.

- We will check the fit of the model using:
  - raw residuals versus fitted values/time (per treatment group)
  - standardized residuals versus fitted values/time (per treatment group)
  - normalized residuals versus fitted values/time (per treatment group)
  - observed versus fitted values
  - QQ-plot of standardized/normalized residuals

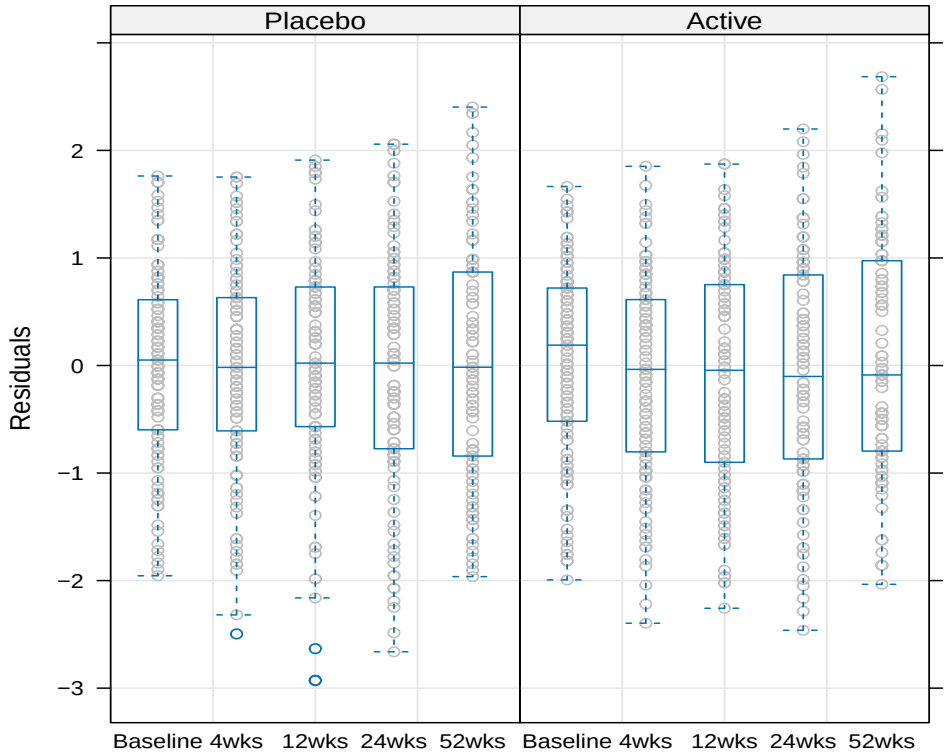
Model diagnostics - R

- Plots of raw residuals (`residuals(model.gls1, type = "response")`):



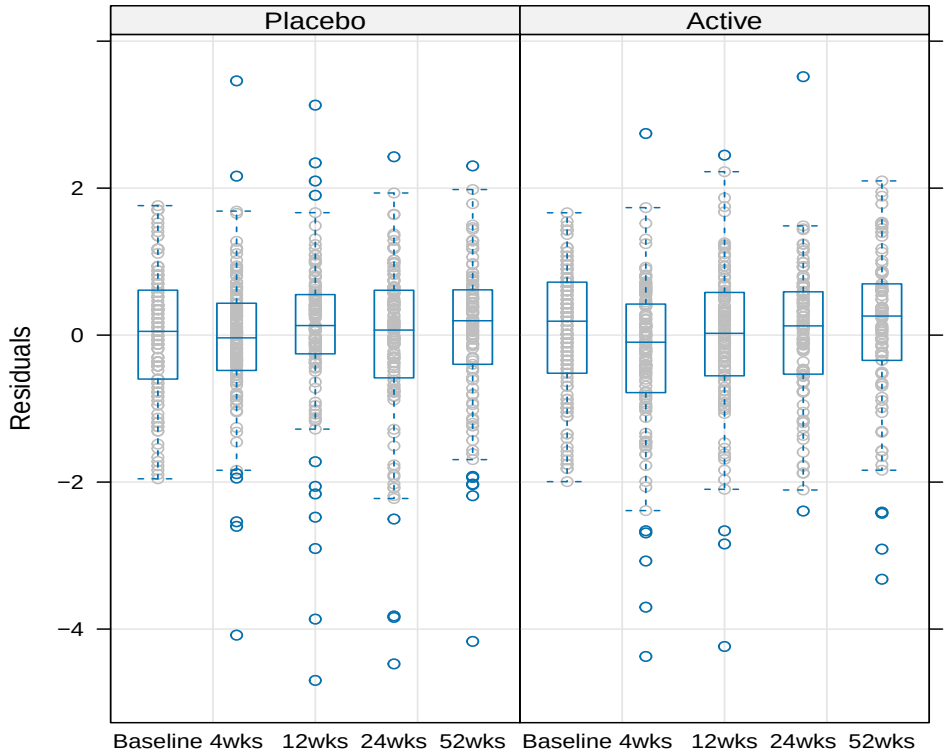
Model diagnostics - R

- Plots of pearson residuals (`residuals(model.gls1, type = "pearson")`):



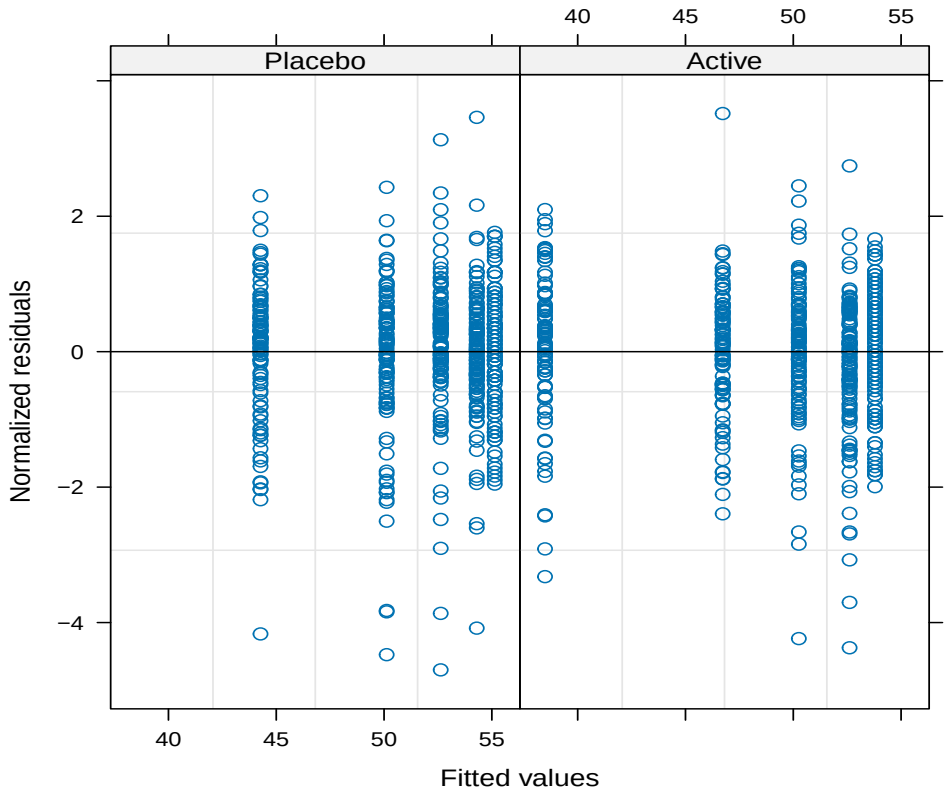
Model diagnostics - R

- Plots of normalized residuals (`residuals(model.gls1, type = "normalized")`):



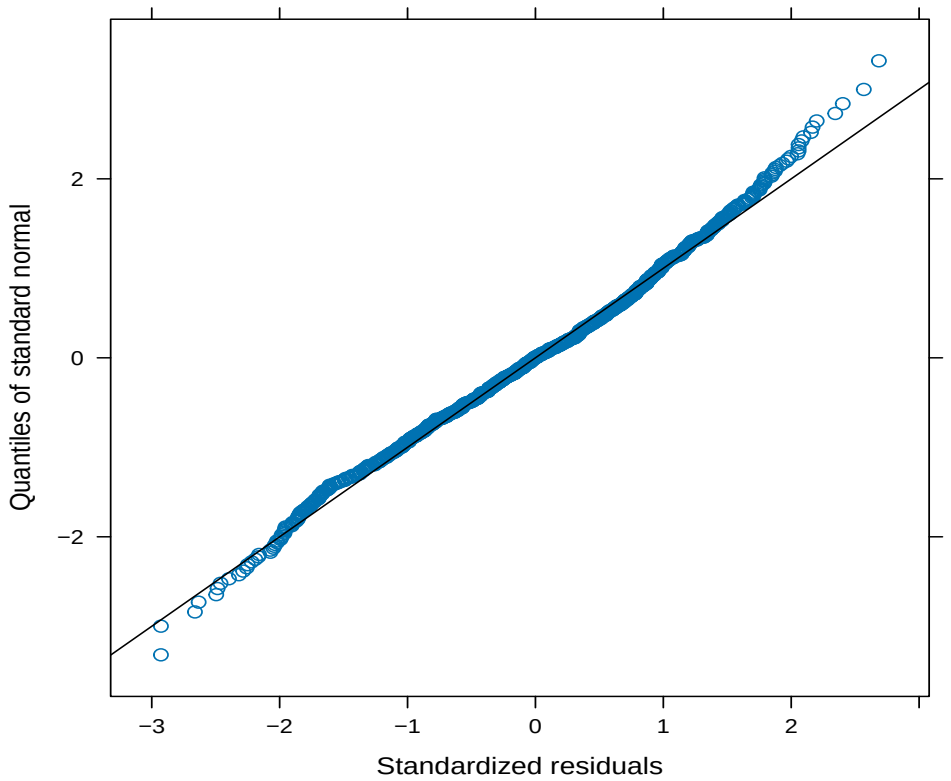
Model diagnostics - R

- Plots of normalized residuals versus fitted values.



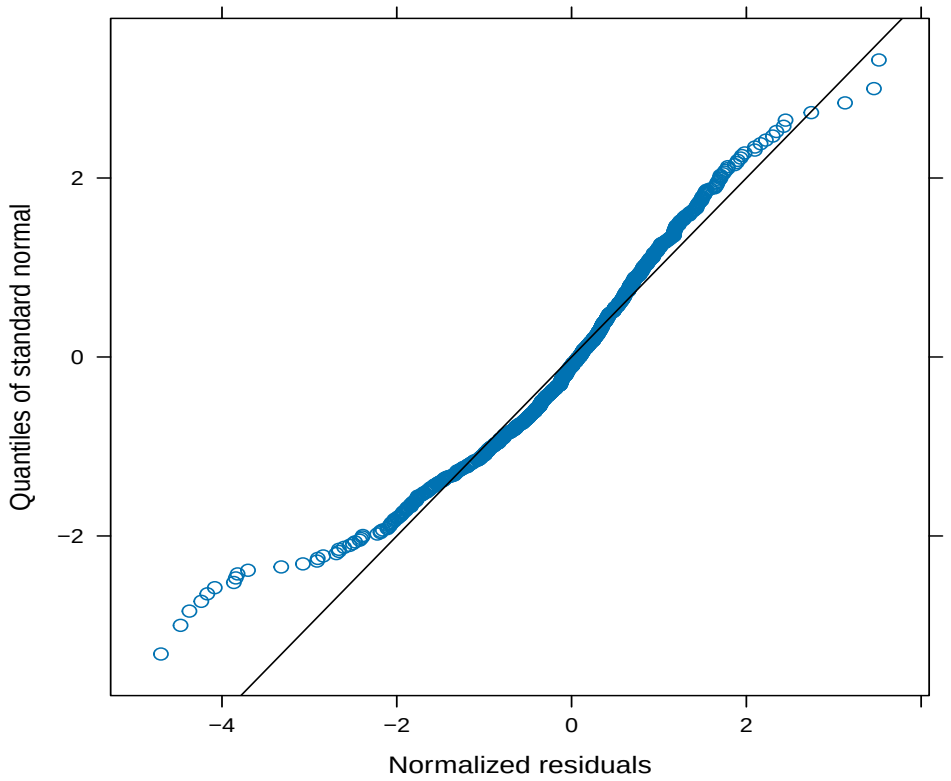
Model diagnostics - R

- Q-Q plot of pearson residuals: `qqnorm(.)`



Model diagnostics - R

- Q-Q plot of normalized residuals.



Chapter 4: Linear mixed-effects models

Master Statistics and Data Science

Outline

Outline

Linear Mixed-Effects Models - Introduction

- So far we modelled the correlation between the repeated measurements using an appropriate variance-covariance matrix $\Rightarrow$ **Marginal models.**

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \varepsilon_i, \quad \varepsilon_i \sim N_{n_i}(\mathbf{0}, \mathbf{V}_i).$$

- **Mixed-effects models** offer an alternative way to model the correlation between repeated measurements on the same subject.

- They model the correlation using **random effects**.

- They contain both fixed effects (population mean) and random effects (covariance).

Linear Mixed-Effects Models - Introduction

- Why do we need to study mixed-effects models?

  Mixed models are very flexible in modelling unbalanced data.

- When do we need mixed-effects models?
  - irregular visit times, i.e. available correlation matrices not possible.
  - long follow-up, i.e. when unstructured not possible or AR-1/compound symmetry unrealistic.
  - individualized predictions, mixed models explicitly identify the individual.

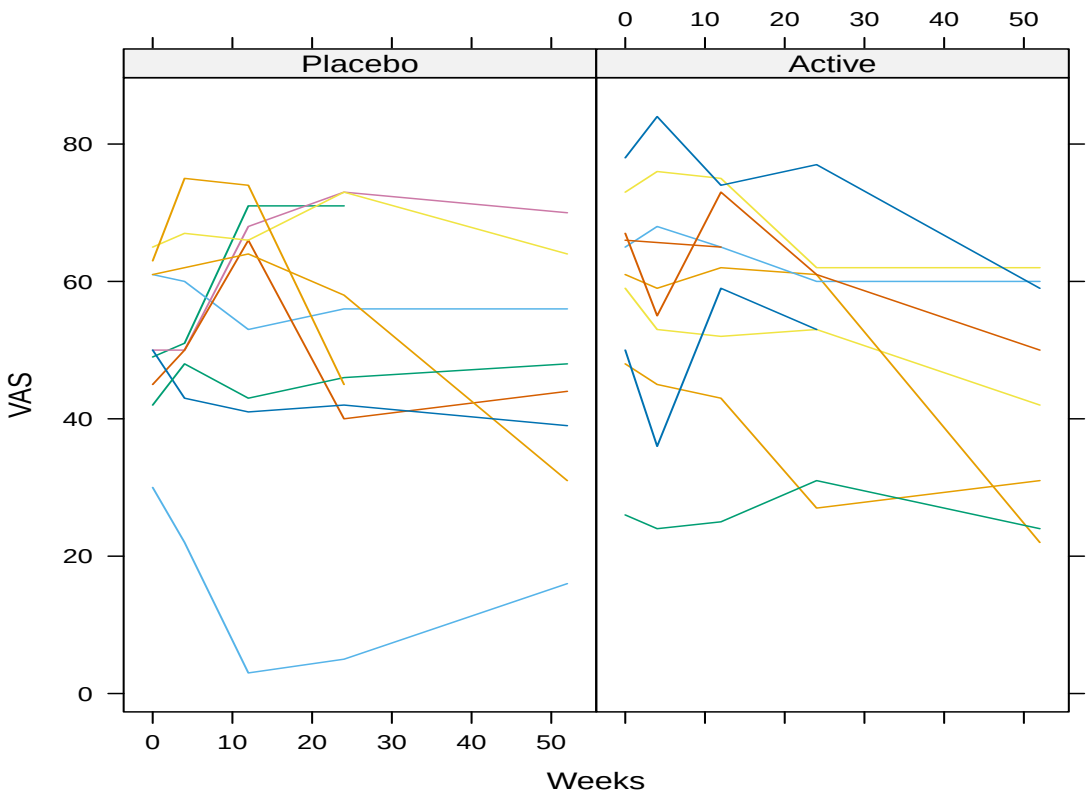
Example: ARMD Trial - Revisit

- Randomized multi-center clinical trial comparing an experimental treatment (interferon-a) versus placebo in 240 patients diagnosed with Age-Related Macular Degeneration Trial (ARMD).

- Response: Visual acuity score based on patient's ability to read lines of letters on standardized vision charts.
- Covariates:
  - Group: Placebo versus the highest dose (6 million units daily) of interferon-α.
  - Follow-up: Baseline and after randomization at 4, 12, 24, and 52 weeks.
- Research question: Assess the impact of treatment on progression of visual acuity score (VAS).

Example: ARMD Trial

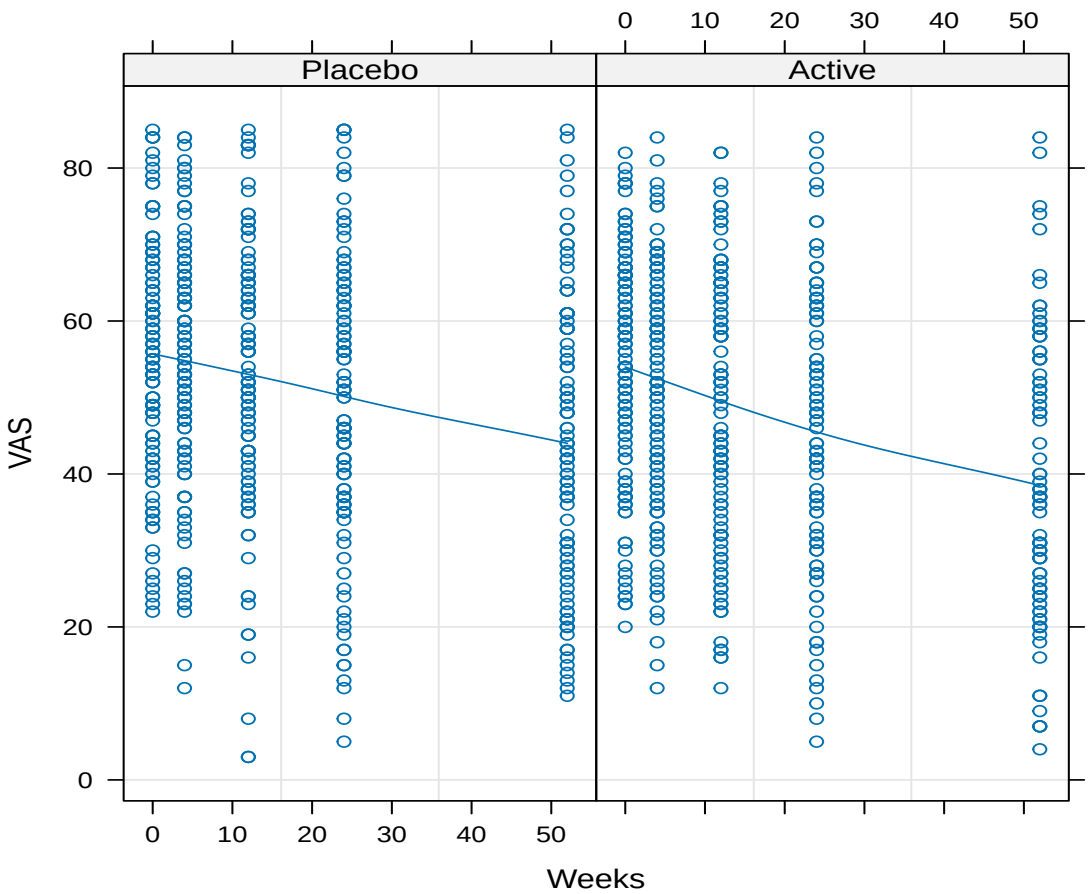
- Sample individual VAS profiles in time per treatment group



- Substantial variability from patient to patient within each treatment group.
- VAS scores of some patients tend to decrease in time, whereas for others they remain relatively constant.

Example: ARMD Trial

- Mean VAS profiles in time per treatment group



Outline

Linear Mixed-Effects Models - Model specification

- **The main idea is that each individual has his own profile which deviates from the population mean profile.**

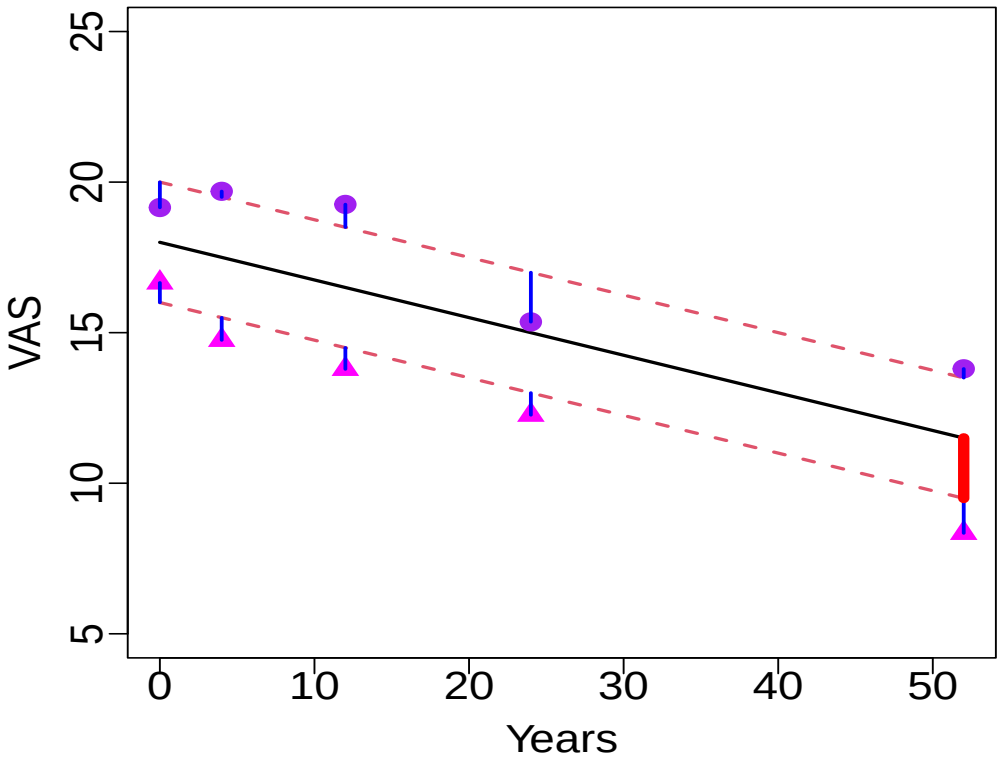
- Example: Consider the ARMD study where the VAS is measured at randomization and at 4, 12, 24, and 52 weeks afterwards in each treatment group.

- Let the population mean $E(y_{ij})$ be modelled as

$$E(y_{ij}) = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i$$

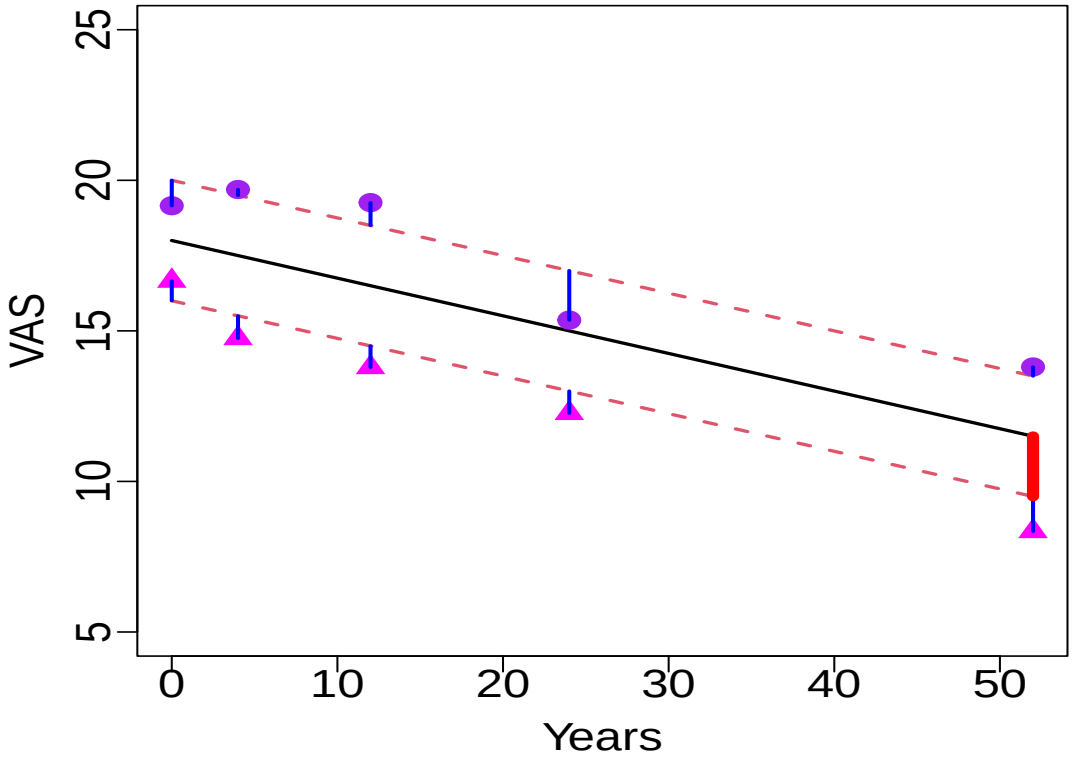
Linear Mixed-Effects Models - Illustration

- For the active treatment group, let the mean VAS score (black solid line).



- Observed measurements of Patient 1 (pink triangles) and Patient 2 (purple bullets).
- Level of patient i (red dashed line): **not observed**.

Linear Mixed-Effects Models - Illustration



$$y_{ij} = \text{population mean} + \boxed{\text{deviation from mean}}$$

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + \boxed{b_i + \varepsilon_{ij}}$$

Linear Mixed-Effects Models - Model specification

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_i + \varepsilon_{ij}$$

- There are two sources of variation:
  - Between-Subject Variation (b_i): subjects have different response levels around the population mean e.g. high, medium, low.

    $$b_i \sim N(0, \sigma_b^2)$$

    σ_b^2 is the between-subjects variance component.

  - Within-Subject Variation (ε_{ij}): The observations of a subject vary randomly and independently around their own mean level due to measurement error (e.g. measuring blood pressure many times at the same visit will not give the same value).

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

    σ^2 is the within-subject variance component.

- Note: b_i and ε_{ij} are assumed to be independent.

Linear Mixed-Effects Models - Model specification

$$\boxed{y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_i + \varepsilon_{ij}}$$

$$b_i \sim N(0, \sigma_b^2) \quad \text{and} \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

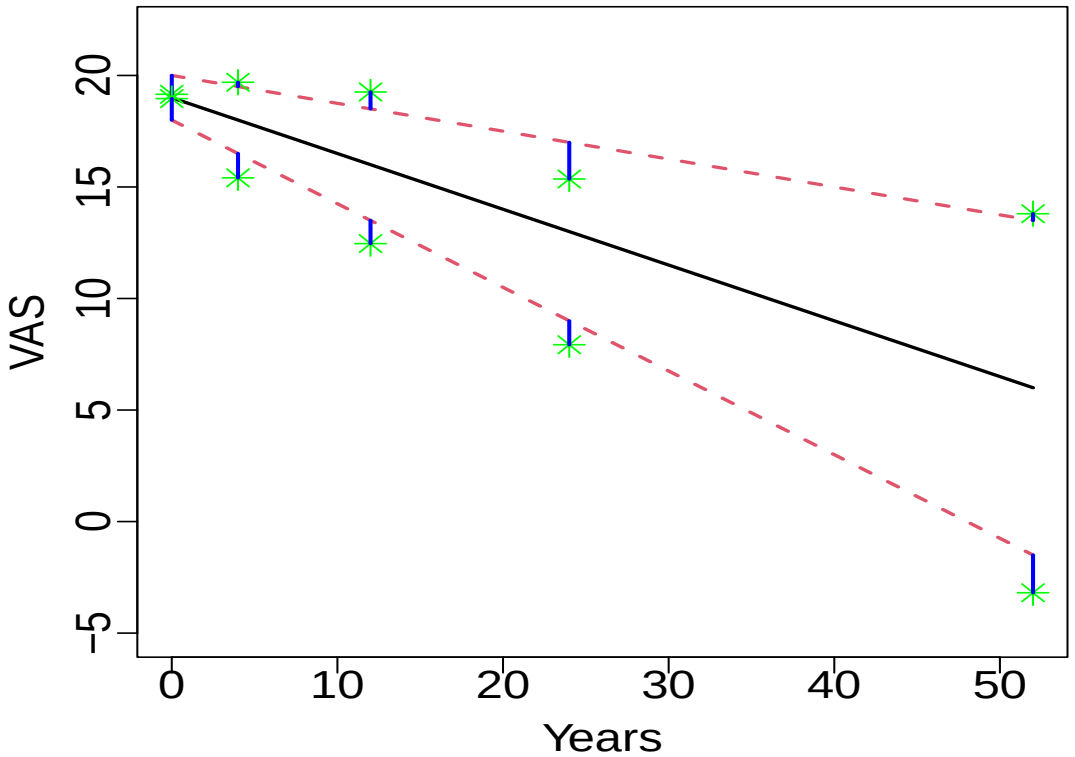
- The model contains both fixed and random effects $\Rightarrow$ **Mixed-effects model**.
- The model above can be re-written as:

$$y_{ij} = (\beta_0 + b_i) + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + \varepsilon_{ij}$$

and thus is called a **random intercepts model**.

Linear Mixed-Effects Models - Model specification

- The random intercepts model can be extended to more random effects:



- Each subject has own level (intercept) and progression (slope).

Linear Mixed-Effects Models - Model specification

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_{i0} + b_{i1} \text{time}_{ij} + \varepsilon_{ij}$$

$$\begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{int} & \sigma_{int,time} \\ \sigma_{int,time} & \sigma^2_{time} \end{bmatrix} \right)$$

and

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

the random effects $\mathbf{b}_i = (b_{i0}, b_{i1})$ and the error terms ε_{ij} are assumed to be independent.

- The model above can be re-written as:

$$y_{ij} = (\beta_0 + b_{i0}) + \beta_1 \text{group}_i + (\beta_2 + b_{i1}) \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + \varepsilon_{ij}$$

and thus is called a **random intercepts and random slopes model**.

- The random effects can be extended to quadratic random slopes, (linear) splines, etc.

Linear Mixed-Effects Models - Interpretation

- The fixed effects parameters β
  - are assumed to be the same for all individuals.
  - have population-averaged interpretation, i.e. β_1 measures the change in the mean response Y for a unit increase in X_1 given that all other covariates are kept fixed.
- The random effects parameters $\mathbf{b}_i$
  - are assumed different for each subject.
  - have subject-specific interpretation i.e. measures the deviation of the profile of each subject i from the population mean.

Example: ARMD Study

- Let's see first the random intercepts model in practice.
- Consider for the ARMD data a mixed model with random intercepts (random effects) and different linear mean VAS evolutions in time per group (fixed effects):

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_i + \varepsilon_{ij}$$

$$b_i \sim N(0, \sigma_b^2) \quad \text{and} \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Example: ARMD Study

- $$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_i + \varepsilon_{ij}$$

```
                  Value Std.Error  DF t-value p-value
(Intercept)      55.144     1.423 865  38.749   0.000
TreatActive      -2.063     2.009 238  -1.027   0.305
Weeks            -0.215     0.020 865 -10.581   0.000
TreatActive:Weeks -0.075     0.030 865  -2.526   0.012
```

- β_0: mean VAS in placebo group at baseline.
- β_1: change in the mean VAS between the two groups at baseline.
- β_2: change in the mean VAS for every week that passes by in the placebo group.
- The change in the mean VAS for every week that passes by in the treatment group is $\beta_2 + \beta_3$.
- The change in the mean VAS between treatment and placebo for every week that passes by is β_3 (= -0.075) (p-value = 0.012).

Example: ARMD Study

- The estimated variance for the random intercepts is: $\sigma_b^2 = 209.69$. It measures the between subjects variability.
- The estimated variance for the errors is: $\sigma^2 = 77.39$. It measures the within subject variability.
- The pairwise correlations are computed by:

$$\rho = \frac{cov(Y_{1i}, Y_{2i})}{\sqrt{Var(Y_{1i})}\sqrt{Var(Y_{2i})}} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2},$$

i.e., $\rho = 0.73$.

Linear Mixed-Effects Models - Error structure

- The error terms ε_{ij} describe the covariance among observations when we focus on the response profile of a specific individual.
- For the error terms ε_{ij} we typically assume $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$
- In this case we assume that the random effects capture all correlations $\Rightarrow$ **Conditional independence assumption**.
- However with few random effects this may not always be realistic e.g. the random intercepts model implies constant variance and constant correlation between any two measurements within subjects.
- In such cases the remaining correlation structure can be accounted for by choosing an appropriate covariance matrix for the errors: $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$.
- Choosing an appropriate matrix for $\mathbf{D}$ and $\boldsymbol{\Sigma}_i$ is done as in the Marginal models.

Linear Mixed-Effects Models - Error structure

- Random effects and serial correlation are in competition since they both model part of the total variability.
- It is often not possible to fit models using both random effects and serial correlation due to numerical problems.
- Model building strategy is needed.

Linear Mixed-Effects Models - Matrix notation

- For each sampling unit i (e.g. patient, school, family) with n_i correlated repeated measurements:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b_i} + \boldsymbol{\varepsilon}_i$$

$$\mathbf{b_i} \sim N(\mathbf{0}, \mathbf{D}), \quad \boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$$

  where

  - $\mathbf{X}_i$ is $n_i \times p$ design matrix for fixed effects $\boldsymbol{\beta}$
  - $\mathbf{Z}_i$ is $n_i \times q$ design matrix for random effects $\mathbf{b_i}$
  - $\mathbf{b_i}$ and $\boldsymbol{\varepsilon}_i$ are assumed independent

- It follows:

$$\mathbf{y}_i \mid \mathbf{b_i} \sim N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b_i},\ \boldsymbol{\Sigma}_i) \quad \mathbf{b_i} \sim N_q(\mathbf{0}, \mathbf{D}).$$

Outline

Linear Mixed-Effects Model - Hierarchical vs Marginal models

- Mixed models are an alternative approach to model flexibly within-subject correlation.
- Random effects can help us in modelling $\mathbf{V}_i$.
- Let the random intercepts model:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_i + \varepsilon_{ij}$$

$$b_i \sim N(0, \sigma_b^2) \text{ and } \varepsilon_{ij} \sim N(0, \sigma^2)$$

- To derive the marginal covariance matrix $\mathbf{V}_i$ we need:

$$\begin{aligned}
\text{Var}(y_{ij}) &= \text{Var}(b_i + \varepsilon_{ij}) = \sigma_b^2 + \sigma^2, \text{ for all } i, j \\
\text{Cov}(y_{i1}, y_{i2}) &= \text{Cov}(y_{i1}, y_{i2}) = \text{Cov}(b_i + \varepsilon_{i1}, b_i + \varepsilon_{i2}) \\
&= \text{Cov}(b_i, b_i) + \text{Cov}(b_i, \varepsilon_{i2}) + \text{Cov}(\varepsilon_{i1}, b_i) + \text{Cov}(\varepsilon_{i1}, \varepsilon_{i2}) \\
&= \sigma_b^2.
\end{aligned}$$

Linear Mixed-Effects Models - Hierarchical vs Marginal models

- The marginal covariance matrix of $(y_{i1}, \ldots, y_{i5})$ is:

$$\text{Cov}(\mathbf{Y}_i) = \begin{bmatrix} \sigma_b^2+\sigma^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2+\sigma^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma^2 \end{bmatrix}$$

$$= (\sigma_b^2+\sigma^2) \begin{bmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{bmatrix}$$

- Constant variance at all visits and constant within-subject pairwise correlations $\rho = \frac{\sigma_b^2}{\sigma_b^2+\sigma^2} \Rightarrow$ Compound Symmetry covariance matrix $\mathbf{V}_i$.

Example: ARMD Study

- For the ARMD data the estimated covariance matrix $\mathbf{V}_i$ is:

$$\begin{bmatrix} \sigma_b^2+\sigma^2 & \sigma_b^2 & \dots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2+\sigma^2 & \dots & \sigma_b^2 \\ \vdots & \vdots & \dots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \dots & \sigma_b^2+\sigma^2 \end{bmatrix} = \begin{bmatrix} 287.1 & 209.7 & \dots & 209.7 \\ 209.7 & 287.1 & \dots & 209.7 \\ \vdots & \vdots & \dots & \vdots \\ 209.7 & 209.7 & \dots & 287.1 \end{bmatrix}.$$

- The pairwise correlations are computed by:

$$\rho = \frac{cov(Y_{1i}, Y_{2i})}{\sqrt{Var(Y_{2i})}\sqrt{Var(Y_{2i})}} = \frac{\sigma_b^2}{\sigma_b^2+\sigma^2},$$

  i.e., $\rho = 0.73$.

- Mixed models imply Marginal Models but not vice-versa.

Linear Mixed-Effects Models - Hierarchical vs Marginal models

- Let the random intercepts and random slopes model:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_{i0} + b_{i1} \text{time}_{ij} + \varepsilon_{ij}$$

$$b_{i0} \sim N(0, \sigma^2_{int}), \qquad b_{i1} \sim N(0, \sigma^2_{time}), \qquad \text{Cov}(b_{i0}, b_{i1}) = \sigma_{int,time}$$

and

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

- The marginal covariance matrix is computed as follows:

$$\begin{aligned} \text{Var}(y_{ij}) &= \sigma^2_{int} + 2\sigma_{int,time}\text{time}_{ij} + \sigma^2_{time}\text{time}^2_{ij} + \sigma^2 \text{ for all } i, j \\ \text{Cov}(y_{i1}, y_{i2}) &= \sigma^2_{int} + \sigma_{int,time}(\text{time}_{i1} + \text{time}_{i2}) + \sigma^2_{time}\text{time}_{i1}\text{time}_{i2}. \end{aligned}$$

- Variance is a quadratic function of time with positive curvature σ^2_{time}.
- Mixed models imply Marginal Models but not vice-versa.

Linear Mixed-Effects Model - Hierarchical vs Marginal models

- The LMM can be written as:

$$\mathbf{y}_i \mid \mathbf{b_i} \sim N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b_i},\ \boldsymbol{\Sigma}_i) \quad \mathbf{b_i} \sim N(\mathbf{0}, \mathbf{D}).$$

- Thus it is called **hierarchical model**.

- Marginally we have:

$$\mathbf{y}_i \sim N_{n_i}(\mathbf{X}_i\boldsymbol{\beta},\ \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i).$$

- **The hierarchical model implies the marginal model but not vice versa.**

- A marginal covariance matrix cannot always be derived from a valid random effects model (hierarchical model).

Linear Mixed-Effects Models - Hierarchical vs Marginal models

- Different hierarchical models can imply the same marginal model.
- Exercise: Consider the case of a longitudinal study with 2 repeated measurements. Derive the marginal covariance matrix when:
  - random intercepts and heterogeneous errors.
  - uncorrelated random intercepts and slopes and homogeneous errors.
- Implication:
  - Hierarchical model requires $\mathbf{D}$ to be positive definite.
  - Marginal model requires $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{\Sigma}_i$ to be positive definite.
- Different statistical software can give negative variance components depending on the model formulation, e.g. SAS vs R.

Outline

Linear Mixed-Effects Models - Estimation

- Estimation of the model parameters $\boldsymbol{\theta}$:
  - fixed effects parameters $\boldsymbol{\beta}$
  - variance parameters $\boldsymbol{\alpha}$ in $\boldsymbol{\Sigma}$ and variance parameters in $\mathbf{D}$

  and the random effects $\mathbf{b}_i$ is done by maximizing the marginal likelihood of the LMM.

- The marginal density for each subject i is:

$$f(\mathbf{y}_i) = \int f(\mathbf{y}_i \mid \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i$$

- Marginal likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} \left\{ (2\pi)^{n_i/2} \mid \mathbf{V}_i \mid^{-1/2} \exp \left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right] \right\}$$

Linear Mixed-Effects Models - Estimation

- For known $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ is estimated by:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{V}_i^{-1} \mathbf{y}_i$$

  where $\mathbf{V}_i$ is replaced by its Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) estimate and thus we get $\hat{\boldsymbol{\beta}}_{ML}$ or $\hat{\boldsymbol{\beta}}_{REML}$, respectively.

- Conditional on $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\beta}}$ follows the multivariate normal distribution with mean $\boldsymbol{\beta}$ and variance-covariance matrix:

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_i \mathbf{X}_i \right)^{-1}$$

Linear Mixed-Effects Models - Estimation

- ML estimation of the model parameters $\boldsymbol{\alpha}$.
  - $\hat{\boldsymbol{\alpha}}_{ML}$ is obtained by maximizing $L_{ML}(\boldsymbol{\alpha}, \hat{\beta})$.
- REML estimation of the model parameters $\boldsymbol{\alpha}$.
  - $\hat{\boldsymbol{\alpha}}_{REML}$ is obtained by maximizing $L_{REML}(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}}) = \mid \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \mid^{-1/2} L_{ML}(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}})$.

Linear Mixed-Effects Models - Estimation: Example

- Consider a random intercepts model for the ARMD data with linear progression in time that differs between the two treatment groups.
- Marginal covariance matrix using ML

```
          1        2        3
1 285.0358 207.8262 207.8262
2 207.8262 285.0358 207.8262
3 207.8262 207.8262 285.0358
```

- Marginal covariance matrix using REML

```
          1        2        3
1 287.0808 209.6936 209.6936
2 209.6936 287.0808 209.6936
3 209.6936 209.6936 287.0808
```

- We observe small differences in this dataset due to the relatively big sample size. In studies with less subjects the differences are more pronounced.
- Also the estimates for fixed effects and their standard errors are similar in this case.

Linear Mixed-Effects Models - Estimation of Random Effects

- The random effects quantify the deviation of each subject's profile from the mean profile over all individuals i.e. $\mathbf{X}_i\boldsymbol{\beta}$.
- The random effects are not observed but they can be estimated.
- The estimates of the random effects $\mathbf{b}_i$ are known as **empirical Bayes** estimates the random effects.
- These estimates are useful for:
  - Detecting outliers.
  - Ranking.
  - Predictions $\Rightarrow$ Individualized predictions.

Linear Mixed-Effects Models - Example: ARMD Study

- Consider the model for the ARMD study:

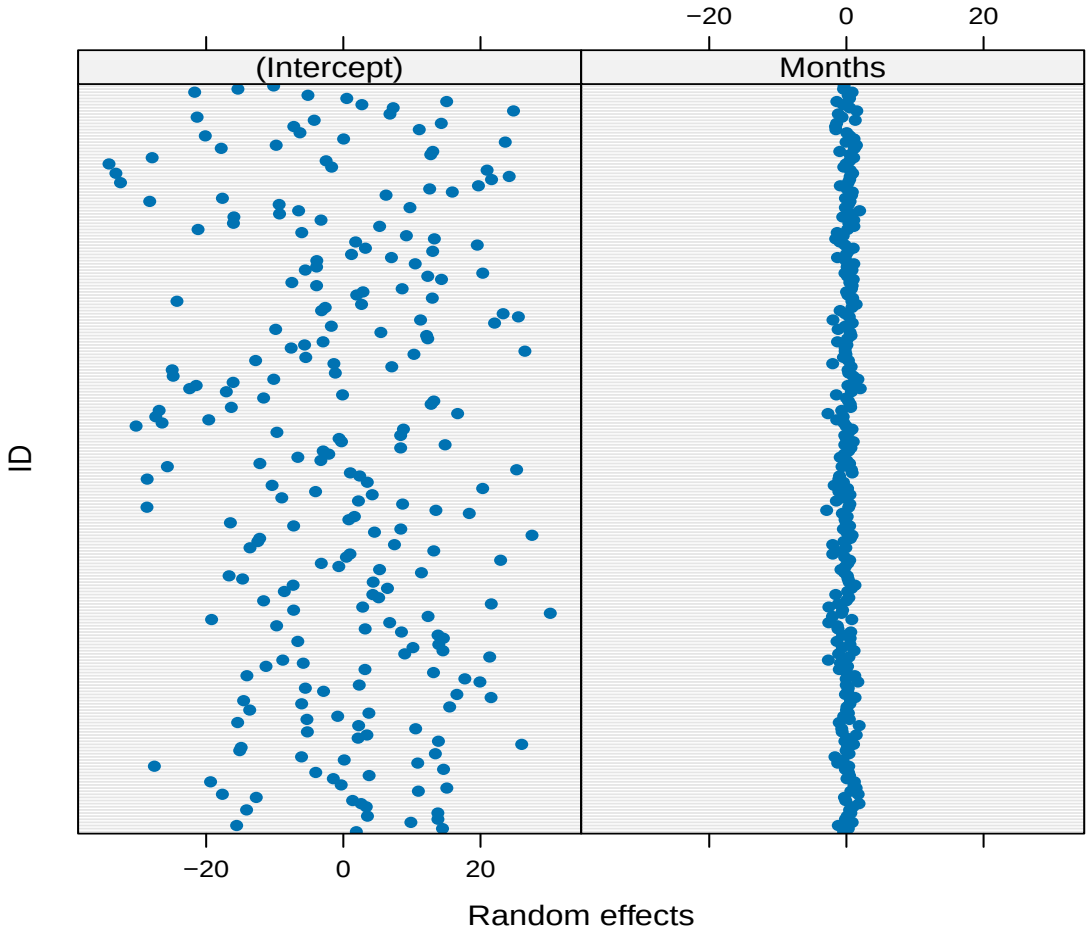
$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_{i0} + b_{i1} \text{time}_{ij} + \varepsilon_{ij}$$

- Histograms and scatterplots of $\hat{b}_{i0}$ and $\hat{b}_{i1}$ can be used to detect outliers i.e. patients with 'exceptional' evolutions in time or violations of model assumptions.

Linear Mixed-Effects Models - Example: ARMD Study

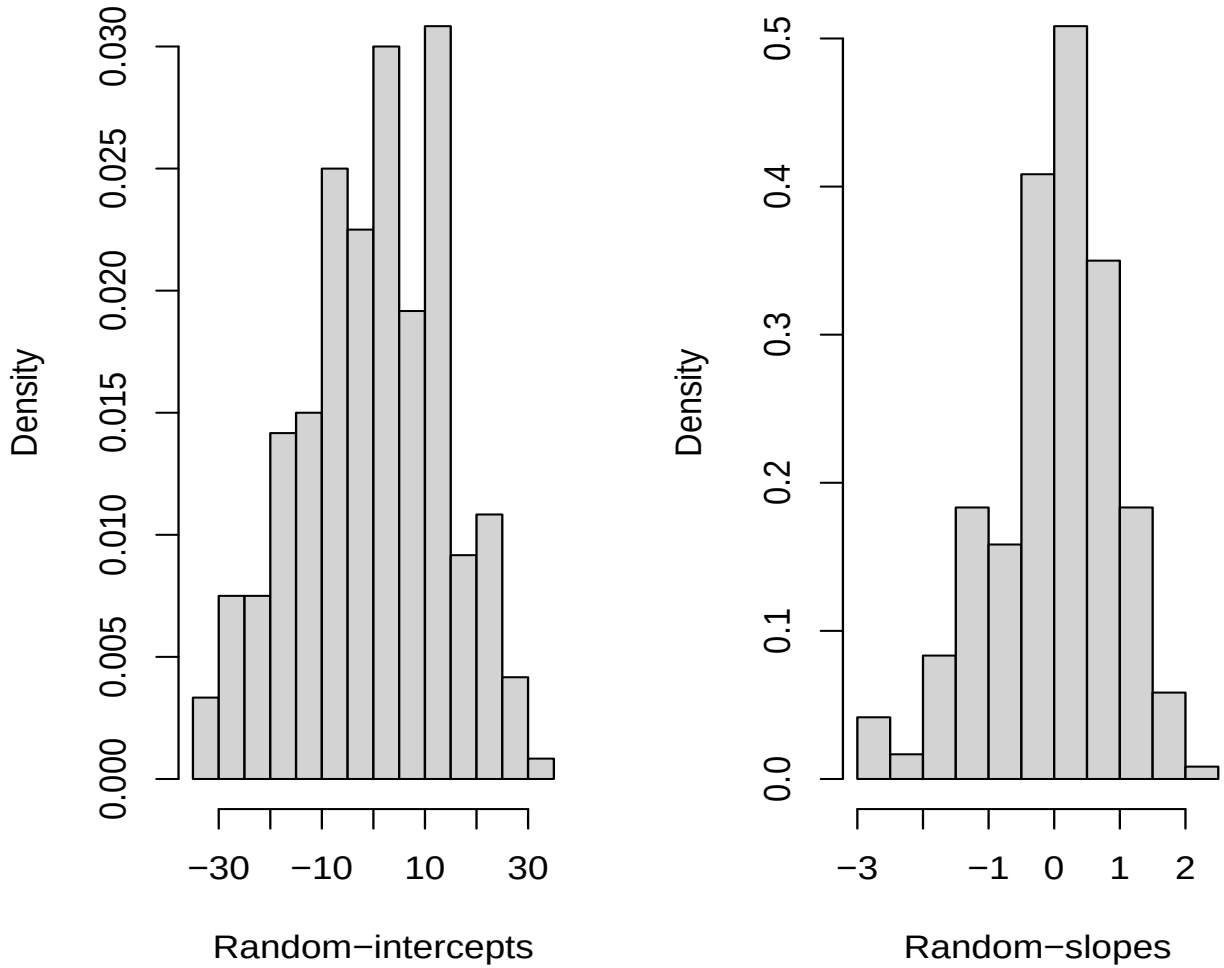
- Consider the model for the ARMD study:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_{i0} + b_{i1} \text{time}_{ij} + \varepsilon_{ij}$$



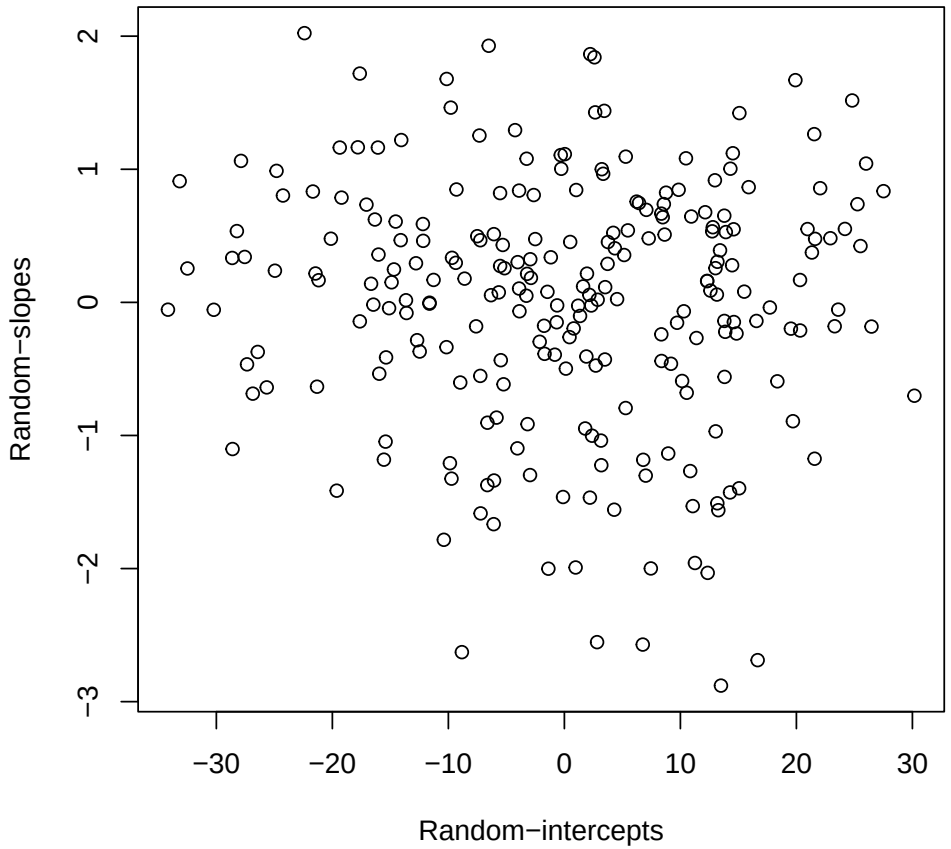
Linear Mixed-Effects Models - Example: ARMD Study

- Histograms of the empirical Bayes estimates:



Linear Mixed-Effects Models - Example: ARMD Study

- Scatterplot of the empirical Bayes estimates:



Linear Mixed-Effects Models - Example: ARMD Study

- The plots do not show outliers or severe deviations from the model assumptions.
- There is not strong correlation between the random intercepts and slopes in accordance with the estimated covariance matrix $\mathbf{D}$:

```
            (Intercept) Months
(Intercept)     214.138 -2.166
Months           -2.166  1.320
```

- or corresponding correlation matrix:

```
            (Intercept) Months
(Intercept)       1.000 -0.129
Months           -0.129  1.000
```

Linear Mixed-Effects Models - Example: ARMD Study

- The random effects estimates can be used for predictions.
- This is a major difference between the random effects models and the marginal models.
- Consider the random effects model for the ARMD study:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_{i0} + b_{i1} \text{time}_{ij} + \varepsilon_{ij}$$

- The predictions based on a marginal model are:

$$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \text{group}_i + \hat{\beta}_2 \text{time}_{ij} + \hat{\beta}_3 \text{time}_{ij} \times \text{group}_i,$$

- whereas the predictions based on a mixed model are:

$$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \text{group}_i + \hat{\beta}_2 \text{time}_{ij} + \hat{\beta}_3 \text{time}_{ij} \times \text{group}_i + \hat{b}_{i0} + \hat{b}_{i1} \text{time}_{ij}$$

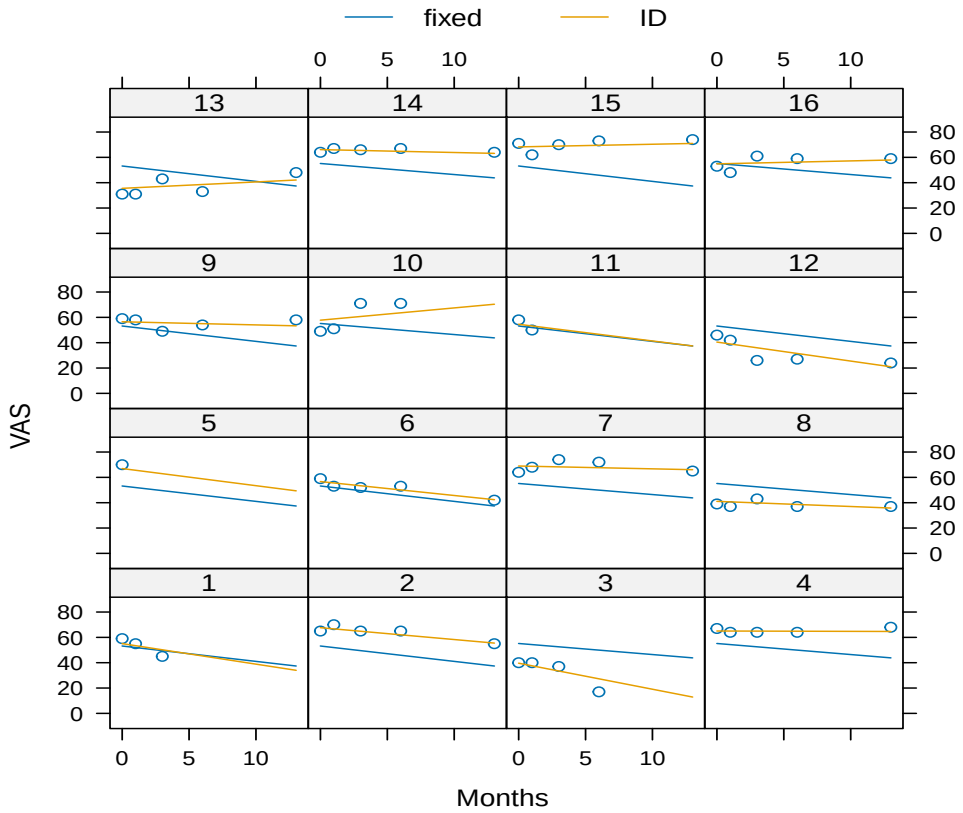
Linear Mixed-Effects Models - Example: ARMD Study

- The difference is the following:
  - from the marginal model we obtain predictions for the average patient having fixed-effects characteristics X_i i.e., group in this example.
  - from the mixed model we obtain predictions for the average patient that has fixed-effects characteristics X_i and observed data y_i (via the random effects) i.e., they individualized predictions.

Linear Mixed-Effects Models - Example: ARMD Study

- Consider the model for the ARMD study:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_{i0} + b_{i1} \text{time}_{ij} + \varepsilon_{ij}$$



- The beauty of random-effects models is that predictions are closer to the observed data.

Outline

Statistical Software to Fit Linear Mixed Models

- Most popular programs in medical applications:
  - R package `nlme` and `lme4`
  - SAS proc MIXED
- STATA
- SPSS
- MLwin (popular in social sciences)
- WINBUGS (Bayesian methodology)
- GENSTAT

Outline

Fitting Linear Mixed-Effects Models in R

There are two main R packages for mixed-effects models:

- `nlme`
  - fits linear and non-linear mixed-effects models, and marginal models for normal data
  - allows for both random effects and correlated error terms
  - allows for nested random-effects designs
  - several options for covariances matrices and variance functions
- `lme4`
  - fits linear, non-linear and generalized mixed-effects models
  - uses only random effects, no serial correlation can be modelled
  - allows for nested and crossed random-effects designs

Fitting Linear Mixed-Effects Models in R

- `nlme`: The basic function is `lme()` with three main arguments:
  - `fixed:` a formula specifying the response vector and the fixed-effects structure.
  - `random:` a formula specifying the random-effects structure.
  - `data:` a data frame containing all the variables.
  - `correlation:` a formula specifying the correlation structure.
  - `weights:` a formula specifying the within-group variance structure.
  - `data:` a data frame containing all the variables.
  - `method:` estimation approach used 'ML' or 'REML'.
- The data should be in long format!

Fitting Linear Mixed-Effects Models in R

- R code for random intercepts and random slopes model:

```
model.lme <- lme(VAS ~ treat + months + months:treat, data = vas, random = ~ months | ID)
```

```
summary(model.lme)
Linear mixed-effects model fit by REML
  Data: armd.
       AIC      BIC   logLik
  8436.921 8476.967 -4210.46

Random effects:
 Formula: ~Months | ID
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 14.633449 (Intr)
Months       1.148949 -0.129
Residual     6.865979

Fixed effects:  VAS ~ Months * Treat
                     Value Std.Error  DF  t-value p-value
(Intercept)        55.17188 1.4006151 865 39.39118  0.0000
Months             -0.87312 0.1262334 865 -6.91675  0.0000
TreatActive        -1.99810 1.9757867 238 -1.01129  0.3129
Months:TreatActive -0.34042 0.1835639 865 -1.85450  0.0640
 Correlation:
                   (Intr) Months TrtAct
Months             -0.207
TreatActive        -0.709  0.147
Months:TreatActive  0.143 -0.688 -0.206

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.96260312 -0.36437094  0.02334052  0.44324588  3.20866457
```

Fitting Linear Mixed-Effects Models in R

- R code for random intercepts model:

```
model.lme <- lme(VAS ~ treat + months + months:treat, data = vas, random = ~ 1 | ID)

summary(model.lme)
Linear mixed-effects model fit by REML
  Data: armd.
       AIC      BIC    logLik
  8583.232 8613.267 -4285.616

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev:     14.4808 8.796997

Fixed effects:  VAS ~ Months * Treat
                      Value Std.Error  DF   t-value p-value
(Intercept)        55.14359 1.4230921 865  38.74914  0.0000
Months             -0.86126 0.0813977 865 -10.58087  0.0000
TreatActive        -2.06324 2.0086064 238  -1.02720  0.3054
Months:TreatActive -0.30100 0.1191406 865  -2.52643  0.0117
 Correlation:
                   (Intr) Months TrtAct
Months             -0.246
TreatActive        -0.708  0.174
Months:TreatActive  0.168 -0.683 -0.242

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-4.14166843 -0.43613934  0.01541633  0.54369700  2.92464503

Number of Observations: 1107
Number of Groups: 240
```

Fitting Linear Mixed-Effects Models in R

- R code for uncorrelated random intercepts and random slopes model:

```
model.lme <- lme(VAS ~ treat + months + months:treat, data = vas, random = list(ID = pdDiag(form = ~
months)))
```

```
summary(model.lme)
Linear mixed-effects model fit by REML
  Data: armd.
       AIC      BIC    logLik
  8437.318 8472.359 -4211.659

Random effects:
 Formula: ~Months | ID
 Structure: Diagonal
        (Intercept)   Months Residual
StdDev:    14.45784 1.129907 6.891205

Fixed effects:  VAS ~ Months * Treat
                      Value Std.Error  DF  t-value p-value
(Intercept)        55.17418 1.3856115 865 39.81937  0.0000
Months             -0.87382 0.1248567 865 -6.99858  0.0000
TreatActive        -1.99853 1.9546927 238 -1.02243  0.3076
Months:TreatActive -0.33944 0.1816824 865 -1.86833  0.0621
 Correlation:
                   (Intr) Months TrtAct
Months             -0.107
TreatActive        -0.709  0.076
Months:TreatActive  0.074 -0.687 -0.109

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.93990817 -0.36412885  0.02997397  0.44841983  3.23930277

Number of Observations: 1107
```

Fitting Linear Mixed-Effects Models in R

- Several positive definite matrices are allowed for the random-effects covariance matrix $\mathbf{D}$:
  - `pdIdent`
  - `pdDiag`
  - `pdCompSymm`, etc

Fitting Linear Mixed-Effects Models in R

- `lme4`: The basic function is `lmer()` with two main arguments:
  - `formula`: a formula specifying the response vector, the fixed-effects and random-effects structure.
  - `data`: a data frame containing all the variables.
- The data should be in long format!
- R code for random intercepts and random slopes model:

```
model.lmer <- lmer(VAS ~ treat + months + months:treat + (months | ID), data = vas)
summary(model.lmer)
```

- R code for random intercepts model:

```
model.lmer <- lmer(VAS ~ treat + months + months:treat + (1 | ID), data = vas)
summary(model.lmer)
```

- R code for uncorrelated random intercepts and random slopes model:

```
model.lmer <- lmer(VAS ~ treat + months + months:treat + (months || ID), data = vas)
summary(model.lmer)
```

Outline

Linear Mixed Effects Models - Model Building

- Mixed effects models consist of three parts:
  - Fixed effects part $\mathbf{X}_i\beta$: describes how the covariates in the model explain the average of the repeated measurements.
  - Random effects part $\mathbf{Z}_i\mathbf{b}_i$: assumed covariance structure between the repeated measurements.
  - Errors terms ε_{ij}: assumed covariance structure between the repeated measurements.
- Scientific interest is often on the fixed effects part.
- However, to obtain valid and efficient inferences for the fixed effects part, the covariance part needs to be correctly specified.

Linear Mixed Effects Models - Model Building

- The recommended strategy is as follows:
  - Consider an elaborate fixed effects part, including possible nonlinear and interaction terms.
  - Then select an appropriate random effects structure that adequately describes the correlations in the repeated measurements.
  - Finally, simplify the fixed effects part by excluding non-significant terms.
- Note that this strategy is followed when we assume independent errors i.e., Σ_i is diagonal or scaled identity.
- For correlated errors the model building strategy involves selecting an appropriate Σ_i as well, which is more complicated.

Outline

Hypothesis testing for covariances

- To find the "optimal" covariance structure between models with the **same mean structure** we can use:
  1. Likelihood ratio test for nested models.
  2. Information criteria for non-nested models:
     - Akaike's Information Criterion (AIC).
     - Schwarz's Bayesian Information Criterion (BIC).

- Note:

  A Model A is nested in Model B if Model A is a "special case" of Model B.

  Loosely speaking, the nested Model A arises by imposing constraints on the parameters in the more general Model B (e.g., certain parameters are set to zero or equal to each other).

Hypothesis testing for covariances - Nested models

- The Likelihood Ratio Test (LRT) can be used to compare two nested models:
  - testing for the need of random effects.
  - choosing the error structure.
- Examples:
  - No random effect versus one random effect e.g. is random intercepts needed?
  - One versus two random effects e.g. is random slopes needed?
  - q versus $q+1$ random effects.

Hypothesis testing for covariances - Nested models

- The LRT is derived as:

$$LRT = \{2\log(L_A)\} - \{2\log(L_0)\} \sim \chi^2_{df}$$

where

  - L_0 is the value of the likelihood function under the null model.
  - L_A is the value of the likelihood function under the alternative model (i.e. more general model).
  - df the difference in the number of parameters between the two models.

- Note: We use REML instead of ML because REML reduces bias in ML estimates of covariance parameters.

Hypothesis testing for covariances - Nested models

- In mixed models the LRT:

$$LRT = \{2\log(L_A)\} - \{2\log(L_0)\} \sim \chi^2_{df}$$

  may not always be valid depending on the null hypothesis.

- When testing if the variance of a random effect is zero, the approximate null distribution is not chi-square.

- The correct approximate distribution is a mixture of chi-squared distributions.

- The standard LR test is too conservative.

Hypothesis testing for covariances - Illustration I

No random effect versus one random effect

Null model: $y_{ij} = \beta_0 + \beta_1 \text{time}_{ij} + \varepsilon_{ij}$

Alternative model: $y_{ij} = \beta_0 + \beta_1 \text{time}_{ij} + b_{i0} + \varepsilon_{ij}, \quad b_{i0} \sim N(0, \sigma^2_{int})$

- Implies testing for

$$H_0 : \sigma^2_{int} = 0 \text{ versus } H_A : \sigma^2_{int} > 0$$

- The asymptotic null distribution of the standard LRT is **not** a chi-squared distribution with 1 df.
- The null hypothesis for σ^2_{int} is on the boundary of its parameter space.
- Instead, LRT under the null hypothesis follows $0.5\chi^2_0 + 0.5\chi^2_1$ distribution.
- If the chi-squared null distribution is used instead, the test is conservative
  $\Rightarrow$ p-value will be overestimated, too simple covariance structure will be selected.

Hypothesis testing for covariances - Illustration II

One versus two random effects

Null model: $y_{ij} = \beta_0 + \beta_1 \text{time}_{ij} + b_{i0} + \varepsilon_{ij}, \quad b_{i0} \sim N(0, \sigma^2_{int})$

Alternative model: $y_{ij} = \beta_0 + \beta_1 \text{time}_{ij} + b_{i0} + b_{i1} \text{time}_{ij} + \varepsilon_{ij},$

$$\begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{int} & \sigma_{int,time} \\ \sigma_{int,time} & \sigma^2_{time} \end{bmatrix}\right)$$

- Implies testing for $H_0 : D = \begin{bmatrix} \sigma^2_{int} & 0 \\ 0 & 0 \end{bmatrix}$ versus $H_A : D = \begin{bmatrix} \sigma^2_{int} & \sigma_{int,time} \\ \sigma_{int,time} & \sigma^2_{time} \end{bmatrix}$
- Implies testing for

$$H_0 : \sigma_{int,time} = \sigma^2_{time} = 0 \text{ versus } H_A : \sigma^2_{time} > 0 \text{ or } \sigma^2_{int,time} \neq 0$$

- The asymptotic null distribution of the standard LRT is **not** a chi-squared distribution with 2 df.
- Instead, it is $0.5\chi^2_1 + 0.5\chi^2_2$ distribution.

Hypothesis testing for covariances - Illustration II

- In general, when testing q versus $q+1$ random effects the asymptotic null distribution of the LRT is

$$LRT \sim 0.5\chi^2_q + 0.5\chi^2_{(q+1)}.$$

Hypothesis testing for covariances - ARMD study

- Consider the ARMD study where we assume linear and quadratic slopes for the fixed and random effects.

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_{i0} + b_{i1} \text{time}_{ij} + b_{i2} \text{time}_{ij}^2 + \varepsilon_{ij}.$$

- The random effects covariance matrix $\mathbf{D}$ in this case is:

$$\mathbf{D} = \begin{bmatrix} \sigma^2_{int} & \sigma_{int,time} & \sigma_{int,time,time2} \\ \sigma_{int,time} & \sigma^2_{time} & \sigma_{int,time} \\ \sigma_{int,time,time2} & \sigma_{int,time} & \sigma^2_{time2} \end{bmatrix}$$

- To test for the significance of the quadratic random slopes, we test the null hypothesis:

$$H_0 : \sigma^2_{time2} = \sigma_{int,time,time2} = \sigma_{int,time} = 0$$

- To test this hypothesis using LRT we need to fit the models that correspond to H_0 and H_A.

Hypothesis testing for covariances - ARMD study

- The random effects covariance matrix under the null is:

$$\begin{bmatrix} \sigma^2_{int} & \sigma_{int,time} \\ \sigma_{int,time} & \sigma^2_{time} \end{bmatrix}$$

- The random effects covariance matrix under the alternative is:

$$\begin{bmatrix} \sigma^2_{int} & \sigma_{int,time} & \sigma_{int,time,time2} \\ \sigma_{int,time} & \sigma^2_{time} & \sigma_{int,time} \\ \sigma_{int,time,time2} & \sigma_{int,time} & \sigma^2_{time2} \end{bmatrix}$$

- $LRT = \{-2\log(L_A)\} - \{-2\log(L_0)\} = (8429.016) - (8365.631) = 63.385$ it should be compared with a $0.5\chi^2_3 + 0.5\chi^2_2$.

Hypothesis testing for covariances - Non-Nested models

- When we have non-nested models we cannot use the LRT.
- Alternatively we can use information criteria: AIC or BIC.
- When we compare two non-nested models we choose the model that has the lowest AIC/BIC value.

Hypothesis testing for fixed effects

- In practice we test hypotheses of the form:

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_A : \beta_3 \neq 0$$

or

$$H_0 : \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_A : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

  namely, the hypotheses involve only one or > 1 parameters.

- Two tests are often used:
  1. t-test (test for 1 parameter) or F-test (test for > 1 parameters).
  2. Likelihood Ratio Test.

Hypothesis testing for fixed effects

- The ML estimate of β is given by:

$$\hat{\beta} = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{V}_i^{-1} \mathbf{y}_i$$

  where $\mathbf{V}_i$ is replaced by its ML or REML estimate.

- The variance-covariance matrix of $\hat{\beta}$ is

$$\text{Var}(\hat{\beta}) = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1}$$

  where again $\mathbf{V}_i$ is replaced by its ML or REML estimate.

- Given $\hat{\mathbf{V}}_i$ it holds $\hat{\beta} \sim N_p \left(\beta, \text{Var}(\hat{\beta}) \right)$.

Hypothesis testing for fixed effects: t-test

- To test hypotheses of the form:

$$H_0 : \beta = 0 \qquad \text{vs} \qquad H_A : \beta \neq 0$$

  i.e. for a single parameter, we use the t-test statistic

$$W = \frac{\hat{\beta}}{s.e.(\hat{\beta})} \sim t_{df}$$

  where $\hat{\beta}$ is the ML estimate of β, $s.e.(\hat{\beta})$ is the standard error of $\hat{\beta}$, and df are the degrees of freedom.

- Note:
  - W does not follow an exact t distribution as in linear regression where $df = n - p$ with p # fixed effects parameters.
  - Instead we use approximate methods to estimate the degrees of freedom to account for the variability introduced by replacing $\mathbf{V}_i^{-1}$ by its estimate (details later).

Hypothesis testing for fixed effects: F-test

- For a null hypothesis with more parameters, e.g.:

$$H_0 : \beta_2 = \beta_3 = 0 \text{ vs } H_A : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

we use the F-test statistic.

- In the general case, for a known $k \times p$ contrasts matrix $\mathbf{L}$ and β the $p \times 1$ vector of fixed effects parameters the set of hypotheses is written as:

$$H_0 : \mathbf{L}\beta = 0 \text{ vs } H_A : \mathbf{L}\beta \neq 0$$

and the F-test statistic is:

$$F = \frac{\hat{\beta}^T \mathbf{L}^T \left[\mathbf{L} \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}^{-1} \mathbf{X}_i \mathbf{L}^T\right]^{-1} \mathbf{L}\hat{\beta}}{\text{rank}(\mathbf{L})} \sim F_{df_1, df_2}$$

where df_1 are the numerator degrees of freedom equal to the rank of matrix $\mathbf{L}$ and df_2 are the denominator degrees of freedom which are estimated from the data.

Hypothesis testing for fixed effects: F-test

- Several methods are available to estimate the denominator degrees of freedom:
  - Containment method
  - Satterthwaite approximation
  - Kenward and Roger approximation, etc.
- In longitudinal studies, all methods typically give similar p-values because of the high number of degrees of freedom.
- More attention should be paid when working with small samples:
  Kenward-Roger method is recommended in these situations.
- R package `lmerTest` allows F-test where the denominator degrees of freedom are estimated with the previously mentioned methods.
- Note: F-tests with small sample corrections are not available in `nlme`.

Example: ARMD Study

- Often hypotheses of interest not automatically computed.
- Consider the model for the ARMD study:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_{i0} + b_{i1} \text{time}_{ij} + \varepsilon_{ij}$$

- To test for different mean trajectories between the two groups, we need to test:

$$H_0 : \beta_1 = \beta_3 = 0 \text{ vs } H_A : \beta_1 \neq 0 \text{ or } \beta_3 \neq 0$$

- In R to test such hypothesis using the F-test we need to use the `anova(.)` method.
- But first we need to understand how to construct the contrasts matrix $\mathbf{L}$, i.e. write $H_0 : \beta_1 = \beta_3 = 0$ in the form $H_0 : \mathbf{L}\beta = 0$ (see discussion on Marginal Models).

Hypothesis testing for fixed effects: Likelihood Ratio Test

- We can use the likelihood ratio test (LRT) to test hypothesis:

$$H_0 : \beta_1 = \beta_3 = 0 \text{ vs } H_A : \beta_1 \neq 0 \text{ or } \beta_3 \neq 0$$

- LRT can be used to compare nested models different mean structures, but equal covariance structure.

- The implied models are:

$$\begin{aligned} H_0 : &\quad y_{ij} = \beta_0 + \beta_2 \text{time}_{ij} + b_{i0} + b_{i1}\text{time}_{ij} + \varepsilon_{ij} \\ H_A : &\quad y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_{i0} + b_{i1}\text{time}_{ij} + \varepsilon_{ij} \end{aligned}$$

  for both models we assume $\varepsilon_{ij} \sim N(0, \sigma^2)$.

- The LRT is given by:

$$LRT = \{-2\log(L_0)\} - \{-2\log(L_A)\} \sim \chi^2_{df}$$

  with L_A and L_0 the values of ML under the alternative and null model, df the difference in the number of parameters between the two models.

Summary of Hypothesis testing for fixed effects

- The likelihood ratio test, and the classical univariate and multivariate Wald tests (i.e., using the χ^2 distribution instead of the t or F distributions) are 'liberal' they give smaller p-values than the ones they should give, especially in small samples.
- Important: The likelihood ratio test for comparing models with different fixed effect parts is only valid when the models have been fitted using ML and not REML.
- REML likelihoods are based on different observations, which makes them no longer comparable.

Outline

Model diagnostics

- **All statistical models make assumptions.**
- For valid statistical inference we need to check if these assumptions hold.
- Assumptions for the linear mixed model:

$$\mathbf{y_i} = \mathbf{X_i}\beta + \mathbf{Z_i}\mathbf{b_i} + \boldsymbol{\varepsilon_i}$$

  - the random effects follow the normal distribution $\mathbf{b}_i \sim N_q(\mathbf{0}, \mathbf{D})$
  - the error terms follow the normal distribution $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$
  - the error terms are independent of the covariates $\mathbf{X}_i$.
- It is also advisable to check if validity of model assumptions sensitive to outliers.

Model diagnostics

- Model diagnostics include:
  - Residual analysis.
  - Empirical Bayes estimates plots.
- Recommendations can be extracted:
  - Do we need to transform response variable and/or covariates?
  - Do we need to change mean model? e.g. add interactions or polynomials?
  - Should we model covariance structure differently?
  - Is the model severely influnced by certain subjects?
  - Is there an important covariate omitted from the $\mathbf{Z}_i$ design matrix of $\mathbf{b}_i$?

Model diagnostics - Residuals

- Two model formulations for the linear mixed model:

1. **Hierarchical model**

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

with $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$ and $\mathbf{b}_i \sim N_q(\mathbf{0}, \mathbf{D})$.

2. **Marginal model**

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

with $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{n_i})$.

Model diagnostics - Residuals

- This implies for the linear mixed model two types of residuals:
  1. Conditional residuals:
  $$\mathbf{e}_i^C = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} - \mathbf{Z}_i\hat{\mathbf{b}}_i$$
  with $\hat{\boldsymbol{\beta}}$ the (RE)ML estimate and $\hat{\mathbf{b}}_i$ the empirical Bayes estimates of the random effects.
  2. Marginal residuals
  $$\mathbf{e}_i^M = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$$
  with $\hat{\boldsymbol{\beta}}$ the (RE)ML estimate.

Model diagnostics - Residuals

- In a correctly specified model $E(\mathbf{e}_i^C) = E(\mathbf{e}_i^M) = \mathbf{0}$ and $\text{cov}(\varepsilon_i) \approx \text{cov}(\mathbf{e}_i) = \hat{\mathbf{V}}_i$.
- By plotting the residuals we can investigate the validity of model assumptions:
  1. Conditional residuals:
     - Misspecification of the mean structure $\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$
     - Validity of the error structure $\sigma^2\mathbf{I}_{n_i}$
  2. Marginal residuals:
     - Misspecification of the mean structure $\mathbf{X}_i\boldsymbol{\beta}$
     - Validity of the error struture $\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{n_i}$

Model diagnostics - Residuals

- As in the marginal models, it is advisable to consider the normalized residuals (if available) instead of the raw or pearson residuals.
- Normality assumption: Normal probability plot (or Q-Q plot)
- Zero mean, heteroskedasticity: Scatterplot of residuals versus fitted values and/or covariates.

Model diagnostics - R

- For the ARMD trial we consider the model:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{time}_{ij} \times \text{group}_i + b_i + \varepsilon_{ij}$$

  where $b_i \sim N(0, \sigma_b^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

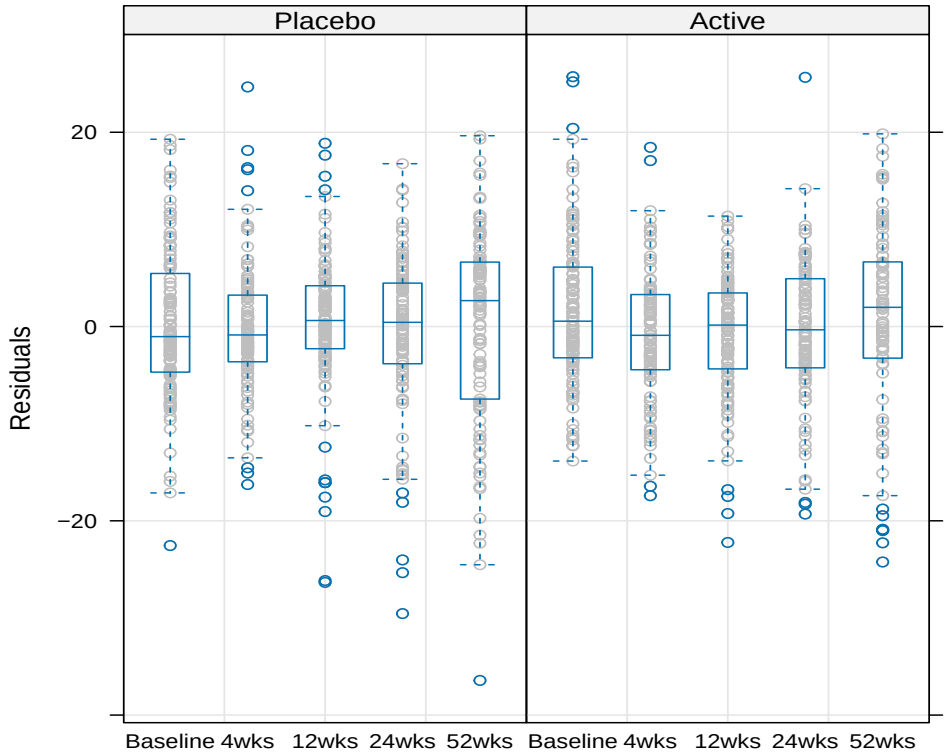
- We will check the fit of the model using:
  - raw residuals versus fitted values/time (per treatment group)
  - standardized residuals versus fitted values/time (per treatment group)
  - normalized residuals versus fitted values/time (per treatment group)
  - observed versus fitted values
  - QQ-plot of standardized/normalized residuals

- All these for both the conditional and marginal residuals.

- In `nlme` both residuals are computed from the fitted object using function `residuals(., level = Q, type = c("response", "pearson", "normalized"))`, with `level = 0` corresponding to population residuals.

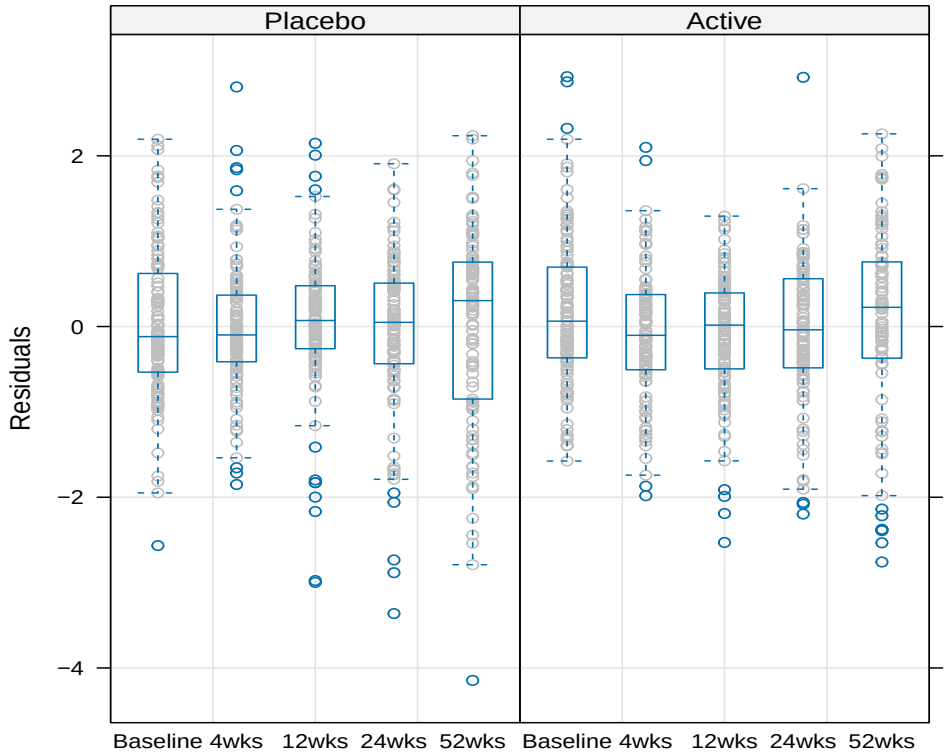
Model diagnostics - R

- Plots of raw residuals (conditional):



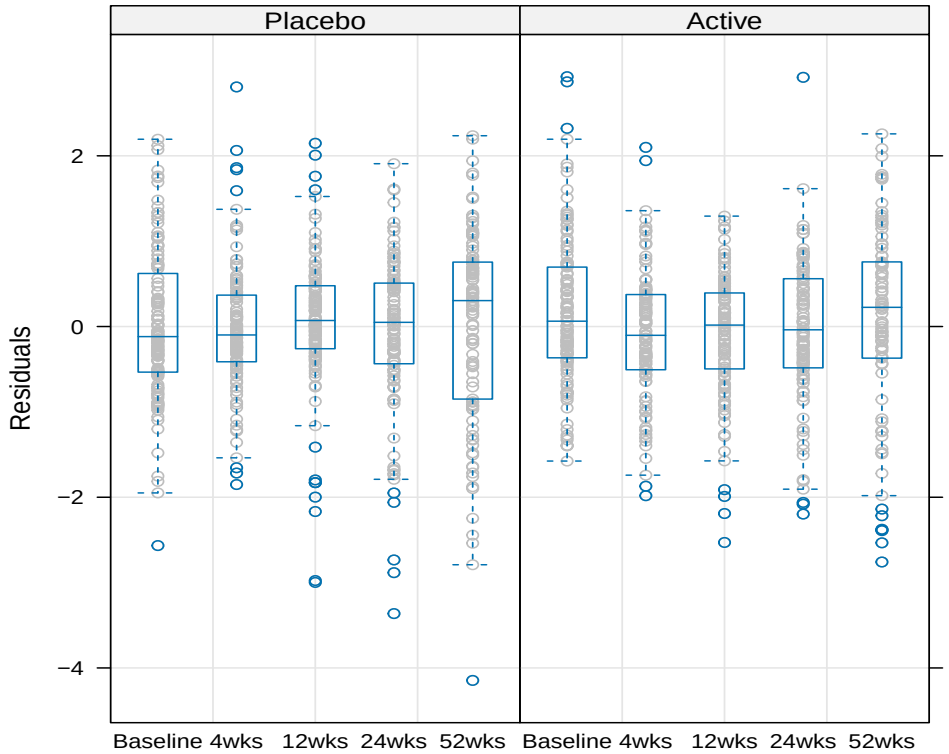
Model diagnostics - R

- Plots of pearson residuals (conditional):



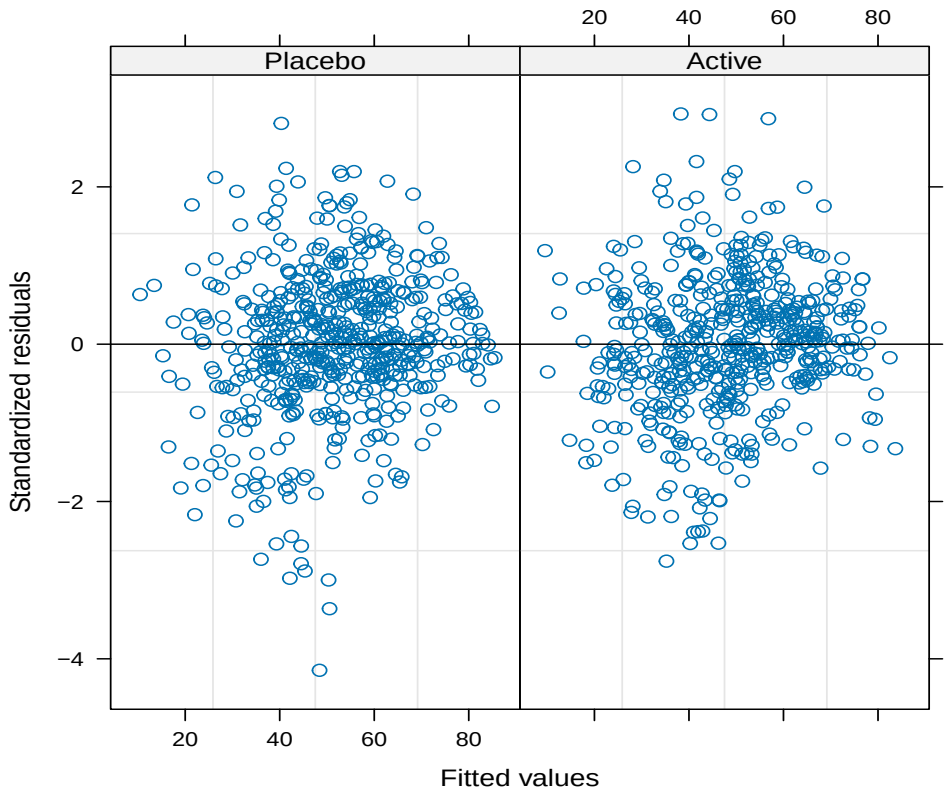
Model diagnostics - R

- Plots of normalized residuals (conditional):



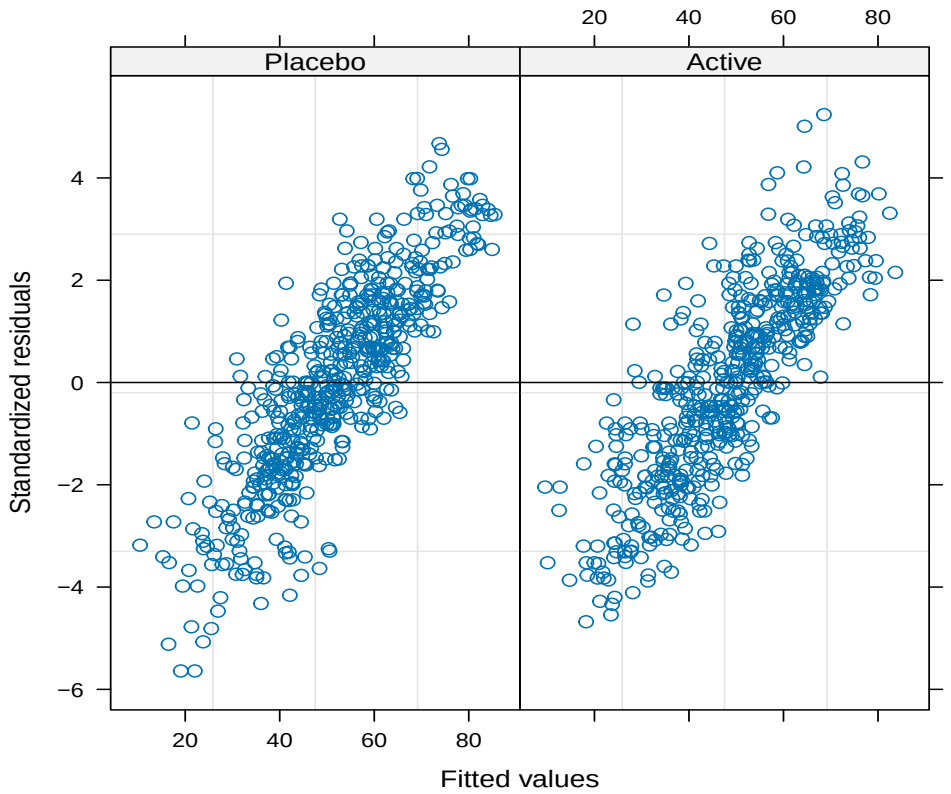
Model diagnostics - R

- Plots of normalized (conditional) residuals versus fitted values.



Model diagnostics - R

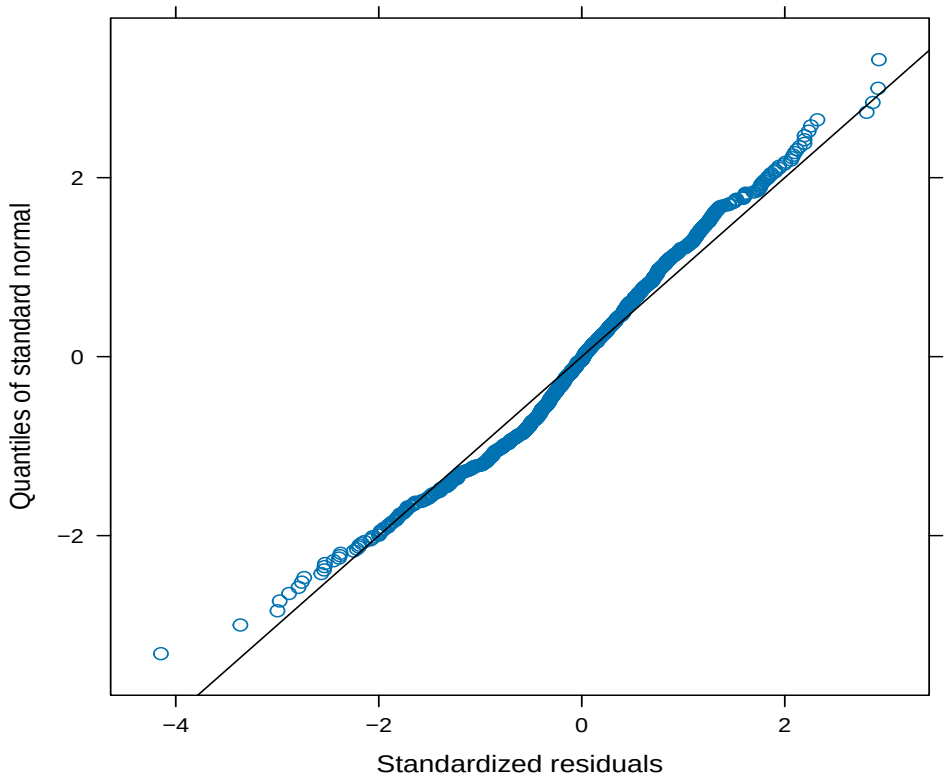
- Plots of raw (marginal) residuals versus fitted values.



- Marginal residuals can be more sensitive to missing data than the conditional ones. In `nlme` normalized marginal residuals are not supported.

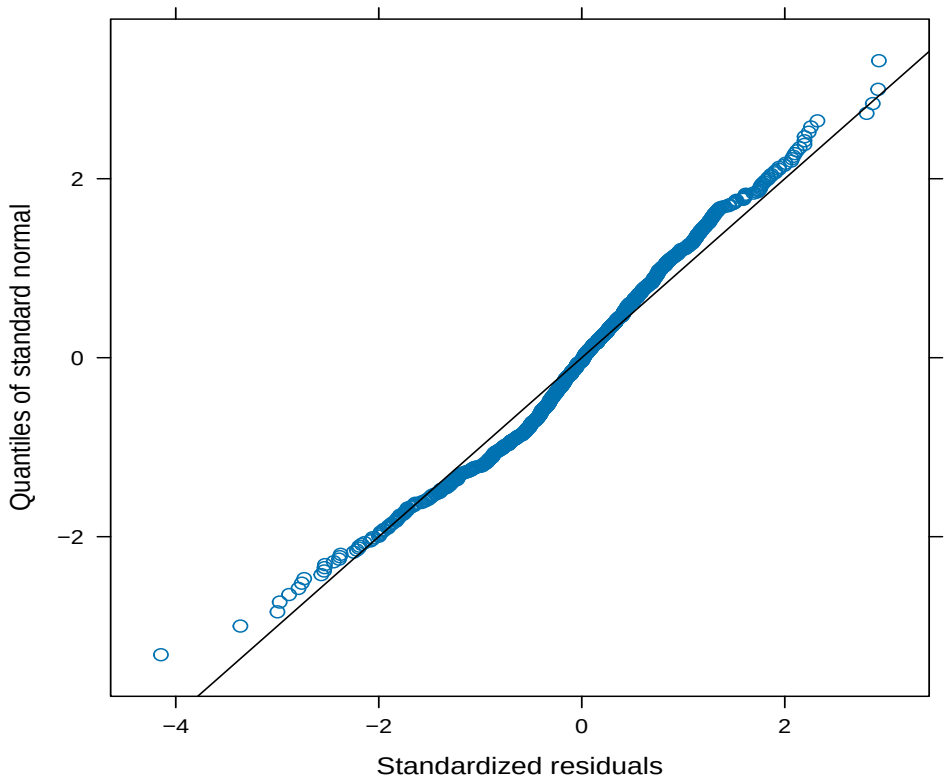
Model diagnostics - R

- Q-Q plot of pearson residuals (conditional).



Model diagnostics - R

- Q-Q plot of normalized residuals.



- Note: no difference between pearson and normalized. Why?

Model diagnostics - R

- In the estimation of the LMM we discussed the the estimation of the random effects $\mathbf{b}_i$.
- Q-Q plots of $\hat{\mathbf{b}}_i$ are not useful in testing the normality assumption for the random effects bue to the shrinkage effect.
- Histograms of $\hat{\mathbf{b}}_i$ can be useful to detect outliers.
- The validity of the normality for the random effects can be evaluated by comparing inference under different distributional assumotions for $\mathbf{b}_i$.

Model diagnostics - Remarks

- The plot of the normalized residuals versus fitted values does not show a systematic trend but some outlying subjects are indicated.
- The QQ-plot of the normalized residuals is not perfect $\Rightarrow$ it suggests that a heavier tailed distribution than the normal may fit the data better.
- Marginal residuals can be more sensitive to missing data than the conditional ones.

Outline

Multilevel Models - Introduction

- So far we have discussed the analysis of longitudinal data: measurements collected on the same subject.
- The examples we have seen so far are a special case of **Multilevel Data**
- Multilevel data arise when there is a natural hierarchy in the sampling design.
  - Studies on schools: data collected on students from different classrooms $\Rightarrow$ districts $\Rightarrow$ cities.
  - Cluster randomized trials: centers randomized to treatments, data collected on individual patients.
  - Family studies: data collected on members of the same family.
- These examples of multilevel data exhibit natural hierarchy with more than 2 levels.
- The analysis of multilevel data is done using mixed models with level-specific random effects.

Multilevel Models - Two level designs

The ARMD study is an example of a 2-level design:

- Patients represent the **level 2** units and the measurements collected on each patient i are the **level 1** units.
- VAS is measured for each patient i at each visit j:
  y_{ij} with $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$.
- Covariates can be measured at both levels e.g. treatment arm (level 2) or age at each visit (level 1).
- The correlation within clusters is modelled using random effects at level 2.

Multilevel Models - Two level designs

- The 2-level model is

$$\mathbf{y}_{ij} = \mathbf{x}_{ij}^T \beta + \mathbf{z}_{ij} \mathbf{b}_i + \varepsilon_{ij}$$

- The random effects are constant for all level 1 units y_{ij}: level 1 units are correlated because they share the same $\mathbf{b}_i$.

- Given $\mathbf{b}_i$ the level 1 units are assumed conditionally independent.

- This LMM has been extensively discussed in the previous sections.

Multilevel Models - Three level designs

- Let us assume that the ARMD study is cluster randomized longitudinal trial.
- Hospitals are randomized to the 2 treatment arms and patients within each hospital are followed up in time.
- The hospitals represent the **level 3** units, the patients represent the **level 2** units and the measurements collected on each patient are the **level 1** units.
- VAS is measured for each patient j of a hospital i at each visit k: y_{ijk} with $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ and $k = 1, \ldots, n_j$.
- Covariates can be measured at any of the 3 levels e.g. treatment arm (level 3), baseline BMI (level 2) or age at each visit (level 1).

Multilevel Models - Three level designs

- Within clusters correlation needs to be taken into account:
  - Within hospitals correlation: patients of the same hospital are correlated.
  - Within patients correlation: measurements of the same patient correlated.
- The correlation within clusters is modelled via the level 2 and level 3 random effects.
- The 3 level mixed model is written as

$$\mathbf{y}_{ijk} = \mathbf{x}_{ijk}^T \beta + \mathbf{z}_{ijk}^{(3)} \mathbf{b}_i^{(3)} + \mathbf{z}_{ijk}^{(2)} \mathbf{b}_j^{(2)} + \varepsilon_{ijk}$$

  where $i = 1, \ldots, N$ denotes the hospital, $j = 1, \ldots, n_i$ the patient within hospital i, $k = 1, \ldots, n_{ij}$ denotes the measurement within patient j and hospital i and

  $\mathbf{b}_i^{(3)} \sim N(\mathbf{0}, \mathbf{D}^{(3)})$, $\mathbf{b}_j^{(2)} \sim N(\mathbf{0}, \mathbf{D}^{(2)})$ and $\varepsilon_{ijk} \sim N(0, \sigma^2)$.
- Given $\mathbf{b}_j^{(2)}$ and $\mathbf{b}_i^{(3)}$ the level 1 units are assumed conditionally independent.

Multilevel Models - Three level designs

- Let us consider the simple case where both the level 3 and level 2 random effects are 1 dimensional:

$$\mathbf{y}_{ijk} = \mathbf{x}_{ijk}^T \beta + b_i^{(3)} + b_j^{(2)} + \varepsilon_{ijk}$$

with $b_i^{(3)} \sim N(0, \sigma^2_{hospital})$, $b_j^{(2)} \sim N(0, \sigma^2_{patient})$ and $\varepsilon_{ijk} \sim N(0, \sigma^2)$.

- Under the assumption of independence between $b_i^{(2)}$ and $b_j^{(3)}$ we have

$$\text{Var}(y_{ijk}) = \sigma^2_{hospital} + \sigma^2_{patient} + \sigma^2.$$

Multilevel Models - Three level designs

$$\mathbf{y}_{ijk} = \mathbf{x}_{ijk}^T \beta + b_i^{(3)} + b_j^{(2)} + \varepsilon_{ijk}$$

with $b_i^{(3)} \sim N(0, \sigma^2_{hospital})$, $b_j^{(2)} \sim N(0, \sigma^2_{patient})$ and $\varepsilon_{ijk} \sim N(0, \sigma^2)$.

- The correlation between observations of the same patient is

$$Cor(y_{ijk}, y_{ijk'}) = \frac{\sigma^2_{hospital} + \sigma^2_{patient}}{\sigma^2_{hospital} + \sigma^2_{patient} + \sigma^2}$$

Multilevel Models - Three level designs

$$\mathbf{y}_{ijk} = \mathbf{x}_{ijk}^T \beta + b_i^{(3)} + b_j^{(2)} + \varepsilon_{ijk}$$

with $b_i^{(3)} \sim N(0, \sigma^2_{hospital})$, $b_j^{(2)} \sim N(0, \sigma^2_{patient})$ and $\varepsilon_{ijk} \sim N(0, \sigma^2)$.

- The correlation between patients of the same hospital is

$$Cor(y_{ijk}, y_{ij'k}) = \frac{\sigma^2_{hospital}}{\sigma^2_{hospital} + \sigma^2_{patient} + \sigma^2}$$

- More complex hierarchical structures can be handled using additional random effects.

Multilevel Models - Nested versus Crossed random effects

- In the ARMD study the hospitals and patients are nested classification factors.
- Patients (level 1) are nested within hospitals (level 2).
- The random effects used to model the 3 sources of variation are called *nested* random effects.
- Let us consider a study on sports activities of school aged children: data are collected on students from different schools who join one of the sport clubs of the city.
- In this case schools and sport clubs are crossed classification factors as children from different shcools may join the same club.
- Further analysis requires random effects at the school and sport club level which are called **crossed** random effects.

Multilevel Models - Estimation and Inference

- Estimation and inference for multilevel models is done as in any LMM we have discussed before.
- Estimation can be more computationally demanding for non-normal responses (to be discussed later on GLMMs).
- Multilevel models can be fitted with any of two main R packages for mixed-effects models: `nlme` or `lme4` as we have seen before for the simple 2-level design.
- `nlme`: proper specification of the `random` argument is needed. It does not handle crossed random effects.
- `lme4`: no special treatment is required to model nested or crossed random effects.

Fitting Linear Mixed-Effects Models in R

- R code for nested random effects in `nlme`:

```
model.nested <- lme(VAS ~ Months * Treat, random = ~ Months|Center/ID, data = armd.)
```

```
summary(model.nested)
Linear mixed-effects model fit by REML
  Data: armd.
       AIC      BIC   logLik
  8439.981 8495.044 -4208.99

Random effects:
 Formula: ~Months | Center
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 2.80316481 (Intr)
Months      0.08691867 -0.027

 Formula: ~Months | ID %in% Center
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 14.403179 (Intr)
Months       1.146962 -0.13
Residual     6.864401

Fixed effects:  VAS ~ Months * Treat
                      Value Std.Error  DF   t-value p-value
(Intercept)        55.45283 1.8003796 865 30.800633  0.0000
Months             -0.87343 0.1311365 865 -6.660502  0.0000
TreatActive        -2.08694 1.9502593 233 -1.070084  0.2857
Months:TreatActive -0.34341 0.1834306 865 -1.872178  0.0615
 Correlation:
                   (Intr) Months TrtAct
Months             -0.160
TreatActive        -0.541  0.143
```

Fitting Linear Mixed-Effects Models in R

- R code for nested random effects in `lme4`:

```
model.nested <- lmer(VAS ~ Months * Treat+ + (Months|Center/ID), data = armd.)
model.nested <- lmer(VAS ~ Months * Treat+ + (Months|ID)+ (Months|Center), data = armd.)
```

```
summary(model.nested)
Linear mixed model fit by REML ['lmerMod']
Formula: VAS ~ Months * Treat + (Months | Center/ID)
   Data: armd.

REML criterion at convergence: 8418

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.9405 -0.3625  0.0313  0.4443  3.2111

Random effects:
 Groups    Name        Variance  Std.Dev. Corr
 ID:Center (Intercept) 2.075e+02 14.40419
           Months      1.315e+00  1.14694 -0.13
 Center    (Intercept) 7.776e+00  2.78849
           Months      7.151e-03  0.08456 0.04
 Residual              4.712e+01  6.86455
Number of obs: 1107, groups:  ID:Center, 240; Center, 6

Fixed effects:
                  Estimate Std. Error t value
(Intercept)        55.4495     1.7966  30.864
Months             -0.8729     0.1309  -6.670
TreatActive        -2.0876     1.9504  -1.070
Months:TreatActive -0.3435     0.1834  -1.873

Correlation of Fixed Effects:
            (Intr) Months TrtAct
```

Fitting Linear Mixed-Effects Models in R

- It is advisable to use in `lme4` the formulation:

```
model.nested <- lmer(VAS ~ Months * Treat+ + (Months|Center/ID), data = armd.)
```

- If patients do not have a unique ID code in the dataframe (i.e. when ID value 1 represents patient 1 in all centers) then using

```
model.nested <- lmer(VAS ~ Months * Treat+ + (Months|ID)+ (Months|Center), data = armd.)
```

  will lead to wrong inference.

Chapter 5: Generalized Estimating Equations

Master Statistics and Data Science

Outline

Repeated measures GLM

- In the previous parts of the course you saw models for normal repeated measures data

  $\Rightarrow$ extensions of linear regression model to repeated measures case.

  1. Marginal models: model the correlation between the repeated measurements of the same subject using various variance-covariance matrices.
  2. Random effects models: model within subjects correlation using random effects.

- For repeatedly measured non-normal data extensions of GLM more difficult.

- Before discussing models for such outcomes we will review

  Generalized Linear Model: single response variable

Outline

Generalized Linear Models - Examples

- We will now review the GLMs via examples.
- Study the interpretation of the regression parameters.
- Analyses in R.

Example: Bronchopulmonary dysplasia study

- Reference: Van Marter, L.J., Leviton, A., Kuban, K.C.K., Pagano, M. and Allred, E.N. (1990). Maternal glucocorticoid therapy and reduced risk of bronchopulmonary dysplasia. Pediatrics, 86, 331-336.

- Development of bronchopulmonary dysplasia (BPD) in a sample of 223 low birth weight infants

ID	BPD	Birth weight (grams)	Gestational age (weeks)	Toxemia
1	1	850	27	0
2	0	1500	33	0
3	1	1360	32	0
⋮	⋮	⋮	⋮	⋮

- Response: Presence of BPD by day 28 of life

$$Y = \begin{cases} 1, \text{ if BPD} \\ 0, \text{ otherwise} \end{cases}$$

Example: Bronchopulmonary dysplasia study

- Covariates:
  - Birth weight of infant in grams
  - Gestational age in weeks
  - Toxemia: 1 if present and 0 otherwise

- Research question: Is the birth weight associated with BPD by day 28 of life after correcting for gestational age and toxemia?

General Linear Model

- For the BPD example the linear regression model is not appropriate

  $\Rightarrow$ the normality assumption is violated

- Response variable Y: Continuous (e.g. blood pressure)

- Goal: relate mean response to covariates

- The model for each subject is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

  with ε the error term $\Rightarrow \varepsilon \sim N(0, \sigma^2)$ and

  - $X_1, X_2, \ldots, X_p$ are the p independent variables (aka covariates)
  - $\beta_0, \beta_1, \ldots, \beta_p$ regression coefficients

General Linear Model

- Y is assumed to have a normal distribution with mean

$$\mu = E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

and variance $\sigma^2 \Rightarrow Y \sim N(\mu, \sigma^2)$

- **Linear** relationship between expected response and covariates

Logistic regression model

- Response variable Y: binary (e.g. presence or absence of a symptom)

$$Y = \begin{cases} 1, & \text{if "success"} \\ 0, & \text{if "failure"} \end{cases}$$

- Goal: Estimate the proportion of successes π and relate it to covariates

Logistic regression model

- The binary response Y is assumed to have the Bernoulli distribution (special case of Binomial distribution)

$$Y \sim Bin(1, \pi), \quad Pr(Y = y) = \pi^y (1-\pi)^{(1-y)}$$

- The mean of binary response Y is the proportion of successes

$$\pi = E(Y) = Pr(Y = 1) = Pr(\text{“}success\text{”})$$

- The variance of Y is $\text{var}(Y) = \pi(1-\pi)$

Logistic regression model

- Say we use a naive strategy and consider a linear regression model for π

$$\pi = E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

- What is the problem?
  - π is a probability and is restricted to values between 0 and 1
  - $\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$ can take values below 0 and above 1, i.e. in $(-\infty, +\infty)$

- We have to bring π in the scale $(-\infty, +\infty)$

$$\begin{aligned} 0 < & \quad \pi & < 1 \quad \Leftrightarrow \\ 0 < & \quad \frac{\pi}{1-\pi} & < +\infty \quad \Leftrightarrow \\ -\infty < & \quad \log \frac{\pi}{1-\pi} & < +\infty \end{aligned}$$

Logistic regression model

- Thus we use the logistic or logit transformation

$$\text{log odds of success} = \log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$$

$$\text{odds of success} = \frac{\pi}{1-\pi} = \exp\left(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p\right)$$

$$\text{probability of success} = \pi = \frac{\exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p)}$$

- Respects the constraint that π takes values between 0 and 1
- Note that
  - the relationship between log odds of success and the covariates is linear
  - the relationship between π and the covariates is non-linear

    $\Rightarrow$ Interpretation of parameters different from the linear regression model

Logistic regression model - Parameter Interpretation

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

- A unit change in X_1 from x to $x+1$ (while all other covariates are held fixed) corresponds to

$$\begin{aligned} \log \frac{\pi}{1-\pi} &= \beta_0 + \beta_1 x + \beta_2 X_2 + \ldots + \beta_p X_p \\ \log \frac{\pi}{1-\pi} &= \beta_0 + \beta_1 (x+1) + \beta_2 X_2 + \ldots + \beta_p X_p \end{aligned}$$

- Thus,

$$\beta_1 = \log \frac{\pi}{1-\pi} - \log \frac{\pi}{1-\pi} = \log \left\{ \frac{\pi}{1-\pi} \Big/ \frac{\pi}{1-\pi} \right\}$$

$$\exp(\beta_1) = \frac{\pi}{1-\pi} \Big/ \frac{\pi}{1-\pi}$$

Logistic regression model - Parameter interpretation

- For binary X_1
  - $\Rightarrow$ β_1 is the log odds ratio of success between the two levels of X_1 given that all other covariates remain constant
  - $\Rightarrow$ $\exp(\beta_1)$ is the odds ratio between the two levels of X_1 given that all other covariates remain constant
- For continuous X_1
  - $\Rightarrow$ β_1 is change in log odds of success for a unit change in X_1 given that all other covariates remain constant

Logistic regression model - Parameter interpretation

- β_1 is called population slope $\Rightarrow$ inference applies on population from which we took sample

- $\exp(\beta_1)$ is the odds ratio or ratio of odds of success for the two possible levels of X_1

$$\exp(\beta_1) = \begin{cases} = 1, & \text{the two odds are the same} \\ > 1, & \text{increased odds of success} \\ < 1, & \text{decreased odds of success} \end{cases}$$

- As π increases
  - $\Rightarrow$ odds of success increases
  - $\Rightarrow$ log odds of success increases

- As π decreases
  - $\Rightarrow$ odds of success decreases
  - $\Rightarrow$ log odds of success decreases

Logistic regression model - Parameter Interpretation

- β_0: population intercept $\Rightarrow$ log odds of success when all covariates are zero
  - centering of continuous covariates
- Changes in the odds from combined effects of different risk factors are multiplicative
  - For the logistic regression model: $\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 \text{treatment} + \beta_2 \text{age} + \beta_3 \text{BMI}$
  - The odds ratio of heart attack between
    - a patient in the placebo group, of age 30 and with BMI 20 and
    - a patient in the treatment group, of age 50 and with BMI 25

    is:
- We will study the parameter interpretation in Example

Logistic regression model - BPD data

- Example: We are interested in the association between birth weight and BPD
- The logistic regression model is

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 \text{Weight}$$

  where $\pi = Pr(BPD = 1)$

- We are interested in β_1
- In R we use:

```
model <- glm(BPD ~ Weight, family = binomial, data = BPDdata)
```

Logistic regression model - Parameter Interpretation

- glm output

```
> summary(model.11)
Call:
glm(formula = BPD ~ BirthWeight, family = binomial, data = BPDdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9916  -0.7993  -0.4096   0.9242   2.4802

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.0342913  0.6957121   5.799 6.68e-09 ***
BirthWeight -0.0042291  0.0006408  -6.600 4.11e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 286.14  on 222  degrees of freedom
Residual deviance: 223.72  on 221  degrees of freedom
AIC: 227.72

Number of Fisher Scoring iterations: 4
```

Logistic regression model - Parameter Interpretation

- The estimated logistic regression (obtained using ML) is

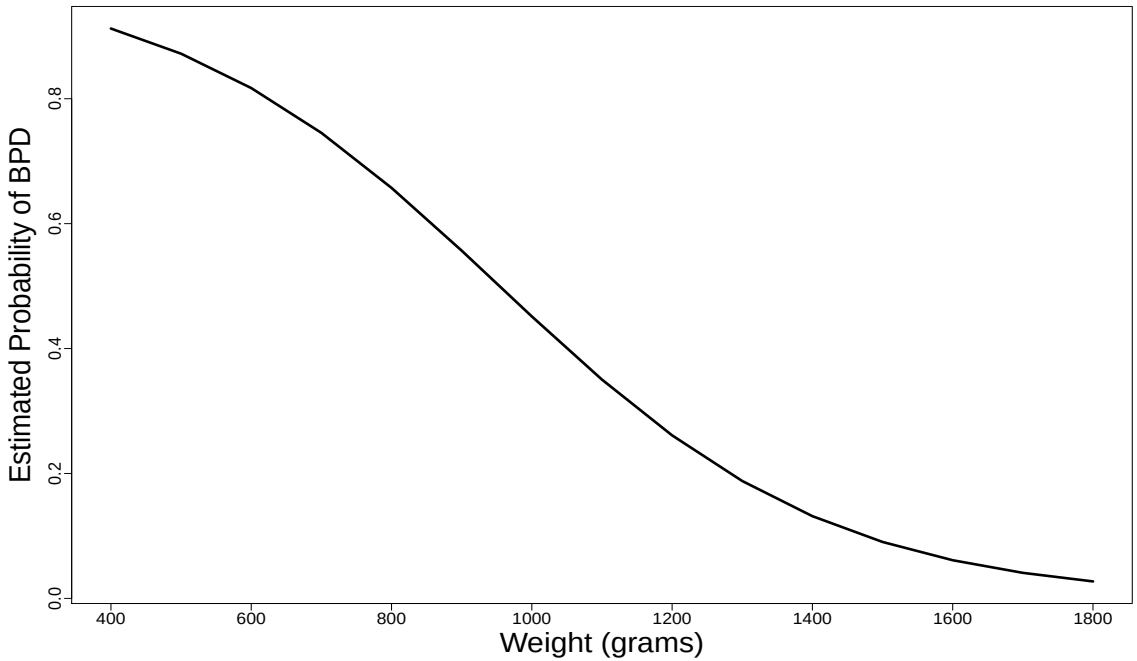
$$\log \frac{\hat{\pi}}{1-\hat{\pi}} = 4.034 - 0.004 \times \text{Weight}$$

- The odds of BPD for an infant weighing 1200 grams is $\exp(4.034 - 0.004 \times 1200) = \exp(-0.766) = 0.465$

- Thus the predicted probability of BPD is: $0.465/(1+0.465) = 0.317$

Logistic regression model - Parameter Interpretation

- For every weight value we obtain the plot of estimated probabilities of BPD



Logistic regression model - Parameter Interpretation

- The coefficient for birth weight is $\beta_1 = -0.004$ with SE $= 0.001$
- Testing for association between weight and BPD corresponds to testing the hypotheses

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned}$$

- Using the Wald test we obtain p-value $< 0.001 \Rightarrow$ the population slope differs from 0 at 5%
- β_1 implies that for every 100 grams increase in birth weight, the log odds of BPD decreases by 0.42

Logistic regression model

- We adjust the analysis for gestational age and toxemia
- The logistic regression model is

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 \text{Weight} + \beta_2 \text{Age} + \beta_3 \text{Toxemia}$$

  where $\pi = Pr(BPD = 1)$

- We are interested in β_1

Logistic regression model

- glm output

```
> summary(model.12)

Call:
glm(formula = BPD ~ BirthWeight + GestationalAge + Toxemia, family = binomial,
    data = BPDdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8400  -0.7029  -0.3352   0.7261   2.9902

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    13.9360826  2.9825508   4.673 2.98e-06 ***
BirthWeight    -0.0026436  0.0008123  -3.254  0.00114 **
GestationalAge -0.3885357  0.1148913  -3.382  0.00072 ***
Toxemia        -1.3437865  0.6075033  -2.212  0.02697 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 286.14  on 222  degrees of freedom
Residual deviance: 203.71  on 219  degrees of freedom
AIC: 211.71

Number of Fisher Scoring iterations: 5
```

Logistic regression model

- β_1 has changed but remains significant (p-value $= 0.001$)
- β_2: a week increase in gestational age results in 0.389 decrease in log-odds of developing BPD for specific toxemia status
- β_3: the odds of developing BPD in infacts of mothers diagnosed with toxemia is 73.9% lower than the risk in infacts of mothers without toxemia for the same gestational age

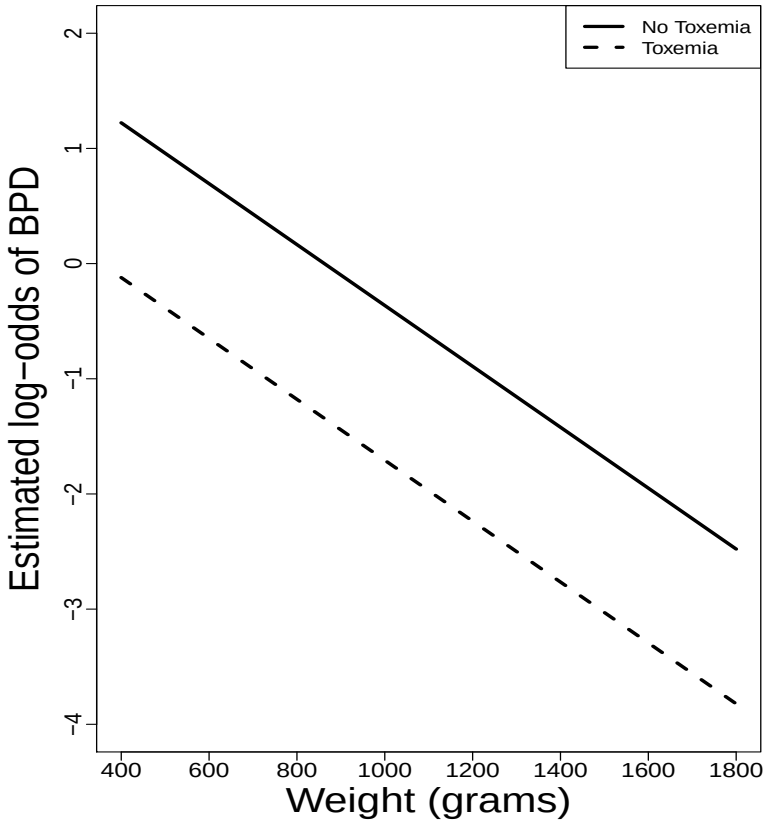
Logistic regression model

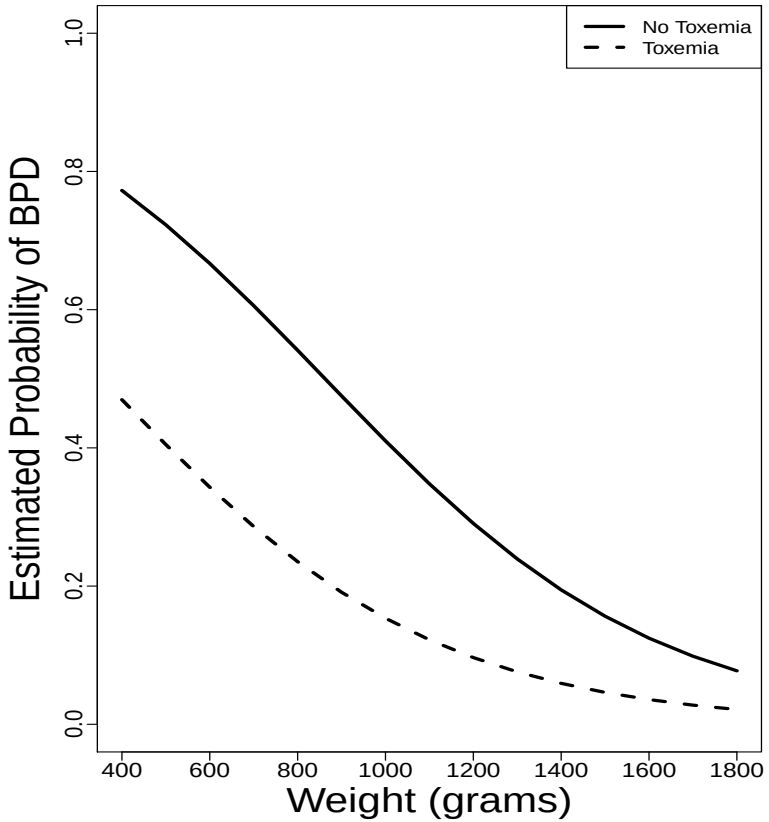
- How does the odds of developing BPD change in infacts born at 32 weeks if their mother is diagnosed with toxemia and are heavier by 200 grams compared to baby's with a mother without toxemia?

  $\Rightarrow$ odds-ratio $= 0.143$, the odds of developing BPD decreases by 85.7%

Logistic regression model - Parameter Interpretation

- For every weight value we obtain the plot of fitted log-odds and probabilities of BPD per toxemia adjusting for gestational age 30 weeks



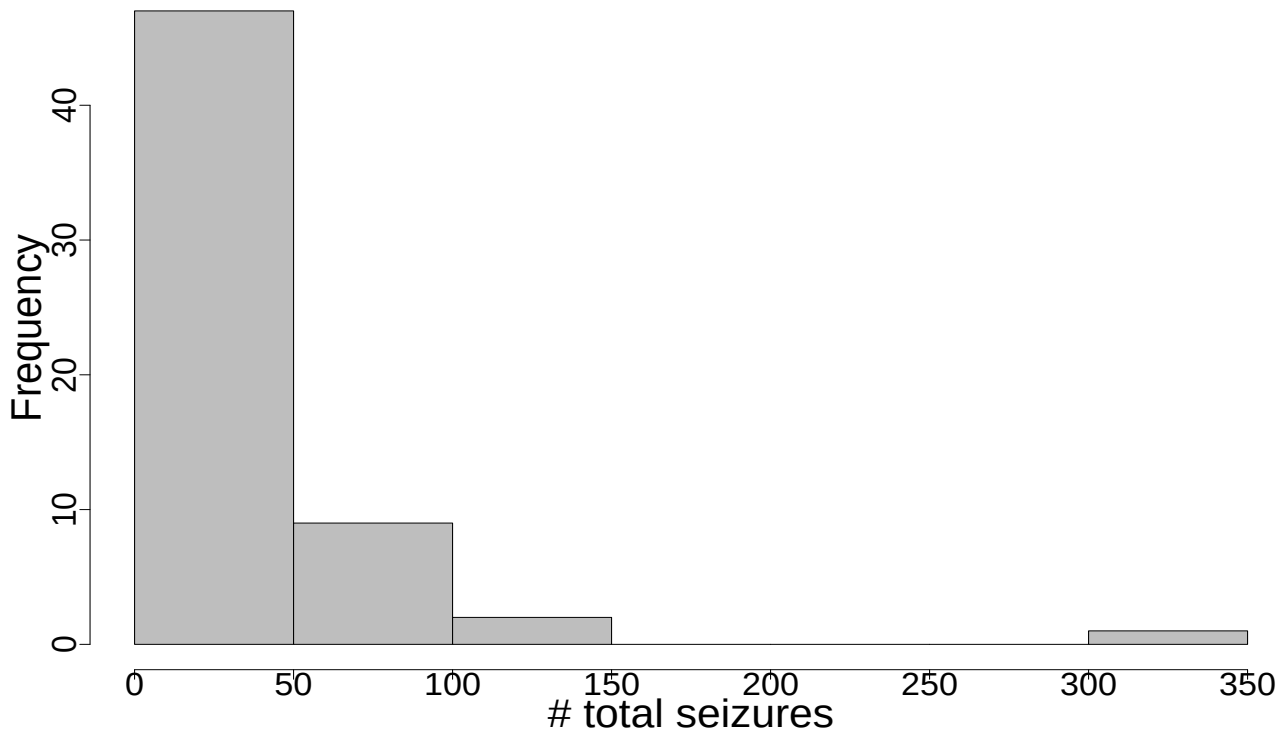


Poisson regression: Anti-epileptic drug study

- Reference: Thall, P.F. and Vail, S.C. (1990). Some covariance models for longitudinal count data with overdispersion. Biometrics, 46, 657-671.
- Randomized placebo-controlled study on seizure counts for 59 epileptics to compare placebo with the anti-epileptic drug progabide
- 28 patients in placebo, 31 in treatment group
- Measurements collected at:
  - Baseline: 8 weeks of baseline period, after which randomization takes place
  - Every 2 weeks up to 8 weeks of follow up
- Response: Total number of seizures a patient experienced after randomization i.e. the sum of all bi-weekly measurements

Example: Anti-epileptic drug study

- Research question: We want to test for a treatment effect on total number of seizures, correcting for the number of seizures during the 8-week baseline phase, prior to the treatment
- Frequency plot of total number of seizures, over all visits, over both treatment groups



Poisson regression model

- We will use Poisson regression, also known as log-linear regression model (when all covariates are discrete)
- Response variable Y: Counts (e.g. # events in a given period of time)
- Goal: to relate mean counts to covariates

Poisson regression model

- The number of events Y is assumed to have a Poisson distribution
- The probability a specific number of events y occurs is

$$Y \sim \text{Poisson}(\lambda), \quad \text{i.e. } Pr(y \text{ events}) = e^{-\lambda}\lambda^{y}/y!$$

  with λ the expected number of events

- Under the assumption of Poisson distribution both mean and variance equal λ i.e. it holds $\text{var}(Y) = \lambda = E(Y)$

Poisson regression model

- Say we use a naive strategy and consider a linear regression model for λ

$$\lambda = E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

- This model is not realistic $\Rightarrow$ λ represents counts that cannot be negative

- Instead we use the logarithmic transformation

$$\log(\lambda) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

- Linear relationship between logarithm of expected counts and covariates

- But non-linear relationship between expected counts and covariates

  $\Rightarrow$ Covariates act multiplicatively on the mean

$$\begin{aligned} \lambda &= \exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p\right) \\ \lambda &= \exp(\beta_0) \times \exp(\beta_1 X_1) \times \exp(\beta_2 X_2) \times \ldots \times \exp(\beta_p X_p) \end{aligned}$$

Poisson regression - Parameter interpretation

$$\boxed{\log(\lambda) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}$$

- A unit change in X_1 from x to $x+1$ (while all other covariates are held fixed) corresponds to

$$\begin{aligned} \log(\lambda) &= \beta_0 + \beta_1 x + \beta_2 X_2 + \ldots + \beta_p X_p \\ \log(\lambda) &= \beta_0 + \beta_1 (x+1) + \beta_2 X_2 + \ldots + \beta_p X_p \end{aligned}$$

- Thus,

$$\beta_1 = \log(\lambda) - \log(\lambda) = \log \frac{\lambda}{\lambda}$$

$$\exp(\beta_1) = \frac{\lambda}{\lambda}$$

Poisson regression - Parameter interpretation

- For binary X_1
  - $\Rightarrow$ β_1 is the log ratio of expected counts between the two levels of X_1 given that all other covariates remain constant
  - $\Rightarrow$ $\exp(\beta_1)$ is the ratio of expected counts between the two levels of X_1 given that all other covariates remain constant
- For continuous X_1
  - $\Rightarrow$ β_1 is change in log expected count for a unit change in X_1 given that all other covariates remain constant

Poisson regression - Parameter interpretation

- β_1 is called population slope

$$\beta_1 = \begin{cases} < 0, \text{ a unit increase in } X_1 \text{ decreases expected count by } \exp(\beta_1) \\ > 0, \text{ a unit increase in } X_1 \text{ increases expected count by } \exp(\beta_1) \end{cases}$$

- β_0: population intercept $\Rightarrow$ log expected count when all covariates are zero

- Changes in the expected counts from combined effects of different risk factors are multiplicative

Poisson regression model - Anti-epileptic drug study

- Example: We want to test for a treatment effect on number of seizures, correcting for the average number of seizures during the 8-week baseline phase, prior to the treatment

- The Poisson regression model is

$$\log(\lambda) = \beta_0 + \beta_1 \text{Treatment} + \beta_2 \text{BaselineRate} + \beta_3 \text{Age}$$

- We are interested in β_1

- In R we use:

```
model <- glm(TotalCounts ~ Treatment + BaseRate + Age, family =
poisson, data = Epilepsydata)
```

Poisson regression model - Anti-epileptic drug study

- glm output

```
> summary(model.11)

Call:
glm(formula = TotalCounts ~ Treatment + BaseRate + Age, family = poisson,
    data = data.epilepsy)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.8483  -2.1312  -1.0035   0.6926  10.9823

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.1544342  0.1198240  17.980  < 2e-16 ***
Treatment   -0.1928933  0.0464638  -4.151 3.30e-05 ***
BaseRate     0.0225023  0.0005121  43.945  < 2e-16 ***
Age          0.0162527  0.0035126   4.627 3.71e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2119.60  on 58  degrees of freedom
Residual deviance:  569.28  on 55  degrees of freedom
AIC: 860.71

Number of Fisher Scoring iterations: 5
```

- We observe strongly significant results

Poisson regression model

- Before reporting these results we have to check the model's assumptions
- The Poisson model assumes that the mean and variance are equal i.e. $E(Y) = \text{var}(Y) = \lambda$
  - $\Rightarrow$ When this assumption does not hold the Poisson regression does not provide a good fit to the data
  - $\Rightarrow$ The phenomenon is called overdispersion when $\text{var}(Y) > \lambda = E(Y)$
- Implications of overdispersion
  - Can lead to wrong inferences $\Rightarrow$ smaller std errors $\Rightarrow$ too narrow CI and too small p-values
  - Model selection strategies will behave poorly

Poisson regression model

- To evaluate goodness-of-fit we use deviance (compares model fit of saturated model to fitted model)

  $\Rightarrow$ Rule of thumb: for a good fitting model: Deviance $\backsimeq$ degrees of freedom

- The model does not fit the data well: Deviance/df $= 10.351$

  $\Rightarrow$ overdispersion

- We would expect Deviance/df $\simeq 1$ for a good fitting model

- We can correct for overdispersion assuming $\text{var}(Y) = \phi\lambda$ with ϕ the scale parameter

- The parameter estimates remain unchanged but the standard errors are adjusted

Poisson regression model - Anti-epileptic drug study

- We refit the model and correct for overdispersion
- In R we use:

```
model <- glm(TotalCounts ~ Treatment + BaseRate + Age, family =
quasipoisson, data = Epilepsydata)
```

- Parameter estimates remain unchaged but standard errors are adjusted

Poisson regression model - Anti-epileptic drug study

- glm output

```
> summary(model.12)
Call:
glm(formula = TotalCounts ~ Treatment + BaseRate + Age, family = quasipoisson,
    data = data.epilepsy)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.8483  -2.1312  -1.0035   0.6926  10.9823

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.154434   0.415910   5.180 3.25e-06 ***
Treatment   -0.192893   0.161276  -1.196    0.237
BaseRate     0.022502   0.001777  12.660  < 2e-16 ***
Age          0.016253   0.012192   1.333    0.188
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 12.04789)

    Null deviance: 2119.60  on 58  degrees of freedom
Residual deviance:  569.28  on 55  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

Poisson regression model - Anti-epileptic drug study

- The estimated Poisson regression (obtained using ML) is

$$\log(\lambda) = 2.154 - 0.193 \text{ Treatment} + 0.180 \text{ BaselineRate} + 0.016 \text{ Age}$$

- Note that the parameter estimates remain unchanged
- The coefficient for treatment is $\beta_1 = -0.193 \Rightarrow$ the expected number of seizures for patients in the anti-epileptic treatment group is a factor $\exp(\beta_1) = 0.825$ or 17.5% lower than in the placebo group

Poisson regression model - Anti-epileptic drug study

- Based on the Wald test there is no evidence of association between the treatment and mean count of seizures (p-value $= 0.232$)

- The expected number of seizures for a patient in the placebo group of age 30 with a baseline rate of 2 is
$\exp(2.154 - 0.193 \times 0 + 0.180 \times 2 + 0.016 \times 30) = \exp(2.994) = 19.965$

Poisson regression model

- Often count data are collected on patients followed for different time periods
  - For example we cannot compare number of seizures in a 4-week period with number of seizures in a 12-month period
- Thus we often use rates instead of counts to have meaningful basis for direct comparisons
- We build Poisson regression for the expected rate λ/t with t a relevant time period (e.g. number of years)

$$\log(\lambda/t) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

- Since $\log(\lambda/t) = \log(\lambda) - \log(t)$

$$\log(\lambda) = \log(t) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

where $\log(t)$ is called offset

- The "coefficient" associated with $\log(t)$ is fixed at 1 $\Rightarrow$ this is supported by all software

Outline

Repeated measures GLM

- They extend GLM to correlated data.
- The various methods presented in this course differ on the way they capture the within-subject correlations.

	Univariate	Repeated
Gaussian	LM	Marginal Model + LMM
Non-Gaussian	GLM	GEE + GLMM

Example: Muscatine Coronary Risk Factor study

- Reference: Woolson, R.F. and Clarke, W.R. (1984). Analysis of categorical incomplete longitudinal data. Journal of the Royal Statistical Society, Series A, 147, 87-99.

- Longitudinal study on obesity in 4856 school-aged children measured in 1977, 1979 and 1981:

ID	Gender	Baseline Age	Age	Obesity	Visit
1	0	6	6	1	1977
1	0	6	8	1	1979
1	0	6	10	1	1981
2	0	6	6	1	1977
2	0	6	8	1	1979
⋮	⋮	⋮	⋮	⋮	⋮

- Response: $Y = 1$ if obese and $Y = 0$ otherwise $\Rightarrow$ binary outcome.

- Covariates: Gender, age at baseline and age at each visit.

Example: Muscatine Coronary Risk Factor study

- Research questions:
  - Does prevalence of obesity increase with age?
  - Are patterns in change of obesity status the same for boys and girls?
- Percentage of children classified as obese in 1977, 1979 and 1981:

Gender	Age Cohort	1977	1979	1981
Males	5-7	7.90	15.40	21.20
	7-9	18.80	20.50	23.70
	9-11	21.20	22.70	22.50
	11-13	24.30	21.80	19.40
	13-15	19.20	21.10	18.20
Females	5-7	14.00	17.20	25.10
	7-9	16.50	24.00	24.90
	9-11	25.40	26.20	22.20
	11-13	23.80	22.10	19.90
	13-15	22.90	25.80	20.90

Example: Muscatine Coronary Risk Factor study

- Caution: Missing data for some children.
- Percentage missingness per gender and year:

Gender	Visit		
	1977	1979	1981
Male	32	30	36
Female	30	29	36

- Percentages computed using available data should be treated with caution:
  - Obesity risk increases from 1st cohort to 4th cohort (i.e. with age).
  - Obesity risk higher for girls than boys in all ages.

Marginal approach for non-gaussian repeated measurements

- General idea:
  1. Specify a model for the mean as you would do in a GLM

     $\Rightarrow$ there is also the time dimension.
  2. Make assumptions about the correlation between time points.

- The term "**marginal**" is used to distinguish these models from the GLMMs.
  - Marginal models make inferences about population averages

    $\Rightarrow$ parameters have the same interpretation as in cross-sectional analyses.
  - Mean and within-subject correlation are modelled separately

    $\Rightarrow$ The nature or magnitude of the correlation does not alter the parameter interpretation.
  - The model for mean response depends only on covariates and not on other responses or random effects.

Generalized Estimating Equations

- So far for LMMs and GLMs maximum likelihood method was used to estimate model parameters.
- For correlated data we need to specify the joint distribution of the repeated measurements.
- For repeated Gaussian data this joint distribution is the multivariate normal.

  $\Rightarrow$ extension of the univariate normal distribution to the multivariate case.
- For non-gaussian data extensions of the univariate distributions to the multivariate case not always straightforward.
- Thus, it is difficult to formulate the likelihood for a marginal model.
- To estimate such models we use Generalized Estimating Equations (GEE).
  (Liang and Zeger, 1986)

GEE approach for non-gaussian repeated measurements

- **GEE is not a likelihood-based approach** $\Rightarrow$ it is an estimation method.
- This has implications for:
  - Hypothesis testing

    $\Rightarrow$ Likelihood ratio test or score test not applicable.

    $\Rightarrow$ The Wald test can be used.
  - Performance when we have missing data.

GEE approach for non-gaussian repeated measurements

- No assumptions for the joint distribution of repeated measurements.
- Three components:
  1. Model for mean response $E(Y)$, e.g., binary data $\pi = E(Y)$:
  $$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p.$$
  2. Variance of $Y \Rightarrow$ follows from GLM assumption for each measurement e.g., binary data:
  $$\text{var}(Y) = \phi\pi(1-\pi)$$
  with ϕ scale parameter that models overdispersion.
  3. Pairwise correlations $\Rightarrow$ we make a "working" assumption that possibly depends on parameters to be estimated.
- Model for the variance and correlation is not supposed to be correct $\Rightarrow$

Interest is primarily in βs, the covariance structure is considered as "nuisance".

GEE approach for non-gaussian repeated measurements

- For a univariate GLM estimated by maximizing the (log) likelihood the score equations were:

$$S(\beta) \quad = \quad \sum_i \frac{\partial \mu_i}{\partial \beta} \; v_i^{-1} \; (Y_i - \mu_i) \; = \; 0$$

  with μ_i the mean of Y_i, v_i the variance of Y_i and i the subject indicator

- In GEE we replace Y_i with a vector of outcomes $\mathbf{Y}_i$, μ_i with the vector $\boldsymbol{\mu}_i$ and v_i by the working covariance matrix $\boldsymbol{\Sigma}_i$ of $\mathbf{Y}_i$.

GEE approach for non-gaussian repeated measurements

- The Generalized Estimating Equations become:

$$S(\beta) = \sum_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \beta} \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0$$

where $\boldsymbol{\mu}_i = E(\mathbf{Y}_i)$ and $\boldsymbol{\Sigma}_i = \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{A}_i^{1/2}$

in which $\boldsymbol{A}_i^{1/2} = \begin{pmatrix} \sqrt{\phi_1 v_{i1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\phi_{n_i} v_{in_i}} \end{pmatrix}$.

- $\boldsymbol{R}_i(\boldsymbol{\alpha})$ is the correlation matrix of $\mathbf{Y}_i$ which depends possibly on a vector $\boldsymbol{\alpha}$ of unknown parameters.

- A choice has to be made for $\boldsymbol{R}_i(\boldsymbol{\alpha})$.

Outline

Properties of $\hat{\boldsymbol{\beta}}$ in GEE

- $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$.
  - $\Rightarrow$ it is consistent even if within-subject associations are misspecified.
  - $\Rightarrow$ mean model needs to be correctly specified.
- If the model for the mean $\boldsymbol{\mu}_i$ is correctly specified:

$$\hat{\boldsymbol{\beta}} \text{ follow approximately } N_p(\boldsymbol{\beta}, \text{Var}(\hat{\boldsymbol{\beta}}))$$

where

$$\boxed{\text{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{V}_0^{-1}\boldsymbol{V}_1\boldsymbol{V}_0^{-1}}$$

with

$$\boldsymbol{V}_0 = \sum_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \quad \text{and} \quad \boldsymbol{V}_1 = \sum_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \boldsymbol{\Sigma}_i^{-1} \text{Var}(\mathbf{Y}_i) \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}.$$

Properties of GEE

- The estimator $\text{Var}(\hat{\boldsymbol{\beta}})$ is known as the **empirical or Sandwich or Robust estimator**.

$$\boxed{\text{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{V}_0^{-1}\boldsymbol{V}_1\boldsymbol{V}_0^{-1} \quad \Rightarrow \quad \text{Sandwich or Robust estimator}}$$

with

$$\boldsymbol{V}_0 = \sum_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \quad \text{and} \quad \boldsymbol{V}_1 = \sum_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \boldsymbol{\Sigma}_i^{-1} \text{Var}(\mathbf{Y}_i) \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}$$

where $\boldsymbol{\beta}$, and ϕ are replaced by their estimates and $\text{Var}(\mathbf{Y}_i) = (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)^T$.

- When $\boldsymbol{\Sigma}_i \equiv \text{Var}(\mathbf{Y}_i)$, i.e. when true variance of Y_i is known :

$$\boxed{\text{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{V}_0^{-1} \quad \Rightarrow \quad \text{Naive or Model-based estimator}}$$

Properties of GEE

- In practice the true variance of Y_i is not known and thus the naive standard errors are not correct.
- A plausible guess has to be made for $\boldsymbol{\Sigma}_i = \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{A}_i^{1/2}$.
- In fact we make an assumption for $\boldsymbol{R}_i(\boldsymbol{\alpha})$.
- Statistical software typically report both the naive and robust standard errors.
- The robust standard errors correct for a potential misspecification of $\boldsymbol{R}_i(\boldsymbol{\alpha})$.
- The asymptotic results suggest that, as long as interest is only in inferences for the mean structure, little effort should be spent in modeling the correlation structure, provided that the data set is sufficiently large.
- This robustness property of GEE makes it very popular for data analyses.

Properties of GEE

- **However, the sandwich estimator has also limitations.**
- The sandwich estimator works better:
  - in large samples
  - balanced designs
  - few repeated measurements and
  - few covariates.
- Otherwise the robust standard errors are biased downwards.
- Reliance on the model-based standard errors is advisable when the above conditions are not satisfied.
- The quality of the model-based standard errors depends on how close to the truth the chosen correlation matrix is.

Properties of GEE

- However, different choices of $\boldsymbol{R}(\boldsymbol{\alpha})$ can lead to substantially different estimated covariance matrices $\Rightarrow$ different statistical inference.

- Appropriate correlation modeling may still be of interest:
  - for the interpretation of random variation in data.
  - for gaining efficiency (i.e. decrease std errors).
  - in the presence of missing data, robust inference only valid under very severe assumptions about the underlying missingness process (see later).

Outline

Examples of a "working" correlation matrix

- We discuss options available in the R package `geepack`.
- Other software (e.g., SAS) support much more options.
- Independence
  - Working assumption: observations of the same individual are independent.
  - For 4 repeated measurements.

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

  - With unadjusted standard errors equivalent to fitting GLM on all measurements.

Examples of a "working" correlation matrix

- **Exchangeable or Compound symmetry**
  - Working assumption: for each subject all pairwise correlations are the same.
  - For 4 repeated measurements

$$\boldsymbol{R} = \begin{pmatrix} 1 & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & 1 & \alpha \\ \alpha & \alpha & \alpha & 1 \end{pmatrix}.$$

  - Not very realistic for longitudinal data $\Rightarrow$ we expect correlations to decrease in time.
  - Perhaps more realistic in multilevel designs e.g., data from students on the same classroom or from members of the same family.

Examples of a "working" correlation matrix

- **Autoregressive (AR-1)**
  - Working assumption: for each subject the correlation depends on the distance in time between two measurements $|t_j - t_k|$.
  - $\text{Corr}(Y_k, Y_j) = \alpha^{|t_j - t_k|}$, $-1 \leq \alpha \leq 1$

    $\Rightarrow$ Repeated measurements have a first-order autoregressive relationship.
  - For 4 repeated measurements collected at equal time points i.e. at $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$

$$\boldsymbol{R} = \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{pmatrix}.$$

  - Because $-1 \leq \alpha \leq 1$ the pairwise correlations decrease with distance in time between the repeated measurements.

Examples of a "working" correlation matrix

- **Autoregressive (AR-1)**
  - If $\alpha = 0.5$

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0.5 & 0.25 & 0.125 \\ 0.5 & 1 & 0.5 & 0.25 \\ 0.25 & 0.5 & 1 & 0.5 \\ 0.125 & 0.25 & 0.5 & 1 \end{pmatrix}.$$

  - If $\alpha = 0.9$

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0.9 & 0.81 & 0.729 \\ 0.9 & 1 & 0.9 & 0.81 \\ 0.81 & 0.9 & 1 & 0.9 \\ 0.729 & 0.81 & 0.9 & 1 \end{pmatrix}.$$

  - Appropriate when we have equally spaced time points at which measurements have been taken.

Examples of a "working" correlation matrix

- **Unstructured**
  - The correlation matrix is left completely free: $\text{Corr}(Y_k, Y_j) = \alpha_{jk}$ with $j \neq k$.
  - For p measurements per subject we have $p(p-1)/2$ correlation parameters.
  - For 4 repeated measurements:

$$\boldsymbol{R} = \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{12} & 1 & \alpha_{23} & \alpha_{24} \\ \alpha_{13} & \alpha_{23} & 1 & \alpha_{34} \\ \alpha_{14} & \alpha_{24} & \alpha_{34} & 1 \end{pmatrix}.$$

  - For long observation periods an unstructured working correlation would require estimation of many correlation parameters.

Examples of a "working" correlation matrix

- M-dependence or Toeplitz: often provided in other software
  - Working assumption:
    - Correlations k occasions apart are the same.
    - Correlations $> M$ occasions apart are zero.
  - For 4 repeated measurements: 1-dependence

$$\boldsymbol{R} = \begin{pmatrix} 1 & \alpha & 0 & 0 \\ \alpha & 1 & \alpha & 0 \\ 0 & \alpha & 1 & \alpha \\ 0 & 0 & \alpha & 1 \end{pmatrix}.$$

Examples of a "working" correlation matrix

- M-dependence or Toeplitz

- For 4 repeated measurements: 2-dependence

$$\boldsymbol{R} = \begin{pmatrix} 1 & \alpha_1 & \alpha_2 & 0 \\ \alpha_1 & 1 & \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_1 & 1 & \alpha_1 \\ 0 & \alpha_2 & \alpha_1 & 1 \end{pmatrix}.$$

Properties of GEE - How to choose $\boldsymbol{R}(\boldsymbol{\alpha})$

- Exploring the correlation structure is often complicated

  $\Rightarrow$ It is not always possible to test with statistical software which matrix fits best to the data.

- Though modifications of AIC measures have been proposed for GEE.

- Some tips:
  - When measurements are collected at the same time points per subject and you have a large sample the "unstructured" correlation matrix is the safest choice.
    - Let p measurements planned per subject, then in the "unstructured" matrix there are $p(p-1)/2$ elements to estimate.
    - If $p(p-1)/2 \gg n$, with n the sample size, then estimation will be unstable.
    - A simpler correlation matrix should be considered.

Properties of GEE - How to choose $\boldsymbol{R}(\boldsymbol{\alpha})$

- Some tips:
  - Choose working correlation that is clinically or biologically most plausible i.e. AR(1) or Toeplitz if realistic.
  - Make sure that the mean has been correctly specified.
- The goal of using the sandwich estimator is to correct std errors such that:
  - type I error is preserved $\Rightarrow$ but we may lose power.
- When some visits are missing $\Rightarrow$ complications discussed later.

Outline

Hypothesis testing

- The GEE approach is not a full likelihood approach $\Rightarrow$ Likelihood Ratio test and Score test cannot be used.
- Instead (generalized) Wald tests are available.
- To test a null hypothesis that involves more than one parameter, a multivariate Wald-type test statistic can be constructed.
- To test $H_0 : \mathbf{L}\beta = 0$, where β is a vector of p regression parameters and $\mathbf{L}$ is the contrasts matrix, the following statistic can be computed.

$$W = (\mathbf{L}\hat{\beta})^T (\mathbf{L}\text{Var}(\hat{\beta})\mathbf{L}^T)^{-1}\mathbf{L}\hat{\beta}$$

Under $H_0 : \mathbf{L}\beta = 0$, W has, approximately, a chi-squared distribution with p degrees of freedom.

Outline

Generalized Estimating Equations in R

- There are two main packages in R to run the GEE approach: `gee` and `geepack`.
- In this course we will use the function `geeglm` from the R package `geepack`.
- The main arguments are:
  - `formula`: An R formula specifying the response variable and the predictors.
  - `family`: a description of the error distribution and link function to be used in the model.
  - `id`: the variable denoting which measurements belong to the same group (e.g., to the same subject).
  - `data`: a data frame that contains all these variables (Note: for longitudinal data analyses, the rows must be ordered with respect to id, and the rows within the same id should be ordered with respect to time).
  - `corstr`: the assumed working correlation matrix: `"independence"`, `"exchangeable"`, `"ar1"`, `"unstructured"`, and `"userdefined"`.
- In R, you can specify your own working correlation matrix, `corstr = "userdefined"`.
- This can be useful when for instance the correlation structure is known e.g. in

GEE in R - Example: Muscatine data

- Model 1: We want to test if obesity increases with age $\Rightarrow$ longitudinal effect.
- We start with the model:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{age}_{ij}.$$

- Marginal probability of obesity is a logistic function of covariates: gender and age at each visit.
- Within-subject correlation:
  - Unstructured
  - Exchangeable
  - AR(1)
  - Independence

GEE in R - Example: Muscatine data / Model 1.1

- We need the data in long format $\Rightarrow$ one row per visit.

```
> library(geepack)
> data.mcrf.new <- data.mcrf[order(data.mcrf$ID, data.mcrf$Visit), ]
> model.11 <- geeglm(Obesity ~ Gender + AgeCat, data = data.mcrf.new,
+                        id = ID, family = binomial("logit"),
+                    corstr = "unstructured")
```

- We first fit the unstructured correlation matrix.

GEE - Muscatine data / Model 1.1

- geeglm output:

```
> summary(model.11)
Call:
geeglm(formula = Obesity ~ Gender + AgeCat, family = binomial("logit"),
    data = data.mcrf.new, id = ID, corstr = "unstructured")

 Coefficients:
             Estimate   Std.err    Wald Pr(>|W|)
(Intercept)  -1.837456  0.108956 284.402  < 2e-16 ***
GenderFemale  0.143719  0.062645   5.263   0.0218 *
AgeCat        0.040233  0.008221  23.953 9.87e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = unstructured
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)   0.9932 0.02606
  Link = identity

Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2   0.5794 0.01983
alpha.1:3   0.4506 0.03107
alpha.2:3   0.5574 0.03181
Number of clusters:   4856  Maximum cluster size: 3
```

GEE - Muscatine data / Model 1.1

- The estimated "working" correlation matrix has upper-diagonal entries:

```
> summary(model.11)$ geese$ correlation
          estimate  san.se  wald p
alpha.1:2   0.5794 0.01983 853.5 0
alpha.1:3   0.4506 0.03107 210.3 0
alpha.2:3   0.5574 0.03181 307.1 0
```

Outline

GEE - Muscatine data / Model 1.1

- The results with unstructured "working" correlation matrix:

```
> summary(model.11)$"coefficients"
             Estimate  Std.err    Wald  Pr(>|W|)
(Intercept)  -1.83746 0.108956 284.402 0.000e+00
GenderFemale  0.14372 0.062645   5.263 2.178e-02
AgeCat        0.04023 0.008221  23.953 9.873e-07
```

- The coefficient for the gender effect is $\beta_1 = 0.144$.
- This means that for children of the same age:
  - the log odds of obesity for girls is higher by 0.144 compared to boys.
  - the odds of obesity of girls is $\exp(\beta_1) = 1.155$

    $\Rightarrow$ girls 15.456% more likely to be classified as obese than boys.

GEE - Muscatine data / Model 1.1

- The coefficient for the age effect is $\beta_2 = 0.04$.
- This means that for children of the same gender:
  - for every year they grow the log odds of obesity increases by 0.04.
  - for every year they grow the odds ratio of obesity ($\exp(\beta_2) = 1.041$) increases by 4%.
- To study association of obesity risk with age we test the hypotheses:

$$\begin{aligned} H_0 &: \beta_2 = 0 \\ H_\alpha &: \beta_2 \neq 0 \end{aligned}$$

- The Wald test is implemented in `geepack`:
  - $\Rightarrow$ $W^2 = 23.953$ with 1 degree of freedom and p-value $= 0$.
  - $\Rightarrow$ the risk of obesity increases with age.

GEE - Muscatine data / Model 1.1

- Note that the βs have a marginal or "population-averaged" interpretation ⇒ Inferences can be extrapolated to the whole population from which the sample was taken:
  - We compare subgroups of patients e.g. males versus females ⇒ cross-sectional effect.
  - We test also within subject-effects i.e. age effects ⇒ longitudinal effect.
- This is in contrast to the GLMM in which interest lies in subject-specific effects (we will see later).

GEE - Muscatine data / Model 1.1

- The estimated "working" correlation matrix has upper-diagonal entries:

```
> summary(model.11)$ geese$ correlation
          estimate  san.se  wald p
alpha.1:2   0.5794 0.01983 853.5 0
alpha.1:3   0.4506 0.03107 210.3 0
alpha.2:3   0.5574 0.03181 307.1 0
```

- We notice that the pairwise correlations are close $\Rightarrow$ the exchangeable or AR(1) seems a reasonable choice.

GEE - Muscatine data / Model 1.2

- We re-do the analysis using the exchangeable matrix:
- geeglm output:

```
> model.12 <- geeglm(Obesity ~ Gender + AgeCat, data = data.mcrf.new,
+                       id = ID, family = binomial("logit"),
+                    corstr = "exchangeable")
> summary(model.12)
Call:
geeglm(formula = Obesity ~ Gender + AgeCat, family = binomial("logit"),
    data = data.mcrf.new, id = ID, corstr = "exchangeable")

 Coefficients:
             Estimate  Std.err   Wald Pr(>|W|)
(Intercept)  -1.82297  0.10881 280.68   <2e-16 ***
GenderFemale  0.15052  0.06263   5.78    0.016 *
AgeCat        0.03906  0.00821  22.61    2e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.991  0.0258
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha     0.54  0.0202
Number of clusters:   4856  Maximum cluster size: 3
```

GEE - Muscatine data / Model 1.3

- We re-do the analysis using the AR(1) matrix.
- `geeglm` output:

```
> model.13 <- geeglm(Obesity ~ Gender + AgeCat, data = data.mcrf.new,
+                       id = ID, family = binomial("logit"),
+                    corstr = "ar1")
> summary(model.13)
Call:
geeglm(formula = Obesity ~ Gender + AgeCat, family = binomial("logit"),
    data = data.mcrf.new, id = ID, corstr = "ar1")

 Coefficients:
             Estimate  Std.err   Wald Pr(>|W|)
(Intercept)  -1.84530  0.10982 282.32  < 2e-16 ***
GenderFemale  0.13381  0.06286   4.53    0.033 *
AgeCat        0.04087  0.00829  24.29  8.3e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = ar1
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.996  0.0266
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.607  0.0197
Number of clusters:   4856  Maximum cluster size: 3
```

GEE - Muscatine data / Model 1.3

- The estimated "working" correlation matrix is:

```
> summary(model.13)$ geese$ correlation
      estimate san.se wald p
alpha    0.607 0.0197  952 0
```

GEE - Muscatine data / Model 1.4

- We re-do the analysis using the independence matrix.
- geeglm output:

```
> model.14 <- geeglm(Obesity ~ Gender + AgeCat, data = data.mcrf.new,
+                       id = ID, family = binomial("logit"),
+                    corstr = "independence")
> summary(model.14)

Call:
geeglm(formula = Obesity ~ Gender + AgeCat, family = binomial("logit"),
    data = data.mcrf.new, id = ID, corstr = "independence")

 Coefficients:
             Estimate  Std.err   Wald Pr(>|W|)
(Intercept)  -1.68949  0.12383 186.15   <2e-16 ***
GenderFemale  0.12576  0.06486   3.76   0.0525 .
AgeCat        0.02734  0.00962   8.07   0.0045 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = independence
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)        1   0.027
Number of clusters:   4856  Maximum cluster size: 3
```

- The independence matrix influences more the estimates and the gender effect gets now bordeline non-significant.

GEE - Example: Muscatine data / Model 2

- Model 2: We want to test if changes in obesity differ between girls and boys $\Rightarrow$ interaction term.
- We start with the model:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{age}_{ij} + \beta_3 \text{gender}_i * \text{age}_{ij}$$

- Marginal probability of obesity is a logistic function of covariates: gender, age at each visit and their interaction.
- Within-subject correlation: Unstructured correlation matrix.

GEE - Muscatine data / Model 2

- geeglm output:

```
> model.2 <- geeglm(Obesity ~ Gender * AgeCat, data = data.mcrf.new,
+                       id = ID, family = binomial("logit"),
+                   corstr = "unstructured")
> summary(model.2)

Call:
geeglm(formula = Obesity ~ Gender * AgeCat, family = binomial("logit"),
    data = data.mcrf.new, id = ID, corstr = "unstructured")

 Coefficients:
                   Estimate  Std.err   Wald Pr(>|W|)
(Intercept)        -1.81495  0.14914 148.09   <2e-16 ***
GenderFemale        0.09927  0.20909   0.23   0.6349
AgeCat              0.03837  0.01179  10.59   0.0011 **
GenderFemale:AgeCat 0.00368  0.01645   0.05   0.8232
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = unstructured
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.993   0.026
  Link = identity

Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2    0.579  0.0198
alpha.1:3    0.451  0.0311
alpha.2:3    0.558  0.0318
Number of clusters:   4856  Maximum cluster size: 3
```

GEE - Muscatine data / Model 2

- For children at the age of 10:
  - the log odds of obesity for girls is higher by 0.136 compared to boys.
  - the odds ratio of obesity between girls and boys is $\exp(\beta_1 + 10\beta_3) = 1.146$.
    $\Rightarrow$ girls 14.572% more likely to be classified as obese than boys.

GEE - Example: Muscatine data / Model 3

- Often polynomial effects need to be included e.g., quadratic age effects.
- We want to study if changes of risk for obesity with age differ between girls and boys.
- Model 3: We start with the model

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{age}_{ij} + \beta_3 \text{age}_{ij}^2 + \beta_4 \text{gender}_i * \text{age}_{ij} + \beta_5 \text{gender}_i * \text{age}_{ij}^2.$$

- Marginal probability of obesity is a logistic function of covariates: we assume quadratic age trend that differ between sexes.
- We choose unstructured correlation matrix.

GEE - Muscatine data / Model 3

- We want to explore whether males and females have different average longitudinal evolutions.
- The hypothesis to be tested is:

$$\begin{aligned} H_0 &: \quad \beta_4 = \beta_5 = 0 \\ H_\alpha &: \quad \text{at least one} \neq 0 \end{aligned}$$

- To test such a hypothesis we need the multivariate Wald test.
- It is implemented in the R package aod.

GEE - Muscatine data / Model 3

- geeglm output:

```
> model.3 <- geeglm(Obesity ~ Gender * (I(AgeCat) + I(AgeCat^2)), data = data.mcrf.new,
+                       id = ID, family = binomial("logit"),
+                   corstr = "unstructured")
> summary(model.3)
Call:
geeglm(formula = Obesity ~ Gender * (I(AgeCat) + I(AgeCat^2)),
    family = binomial("logit"), data = data.mcrf.new, id = ID,
    corstr = "unstructured")

 Coefficients:
                        Estimate  Std.err Wald Pr(>|W|)
(Intercept)             -4.27211  0.49038 75.9  < 2e-16 ***
GenderFemale             0.60888  0.68124  0.8     0.37
I(AgeCat)                0.47032  0.08213 32.8  1.0e-08 ***
I(AgeCat^2)             -0.01793  0.00339 28.0  1.2e-07 ***
GenderFemale:I(AgeCat)  -0.08792  0.11337  0.6     0.44
GenderFemale:I(AgeCat^2) 0.00388  0.00464  0.7     0.40
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = unstructured
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.992  0.0275
  Link = identity

Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2    0.579  0.0204
alpha.1:3    0.462  0.0319
alpha.2:3    0.559  0.0321
```

GEE - Muscatine data / Model 3

- Multivariate Wald test:

```
> library(aod)
> t1 <- wald.test(b = model.3$"coefficients",
+                 Sigma = model.3$"geese"$"vbeta",
+                 L = rbind(c(rep(0, 4), 1, 0), c(rep(0, 5), 1)))
```

- The test produces a Wald statistic 0.795 with 2 degrees of freedom $\Rightarrow$ p-value $= 0.672$.

- There is no evidence that the average evolutions have a different shape between males and females (interaction terms not significant).

Outline

Example: Leprosy study

- Reference: Snedecor, G.W. and Cochran, W.G. (1967). Statistical Methods, (6th edn). Ames, Iowa: Iowa State University Press

- Randomized trial for effect of two antibiotics (drug A and B) and placebo (drug C) on treatment of leprosy bacilli in a sample of 30 patients

ID	Drug	Pre-Treatment Counts	Post-Treatment Counts
1	A	11	6
2	B	6	0
3	C	16	13
4	A	8	0
5	B	6	2
⋮	⋮	⋮	⋮

- Response: Y = number of leprosy bacilli in six sites of the body $\Rightarrow$ counts

- Covariate: Treatment group and time period

- Research question: Do drugs A and B reduce abundance of leprosy bacilli versus placebo?

Example: Leprosy study

- Compute mean and variance per treatment group and period

Drug	Pre-Treatment (mean/var)	Post-Treatment (mean/var)
A	9.30 (22.68)	5.30 (21.57)
B	10.00 (27.56)	6.10 (37.88)
C	12.90 (15.66)	12.30 (51.12)

- We observe greater variances than means
- Under Poisson distribution assumption $\Rightarrow$ overdispersion

GEE - Example 10: Leprosy data / Model 1

- Model 1: We want to test if treatment with drug A and B reduces the abundance of leprosy bacilli compared to drug C
- We start with the model for the mean counts: Poisson regression

$$\log E(Y) = \log \lambda = \beta_0 + \beta_1 \text{time} + \beta_2 \text{time drug}_A + \beta_3 \text{time drug}_B$$

where

$\text{drug}_A = 1$ if patients takes drug A, otherwise 0

$\text{drug}_B = 1$ if patients takes drug B, otherwise 0

$\text{time} = 0$ is the pre-treatment period and $\text{time} = 1$ is the post-treatment period

- Note that main effects of drug_A and drug_B are excluded $\Rightarrow$ randomized study
- Expected number of leprosy bacilli is a logarithmic function of covariates: drug type and treatment period

GEE - Example: Leprosy data / Model 1

- Within-subject correlation:
  - Unstructured correlation matrix or exchangeable $\Rightarrow$ coincide only 2 repeated measurements
- R code: Model 1

```
> library(geepack)
> model.11 <- geeglm(Counts ~ Time + (DrugA + DrugB):Time, data = data.,
+                    id = ID, family = poisson("log"),
+                    corstr = "exchangeable")
>
```

GEE - Leprosy data / Model 1

- geeglm output

```
> summary(model.11)
Call:
geeglm(formula = Counts ~ Time + (DrugA + DrugB):Time, family = poisson("log"),
    data = data., id = ID, corstr = "exchangeable")

 Coefficients:
            Estimate  Std.err   Wald Pr(>|W|)
(Intercept)  2.37335  0.08014 877.10   <2e-16 ***
Time        -0.00288  0.15701   0.00    0.985
Time:DrugA  -0.56257  0.22198   6.42    0.011 *
Time:DrugB  -0.49528  0.23420   4.47    0.034 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)     3.21     0.5
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.738  0.0815
Number of clusters:   30  Maximum cluster size: 2
```

- The estimated Poisson regression is

$$\log E(Y) = \log \lambda = 2.373 - 0.003\text{time} - 0.563\text{time} \times \text{drug}_A - 0.495\text{time} \times \text{drug}_B$$

GEE - Leprosy data / Model 1

- For drug C
  - $\exp(2.373) = 10.73$: expected counts of leprocy bacilli pre-treatment
  - $\exp(2.373 - 0.003) = 10.70$: expected counts of leprocy bacilli post-treatment
  - $\exp(-0.003) = 0.997$: rate ratio of leprocy bacilli between post-treatment and pre-treatment period
- For drug A
  - expected counts of leprocy bacilli pre-treatment $\exp(2.373) = 10.73 \Rightarrow$ due to randomization
  - $\exp(2.373 - 0.003 - 0.563) = 6.092$: expected counts of leprocy bacilli post-treatment
  - $\exp(-0.003 - 0.563) = 0.568$: rate ratio of leprocy bacilli between post-treatment and pre-treatment period
  - $\exp(-0.563) = 0.569$: rate ratio of leprocy bacilli at post-treatment between drug A and C $\Rightarrow$ p-value $= 0.011$
- Similarly for drug B

GEE - Leprosy data / Model 1

- To test if drugs A, B and C are equally effective we test

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_1 : \text{at least one } \neq 0$$

using multivariate Wald test:

```
> library(aod)
> t1 <- wald.test(b = model.11$"coefficients",
+                 Sigma = model.11$"geese"$"vbeta",
+                 L = rbind(c(rep(0, 2), 1, 0), c(rep(0, 3), 1)))
```

- The test produces a Wald statistic 7.34 with 2 degrees of freedom $\Rightarrow$ p-value $= 0.025$
- The β have a marginal or "population-averaged" interpretation
  - We compare subgroups of patients e.g. patients in treatment A vesus C $\Rightarrow$ cross-sectional effect
  - We test also within subject effects i.e. pre- and post- treatment effects $\Rightarrow$ longitudinal effect
- This is in contrast to the GLMM in which the subject-specific odds ratio is modeled

GEE - Leprosy data / Model 1

- The estimated "working" correlation matrix is

```
> summary(model.11)$ geese$ correlation
      estimate san.se wald p
alpha    0.738 0.0815 82.1 0
```

- The pairwise correlation is 0.738 $\Rightarrow$ substantial among patients

GEE for normal data

- GEE score equations can be derived for normal data, as well.
- For the the known correlation matrices, parameter estimates and model-based standard errors coincide those of the linear regression model with correlated errors (or marginal model).
- The use of the robust estimator can protect from violations of the covariance matrix and violations of the normality assumption.

Chapter 6: Missing Data in Longitudinal Studies

Master Statistics and Data Science

Outline

Outline

Missing Data in Longitudinal Studies

- A major challenge for the analysis of longitudinal data is the problem of missing data:
  - studies are designed to collect data on every subject in the sample at a set of pre-specified follow-up times.
  - often subjects miss some of their planned measurements for a variety of reasons.
- We can have different patterns of missing data.

Missing Data in Longitudinal Studies

Subject	Visits				
	1	2	3	4	5
1	x	x	x	x	x
2	x	x	x	?	?
3	?	x	x	x	x
4	?	x	?	x	?

- Subject 1: Completer
- Subject 2: dropout
- Subject 3: late entry
- Subject 4: intermittent

- In observational studies, it is difficult to define missingness: instead visiting process.

Missing Data in Longitudinal Studies

- Implications of missingness:
  - not all subjects have the same number of measurements ⇒ unbalanced datasets.
  - we collect less data than originally planned ⇒ loss of efficiency.
  - missingness may depend on outcome ⇒ potential bias.

Missingness cannot be ignored and should be taken into account in the analysis.

Missing Data in Longitudinal Studies

- For the handling of missing data, we introduce the missing data indicator:

$$r_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$$

- We obtain a partition of the **complete response vector**; $\mathbf{y}_i$
  - observed data $\mathbf{y}_i^o$, containing those y_{ij} for which $r_{ij} = 1$.
  - missing data $\mathbf{y}_i^m$, containing those y_{ij} for which $r_{ij} = 0$.

- If there are no missing data, we observe the complete response vector $\mathbf{y}_i$.

- Otherwise, we get to observe $\mathbf{y}_i^o$.

- In case of dropout we define the dropout indicator:

$$D_i = 1 + \sum_{j=1}^{n_i} r_{ij}$$

  D_i is the time of dropout, i.e., the first occasion for which no observation is available.

- Completers have observed value $d_i = n_i + 1$.

Outline

Missing Data Mechanisms

- To describe the probabilistic relation between the measurement and missingness processes Rubin (1976, Biometrika) has introduced three mechanisms.
- Missing Completely At Random (MCAR): The probability that responses are missing is unrelated to both $\mathbf{y}_i^o$ and $\mathbf{y}_i^m$

$$p(\mathbf{r}_i \mid \mathbf{y}_i^o, \mathbf{y}_i^m) = p(\mathbf{r}_i)$$

- Examples:
  - patients lost to follow up because of relocation.
  - laboratory measurements are lost due to equipment malfunction.

Missing Data Mechanisms

- Features of MCAR:
- The observed data $\mathbf{y}_i^o$ can be considered a random sample of the complete data $\mathbf{y}_i$.
- We can use any statistical procedure that is valid for complete data:
  * sample averages per time point.
  * t-test at the last time point.
  * GEE gives valid inferences.

Missing Data Mechanisms

- Missing At Random (MAR): The probability that responses are missing is related to $\mathbf{y}_i^o$, but is unrelated to $\mathbf{y}_i^m$

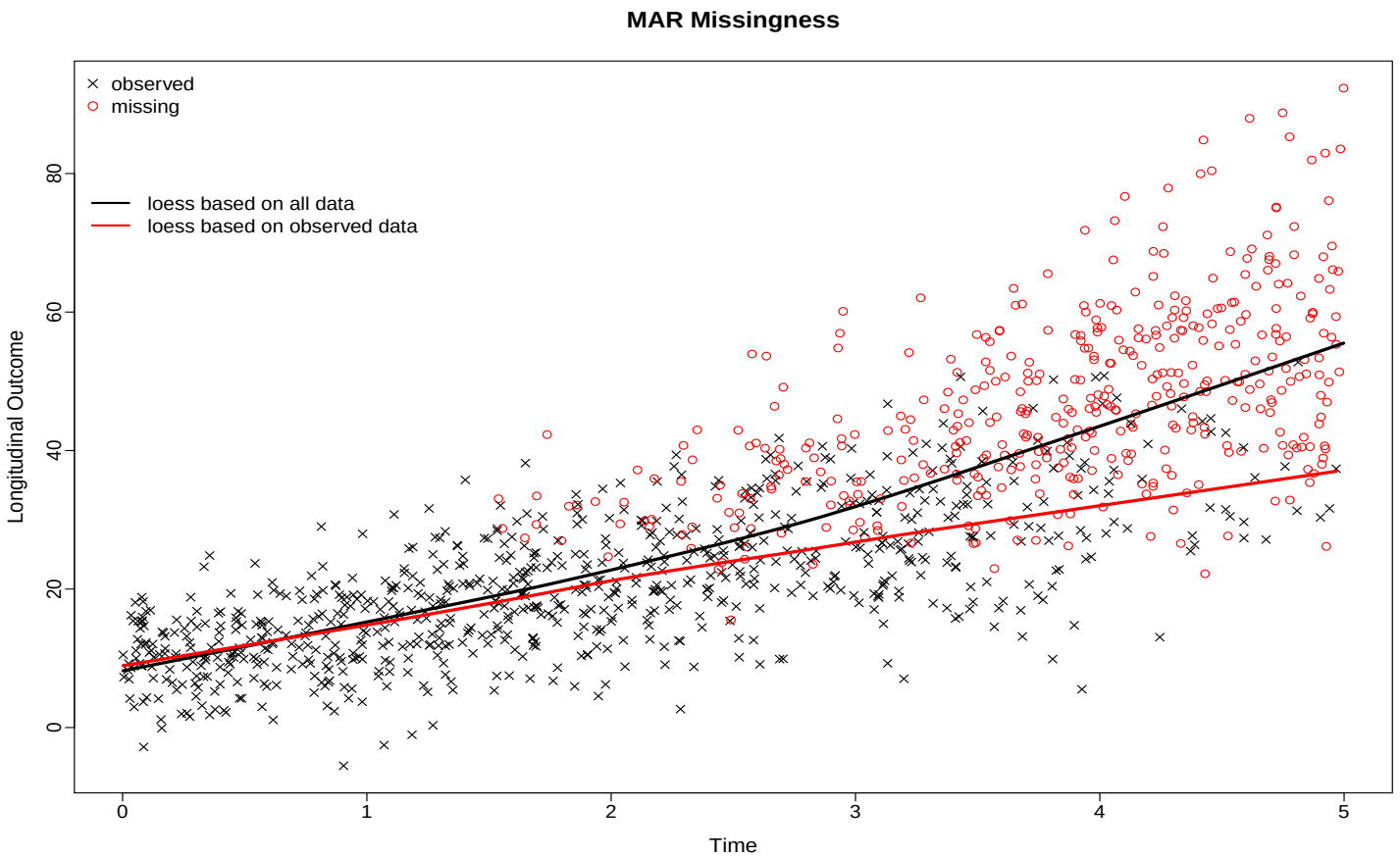
$$p(\mathbf{r}_i \mid \mathbf{y}_i^o, \mathbf{y}_i^m) = p(\mathbf{r}_i \mid \mathbf{y}_i^o)$$

- Examples:
  - study protocol requires patients whose response value exceeds a threshold to be removed from the study.
  - physicians give rescue medication to patients who do not respond to treatment.

Missing Data Mechanisms

- Features of MAR:
  - The observed data cannot be considered a random sample from the target population.
  - Not all statistical procedures provide valid results.
  - GEE is biased under MAR $\Rightarrow$ to derive unbiased estimates Weighted GEE is used (we will not discuss this further).
  - Likelihood based models for correlated data e.g., mixed models give valid results.

Missing Data Mechanisms



Missing Data Mechanisms

- Under MAR, the likelihood contribution of subject i equals

$$f(\mathbf{y}_i^o, \mathbf{r}_i \mid \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) \quad = \quad f(\mathbf{y}_i^o \mid \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{r}_i \mid \mathbf{y}_i^o, \mathbf{X}_i, \boldsymbol{\psi}).$$

- The measurement model $f(\mathbf{y}_i^o \mid \mathbf{X}_i, \boldsymbol{\theta})$ and the missingness model $f(\mathbf{r}_i \mid \mathbf{y}_i^o, \mathbf{X}_i, \boldsymbol{\psi})$ can be fitted completely separately, provided that the parameters in both parts are completely separated.

- Thus, under MAR, and with likelihood-based inference, the missingness model can be ignored completely (unless interest is also in $f(\mathbf{r}_i \mid \mathbf{y}_i^o)$).

- Note that this also holds under MCAR. Hence, under MCAR and MAR, with likelihood-based inference, the missingness process is **ignorable**.

Likelihood based methods give valid results, under MAR, provided the correct model was fitted.

Missing Data Mechanisms

- Missing Not At Random (MNAR): The probability that responses are missing is related to $\mathbf{y}_i^m$, and possibly also to $\mathbf{y}_i^o$

$$p(\mathbf{r}_i \mid \mathbf{y}_i^m) \quad \text{or} \quad p(\mathbf{r}_i \mid \mathbf{y}_i^o, \mathbf{y}_i^m)$$

- Examples:
  - in studies on drug addicts, people who return to drugs are less likely than others to report their status.
  - in longitudinal studies for quality-of-life, patients may fail to complete the questionnaire at occasions when their quality-of-life is compromised.

Missing Data Mechanisms

- Features of MNAR:
  - The observed data cannot be considered a random sample from the target population.
  - Only procedures that explicitly model the joint distribution $\{\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{r}_i\}$ provide valid inferences $\Rightarrow$ analyses which are valid under MAR will not be valid under MNAR.
  - GEE gives biased results in this case.
  - Correct analysis under MNAR cannot be implemented with standard software (except shared parameter models, e.g., R package JM).
  - Usually, a sensitivity analysis is recommended.

Missing Data Mechanisms: Summary

- Missing data are often not under the control of the investigator.
- Missing data mechanism can be: MCAR/MAR/MNAR.
- Analysis with methods that are valid under MAR must be done.
- Sensitivity analysis is recommended.

Outline

Complete Cases Analysis

- General idea: Restrict analyses to only those subjects for which all measurements are observed.
- Examples: Repeated measures ANOVA or paired t-test.
- Advantages:
  * very simple to implement.
  * standard software can be used for further analyses, e.g., mixed models, GEE, Repeated measures ANOVA or paired t-test, etc.
- Disadvantages:
  * substantial loss of information.
  * valid inferences only when missingness is completely unrelated to the outcome (i.e., MCAR).

Methods that use all available data and make less strict assumptions about missingness should be prefered.

Last Observation Carried Forward (LOCF)

- General idea: Any missing value is replaced by the last observed value.
- Advantages:
  * very simple to implement.
  * standard software can be used for further analyses, e.g. mixed modedls, GEE, etc.
- Disadvantages:
  * extremely strong assumption that a subject's measurement stays at the same level as soon as he/she is not observed.
  * even if the mechanism is MCAR, LOCF may not provide valid results.
  * overestimates precision.

Unconditional Mean Imputation

- General idea: Each missing outcome y_{ij}^m is replaced by the average of the observed measurements at the j-th occasion.

- Advantages:
  * very simple to implement.
  * standard software can be used.

- Disadvantages:
  * can only be implemented with balanced designs.
  * it provides valid results only under MCAR.
  * overestimates precision.

Conditional Mean Imputation

- General idea: The vector $\mathbf{y}_i^m$ of missing observations for the i-th subject is replaced by its prediction, conditional on the vector $\mathbf{y}_i^o$ of observed observations for that subject.
  - we specify a model for $\mathbf{y}_i^m$ conditional on $\mathbf{y}_i^o$ and parameters $\boldsymbol{\theta}$.
  - we fit the model to the completers and obtain estimates $\widehat{\boldsymbol{\theta}}$ for the parameters.
  - based on this fitted model we can calculate predictions for the missing observations, i.e.,

$$\widehat{\mathbf{y}}_i^m = E(\mathbf{y}_i^m \mid \mathbf{y}_i^o, \widehat{\boldsymbol{\theta}}).$$

Example of Conditional Mean Imputation

- Let $Y_i = (Y_i^o, Y_i^m)$ follows a normal distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{V}$:

$$\begin{pmatrix} Y_i^o \\ Y_i^m \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}^o \\ \boldsymbol{\mu}^m \end{pmatrix}, \begin{pmatrix} \mathbf{V}^{oo} & \mathbf{V}^{om} \\ \mathbf{V}^{mo} & \mathbf{V}^{mm} \end{pmatrix}\right].$$

- It follows that:

$$Y_i^m \mid Y_i^o = \mathbf{y}_i^o \sim N\left[\boldsymbol{\mu}^m + \mathbf{V}^{mo}(\mathbf{V}^{oo})^{-1}(\mathbf{y}_i^o - \boldsymbol{\mu}^o), \mathbf{V}^{mm} - \mathbf{V}^{mo}(\mathbf{V}^{oo})^{-1}\mathbf{V}^{om}\right].$$

- The vector $\boldsymbol{\theta}$ consists of all regression parameters in the mean $\boldsymbol{\mu}^m + \mathbf{V}^{mo}(\mathbf{V}^{oo})^{-1}(\mathbf{y}_i^o - \boldsymbol{\mu}^o)$ and all parameters in $\mathbf{V}^{mm} - \mathbf{V}^{mo}(\mathbf{V}^{oo})^{-1}\mathbf{V}^{om}$.

Example of Conditional Mean Imputation

- $\boldsymbol{\theta}$ is estimated by fitting the regression model

$$Y_i^m = \left[\boldsymbol{\mu}^m - V^{mo}(V^{oo})^{-1}\boldsymbol{\mu}^o\right] + \mathbf{V}^{mo}(V^{oo})^{-1}\mathbf{y}_i^o + \boldsymbol{\varepsilon}_i^m$$

  to the completers.

- The missing observations $\mathbf{y}_i^m$ are then replaced by their predictions

$$\widehat{\mathbf{y}}_i^m = \left[\boldsymbol{\mu}^m - \mathbf{V}^{mo}(\mathbf{V}^{oo})^{-1}\boldsymbol{\mu}^o\right] + \mathbf{V}^{mo}(\mathbf{V}^{oo})^{-1}\mathbf{y}_i^o$$

  where we plug in the estimates $\widehat{\boldsymbol{\theta}}$.

Conditional Mean Imputation

- Advantages:
  - less strict assumptions that the previously mentioned approaches.
- Disadvantages:
  - requires programming for its implementation.
  - overestimates precision.

Multiple Imputation

- A common issue in the previous imputation techniques is the overestimation of precision $\Rightarrow$ no correction was made for the uncertainty introduced from imputing the missing observations.

- General idea: To account for this uncertainty we impute not only once but multiple times from the conditional distribution $f(\mathbf{y}_i^m \mid \mathbf{y}_i^o, \widehat{\boldsymbol{\theta}})$:
  * M completed datasets are formed.
  * we perform the same analysis in each.
  * we pool the estimated parameters using Rubin's formula's.

Multiple Imputation

- Suppose M completed data sets are then obtained.
- The analysis of each data set yields an estimate $\widehat{\beta}^{(\nu)}$, $\nu = 1, \ldots, M$.
- The variation in the $\widehat{\beta}^{(\nu)}$ clearly reflects the variability introduced from imputing the missing observations.
- Note that there is also variability introduced from replacing $\boldsymbol{\theta}$ in $f(Y_i^m \mid Y_i^o, \boldsymbol{\theta})$ by an estimate.
- However, we usually have an estimate for the variation in $\widehat{\boldsymbol{\theta}}$:

$$\widehat{\boldsymbol{\theta}} \approx N(\boldsymbol{\theta}, \widehat{\Sigma}_{\boldsymbol{\theta}}).$$

- Drawing $\boldsymbol{\theta}$ from $N(\widehat{\boldsymbol{\theta}}, \widehat{\Sigma}_{\boldsymbol{\theta}})$ accounts for this additional variation.

Multiple Imputation

- The imputation algorithm is as follows:
  1. Draw $\boldsymbol{\theta}^{(\nu)}$ from $N(\widehat{\boldsymbol{\theta}}, \widehat{\Sigma}_{\boldsymbol{\theta}})$.
  2. **Imputation step:** Draw $\mathbf{y}_i^{m(\nu)}$ from $f(Y_i^m \mid Y_i^o = \mathbf{y}_i^o, \boldsymbol{\theta}^{(\nu)})$.
  3. **Analysis step:** Using the completed data $(\mathbf{y}_i^o, \mathbf{y}_i^{m(\nu)})$, calculate an estimate $\widehat{\beta}^{(\nu)}$ for the parameter β of interest, as well as its covariance matrix $U^{(\nu)}$.

     Repeat steps 1 - 3 M times.
- $U^{(\nu)}$ reflects the sampling uncertainty, i.e., the uncertainty in the estimates of β due to the fact that only a finite sample is available.

Multiple Imputation

- We can now obtain inferences for β from pooling the estimates:

$$\boxed{\widehat{\beta} = \frac{1}{M}\sum_{\nu=1}^{M}\widehat{\beta}^{(\nu)}}$$

- The covariance matrix of $\widehat{\beta}$ equals:

$$\boxed{\text{var}(\widehat{\beta}) = \widehat{W} + \left(\frac{M+1}{M}\right)\widehat{B}}$$

with $\widehat{W} = \dfrac{\sum_{\nu=1}^{M} U^{(\nu)}}{M}$ and $\widehat{B} = \dfrac{\sum_{\nu=1}^{M}(\widehat{\beta}^{(\nu)} - \widehat{\beta})(\widehat{\beta}^{(\nu)} - \widehat{\beta})'}{M-1}$.

- $\widehat{W}$ represents the within-imputation variance, representing sampling uncertainty.
- $\widehat{B}$ represents the between-imputation variance, representing the uncertainty in imputing the missing observations as well as the uncertainty in the estimation of β.
- Typically, M will be small: $M = 2, 3$ already yields a major improvement over single imputation.

Multiple Imputation

- Advantages:
  - correctly propagates uncertainty due to incomplete data.
  - valid under MAR.
  - imputation model and analysis model do not need to be the same: allows for different types of analysis (e.g., concentrate at a specific time point – cross-sectional analysis).
- Disadvantages:
  - not available for grouped/clustered data in all software.

Multiple Imputation

- Multiple Imputation has been proposed in the '90s when mixed models were not yet broadly available with standard statistical software.
- Longitudinal data analysis was possible via software for mainly balanced data e.g., Repeated Measures ANOVA.
- Multiple Imputation was a viable solution to deal with the missing data problem in this context at that time.
- Nowadays with with the availability of software to fit mixed models the use of Multiple Imputation is not necessary to deal with missingness in the outcome.
- Mixed models are more flexible and valid under MAR.
- Nonetheless, it may though be useful when we have missing data in covariates.
- Or in combination with GEE $\Rightarrow$ MI-GEE approach.

Multiple Imputation

- Several R packages are available:
  - `mice` the basic R package for multiple imputation with chained equations.
  - `smcfcs` is an R package for multiple imputation based on a more general framework than the mice package. This would be better suited when interactions or nonlinear effects are of interest.
  - `jomo` is an R package for multilevel joint modelling multiple imputation that can also handle categorical data.
  - `pan` is an R package for multiple imputation for multivariate panel or clustered data.
  - `mitools` is an R package with tools to perform analyses and combine results from multiple-imputation datasets.

Full Specification of the Outcome Distribution

- General idea: Use a model for the joint distribution of the responses, e.g. Mixed models, Marginal Regression Models (but not the GEE approach).

- Advantages:
  * no requirement to impute data.
  * available in all standard software.
  * valid results under MCAR and MAR.

- Disadvantages:
  * not valid results under MNAR.

Missing Not At Random Models

- General idea: When the missing data mechanism is MNAR, we need to define a model for the joint distribution of the longitudinal outcome $\{\mathbf{y}_i^o, \mathbf{y}_i^m\}$ and the missingness outcome $\mathbf{r}_i$.

- Three model families have been proposed:
  1. Selection models.
  2. Pattern mixture models.
  3. Shared parameter models.

Missing Not At Random Models

- Selection models use the decomposition:

$$p(\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{r}_i) = p(\mathbf{y}_i^o, \mathbf{y}_i^m)\; p(\mathbf{r}_i \mid \mathbf{y}_i^o, \mathbf{y}_i^m).$$

- These models postulate that the probability of dropping out is directly related on the missing longitudinal outcomes.

- Computation of the observed data likelihood requires integration over the missing data, which is rather computationally intensive and no established software is available.

Missing Not At Random Models

- Pattern mixture models use the decomposition:

$$p(\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{r}_i) = p(\mathbf{y}_i^o, \mathbf{y}_i^m \mid \mathbf{r}_i)\, p(\mathbf{r}_i)$$

- These models postulate that we have a different specification of the longitudinal model per dropout pattern (e.g., completers show different average evolutions than subjects who dropout earlier on).

- To make inference we combine/mix information from all missing data patterns with mixing weights $p(\mathbf{r}_i)$.

Missing Not At Random Models

- Shared parameter models use the decomposition:

$$p(\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{r}_i) = \int p(\mathbf{y}_i^o, \mathbf{y}_i^m \mid \mathbf{b}_i)\, p(\mathbf{r}_i \mid \mathbf{b}_i)\, p(\mathbf{b}_i)\, d\mathbf{b}_i$$

- These models postulate that the characteristics of the longitudinal profile of a subject (described by the random effects) dictate the chance of dropping out.

- For the computation of the observed data likelihood, the integration over the missing data is done automatically under the conditional independence assumption.

- Shared parameter models can be fitted using `JM` and `JMbayes`.

Missing Not At Random Models

- All three approaches give valid results under MNAR, but MNAR is specified differently in each one.
- So we need to carefully consider which is the most appropriate framework in each case.
- Advantages:
  - no requirement to impute data.
  - provide valid results under MNAR.
- Disadvantages:
  - only some of them available in software.
  - difficult to fit.
  - require sensitivity analysis.

Summary

- Missing data are often not under the control of the investigator.
- Reasons of missing data are described as MCAR, MAR, MNAR.
- Assumptions about the missing data are needed.
- The validity of the analyses results depends on the validity of the assumptions.
- Assumptions about the missing data cannot be checked using observed data: Sensitivity analysis is necessary.
- Methods valid under MAR should be the default analysis.

Chapter 7: Generalized Mixed Effects Models

Outline

Outline

Generalized Linear Mixed Effects Models

- Second extension of GLM to correlated data after GEE

  $\Rightarrow$ Generalized Linear Mixed Effects Models (GLMM).

- GLMM introduces random effects to capture within subject correlations.

- Main idea: **Each subject has his own profile in time**

  $\Rightarrow$ some regression coefficients vary from person to person, e.g.,

  - The disease risk for each patient at baseline may deviate considerably from the average patient $\Rightarrow$ intercept varies.
  - The disease risk for each patient may evolve differently in time $\Rightarrow$ slope varies.

- We assume that the coefficients vary according to a distribution $\Rightarrow$ random effects distribution (typically Normal).

Generalized Linear Mixed Effects Models

- GLMM in the same spirit as the Linear Mixed Effects Model for continuous data.
- However GLMMs are conceptually and mathematically more complex.
- Before discussing GLMM we will revisit the simplest case i.e., the LMM.

Outline

Generalized Linear Mixed Effects Models

- As in GEE, we will study GLMMs for:
  - Binary data $\Rightarrow$ Binomial distribution.
  - Count data $\Rightarrow$ Poisson distribution.
- Mixed Effects Logistic Regression.
- Mixed Effects Poisson Regression.

Example: Onychomycosis Study

- Reference: De Backer, M., De Vroey, C., Lesaffre, E., Scheys, I., and De Keyser, P. (1998). Twelve weeks of continuous oral therapy for toenail onychomycosis caused by dermatophytes: A double-blind comparative trial of terbinafine 250 mg/day versus itraconazole 200 mg/day. Journal of the American Academy of Dermatology, 38, 57-63.

- Randomized, double-blind, parallel-group, multicenter study on 294 patients to compare two oral treatments (A and B) for toe-nail infection.

ID	Time	Treat	Response
2	0	0	0
2	1	0	0
2	2	0	1
2	3	0	1
2	6	0	0
⋮	⋮	⋮	⋮

- Response: $Y = 1$ if severe infection and $Y = 0$ otherwise $\Rightarrow$ binary

- Covariates:
  - Treatment: 0 if treatment A and 1 if treatment B
  - Follow-up: baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48 thereafter

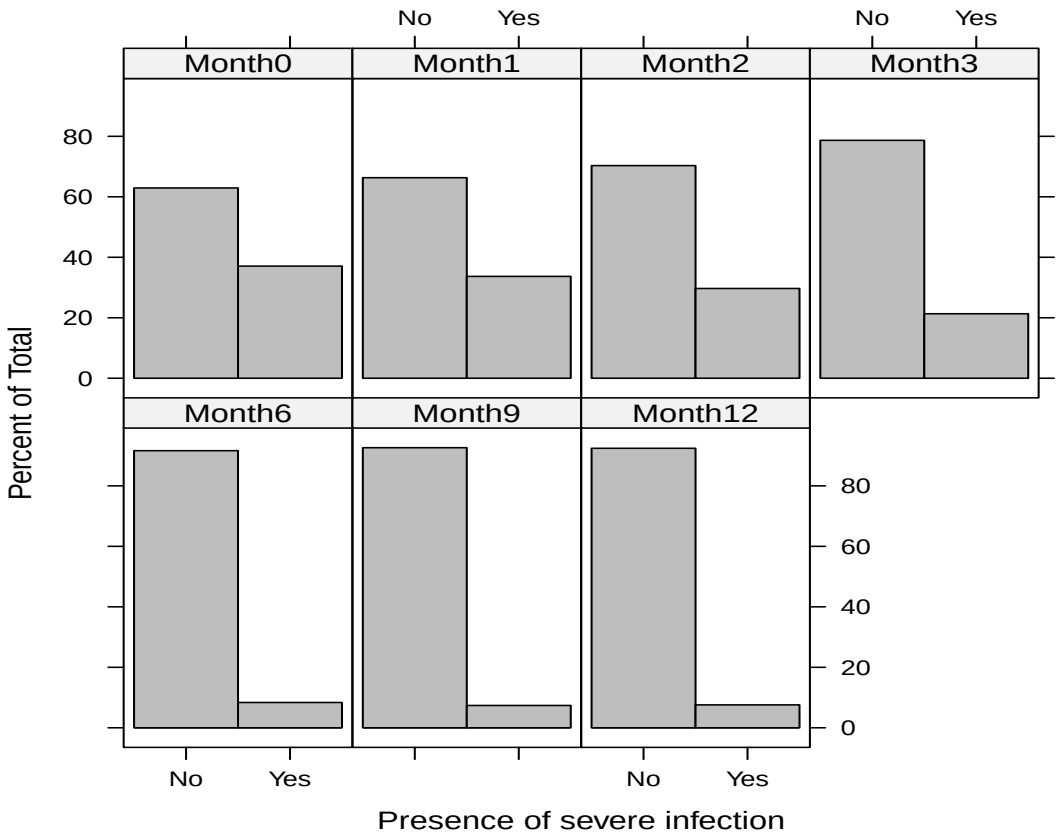
Example: Onychomycosis Study

- Research questions:
  - Are both treatments equally effective for the treatment of onychomycosis ?
- Complication:
  - Fixed number of measurements scheduled per subject, but not all measurements available due to dropout (24%), for unknown reasons

	# Observations		
Time (months)	Treatment A	Treatment B	Total
0	150	148	298
1	149	142	291
2	146	138	284
3	140	131	271
6	131	124	255
9	120	109	229
12	118	108	226

Example: Onychomycosis Study

- Percentages of severity of infection per visit (both treatments):



- Overall, there seems to be a trend toward less severe infections
- Note that, again, there is a problem of missing observations

GLMM - Mixed Effects Logistic Regression

- General definition:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i$$

- We extend the standard Logistic Regression model by introducing random effects to model the mean response $\boldsymbol{\pi}$

- As in the LMM
  - $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$
  - $\mathbf{X}_i$ design matrix for the fixed effects $\boldsymbol{\beta}$
  - $\mathbf{Z}_i$ design matrix for the random effects $\mathbf{b}_i$

GLMM - Mixed Effects Logistic Regression

- Given $\mathbf{b}_i$ the measurements y_{ij} of each subject i are independent (Conditional Independence Assumption) with density

$$\boxed{\mathrm{f}(\mathrm{y}_{ij} \mid \theta_{ij}, \phi) = \exp\left\{ \frac{y_{ij}\theta_{ij} - \alpha(\theta_{ij})}{\phi} + c(y_{ij}, \phi) \right\}}$$

  where

$$\theta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i \quad \text{and} \quad \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

- For binary data:

$$y_{ij} \mid \mathbf{b}_i \sim \text{Bernoulli}(\pi_{ij})$$

  - $E(y_{ij} \mid \mathbf{b}_i) = \pi_{ij}$
  - $\text{var}(y_{ij} \mid \mathbf{b}_i) = \pi_{ij}(1 - \pi_{ij})$

- This is the hierarchical model formulation.

GLMM - Mixed Effects Logistic Regression

- As in LMM, the marginal model formulation needs integrating out the $\mathbf{b}_i$.

- The implied marginal density is:

$$\begin{aligned} f(\mathbf{y}_i \mid \boldsymbol{\beta}, \mathbf{D}, \phi) &= \int f(\mathbf{y}_i, \mathbf{b}_i \mid \boldsymbol{\beta}, \phi, \mathbf{D}) d\mathbf{b}_i \\ &= \int f(\mathbf{y}_i \mid \mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i \mid \mathbf{D}) d\mathbf{b}_i \\ &= \int \prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i \mid \mathbf{D}) d\mathbf{b}_i \end{aligned}$$

- We also have:
  - $E(y_{ij}) = E_b\left[E(y_{ij} \mid \mathbf{b}_i)\right]$.
  - $\text{var}(y_{ij}) = E_b\left[\text{var}(y_{ij} \mid \mathbf{b}_i)\right] + \text{var}\left\{E_b\left[E(y_{ij} \mid \mathbf{b}_i)\right]\right\}$.

- In contrast to GEE:
  - GLMM fully specify the joint distribution $\mathbf{y}_i$.
  - Mean and correlation are modelled at the same time $\Rightarrow$ implication for interpretation of parameters.

Outline

GLMM - Mixed Effects Logistic Regression

- Example:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 t_{ij} + b_i, \quad b_i \sim N(0, \sigma_b^2)$$

where

- π_{ij} is the probability of severe onychomycosis infection for subject i in visit j.
- t_{ij} denotes the visit times for subject i.
- β_0 intercept $\Rightarrow$ baseline log-odds.
- b_i deviation of the intercept of subject i from the baseline log-odds β_0.

  $\Rightarrow$ if $b_i > 0 (< 0)$ the subject more prone (less) prone to e.g., obesity.
- $\beta_0 + b_i$ is the subject-specific intercept.
- σ_b^2 heterogeneity/variability between subjects on the log-odds. scale $\Rightarrow$ the highest the stronger the correlations.
- β_1 measures the change in the log odds of severe infection for every month that passes by, for a specific individual having an underlying propensity to severe infection, b_i.

GLMM - Mixed Effects Logistic Regression

- Example:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 t_{ij} + b_i$$

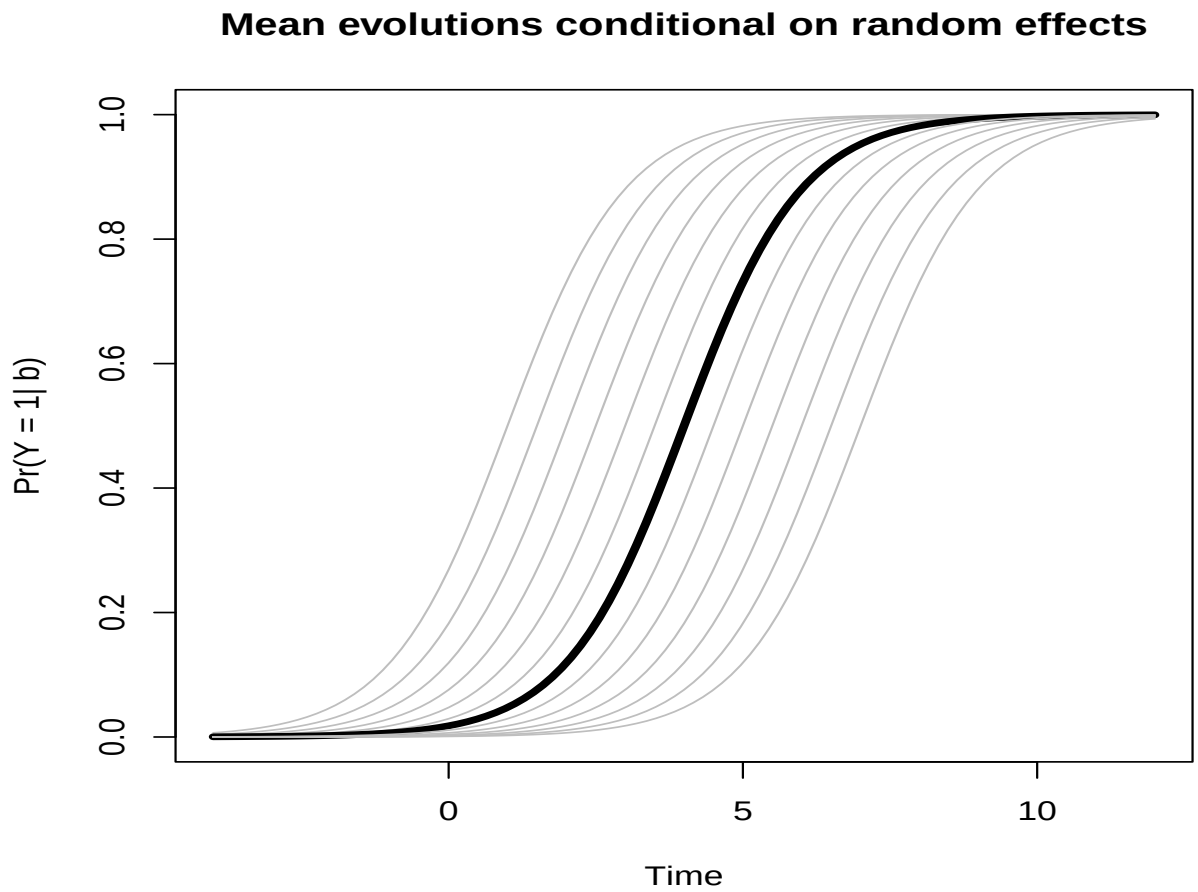
- The conditional on the random effects probabilities are:

$$\pi_{ij} = \frac{\exp(\beta_0 + \beta_1 t_{ij} + b_i)}{1 + \exp(\beta_0 + \beta_1 t_{ij} + b_i)}.$$

- Each subject has his own probability of severe infection in time.

GLMM - Interpretation

- Graphical representation:



GLMM - Mixed Effects Logistic Regression

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 t_{ij} + b_i$$

- The conditional on the random effects probabilities are:

$$\pi_{ij} = \frac{\exp(\beta_0 + \beta_1 t_{ij} + b_i)}{1 + \exp(\beta_0 + \beta_1 t_{ij} + b_i)}$$

- This is to be contrasted with the GEE approach where the probability of severe infection is:

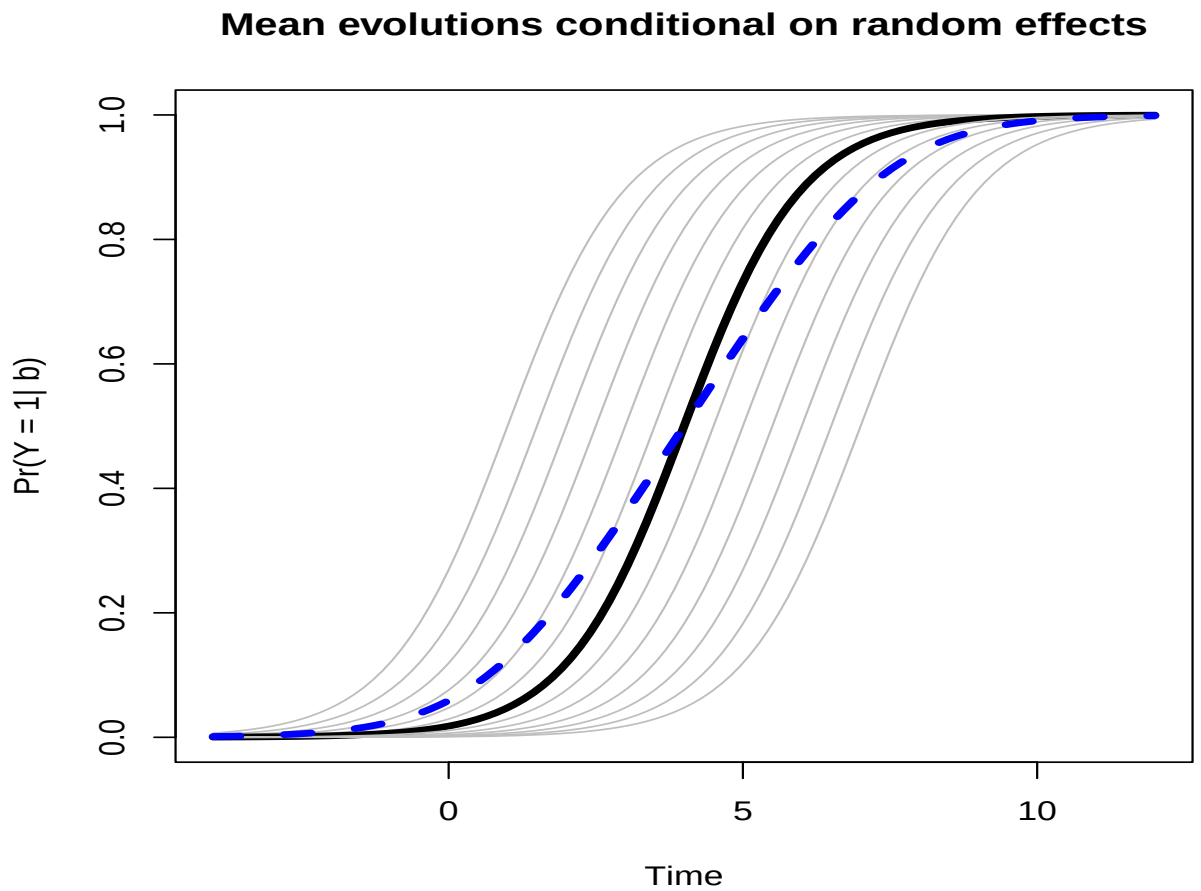
$$\pi_{ij} = \frac{\exp(\beta_0 + \beta_1 t_{ij})}{1 + \exp(\beta_0 + \beta_1 t_{ij})}$$

- Thus under GEE the probability of infection has a marginal and not a subject-specific interpretation.

GLMM - Interpretation

- Graphical representation:

  Subject-specific probabilities (GLMM) versus marginal probabilities (GEE) (blue dashed line).



GLMM - Mixed Effects Logistic Regression

- We distinguish the two probabilities.
- The conditional on the random effects probabilities from GLMM:

$$\pi_{ij}^{SS} = \frac{\exp(\beta_0^{SS} + \beta_1^{SS} t_{ij} + b_i)}{1 + \exp(\beta_0^{SS} + \beta_1^{SS} t_{ij} + b_i)}$$

  where SS denotes **S**ubject-**S**pecific parameters.

- The marginal probabilities from GEE approach:

$$\pi_{ij}^{M} = \frac{\exp(\beta_0^{M} + \beta_1^{M} t_{ij})}{1 + \exp(\beta_0^{M} + \beta_1^{M} t_{ij})}$$

  where M denotes **M**arginal parameters.

- Setting $b_i = 0$ in π_{ij}^{SS} does not lead to π_{ij}^{M}.

GLMM - Mixed Effects Logistic Regression

- Setting $b_i = 0$ in π_{ij}^{SS} leads to:

$$\pi_{ij}^{SS} = \frac{\exp(\beta_0^{SS} + \beta_1^{SS} t_{ij})}{1 + \exp(\beta_0^{SS} + \beta_1^{SS} t_{ij})}.$$

- This is the subject-specific probabilities for the average subject not the average or marginal probabilities π_{ij}^{M}.

- **Probability for the average subject $\neq$ average of probabilities of all subjects.**

GLMM - Mixed Effects Logistic Regression

- This carries over to regresssion parameters, i.e. subject-specific odds ratio's and not marginal odds ratio's.
- Let us consider the mixed effects logistic regression:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 \text{treat}_{ij} + b_i$$

- We are interested on marginal odds ratio's for 'treat', i.e. we want to measure change in the odds between the 2 groups.
- In such a case we need to compute odds over all subjects at group 1 and at group 0 and take the ratio.
- We will check if that would be possible under the GLMM.

GLMM - Mixed Effects Logistic Regression

- We write the mixed effects logistic regression for both groups:
- Logistic regression for subject i:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 \cdot 0 + b_i$$

- Logistic regression for subject i':

$$\log \frac{\pi_{i'j}}{1-\pi_{i'j}} = \beta_0 + \beta_1 \cdot 1 + b_{i'}$$

- β_1 is the log-odds ratio given the random effects:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} - \log \frac{\pi_{i'j}}{1-\pi_{i'j}} = \beta_1 + (b_i - b_{i'})$$

- The conditional on the random effects odds ratio is:

$$\frac{\pi_{ij}}{1-\pi_{ij}} / \frac{\pi_{i'j}}{1-\pi_{i'j}} = \exp\{\beta_1 + (b_i - b_{i'})\} \neq \exp\{\beta_1\}$$

GLMM - Mixed Effects Logistic Regression

- β_1^{SS} is the log-odds ratio given the random effects:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} - \log \frac{\pi_{i'j}}{1-\pi_{i'j}} = \beta_1^{SS} + (b_i - b_{i'})$$

- The conditional on the random effects odds ratio is:

$$\frac{\pi_{ij}}{1-\pi_{ij}} / \frac{\pi_{i'j}}{1-\pi_{i'j}} = \exp\{\beta_1^{SS} + (b_i - b_{i'})\} \neq \exp\{\beta_1^{M}\}$$

- Integrating out the random effects in $\exp\{\beta_1^{SS} + (b_i - b_{i'})\}$ does not lead to $\exp\{\beta_1^{M}\}$.

- It is often mistakenly believed that for within subject effects β_1 has a marginal interpretation.

- That was not the case in the LMM.

GLMM - Interpretation

- To make the difference between β^{SS} and β^{M} more clear we consider the following hypothetical example.

- Assume epileptic patients are treated with 2 drugs (A and B) in various hospitals:
  - Under GEE interest on effect of treatment on risk of epilepsy over all hospitals $\Rightarrow$ we compare groups of patients in treatment A and treatment B (cross-sectional effect): β^{M}.
  - Under GLMM interest on effect of treatment on risk of epilepsy for a specific hospital $\Rightarrow$ we compare groups of patients in treatment A and treatment B at a specific hospital (subject-specific effect): β^{SS}.

GLMM - Interpretation

- The basis for the distinction between GEE and GLMMs.
  - In GEE the mean and correlation structure are treated separately.
  - In GLMM the mean and correlation structure are treated at the same time via the introduction of random effects.

- This affects the interpretation of the parameters $\Rightarrow$ The problem arises from the fact that the covariates are related to the mean response non-linearly (e.g. logistic transformation).

- Even though random effects are used also in LMM, the parameters there have both marginal and subject-specific interpretation.

- So, whenever the random effects enter the mean in a non-linear way, the regression parameters in the marginal model need to be interpreted differently from the regression parameters in the mixed model.

Interpretation of GLMM Parameters

- From the GLMM parameters we can obtain approximately parameters with marginal interpretation in special cases.
  - In logistic regression with random intercepts only:

$$\beta^M \approx \frac{\beta^{SS}}{\sqrt{1+0.346\sigma_b^2}}.$$

  - In Poisson regression with random intercepts only:
    - β_1^M and β_1^{SS} are equal.
    - But for the intercept: $\beta_0^M = \beta_0^{SS} + \sigma_b^2/2$.

- Subject-specific parameters are larger than marginal parameters in absolute value.

- Such formula's not available for higher dimensional random effects (e.g. random intercepts and random slopes) or any other non-Gaussian outcome.

Interpretation of GLMM Parameters

- For general random effects structures and response variables:
  - Marginalized Mixed Models.
  - not available with standard software.

- An alternative solution has been recently proposed (Hedeker *et al.*2018) .

- The marginal coefficients β^M are obtained as the solution to the

$$\beta^M = (X^T X)^{-1} X^T \text{logit}(\pi^M)$$

where $\pi_i^M = \int_b \text{logit}^{-1}(X_i^T \beta^{SS} + Z_{ij}^T b_i) p(b_i) db_i$.

- Available in `marginal_coefs(.)` of `GLMMadaptive`.

Outline

GLMM - Estimation

- An issue of generalized linear mixed models is their computational burden.
- Contrary to GEE estimation of parameters can be done using the maximum likelihood method.
- This has implications:
  - GLMMs algebraically and numerically more difficult.
  - Hypothesis testing: Likelihood ratio test, score test and Wald test are all appropriate.
  - Missing data: Valid inference under less strict assumptions compared to GEE because it is a likelihood-based method.

GLMM - Estimation

- Because, in general, there is no simple closed-form solution for the marginal likelihood, numerical integration techniques are required.

- The likelihood function for β, $\mathbf{D}$, and ϕ now equals:

$$\begin{aligned} L(\beta, \mathbf{D}, \phi) &= \prod_{i=1}^{n} f(\mathbf{y}_i; \beta, \mathbf{D}, \phi) \\ &= \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{b}_i; \beta, \phi)\ f(\mathbf{b}_i \mid \mathbf{D})\ d\mathbf{b}_i \end{aligned}$$

- Under the normal linear model, the integral can be worked out analytically.

GLMM - Estimation

- Estimation of parameters of a GLMM is possible with various statistical software e.g., SPSS, SAS, R, etc.
- Parameter estimates for the same dataset from various software can lead to very different results because each software uses a different method to fit GLMMs.
- Thus it is important to be aware of which method each software uses $\Rightarrow$ understand their limitations and strengths.
- The various software differ mainly on how they deal with these integrations.
- We will discuss which methods are available and which are supported by statistical software.

GLMM - Evaluate likelihood

- In general approximations are required:
  1. Approximation of integrand $\Rightarrow$ Laplace approximations.
  2. Approximation of data $\Rightarrow$ Penalized quasi-likelihood & Marginal quasi-likelihood.
  3. Approximation of integral $\Rightarrow$ Non-adaptive or Adaptive Gaussian quadrature methods.

GLMM - Approximation of integrand

- Laplace approximations:
  - Roughly speaking, $\prod_j f_{ij}(y_{ij}|b_i, \beta, \phi)$ is approximated by a normal density.
  - The integral can be calculated analytically, as in the LMM.
  - Good approximation in case of many repeated measures per subject.
  - Implemented by R package `lme4` in R. item Implemented by `PROC NLMIXED` in SAS.

GLMM - Approximation of data

- In Penalized quasi-likelihood & Marginal quasi-likelihood:
  - Roughly speaking, the LMM is applied on pseudo data.
  - Perform reasonably well when random-effects variance is (very) small.
  - Perform bad for binary outcomes with few repeated measurements per cluster.
  - Implemented by function `glmmPQL` in R package `MASS`.
  - Implemented by `GENLINMIXED` in SPSS.
  - Implemented by `PROC GLIMMIX` in SAS.

GLMM - Approximation of Integral

- Gaussian quadrature methods:
  - The likelihood contribution of every subject is of the form $\int f(z)\phi(z)dz.$ where $\phi(z)$ is the density of the (multivariate) normal distribution
  - Gaussian quadrature methods replace the integral by a weighted sum:

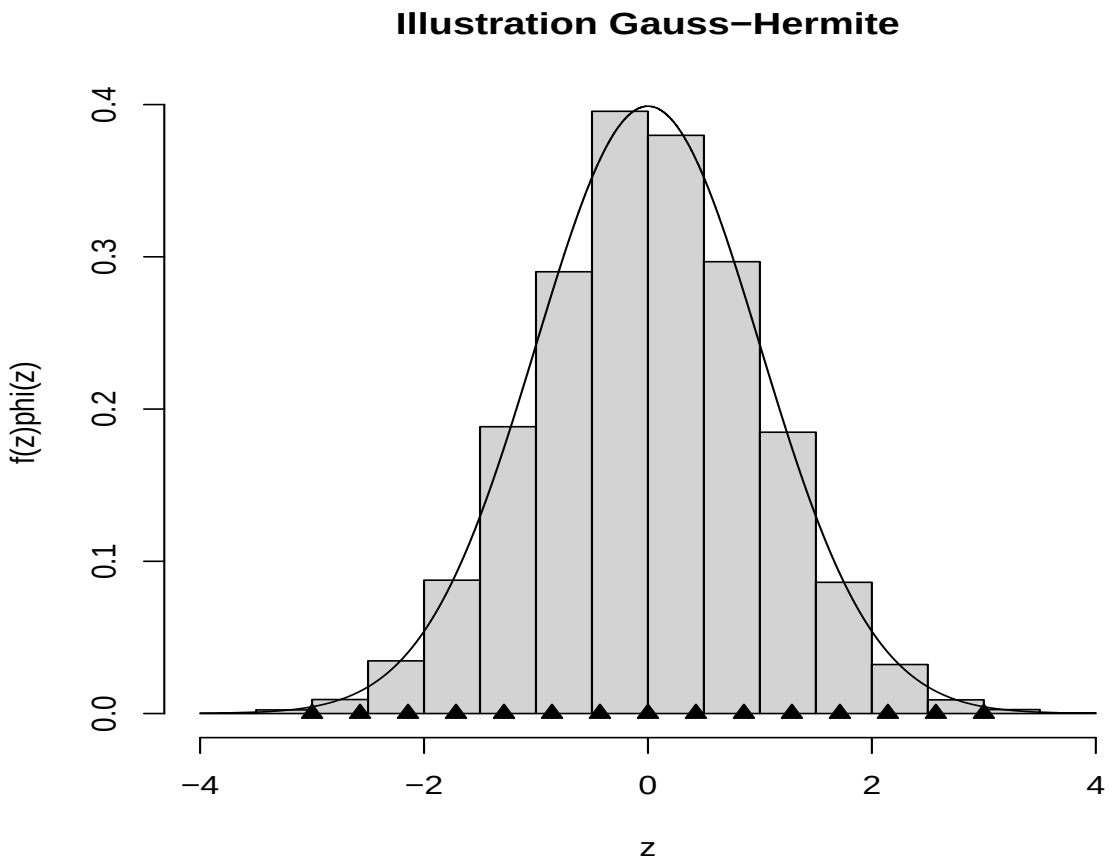
$$\int f(z)\phi(z)dz \quad \approx \quad \sum_{q=1}^{Q} w_q f(z_q)$$

  - Q is the order of the approximation $\Rightarrow$ The higher Q the more accurate the approximation will be.

GLMM - Approximation of Integral

- In case of univariate integration, the approximation consists of subdividing the integration region in intervals, and approximating the surface under the integrant by the sum of surfaces of the so-obtained approximating rectangles.

- Graphically ($Q = 15$):



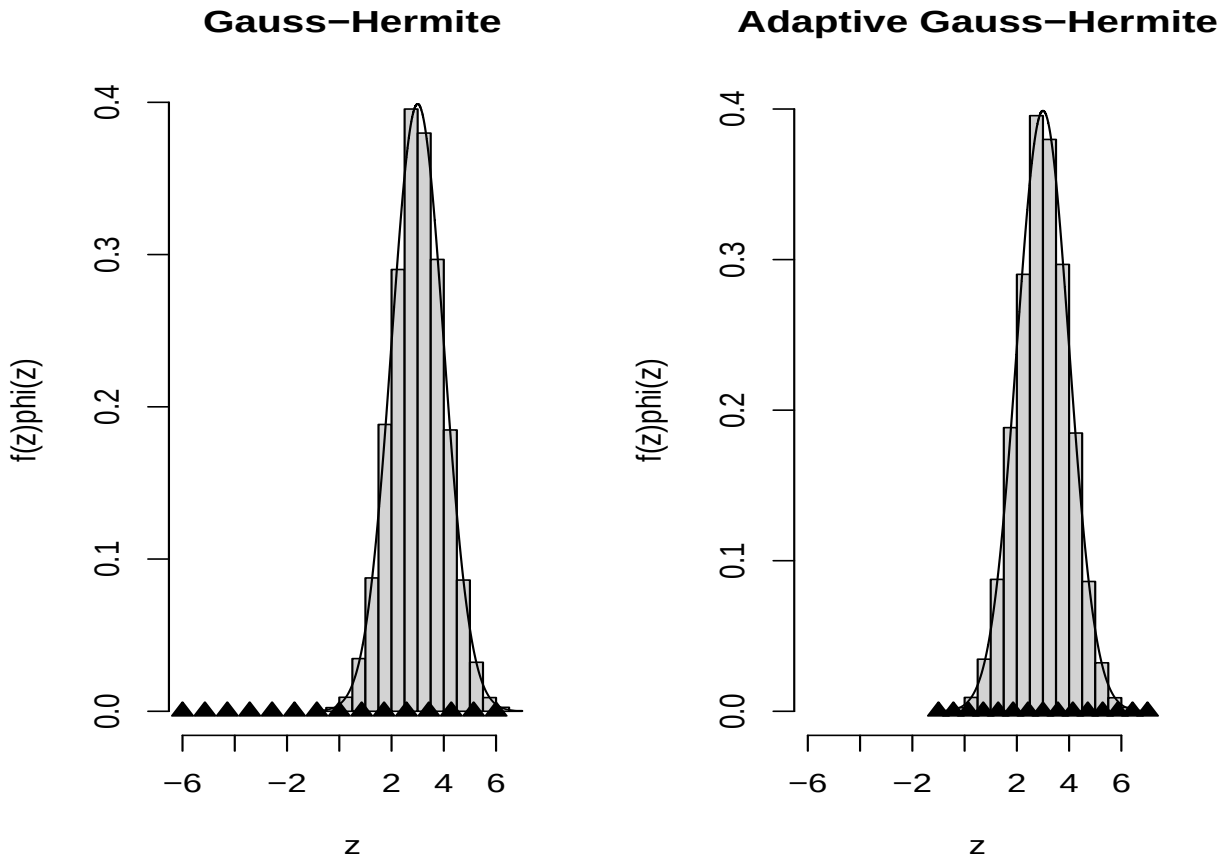
Approximation of Integral

- Two frequently used procedures for approximating integrals of the above form are.
  - Gaussian quadrature $\Rightarrow$ the nodes and weights are fixed, independent of $f(z)\phi(z)$.
  - Adaptive Gaussian quadrature $\Rightarrow$ the nodes and weights are adapted to the 'support' of $f(z)\phi(z)$.

- Adaptive Gaussian quadrature rescales and shifts the quadrature points such that more quadrature points lie in the region of interest.

Approximation of Integral

- Graphically ($Q = 15$):



- Adaptive Gaussian quadrature needs less quadrature points than non-adaptive Gaussian quadrature.
- On the other hand, adaptive Gaussian quadrature is much more time consuming.

GLMM - Evaluate likelihood - To sum up

- Recap: Availability in statistical software
  1. Approximation of integrand: R `lme4` with option '`AGH` $= \mathbf{1}$', SAS `PROC NLMIXED`.
  2. Approximation of data: R: `MASS - glmmPQL`, SAS: `PROC GLIMMIX`, SPSS.
  3. Approximation of integral: SAS `PROC NLMIXED`, R. `GLMMadaptive` and `lme4` (only for random intercepts).

- Pros & Cons
  - Laplace, MQL and PQL perform poorly with binary data and require many repeated measurements per subject.
  - Approximation of integral methods should be preferred (i.e., R: `GLMMadaptive` and `lme4`, SAS: `PROC NLMIXED`).
  - No free lunch:
    - They can be more intensive computationally compared to the other methods.
    - They can be sensitive to starting values (in some software you can adjust it).
    - You need to choose number of quadrature points .
    - Adaptive gaussian quadrature requires less points than simple gaussian quadrature but is more intensive.

Outline

Example: Onychomycosis Study / Model 1

- Model 1: We start with logistic regression model with only random intercepts b_i

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

  where

  - T_i: treatment indicator for subject i.
  - t_{ij}: time point at which jth measurement is taken for ith subject.

- Even though randomized study we include main effect of treatment for illustrative reasons.

- More complex mean models can be considered as well (e.g., polynomial time effects, covariates or other random effects).

- Within-subject correlation: random intercepts.

GLMM - Example: Onychomycosis Study / Model 1

- We will first analyse the data using function `glmer(.)` from R package `lme4` in R.
- Main arguments in `glmer(.)`:
  - `formula`: a formula specifying the response vector, the fixed- and random-effects structure.
  - `data`: a data frame containing all the variables.
  - `family`: a description of the error distribution and link function to be used in the model.
  - `nAGQ`: the number of quadrature points.

GLMM - Example: Onychomycosis Study / `glmer(.)`

- We will first analyse the data using function `glmer(.)` from R package `lme4` in R.
- We need the data in long format $\Rightarrow$ one row per visit.

```
> #file. <- "..." # specify the directory with data
> library(lme4)
> data. <- read.csv(file.path(file.,  "ToenailBinaryLong.csv"),
+                           header = TRUE)[, -c(1, 2)]
> data.$Treat <- factor(data.$Treat, levels = 0:1,
+                               labels = c("A", "B"))
> model.1 <- glmer(Response ~ Treat + Time + Treat:Time + (1|ID),
+                  data = data., family =  "binomial",
+                  nAGQ = 1, control = glmerControl(optimizer="bobyqa"))
```

GLMM - Example: Onychomycosis Study / Model 1

- glmer output

```
> print(summary(model.1), corr = FALSE)
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: Response ~ Treat + Time + Treat:Time + (1 | ID)
   Data: data.
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
    1263     1290     -626     1253     1903

Scaled residuals:
   Min     1Q Median     3Q    Max
 -3.02  -0.15  -0.07   0.00  41.35

Random effects:
 Groups Name        Variance Std.Dev.
 ID     (Intercept) 21.1     4.59
Number of obs: 1908, groups:  ID, 294

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.5467     0.7824   -3.26   0.0011 **
TreatB       -0.2580     0.6913   -0.37   0.7090
Time         -0.4137     0.0486   -8.52   <2e-16 ***
TreatB:Time  -0.1624     0.0736   -2.21   0.0273 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

GLMM - Example: Onychomycosis Study / Model 1

- The estimated variance of the random intercepts is relatively large: $\sigma_b^2 = 21.05 \Rightarrow$ heterogeneity should not be ignored.
- β_1 is the log-odds ratio of severe infection at baseline between treatment A and treatment B for the same patient.
  - The odds of severe infection at baseline for a patient at treatment B is 22.742% lower than the odds if the same patient was assigned to treatment A.
  - However this difference is not statistically significant (p-value $= 0.709$) $\Rightarrow$ as expected due to randomization.
- The odds of severe infection at week 1 for a patient at treatment B is 0.65 times the odds of the same patient assigned to treatment A at week 1.
- The parameters clearly have subject-specific interpretation.

GLMM - Example: Onychomycosis Study / Model 1

- Since we have included a single random intercepts we can derive parameters with marginal interpretation using the formula we saw before
  - $\beta_0^M \approx -0.885$
  - $\beta_1^M \approx -0.09$
  - $\beta_2^M \approx -0.144$
  - $\beta_3^M \approx -0.06.$
- Marginal parameters are smaller in absolute value than the subject-specific parameters.
- However computation of standard errors for the derived β^M is not straightforward $\Rightarrow$ we need the Delta method (we will see an example in the practicals).
- The $\beta_0^M, \beta_1^M, \ldots$ differ from those in GEE: perhaps due to missing data.
- Comparing different models using LRT, AIC or BIC is correct in this case.
- For the PQL it is not correct, because for each model different pseudo-data are created on which the model is fitted.

GLMM - Example: Onychomycosis Study / Model 2

- We have explained that the Laplace method can behave poorly with binary data and few repeated measurements.
- We will analyse the same data using adaptive Gaussian quadrature.
- In ‘glmer(.)‘ this is only possible for the random-intercepts.
- R code:

```
> #file. <- "..." # specify the directory with data
> model.2 <- glmer(Response ~ Treat + Time + Treat:Time + (1|ID),
+                  data = data., family =  "binomial",
+                  nAGQ = 11, control = glmerControl(optimizer="bobyqa"))
```

GLMM - Example: Onychomycosis Study / Model 2

- glmer output

```
> print(summary(model.2), corr = FALSE)
Generalized linear mixed model fit by maximum likelihood (Adaptive
  Gauss-Hermite Quadrature, nAGQ = 11) [glmerMod]
 Family: binomial  ( logit )
Formula: Response ~ Treat + Time + Treat:Time + (1 | ID)
   Data: data.
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
    1257     1285     -624     1247     1903

Scaled residuals:
   Min     1Q Median     3Q    Max
 -2.95  -0.19  -0.09  -0.01  38.10

Random effects:
 Groups Name        Variance Std.Dev.
 ID     (Intercept) 16.1     4.02
Number of obs: 1908, groups:  ID, 294

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.6415     0.4342   -3.78  0.00016 ***
TreatB       -0.1171     0.5871   -0.20  0.84189
Time         -0.4039     0.0458   -8.81  < 2e-16 ***
TreatB:Time  -0.1611     0.0718   -2.24  0.02490 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

GLMM - Example: Onychomycosis Study / Model 2

- The interpretation of the parameters remains the same but notice the difference in the size of the estimates.
- In this example the inference does not change.

GLMM - Example: Onychomycosis Study / Model 3

- The inclusion of a random slope can be specified as follows.

```
> model.3 <- glmer(Response ~ Treat + Time + Treat:Time + (Time|ID),
+                  data = data., family =  "binomial",
+                  nAGQ = 1, control = glmerControl(optimizer="bobyqa"))
```

- Should we add the random slopes in our model?
- In general we can test for this using the likelihood ratio test, i.e., we compare the log-likelihoods obtained from Model 1 and Model 3.
- LRT:

```
> anova(model.1, model.3)
Data: data.
Models:
model.1: Response ~ Treat + Time + Treat:Time + (1 | ID)
model.3: Response ~ Treat + Time + Treat:Time + (Time | ID)
        npar  AIC  BIC logLik deviance Chisq Df Pr(>Chisq)
model.1    5 1263 1290   -626     1253
model.3    7  997 1036   -492      983   269  2     <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

GLMM - Example: Onychomycosis Study / `GLMMadaptive`

- The function that fits GLMMs in `GLMMadaptive` is `mixed_model()`; its basic arguments are:
  - `fixed`: a formula specifying the response vector, and the fixed-effects part of the model.
  - `random`: a formula specifying the random-effects part.
  - `data`: a data frame containing all the variables.
  - `family`: a family object specifying the distribution of the outcome and the link function.
  - `nAGQ`: the number of quadrature points.

GLMM - Example: Onychomycosis Study / Model 1

- To fit the same model for the Onychomycosis Study as we did above with `glmer()` the code is:

```
glmmFit <- mixed_model(fixed = Response ~ Treat + Time + Treat:Time,
random = ~ Time|ID, family = binomial(), data = data., nAGQ = 15)
summary(glmmFit)
```

GLMM - GLMMadaptive vs glmer(.)

- Differences between `glmer()` (package `lme4`) and `mixed_model()` (package `GLMMadaptive`).
- `glmer()` only provides the adaptive Gaussian quadrature rule for the random intercepts case, whereas `mixed_model()` uses this integration method with several random terms.
- `mixed_model()` currently only handles a single grouping factor for the random effects, i.e., you cannot fit nested or crossed random effects, whereas such designs can be fitted with `glmer()`.
- `mixed_model()` can fit zero-inflated Poisson and negative binomial data, allowing for random effects in the zero part.
- In `GLMMadaptive` parameters with marginal interpretation are estimated as well.

GLMM - Example: Onychomycosis Study - Quadrature points

- In order to investigate the accuracy of the numerical integration method, the model was refitted, for varying numbers of quadrature points, for adaptive quadrature method.
- The evaluations have been done using the R package `lme4`

	Adaptive Gaussian quadrature				
	$Q = 3$	$Q = 5$	$Q = 10$	$Q = 20$	$Q = 50$
β_0	-1.86(0.39)	-1.85(0.39)	-1.96(0.40)	-1.95(0.40)	-1.95(0.40)
β_1	-0.16(0.56)	-0.16(0.55)	-0.18(0.57)	-0.18(0.57)	-0.18(0.57)
β_2	-0.39(0.04)	-0.39(0.04)	-0.40(0.04)	-0.40(0.04)	-0.40(0.04)
β_3	-0.15(0.07)	-0.15(0.07)	-0.15(0.07)	-0.15(0.07)	-0.15(0.07)
σ_b^2	14.737	14.449	15.535	15.428	15.429
Log-Likelihood	-688.66	-691.09	-690.16	-690.2	-690.2

GLMM - Example: Onychomycosis Study - Quadrature points

- (Log-)likelihoods are not comparable.
- Different Q can lead to considerable differences in estimates and standard errors, but this is not the case for the Onychomycosis study.
- The higher Q is, the better the approximation in the integral.
- But the higher the Q, the more computationally intensive the estimation can get.
- The parameter estimates stabilize after $Q = 10 \Rightarrow$ this is a reasonable choice for further analysis.

Interpretation of GLMM Parameters: Onychomycosis Study / Model 1

- We compare our GLMM results for the toenail data with those from fitting GEE's (exchangeable working correlation)

	GLMM	GEE
Parameter	Estimate (s.e.)	Estimate (s.e.)
Intercept group A	−1.950 (0.399)	−0.584 (0.173)
Intercept group B	−2.127 (0.407)	−0.572 (0.196)
Slope group A	−0.397 (0.045)	−0.117 (0.031)
Slope group B	−0.552 (0.059)	−0.266 (0.048)

- We observe considerable differences in the size of the estimates $\Rightarrow$ inference does not change.

- **It is not that one is right and the other wrong $\Rightarrow$ they are just not comparable.**

Interpretation of GLMM Parameters: Onychomycosis Study

- The strong differences can be explained because the covariates are related to the mean response non-linearly.
- Again we can derive the marginal parameters using the formulas
  - $\beta_0^M \approx -0.672$
  - $\beta_1^M \approx -0.025$
  - $\beta_2^M \approx -0.001$
  - $\beta_3^M \approx -0.031.$
- We observe that the derived marginal parameters get close to the those obtained under GEE.
- There are small differences that can be due to the fact that:
  - The formulas are approximate not exact.
  - We have missing data and the two methods are valid under different assumptions about the reasons for missingness.

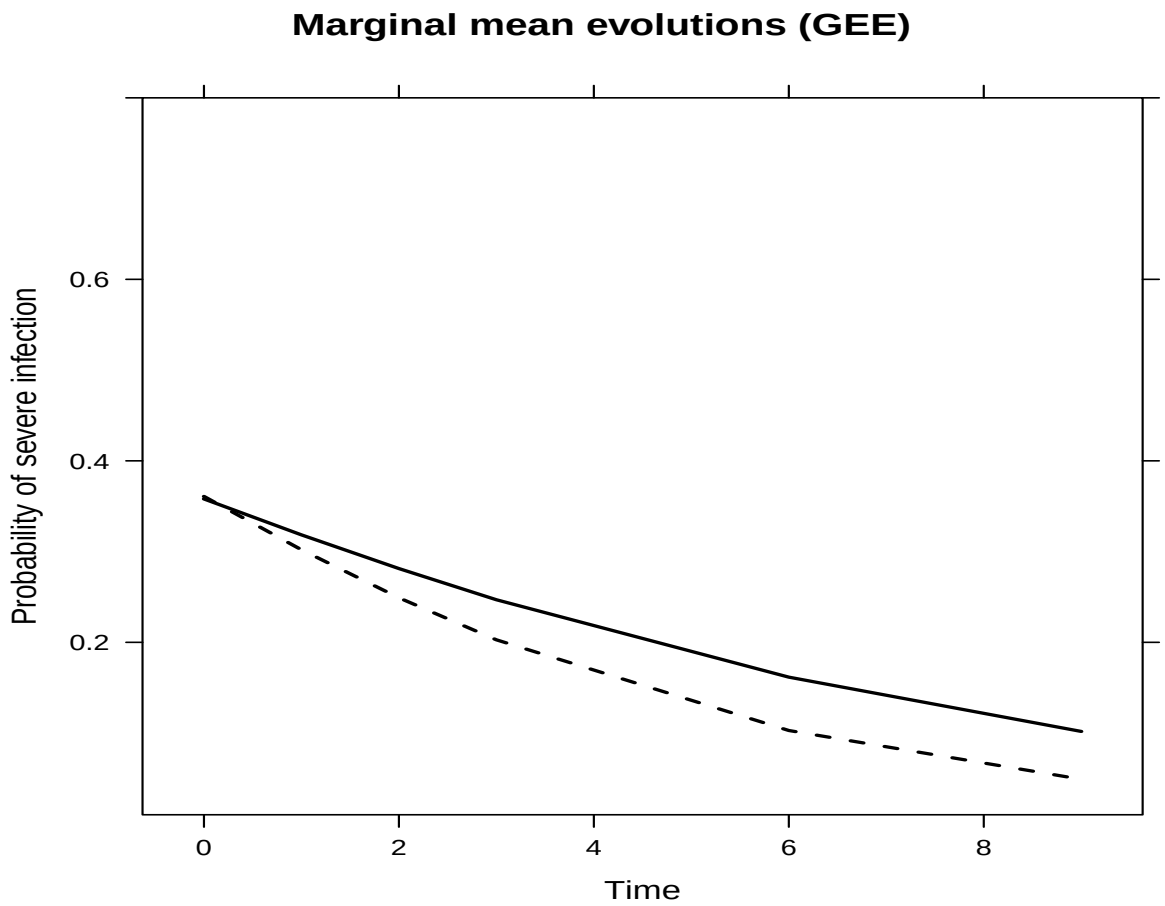
Interpretation of GLMM Parameters: Onychomycosis Study

- Let us have a look at the fitted probabilities of severe infection in time per group under GEE and GLMM.
- Average evolutions obtained from the GEE analyses:

$$P(Y_{ij}=1) = \begin{cases} \dfrac{\exp(-0.584-0.177t_{ij})}{1+\exp(-0.584-0.177t_{ij})}, & \text{Treatment A} \\ \dfrac{\exp(-0.572-0.266t_{ij})}{1+\exp(-0.572-0.266t_{ij})}, & \text{Treatment B} \end{cases}$$

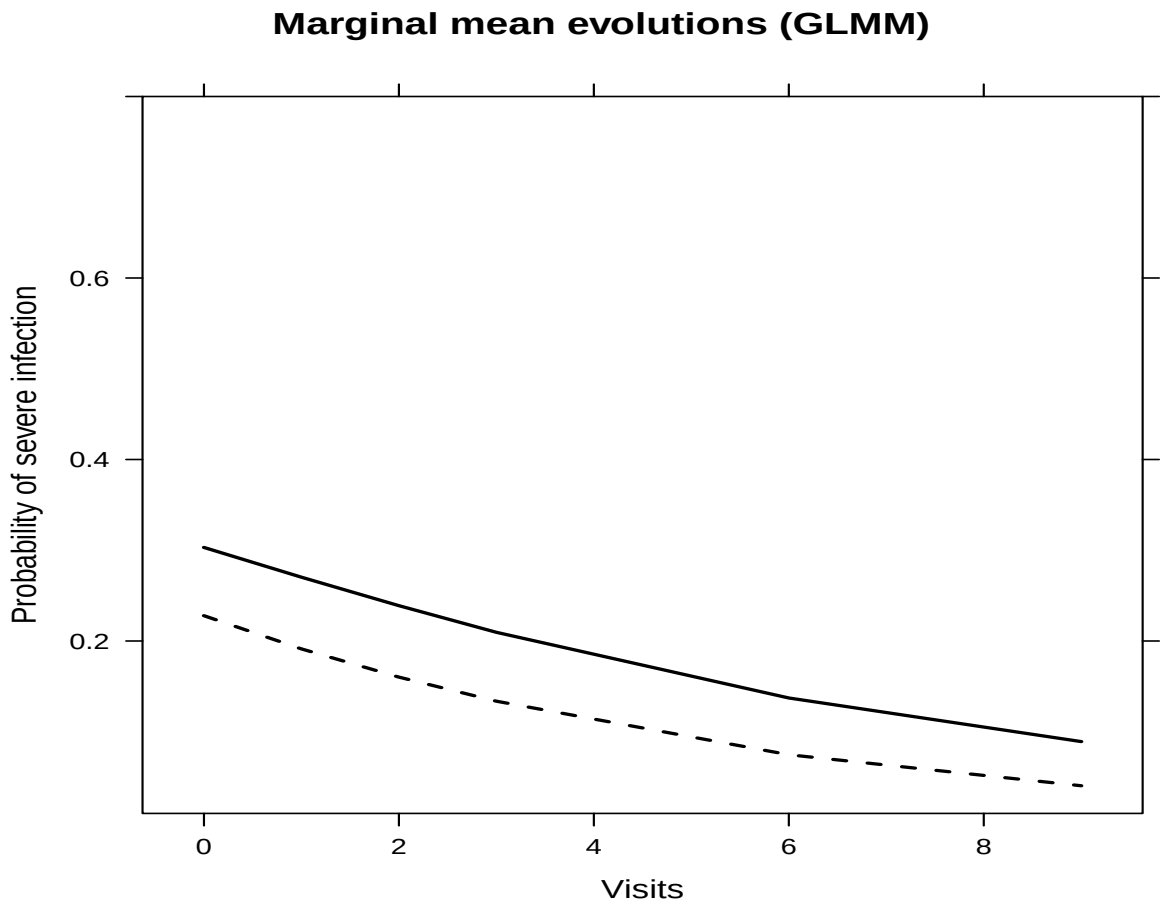
Interpretation of GLMM Parameters: Onychomycosis Study

- Graphical representation:



Interpretation of GLMM Parameters: Onychomycosis Study

- In a GLMM context the marginal averages can be derived in a more complex manner.
- Graphical representation:



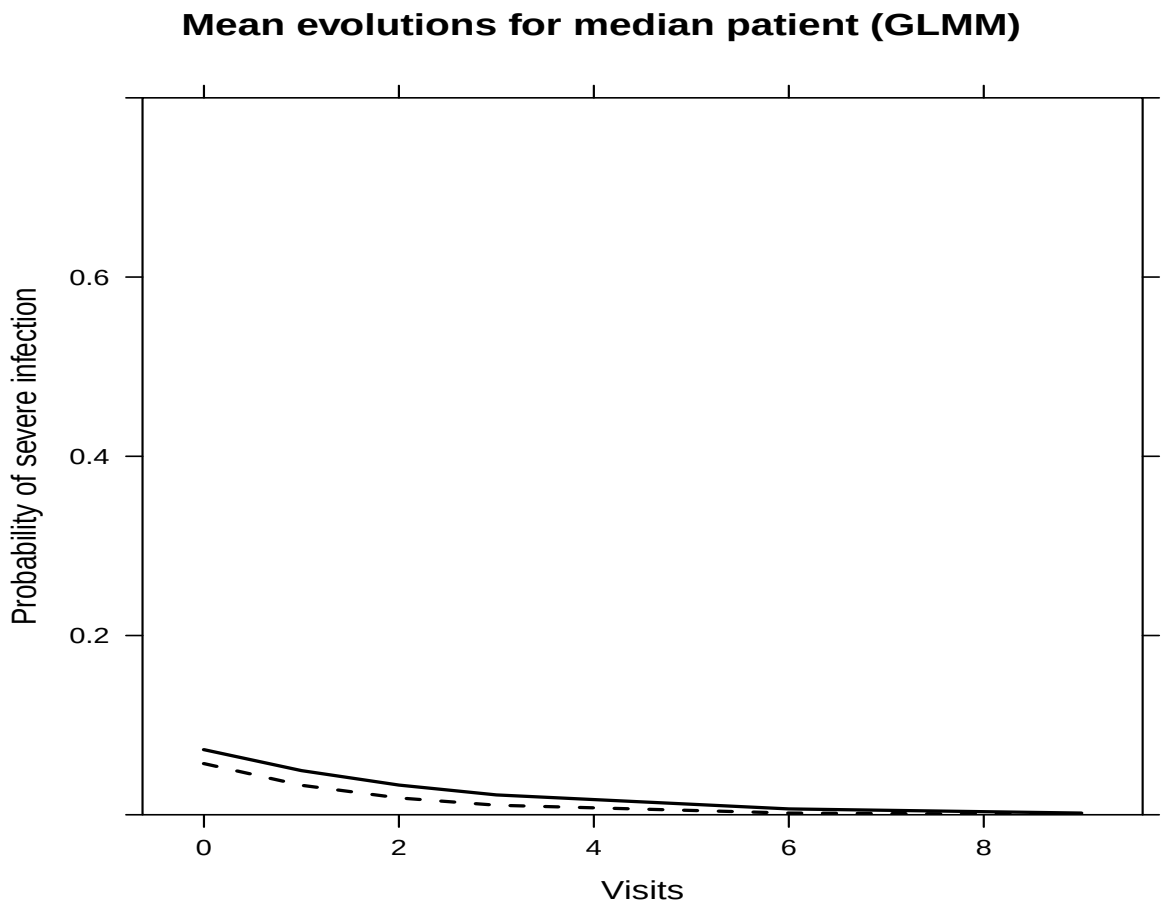
Interpretation of GLMM Parameters: Onychomycosis Study

- In a GLMM context, rather than plotting the marginal averages, one can also plot the profile for an 'average' subject, i.e., a subject with random effect $b_i = 0$.

$$P(Y_{ij} = 1 \mid b_i = 0) = \left\{ \begin{array}{ll} \dfrac{\exp(-2.547 - 0.414 t_{ij})}{1 + \exp(-2.547 - 0.414 t_{ij})}, & \text{Treatment A} \\ \dfrac{\exp(-2.805 - 0.576 t_{ij})}{1 + \exp(-2.805 - 0.576 t_{ij})}, & \text{Treatment B} \end{array} \right\}$$

Interpretation of GLMM Parameters: Onychomycosis Study

- Graphical representation:



Example: Leprosy study - Revisited

- We have analysed this dataset before using the GEE approach.
- Randomized trial for effect of two antibiotics (drug A and B) and placebo (drug C) on treatment of leprosy bacilli in a sample of 30 patients.
- Response: Y = number of leprosy bacilli in six sites of the body $\Rightarrow$ counts.
- Covariate: Treatment group and time period.
- Research question: Do drugs A and B reduce abundance of leprosy bacilli versus placebo?

GLMM - Example: Leprosy data - Revisited

- We perform similar analyses for the lepropsy data.
- We fit the model with the R package lme4 and get

```
> #file. <- "..." # specify the directory with data
> data. <- read.csv(file.path(file.,  "LeprosyLong.csv"), header = TRUE)[, -1]
> data.$Drug <- relevel(as.factor(data.$Drug), ref = "C")
> library(lme4)
> data. <- data.[order(data.$ID, data.$Time), ]
> model.1 <- glmer(Counts ~ Drug * Time - Drug + (1 | ID), data = data.,
+                  family = poisson, nAGQ = 15)
```

GLMM - Example: Leprosy data

- glmer output

```
> print(summary(model.1), corr = FALSE)
Generalized linear mixed model fit by maximum likelihood (Adaptive
  Gauss-Hermite Quadrature, nAGQ = 15) [glmerMod]
 Family: poisson  ( log )
Formula: Counts ~ Drug * Time - Drug + (1 | ID)
   Data: data.

     AIC      BIC   logLik deviance df.resid
   136.8    147.3    -63.4    126.8       55

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8654 -0.6068  0.0646  0.4641  1.8851

Random effects:
 Groups Name        Variance Std.Dev.
 ID     (Intercept) 0.262    0.512
Number of obs: 60, groups:  ID, 30

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.5976     0.1624   16.00  < 2e-16 ***
Time         -0.0739     0.1040   -0.71  0.47748
DrugA:Time   -0.4626     0.1383   -3.35  0.00082 ***
DrugB:Time   -0.4031     0.1352   -2.98  0.00287 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The estimated variance of the random intercepts is relatively low: $\sigma_b^2 = 0.262$

GLMM - Example: Leprosy data

- We compare our GLMM results for the Leprosy data with those from fitting GEE's (exchangeable working correlation):

	GLMM	GEE
Parameter	Estimate (s.e.)	Estimate (s.e.)
Intercept	2.600 (0.162)	2.373 (0.080)
Time	−0.074 (0.104)	−0.014 (0.157)
TreatA × Time	−0.462 (0.138)	−0.541 (0.219)
TreatB × Time	−0.403 (0.135)	−0.479 (0.228)

- We observe smaller differences in the size of the estimates compared to the Onychomycosis study $\Rightarrow$ the estimated random-effects variance is rather low.

Marginal Versus Random-effects Models

- The two approaches differ on the way they model the within-subjects correlations, the assumptions about missing data and interpretation of parameters.
- Choice between marginal and random-effects model depends on inferential goal $\Rightarrow$ they reply to different questions.
  - Marginal models are used to make inferences about population averages.
  - For mixed models the main focus is on the individual and the influence of covariates on the individual.
- This leads to differences in parameter interpretation.
- Missing data:
  - GLMM valid inferences under MCAR and MAR, whereas GEE only under MCAR.