Linear mixed effect models

Ibuprofens influence on fractured wrists compared to other painkillers

Masters thesis

Gruppe 5.220d

31. May 2018

Aalborg University
Department of Mathematical Sciences
Skjernvej 4A
DK-9220 Aalborg Ø

This master thesis is written in LaTeX

School of Engineering and Science
Department of Mathematical Sciences
Skjernvej 4A
DK-9220 Aalborg
http://www.math.aau.dk

AALBORG UNIVERSITY

STUDENT REPORT

Title:
Linear mixed effect models

Subject:
Ibuprofens influence on fractured wrists compared to other painkillers

Project period:
Spring 2018

Project group:
Group 5.220d

Participants:
Nikolaj Beck Nielsen

Supervisor:
Jakob Gulddahl Rasmussen

Printed copies: 2

Number of pages: 44

Date of hand in:
May 31 2018

Synopsis:

The purpose of this report is to set up a model which can determine, whether the current used painkiller treatment after operation on a broken wrist is better than other alternatives. The report starts with a chapter where the given datasets are described, and how missing values are handled. Hereafter is a theoretical chapter where firstly random effect model are introduced, to better understand mixed effect model which are introduced afterwards, and is the model used in this report. Two models are set up in this report, one with the improvement in moveability 6 weeks, 3 months and a year after the operation on the broken wrist as response. And one with the pain felt in the 14 days after the operation as response. In both models we conclude that there is no difference between the current used painkiller treatment and the other two options, hence the doctors can use one of the other treatments without affecting the patients.

Preface

This report is written by a masters student on Mathematics and statistics at Aalborg university. The report is written as documentation for my master thesis. The subject of the report is linear mixed effect models. The aim of the report is to determine whether the current used painkiller treatment is better, that other alternatives on patients who had broken one of their wrists. It is assumed the reader have knowledge about fixed effect model.

I would like to thank Jakob Gulddahl Rasmussen for constructive criticism and supervision throughout the project period. Further I would like to thank Aalborg university hospital for supplying the dataset.

Resumé

I dette projekt fik jeg udleveret et datasæt med målinger på 83 patienter, som havde brækket et af deres håndled. Hver patients bevægelighed er målt 3 gange over tid. Og deres smerte er målt i 14 dage efter operationen. Målet med rapporten er at finde ud af, hvilken af tre typer smertebehandling der er bedst. For at finde ud af hvilken af behandlingerne der er bedst ser jeg på, hvor meget bevæglighed patienterne har fået tilbage efter hhv. 6 uger, 3 måneder og 1 år, og på hvor meget smerte en patient havde i 14 dage efter operation på håndledet.

Først er der et kapitel, hvor det givne datasæt bliver beskrevet, og hvordan manglende værdier imputeres. Derefter er der et kapitel, hvori teori omkring tilfældig effekt modeller beksrives, og hvordan disse sammen med fixed effekt bliver til mixed effekt modeller. Modeltjek laves med visuel residual analyse, hvorfor der laves et simulation studie sådan, at vi kan få en idé omkring hvordan vi kan forvente residualerne ser ud når modellen er specificeret korrekt. Dernæst er der lavet en foranalyse, hvor der er lavet nogle simple plots af data for at få en idé omkring, hvad vi kan forvente af dataanalysen. Og tilsidst er der en et kapitel med behandling af det givne datasæt.

Reading guide

All plots and calculations are made in the statistical program R. Some functions have been implemented into R, and the code for some of these can be found in the Appendix. Further there are some figures in the appendix which are not essential for the report, but might be interesting for the reader to see.

All chapters have a short introduction explaining the purpose of the chapter. Each chapter can consist of sections and subsections, which also have a short introduction explaining the purpose of these. Further any literature used in each chapter, section or subsection will be referred to in the introductions.

The notation used in this report will be explained below.

- Stochastic vectors and matrices are noted with bold capital letters. It can be understood from the context whether it is a matrix or a stochastic vector.
- All vectors are column vectors.
- In calculations references to equations can occur as follows:

$$a \overset{(4.10)}{=} b^c + d,$$

  here it means that equation (4.10) is used to rewrite the left side of the equation to the right side.

- All proofs are ended with a ■

Contents

1 Introduction **1**

2 Data description **2**

3 Mixed models **5**
3.1 One-way random model 5
3.1.1 Distribution of the random model 6
3.1.2 Parameter estimation 7
3.1.3 Parameter estimation using orthogonal projection 10
3.2 Linear mixed effect models 12
3.2.1 Fixed effects and variance parameter estimation 12
3.2.2 Restricted maximum likelihood estimate 13
3.2.3 Random effects estimate 15
3.2.4 Estimation of $\boldsymbol{\beta}$ and $\mathbf{u}$ simultaneously 16
3.3 Model comparison 18
3.3.1 Fixed effects 18
3.3.2 Random effects 19
3.4 Residuals 20
3.4.1 Marginal residuals 21
3.4.2 Conditional residuals 21
3.4.3 BLUPs 22

4 Simulation Study **23**

5 Preliminary analysis **25**
5.1 Rotation response 25
5.2 Pain response 28

6 Data analysis **30**
6.1 Moveability 30
6.2 Pain 34

7 Discussion **36**

8 Conclusion **37**

Appendices **38**

A Extra figures **38**

B R code **41**

1 Introduction

I was given a dataset containing measurements on some patients, who had surgery after they had broken one of their wrists. Each patient was measured three times, once 6 weeks after the operation, once 3 months after the operation and once one year after the operation. This was done so in order to see the improvement in moveability of the broken wrist. Further each patient filled out a pain journal, where they had to note on a scale of 1 - 10 how much pain they felt each day in the 14 days following the operation. The purpose of this project is to see whether the current used painkiller treatment is better than other alternatives, or if the treatments have the same effects on the moveability and on the pain. The doctors at least hope that the treatments have the same effects, since the current used treatments can have some side effects which the others does not have.

To analyse the data I will use linear mixed models, which is a useful class of models when datasets contain repeated measurements on the same subject.

Below is a short overview of the content of this report:

- Chapter 2: The given dataset will be described.
- Chapter 3: Theory about linear mixed models will be described.
- Chapter 4: A simulation study to get an idea about what to expect from the residuals of the models.
- Chapter 5: A preliminary analysis of the given dataset, in order to get an idea about how the data behaves.
- Chapter 6: The theory in chapter 3 will be used to analyse the given dataset.
- Chapter 7: The approaches and choices made throughout the report is discussed here.
- Chapter 8: Conclusion of the report.
- Appendix A: Some extra figures which are not important to report, but the reader might want to see.
- Appendix B: The R code to some of the implemented functions.

2 Data description

In this project I was given two datasets containing measurements on some who had broken one of their wrists, and undergone an operation. All the wrists had the same break and the same operation was performed on all of them. After the operation the patients were randomly selected to receive one of three type of painkiller medicine. The first group, group 1, was prescribed 600mg $\times$ 3 Ibuprofen each day for seven days, the second group, group 2, was prescribed 600mg $\times$ 3 Ibuprofen each day for three days and three placebo pill each day the remaining four days and the last group, group 3, was prescribed three placebo pills each day for seven days. Note in this report placebo is another type of painkiller. The purpose of this project is to see if the current used painkiller treatment is better or worse than other alternatives. One way to determine this if by the moveability improvement of the broken wrist which is measured 6 weeks, 3 months and 1 year after the operation, the first dataset contains these measurements. Some people are more flexible than others hence the healthy wrist of each patient is used as baseline. There are three pairs of movements supination and pronation which can be seen in figure 1, dorsal and volar flexion which can be seen in figure 2 and radial and ulnar deviation which can be seen in figure 3. The measured degrees of movability are all measured with five degrees of accuracy. The sum of each pair of movements, and how much in percentage the patient is able to flex in each direction compared to the baseline is also given.

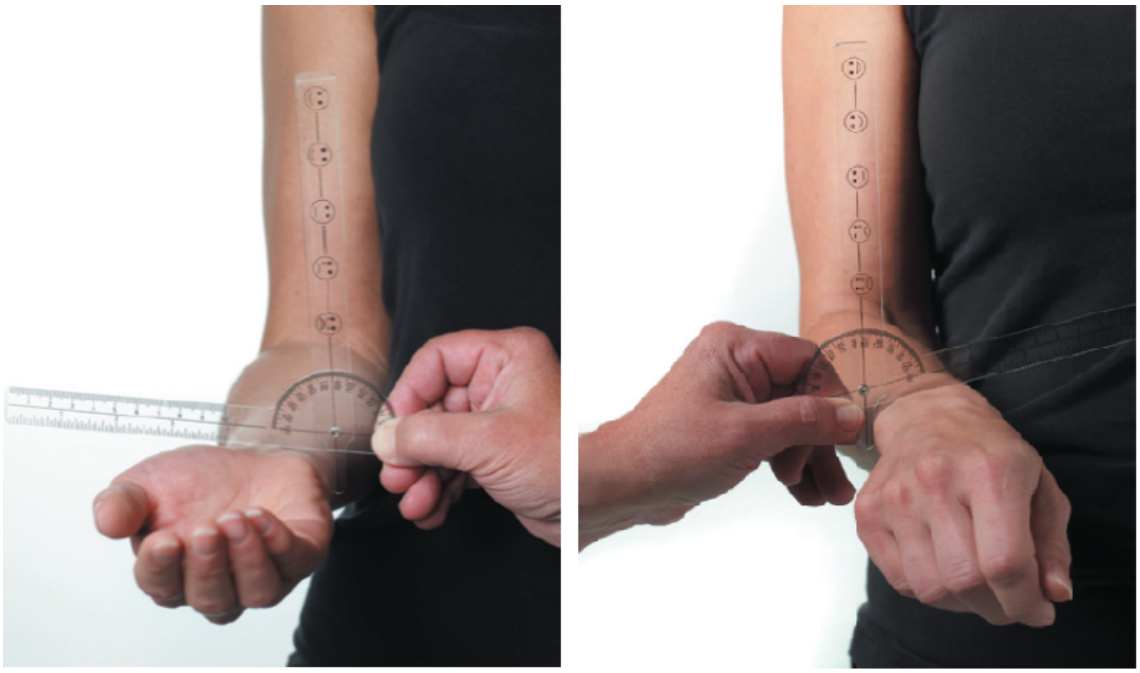

Figure 1: Left picture: Fully supination flexed. Right picture: Fully pronation flexed.

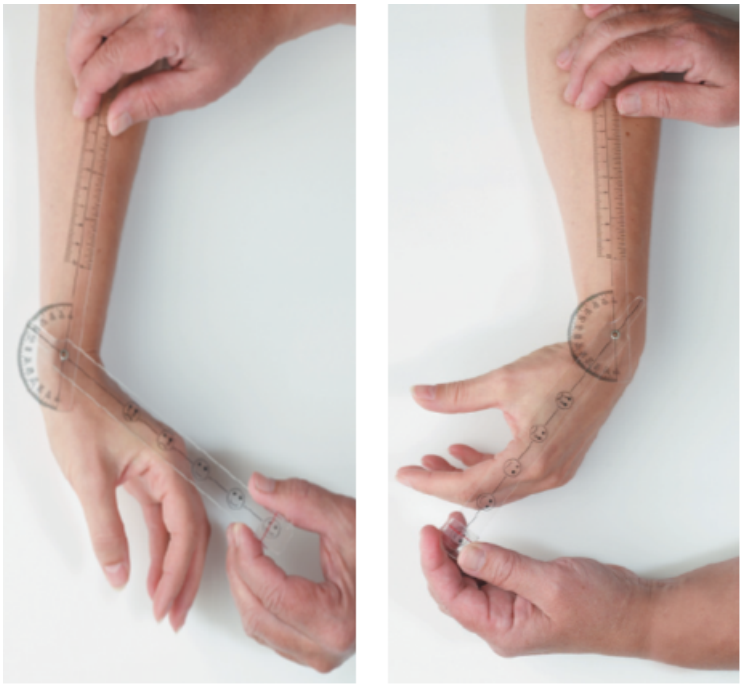

Figure 2: Left pricture: Fully dorsal flexed. Right picture: Fully volar flexed.

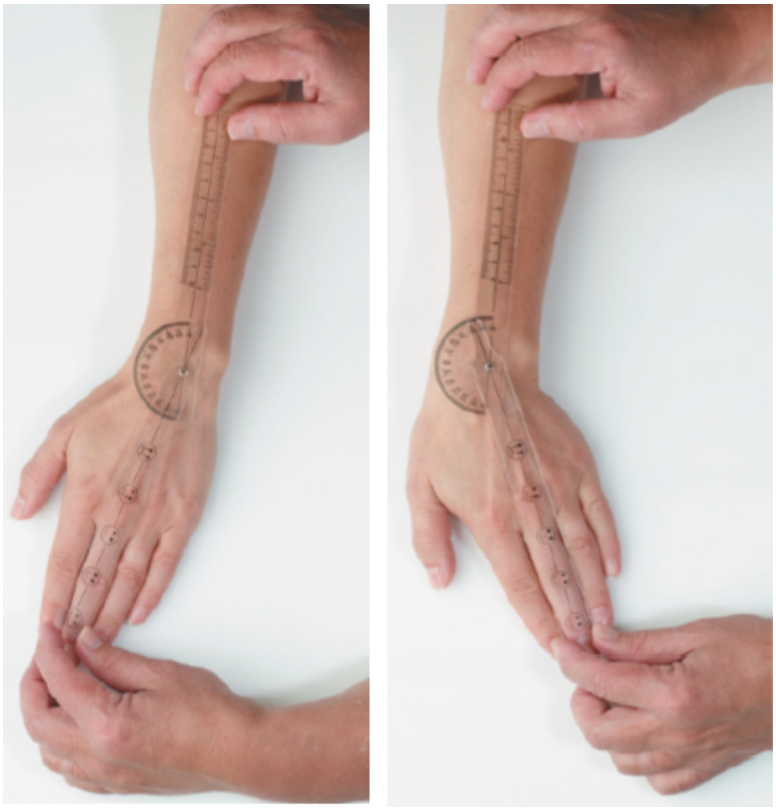

Figure 3: Left picture: Fully radial flexed. Right picture: Fully ulnar flexed.

Figure 1, 2 and 3 are taken from [Aliuskevicius, 2017].

Another approach to see which group is best is to see how much pain the patients felt, i.e. which group experience the least amount of pain. In the 14 days after the operation the patients had to write down, in a pain journal, how much pain, on a scale of 1-10, they were in each day, this is also given in the second dataset.

There are 83 patients in both datasets, 28 in the first treatment group, 28 in the second treatment group and 27 in the third treatment group. We also have the age, sex and KIT.Nr for the patients in both datasets, KIT.Nr is the patients observation number. All the patients are in the age group 40-85 years. Most of the patients are the same in both dataset, but there are a few patients who differs.

In case a patient was in too much pain and the prescribed medicine was not enough, they could take Tradolan. The patient had to write down how many Tradolan they took each day for 14 days after the operation, this is given in the second data set. And the total number of Tradolan pills a patient took is given in the first dataset

Note that when we use the rotations as response we have 3 measurements pr. subject. Whereas when we use pain as the response we have 14 measurements pr. subject.

There are a few missing values and these have been imputed differently, depending on which response the covariate belongs to. There are three missing values in total Tradolan usage in the first dataset. These missing values have been imputed simply by the group average of the group the patient belongs to. And there are three missing values in the pain felt on day 14 in the second data set. These values have been imputed by the average pain the patient felt from day 7 to day 14.

3 Mixed models

Mixed models also called mixed effect models is a class of models for analysing *grouped* data. A form of grouping is repeated measures, where an observation is measured multiple times. The set-up could be as follows

Group	Observations
1	$\mathbf{Y}_{11}, \mathbf{Y}_{12}, \ldots, \mathbf{Y}_{1n_1}$
2	$\mathbf{Y}_{21}, \mathbf{Y}_{22}, \ldots, \mathbf{Y}_{2n_2}$
$\vdots$	$\vdots \quad , \quad \vdots$
k	$\mathbf{Y}_{k1}, \mathbf{Y}_{k2}, \ldots, \mathbf{Y}_{kn_k}$

where k is the classification into the groups and n_i is the number of observations in each group. If there are the same number of observations in each group, i.e. $n_1 = n_2 = \ldots = n_k = n$, we say the experiment is *balanced*. The name mixed models comes from the fact that the models consists of fixed and random effects, note the above set-up is not a mixed set-up. In order to understand random effects we will in the next section look at the one-way random model. This chapter is largely based on [Madsen & Thyregod, 2011] and lecture notes from [Waagepetersen, 2018(1)], with extra details from [Pawitan, 2013].

3.1 One-way random model

Assuming we have $i = 1, \ldots, k$ groups and $j = 1, \ldots, n_i$ observations per group, the fixed effect model will look like

$$\mathbf{Y}_{ij} = \mu + \alpha_i + \boldsymbol{\epsilon}_{ij},$$

where $\boldsymbol{\epsilon}_{ij} \sim \mathcal{N}(0, \sigma^2)$. In this case there are many parameters $(\mu, \sigma^2, \alpha_2, \ldots, \alpha_k)$, where the parameters $\alpha_2, \ldots, \alpha_k$ might not be of interest individually but the variation between them is. Note that α_1 is set as reference and is included in μ. However, if we specify the model

$$\mathbf{Y}_{ij} = \mu + \mathbf{U}_i + \boldsymbol{\epsilon}_{ij},$$

where $\mathbf{U}_i \sim \mathcal{N}(0, \sigma_u^2)$, we only have three parameters (μ, σ^2 and $\text{Var}[\mathbf{U}_i] = \sigma_u^2$). This is called the random effect model.

Definition 3.1 (One-way random model) Assuming we have the random variables $\mathbf{Y}_{ij}$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, n_i$. A one-way random model for $\mathbf{Y}_{ij}$ is defined as

$$\mathbf{Y}_{ij} = \mu + \mathbf{U}_i + \boldsymbol{\epsilon}_{ij},$$

where $\boldsymbol{\epsilon}_{ij} \sim \mathcal{N}(0, \sigma^2)$ are mutually independent, $\mathbf{U}_i \sim \mathcal{N}(0, \sigma_u^2)$ are mutually independent and $\mathbf{U}_i$ is independent of $\boldsymbol{\epsilon}_{ij}$. Further more, we let

$$N = \sum_{i=1}^{k} n_i.$$

If the experiment is balanced, i.e. $n_1 = n_2 = \ldots = n_k$, we say the model is balanced.

Sometimes we use $\gamma = \sigma_u^2/\sigma^2$ instead of σ_u^2, which we call signal to noise ratio. The parameter γ describes the inhomogeneity between groups relative to the variation in the groups.

As mentioned in the beginning of this section random effect models are used in situations, where the interest is not the covariates at hand, but where the covariates are considered as a random sample from a larger population. Hence, it is the variation between groups, σ_u^2, which is of interest, and not the results of the individual groups, α_i, as in fixed effect models. In the next section we'll see how $\mathbf{Y}_{ij}$ is distributed.

3.1.1 Distribution of the random model

In this subsection the marginal distribution of the random effects model will be specified in the one-way ANOVA case

Theorem 3.2 (Marginal distribution of the random effects model) The marginal distribution of $\mathbf{Y}_{ij}$ is normal with mean and variance:

$$\begin{aligned} \mathrm{E}[\mathbf{Y}_{ij}] &= \mu \\ \mathrm{Cov}[\mathbf{Y}_{ij}, \mathbf{Y}_{hl}] &= \begin{cases} \sigma_u^2 + \sigma^2, & \text{for } i = h, \quad j = l \\ \sigma_u^2, & \text{for } i = h, \quad j \neq l \\ 0, & \text{for } i \neq h. \end{cases} \end{aligned}$$

Proof:
The expected value of $\mathbf{Y}_{ij}$ follows trivially from definition 3.1 hence it is omitted. We will look at the three cases of the covariance individually. For $i = h$ and $j = l$

$$\begin{aligned} \mathrm{Cov}[\mathbf{Y}_{ij}, \mathbf{Y}_{hl}] &= \mathrm{E}[(\mathbf{Y}_{ij} - \mathrm{E}[\mathbf{Y}_{ij}])\,(\mathbf{Y}_{ij} - \mathrm{E}[\mathbf{Y}_{ij}])] = \mathrm{E}[(\mathbf{Y}_{ij} - \mathrm{E}[\mathbf{Y}_{ij}])^2] \\ &= \mathrm{Var}[\mathbf{Y}_{ij}] = \mathrm{Var}[\mathbf{U}_i] + \mathrm{Var}[\boldsymbol{\epsilon}_{ij}] = \sigma_u^2 + \sigma^2. \end{aligned}$$

For $i = h$ and $j \neq l$

$$\begin{aligned} \mathrm{Cov}[\mathbf{Y}_{ij}, \mathbf{Y}_{hl}] &= \mathrm{Cov}[\mu + \mathbf{U}_i + \boldsymbol{\epsilon}_{ij}, \mu + \mathbf{U}_i + \boldsymbol{\epsilon}_{il}] = \mathrm{Cov}[\mathbf{U}_i + \boldsymbol{\epsilon}_{ij}, \mathbf{U}_i + \boldsymbol{\epsilon}_{il}] \\ &= \mathrm{Cov}[\mathbf{U}_i, \mathbf{U}_i] + \mathrm{Cov}[\mathbf{U}_i, \boldsymbol{\epsilon}_{il}] + \mathrm{Cov}[\boldsymbol{\epsilon}_{ij}, \mathbf{U}_i] + \mathrm{Cov}[\boldsymbol{\epsilon}_{ij}, \boldsymbol{\epsilon}_{il}] \\ &= \mathrm{Var}[\mathbf{U}_i] = \sigma_u^2, \end{aligned}$$

where the rest of the terms are zero due to independence. For $i \neq h$

$$\mathrm{Cov}[\mathbf{Y}_{ij}, \mathbf{Y}_{hl}] = \mathrm{E}[\mathbf{Y}_{ij}\mathbf{Y}_{hl}] - \mathrm{E}[\mathbf{Y}_{ij}]\mathrm{E}[\mathbf{Y}_{hl}] = \mathrm{E}[\mathbf{Y}_{ij}]\mathrm{E}[\mathbf{Y}_{hl}] - \mathrm{E}[\mathbf{Y}_{ij}]\mathrm{E}[\mathbf{Y}_{hl}] = 0.$$

■

From theorem 3.2 it is easily seen that

$$\mathrm{Corr}[\mathbf{Y}_{ij}, \mathbf{Y}_{hl}] = \begin{cases} 1, & \text{for } i = h, \quad j = l \\ \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2} & \text{for } i = h, \quad j \neq l \\ 0, & \text{for } i \neq h. \end{cases}$$

we can now find the distribution of the individual group averages

$$\begin{aligned}\mathrm{E}[\bar{\mathbf{Y}}_i.] &= \mu \\ \mathrm{Cov}[\bar{\mathbf{Y}}_i., \bar{\mathbf{Y}}_h.] &= \begin{cases}\sigma_u^2 + \sigma^2/n_i, & \text{for } \ i = h \\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

where we note that the group averages are mutually independent, and the variance of the group averages depends on the variance of the random effect as well as the variance of the residuals scaled with the number of observations in the groups.

There are three types of sum of squares in random effect models, there are the error within groups noted SSE, the error between groups noted SSB, and the total error noted SST.

$$\begin{aligned}\mathrm{SSE} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i.)^2 \\ \bar{\mathbf{y}}.. &= \sum_{i=1}^{k} n_i\bar{\mathbf{y}}_i./N \\ \mathrm{SSB} &= \sum_{i=1}^{k} n_i(\bar{\mathbf{y}}_i. - \bar{\mathbf{y}}..)^2 \\ \mathrm{SST} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}..)^2 = \mathrm{SSB} + \mathrm{SSE},\end{aligned}$$

In the balanced case the error have the following distribution

$$\mathrm{SSE} \sim \sigma^2\chi^2(k(n-1)) \quad \text{and} \quad \mathrm{SSB} \sim \left(\frac{\sigma^2}{\frac{1}{1+n\gamma}}\right)\chi^2(k-1).$$

These error distributions will be derived in subsection 3.1.3, when the necessary notation have been introduced. We will now continue on to parameter estimation.

3.1.2 Parameter estimation

The parameters needed to be estimated are the fixed parameters, μ, σ^2 and σ_u^2 or γ. We will in this subsection see how to estimate these.

The given dataset described in chapter 2 is balanced since we have the same number of measurements for each patient. Hence we will in the reset of this report only focus on theory for balanced experiments.

Theorem 3.3 (Estimation of parameters in random models) The moment estimates of μ, σ^2 and σ_u^2 are

$$\begin{aligned}\hat{\mu} &= \bar{\mathbf{Y}}..\\ \hat{\sigma^2} &= \text{SSE}/(N-k)\\ \hat{\sigma_u^2} &= \frac{\text{SSB}/(k-1) - \hat{\sigma}^2}{n}.\end{aligned}$$

These estimates are unbiased.

Proof (of unbiasness):

$$\text{E}[\text{SSE}/(N-k)] = \text{E}\left[\frac{\sum_{i=1}^k \sum_{j=1}^n (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i}.)^2}{N-k}\right] = \frac{\text{E}\left[\sum_{i=1}^k \sum_{j=1}^n (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i}.)^2\right]}{N-k}$$

We will ignore the constant for now and focus on the numerator

$$\begin{aligned}
&\text{E}\left[\sum_{i=1}^k \sum_{j=1}^n (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i}.)^2\right] = \text{E}\left[\sum_{i=1}^k \sum_{j=1}^n (\mathbf{Y}_{ij}^2 + \bar{\mathbf{Y}}_{i}.^2 - 2\mathbf{Y}_{ij}\bar{\mathbf{Y}}_{i}.)\right]\\
&= \sum_{i=1}^k \sum_{j=1}^n \left(\text{E}\left[\mathbf{Y}_{ij}^2\right] + \text{E}\left[\bar{\mathbf{Y}}_{i}.^2\right] - 2\text{E}\left[\mathbf{Y}_{ij}\bar{\mathbf{Y}}_{i}.\right]\right)\\
&= \sum_{i=1}^k \sum_{j=1}^n \left(\text{Var}[\mathbf{Y}_{ij}] + \text{E}[\mathbf{Y}_{ij}]^2 + \text{Var}[\bar{\mathbf{Y}}_{i}.] + \text{E}[\bar{\mathbf{Y}}_{i}.]^2 - 2\text{E}\left[\mathbf{Y}_{ij}\frac{1}{n}\sum_{l=1}^n \mathbf{Y}_{il}\right]\right)\\
&= \sum_{i=1}^k \sum_{j=1}^n \left(\text{Var}[\mathbf{Y}_{ij}] + \text{E}[\mathbf{Y}_{ij}]^2 + \text{Var}[\bar{\mathbf{Y}}_{i}.] + \text{E}[\bar{\mathbf{Y}}_{i}.]^2 - 2\frac{1}{n}\text{E}\left[\mathbf{Y}_{ij}^2 + \mathbf{Y}_{ij}\sum_{l\neq j} \mathbf{Y}_{il}\right]\right)\\
&= \sum_{i=1}^k \sum_{j=1}^n \left(\sigma^2 + \sigma_u^2 + \mu^2 + \sigma_u^2 + \frac{\sigma^2}{n} + \mu^2 - 2\frac{1}{n}\text{E}[\mathbf{Y}_{ij}^2] - 2\frac{1}{n}\text{E}\left[\mathbf{Y}_{ij}\sum_{l\neq j} \mathbf{Y}_{il}\right]\right)\\
&= \sum_{i=1}^k \sum_{j=1}^n \left(\sigma^2(1+1/n) + 2\sigma_u^2 + 2\mu^2 - \frac{2}{n}\text{Var}[\mathbf{Y}_{ij}] - \frac{2}{n}\text{E}[\mathbf{Y}_{ij}]^2 - \frac{2}{n}\sum_{l\neq j}\text{Cov}[\mathbf{Y}_{ij}, \mathbf{Y}_{il}]\right.\\
&\qquad\qquad \left. - \frac{2}{n}\sum_{l\neq j}\text{E}[\mathbf{Y}_{ij}]\text{E}[\mathbf{Y}_{il}]\right)\\
&= \sum_{i=1}^k \sum_{j=1}^n \left(\sigma^2(1+1/n) + 2\sigma_u^2 + 2\mu^2 - \frac{2}{n}(\sigma_u^2 + \sigma^2) - \frac{2}{n}\mu^2 - \frac{2}{n}\sum_{l\neq j}\sigma_u^2 - \frac{2}{n}\sum_{l\neq j}\mu^2\right)\\
&= \sigma^2(N-k) + 2\sigma_u^2(N-k) + 2\mu^2(N-k) - 2\sum_{i=1}^k \sum_{j=1}^n \left(\frac{1}{n}\sum_{l\neq j}\sigma_u^2\right) - 2\sum_{i=1}^k \sum_{j=1}^n \left(\frac{1}{n}\sum_{l\neq j}\mu^2\right)
\end{aligned}$$

$$= \sigma^2(N-k) + 2\sigma_u^2(N-k) + 2\mu^2(N-k) - 2\sum_{i=1}^{k}\left(\sum_{l=1}^{n}(\sigma_u^2) - \sigma_u^2\right) - 2\sum_{i=1}^{k}\left(\sum_{h=1}^{n}(\mu^2) - \mu^2\right)$$
$$= \sigma^2(N-k) + 2\sigma_u^2(N-k) + 2\mu^2(N-k) - 2\sigma_u^2(N-k) - 2\mu^2(N-k) = \sigma^2(N-k)$$

Hence $\hat{\sigma}^2 = \text{SSE}/(N-k)$ is an unbiased estimate for σ^2. And now we will see that $\text{E}[\text{SSB}/(k-1)] = \sigma^2 - n\sigma_u^2$, is also an unbiased estimate of σ^2 under $\mathcal{H}_0 : \sigma_u^2 = 0$.

$$\text{E}[SSB/(k-1)] = \text{E}\left[\sum_{i=1}^{k} n(\bar{\mathbf{Y}}_{i\cdot} - \bar{\mathbf{Y}}_{\cdot\cdot})^2/(k-1)\right] = \frac{n}{k-1}\text{E}\left[\sum_{i=1}^{k}(\bar{\mathbf{Y}}_{i\cdot} - \bar{\mathbf{Y}}_{\cdot\cdot})^2\right].$$

We will again ignore the constant and focus on the expected value.

$$\text{E}\left[\sum_{i=1}^{k}(\bar{\mathbf{Y}}_{i\cdot} - \bar{\mathbf{Y}}_{\cdot\cdot})^2\right] = \text{E}\left[\sum_{i=1}^{k}(\bar{\mathbf{Y}}_{i\cdot}{}^2 + \bar{\mathbf{Y}}_{\cdot\cdot}{}^2 - 2\bar{\mathbf{Y}}_{i\cdot}\bar{\mathbf{Y}}_{\cdot\cdot})\right]$$
$$= \sum_{i=1}^{k}(\text{E}[\bar{\mathbf{Y}}_{i\cdot}{}^2] + \text{E}[\bar{\mathbf{Y}}_{\cdot\cdot}{}^2] - 2\text{E}[\bar{\mathbf{Y}}_{i\cdot}\bar{\mathbf{Y}}_{\cdot\cdot}])$$
$$= \sum_{i=1}^{k}(\text{Var}[\bar{\mathbf{Y}}_{i\cdot}] + \text{E}[\bar{\mathbf{Y}}_{i\cdot}]^2 + \text{Var}[\bar{\mathbf{Y}}_{\cdot\cdot}] + \text{E}[\bar{\mathbf{Y}}_{\cdot\cdot}]^2 - 2\text{E}[\bar{\mathbf{Y}}_{i\cdot}\bar{\mathbf{Y}}_{\cdot\cdot}])$$
$$= \sum_{i=1}^{k}\left(\sigma_u^2 + \sigma^2/n + \mu^2 + \sigma_u^2/k + \sigma^2/N + \mu^2 - 2\text{E}\left[\bar{\mathbf{Y}}_{i\cdot}\frac{1}{k}\sum_{h=1}^{k}\bar{\mathbf{Y}}_{h\cdot}\right]\right)$$
$$= \sum_{i=1}^{k}\left(\sigma^2(1/n + 1/N) + \sigma_u^2(1 + 1/k) + 2\mu^2 - \frac{2}{k}\text{E}\left[\bar{\mathbf{Y}}_{i\cdot}{}^2 + \bar{\mathbf{Y}}_{i\cdot}\sum_{h\neq i}^{k}\bar{\mathbf{Y}}_{h\cdot}\right]\right)$$
$$= \sum_{i=1}^{k}\left(\sigma^2(1/n + 1/N) + \sigma_u^2(1 + 1/k) + 2\mu^2 - \frac{2}{k}\text{E}\left[\bar{\mathbf{Y}}_{i\cdot}{}^2\right] - \frac{2}{k}\sum_{h\neq i}^{k}\text{E}\left[\bar{\mathbf{Y}}_{i\cdot}\bar{\mathbf{Y}}_{h\cdot}\right]\right)$$
$$= \sum_{i=1}^{k}\left(\sigma^2(1/n + 1/N) + \sigma_u^2(1 + 1/k) + 2\mu^2 - \frac{2}{k}\text{Var}\left[\bar{\mathbf{Y}}_{i\cdot}\right] - \frac{2}{k}\text{E}\left[\bar{\mathbf{Y}}_{i\cdot}\right]^2 \right.$$
$$\left. - \frac{2}{k}\sum_{h\neq i}^{k}\text{Cov}\left[\bar{\mathbf{Y}}_{i\cdot}, \bar{\mathbf{Y}}_{h\cdot}\right] - \frac{2}{k}\sum_{h\neq i}^{k}\text{E}[\bar{\mathbf{Y}}_{i\cdot}]\text{E}[\bar{\mathbf{Y}}_{h\cdot}]\right)$$
$$= \sum_{i=1}^{k}\left(\sigma^2(1/n + 1/N) + \sigma_u^2(1 + 1/k) + 2\mu^2 - \frac{2}{k}(\sigma_u^2 + \sigma^2/n) - \frac{2}{k}\mu^2 - \frac{2(k-1)}{k}\mu^2\right)$$
$$= \sum_{i=1}^{k}\left(\sigma^2\left(\frac{1}{n} + \frac{1}{N} - \frac{2}{N}\right) + \sigma_u^2\left(1 + \frac{1}{k} - \frac{2}{k}\right)\right) = \sum_{i=1}^{k}\left(\sigma^2\frac{k-1}{N} + \sigma_u^2\frac{k-1}{k}\right).$$

Now remembering the constant from the beginning we get

$$\frac{n}{k-1}\sum_{i=1}^{k}\left(\sigma^2\frac{k-1}{N} + \sigma_u^2\frac{k-1}{k}\right) = \sigma^2 + n\sigma_u^2$$

■

3.1.3 Parameter estimation using orthogonal projection

We can also base our maximum likelihood estimate on orthogonal projection. We'll start by looking at the general linear model setup. Suppose $\mathbf{Y} \sim \mathcal{N}_N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ with $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$. Let $\mathbf{P}$ denote the orthogonal projection on $\text{span}\{\mathbf{X}\}$, assuming $\mathbf{X}$ has full rank then $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. By Pythagoras $\|\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\|^2 = \|\mathbf{Y}-\mathbf{P}\mathbf{Y}\|^2 + \|\mathbf{P}\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\|^2$. From this we can estimate $\hat{\boldsymbol{\mu}} = \mathbf{P}\mathbf{y}$, $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ and $\hat{\sigma}^2 = \|\mathbf{Y} - \mathbf{P}\mathbf{Y}\|^2/N = \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2/N$.

If we consider the One-way random model from definition 3.1 in matrix form

$$\mathbf{Y} = \mu\mathbf{1}_N + \mathbf{Z}_F\mathbf{U} + \boldsymbol{\epsilon},$$

where N is the total number of observations and F is a factor which assigns observations into k groups with $\mathbf{Z}_F$ the design matrix corresponding to F, i.e. the ij,q'th entry is 1 if $\mathbf{Y}_{ij}$ is in the q'th group and zero otherwise. The orthogonal projection $\mathbf{P}_F$ on $L_F = \text{span}\{\mathbf{Z}_F\}$ is

$$\mathbf{P}_F = \mathbf{Z}_F(\mathbf{Z}_F^T\mathbf{Z}_F)^{-1}\mathbf{Z}_F^T = \frac{1}{n}\mathbf{Z}_F\mathbf{Z}_F^T.$$

Where $(\mathbf{Z}_F^T\mathbf{Z}_F)^{-1} = (1/n)\mathbf{I}_k$. We have two special factors; the unit factor $\mathbf{I}$ which has a unique level for each observation, hence $L_\mathbf{I} = \mathbb{R}^N$ and $\mathbf{P}_\mathbf{I} = \mathbf{I}$, note that $\mathbf{I}$ is used both for the factor and the identity matrix. And the factor 0 which assign all observations to the same group, hence $L_0 = \text{span}\{\mathbf{1}_N\}$ and $\mathbf{P}_0 = \mathbf{1}_N\mathbf{1}_N^T/N$. It can be noted that $L_0 \subset L_F \subset L_\mathbf{I}$. We can now make an orthogonal decomposition of $\mathbb{R}^N$

$$\mathbb{R}^N = V_0 \oplus V_F \oplus V_\mathbf{I},$$

where $V_0 = L_0$, $V_F = L_F \ominus V_0$ and $V_\mathbf{I} = \mathbb{R}^N \ominus L_F$. The dimensions of V_0, V_F and $V_\mathbf{I}$ is 1, $k-1$ and $N-k$ respectively. The orthogonal projections on V_0, V_F and $V_\mathbf{I}$ are $\mathbf{Q}_0 = \mathbf{P}_0$, $\mathbf{Q}_F = \mathbf{P}_F - \mathbf{P}_0$ and $\mathbf{Q}_\mathbf{I} = \mathbf{I} - \mathbf{P}_F$ respectively. The covariance of $\mathbf{Y}$ is given by

$$\text{Cov}[\mathbf{Y}] = \mathbf{Z}_F\sigma_u^2\mathbf{Z}_F^T + \sigma^2\mathbf{I} = \sigma_u^2 n\mathbf{P}_F + \sigma^2\mathbf{I} = \lambda\mathbf{P}_F + \sigma^2\mathbf{Q}_\mathbf{I},$$

where $\lambda = n\sigma_u^2 + \sigma^2$ and $\mathbf{I} = \mathbf{P}_F + \mathbf{Q}_\mathbf{I}$. We note that $\mathbf{P}_F\mathbf{1}_N = \mathbf{Q}_0\mathbf{1}_N = \mathbf{1}_N$, $\mathbf{Q}_\mathbf{I}\mathbf{1}_N = 0$ and $\mathbf{Q}_\mathbf{I}\mathbf{P}_F = 0$, which gives the following distribution

$$\begin{bmatrix}\mathbf{P}_F \\ \mathbf{Q}_\mathbf{I}\end{bmatrix}\mathbf{Y} = \mathcal{N}\left(\begin{pmatrix}\mathbf{1}_N\mu \\ \mathbf{0}_N\end{pmatrix}, \begin{bmatrix}\lambda\mathbf{P}_F & 0 \\ 0 & \sigma^2\mathbf{Q}_\mathbf{I}\end{bmatrix}\right).$$

Thus we can base MLE of μ and λ on $\mathbf{P}_F\mathbf{Y}$ and MLE of σ^2 on $\mathbf{Q}_\mathbf{I}\mathbf{Y}$. Specifically we get the likelihood (ignoring the constants)

$$\begin{aligned}&|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{1}_N\mu)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{1}_N\mu)\right) \\ &= \lambda^{-k/2}\exp\left(-\frac{1}{2\lambda}||\mathbf{P}_F\mathbf{y}-\mathbf{1}_N\mu||^2\right)\cdot(\sigma^2)^{-k(n-1)/2}\exp\left(-\frac{1}{2\sigma^2}||\mathbf{Q}_\mathbf{I}\mathbf{y}||^2\right),\end{aligned}$$

where $\boldsymbol{\Sigma}^{-1} = \sigma^{-2}\mathbf{Q}_\mathbf{I} + \lambda^{-1}\mathbf{P}_F$ and $|\boldsymbol{\Sigma}| = \lambda^k(\sigma^2)^{N-k}$. If we consider one of the factors in the above likelihood e.g $\lambda^{-k/2}\exp\left(-\frac{1}{2\lambda}||\mathbf{P}_F\mathbf{y}-\mathbf{1}_N\mu||^2\right)$ we see that it only involves the

parameters μ and λ. To find the MLEs of these parameters we can do the same as in the beginning of this subsection where we considered the likelihood of $\mathcal{N}_N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$.

Random effect models have now been introduced, and we will in next section see how to mix these with the fixed effect models, in order to obtain the mixed effect models. However before we introduce mixed models we will derive the distribution of SSE and SSB.

The distribution of SSE and SSB will be derived using orthogonal projections. We have that

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \sigma^2\mathbf{Q_I} + \lambda\mathbf{P}_F$. We have that $\mathbf{Q_I}\mathbf{Q}_F = 0$, where $\mathbf{Q}_F = \mathbf{P}_F - \mathbf{P}_0$. It can be shown that

$$\begin{aligned}
&\text{SSE} = \|\mathbf{Y} - \mathbf{P}_F\mathbf{Y}\|^2 = \|\mathbf{Q_I}\mathbf{Y}\|^2 \\
&\text{and} \\
&\text{SSB} = \|\mathbf{P}_F\mathbf{Y} - \mathbf{P}_0\mathbf{Y}\|^2 = \|\mathbf{Q}_F\mathbf{Y}\|^2.
\end{aligned}$$

We first show that SSE and SSB are independent.

$$\begin{aligned}
\text{Cov}[\mathbf{Q_I}\mathbf{Y}, \mathbf{Q}_F\mathbf{Y}] &= \mathbf{Q_I}\text{Cov}[\mathbf{Y}]\mathbf{Q}_F^T = \mathbf{Q_I}\boldsymbol{\Sigma}\mathbf{Q}_F^T = \mathbf{Q_I}(\sigma^2\mathbf{Q_I} - \lambda\mathbf{P}_F)\mathbf{Q}_F^T \\
&= \sigma^2\mathbf{Q_I}\mathbf{Q}_F^T - \lambda\mathbf{Q_I}\mathbf{P}_F\mathbf{Q}_F^T \\
&= \sigma^2(\mathbf{I} - \mathbf{P}_F)(\mathbf{P}_F - \mathbf{P}_0) - \lambda(\mathbf{I} - \mathbf{P}_F)\mathbf{P}_F(\mathbf{P}_F - \mathbf{P}_0) \\
&= \sigma^2(\mathbf{P}_F - \mathbf{P}_0 - \mathbf{P}_F + \mathbf{P}_0) - \lambda(\mathbf{P}_F - \mathbf{P}_F)(\mathbf{P}_F - \mathbf{P}_0) = 0,
\end{aligned}$$

hence SSE and SSB are independent. To shown they are χ^2-distributed we need the following result:

> **Result 1** Assume $\mathbf{X} \sim \mathcal{N}(0, \sigma^2\mathbf{P})$, where $\mathbf{P}$ is a projection on some space onto a space L of dimension d. Then $\|\mathbf{X}\|^2 \sim \sigma^2\chi^2(d)$.

From earlier we have that $Q_\mathbf{I}\mathbf{Y} \sim \mathcal{N}(0, \sigma^2\mathbf{Q_I})$, then by Result 1 we get that $\|\mathbf{Q_I}\mathbf{Y}\|^2 \sim \sigma^2\chi^2(N-k)$. We will now find the distribution of $\mathbf{Q}_F\mathbf{Y}$.

$$\begin{aligned}
\text{E}[\mathbf{Q}_F\mathbf{Y}] &= \mathbf{Q}_F\mathbf{1}_N\mu = \mathbf{P}_F\mathbf{1}_N\mu - \mathbf{P}_0\mathbf{1}_N\mu = \mathbf{0} \\
\text{Var}[\mathbf{Q}_F\mathbf{Y}] &= \mathbf{Q}_F\text{Var}[\mathbf{Y}]\mathbf{Q}_F = \mathbf{Q}_F(\lambda\mathbf{P}_F + \sigma^2\mathbf{Q_I})\mathbf{Q}_F = \lambda\mathbf{Q}_F\mathbf{P}_F\mathbf{Q}_F + \sigma^2\mathbf{Q}_F\mathbf{Q_I}\mathbf{Q}_F \\
&= \lambda(\mathbf{P}_F - \mathbf{P}_0)\mathbf{P}_F(\mathbf{P}_F - \mathbf{P}_0) + \sigma^2(\mathbf{P}_F - \mathbf{P}_0)(\mathbf{I} - \mathbf{P}_F)(\mathbf{P}_F - \mathbf{P}_0) \\
&= \lambda(\mathbf{P}_F - \mathbf{P}_0 - \mathbf{P}_0 + \mathbf{P}_0) + \sigma^2(\mathbf{P}_F - \mathbf{P}_F - \mathbf{P}_0 + \mathbf{P}_0)(\mathbf{P}_F - \mathbf{P}_0) \\
&= \lambda\mathbf{Q}_F.
\end{aligned}$$

Hence we get that $\mathbf{Q}_F\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \lambda\mathbf{Q}_F)$. We can again use Result 1 and get that $\|\mathbf{Q}_F\mathbf{Y}\| \sim \lambda\chi^2(k-1)$. Lastly it is easily seen that

$$\lambda = \frac{\sigma^2}{\frac{1}{1+n\gamma}}.$$

And we have now shown the distributions of SSE and SSB.

3.2 Linear mixed effect models

In this subsection we will consider a more general class of models namely the *linear mixed effects models*. The random effects models and the fixed effects models are special cases of the linear mixed effects models.

Definition 3.4 (Linear mixed effects model) The model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\epsilon}, \tag{3.1}$$

where $\mathbf{X}$ is a $N \times p$ design matrix and $\mathbf{Z}$ is a $N \times k$ design matrix, $\mathbf{U} \sim \mathcal{N}_k(0, \boldsymbol{\Psi})$, $\boldsymbol{\epsilon} \sim \mathcal{N}_N(0, \boldsymbol{\Sigma})$ and $\mathbf{U}$ and $\boldsymbol{\epsilon}$ are independent, is called a *mixed general linear model*. The covariance matrices, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$, may depend on some unknown parameters, $\boldsymbol{\psi}$, which also needs to be estimated. The parameters $\boldsymbol{\beta}$ are called the fixed effects and the quantities $\mathbf{U}$ are called the random effects.

The model from equation (3.1) has a multivariate normal distribution

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta},\ \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T + \boldsymbol{\Sigma}). \tag{3.2}$$

To ease notation we let V denote the covariance of $\mathbf{Y}$, i.e. $\mathbf{V} = \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T + \boldsymbol{\Sigma}$.

3.2.1 Fixed effects and variance parameter estimation

We can estimate the fixed effect parameters, $\boldsymbol{\beta}$, and the variance parameters, $\boldsymbol{\psi}$, from the marginal distribution of $\mathbf{Y}$ given by equation (3.2). Since $\mathbf{Y}$ is multivariate normal distributed we have the likelihood

$$L(\boldsymbol{\beta}, \boldsymbol{\psi}; \mathbf{y}) = \frac{1}{\sqrt{2\pi|\mathbf{V}|}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right).$$

The log-likelihood is then given by

$$\ell(\boldsymbol{\beta}, \boldsymbol{\psi}; \mathbf{y}) \equiv -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

We can now derive the maximum likelihood estimate for $\boldsymbol{\beta}$

$$\ell'(\boldsymbol{\beta}; \boldsymbol{\psi}) = \mathbf{X}^T(\mathbf{V}^{-1}\mathbf{y} - \mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}),$$

hence we get a *weighted least squares* estimate for $\boldsymbol{\beta}$, given $\boldsymbol{\psi}$ is fixed

$$\begin{aligned}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})\boldsymbol{\beta} &= \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}.\\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}\end{aligned} \tag{3.3}$$

To get an estimate for the variance of $\hat{\boldsymbol{\beta}}$ we see that the observed Fisher information for $\boldsymbol{\beta}$ is

$$I(\hat{\boldsymbol{\beta}}) = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$$

from this we get the estimate

$$\text{Var}[\hat{\boldsymbol{\beta}}] = I(\hat{\boldsymbol{\beta}})^{-1} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$

As mentioned in definition 3.4 the covariance matrices might depend on the unknown variance parameters, $\boldsymbol{\psi}$, hence $\hat{\boldsymbol{\beta}}$ might depend on $\boldsymbol{\psi}$. The profile log-likelihood for $\boldsymbol{\psi}$ is

$$\ell(\boldsymbol{\psi}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

where $\hat{\boldsymbol{\beta}}$ is the estimate found in equation (3.3). But we need to modify the profile log-likelihood for $\boldsymbol{\psi}$ to compensate for the estimate of $\boldsymbol{\beta}$. Hence we get the log-likelihood:

$$\ell_m(\boldsymbol{\psi}) = -\frac{1}{2}\log(|\mathbf{V}|) - \frac{1}{2}\log(|\mathbf{X}^T\mathbf{V}\mathbf{X}|) - \frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{3.4}$$

But if $\hat{\boldsymbol{\beta}}$ depend on $\boldsymbol{\psi}$ we need to iterate to find the solutions of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\psi}}$.

The profile log-likelihood in equation (3.4) equals the *restricted* or *residual maximum likelihood* (*REML*)-method. In the next subsection we'll go into further details about REML and how to derive equation (3.4).

3.2.2 Restricted maximum likelihood estimate

In this subsection we will derive the REML-function. In REML one transforms the data such that nuisance parameters are removed. This subsection is based on [LaMotte, 2007] and [Courrieu, 2009].

In REML one transforms the data vector $\mathbf{y}$ by a matrix $\mathbf{A}^T$ such that $\mathbf{A}^T\mathbf{X} = 0$, hence we get the data vector $\tilde{\mathbf{y}} = \mathbf{A}^T\mathbf{y}$.

Theorem 3.5 Let the data vector $\mathbf{Y}$ be normally distributed with mean $\mathbf{X}\boldsymbol{\beta}$ and variance $\mathbf{V}(\boldsymbol{\psi})$, then the likelihood for $\tilde{\mathbf{Y}}$ is

$$L_{\text{REML}} = (2\pi)^{-(n-p)/2}|\mathbf{A}^T\mathbf{V}\mathbf{A}|^{-1/2}\exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{V}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}\right), \tag{3.5}$$

to ease notation we just write $\mathbf{V}(\boldsymbol{\psi})$ as $\mathbf{V}$. Equation (3.5) can be written as:

$$L_{\text{REML}} = (\text{Const.})|\mathbf{V}|^{-1/2}|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\right), \tag{3.6}$$

where Const. is just some constant.

Note that equation (3.4) is log of equation (3.6). In order to prove theorem 3.5 proportion 3.6 is needed.

Proposition 3.6 If $\mathbf{V}$ is an $n \times n$ positive-definite matrix and $\mathbf{X}$ and $\mathbf{A}$ are defined as above, then

$$\mathbf{V}^{-1} = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1} + \mathbf{A}(\mathbf{A}^T\mathbf{V}\mathbf{A})^{-1}\mathbf{A}^T. \tag{3.7}$$

Proof:
Note that

$$\begin{aligned}\begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}^{-1}\mathbf{V}^{-1}\left(\begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}^T\right)^{-1} &= \left[\begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}^T\mathbf{V}\begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}\right]^{-1} \\ &= \begin{bmatrix}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} & 0 \\ 0 & \mathbf{A}^T\mathbf{V}\mathbf{A}\end{bmatrix}^{-1}\end{aligned}$$

Then by isolating $\mathbf{V}^{-1}$ in the above equation one gets

$$\begin{aligned}\mathbf{V}^{-1} &= \begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}\left[\begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}^T\mathbf{V}\begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}\right]^{-1}\begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}^T \\ &= \begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}\begin{bmatrix}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} & 0 \\ 0 & (\mathbf{A}^T\mathbf{V}\mathbf{A})^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{V}^{-1}\mathbf{X} & \mathbf{A}\end{bmatrix}^T \\ &= \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1} + \mathbf{A}(\mathbf{A}^T\mathbf{V}\mathbf{A})^{-1}\mathbf{A}^T.\end{aligned}$$

■

From proportion 3.6 one gets

$$\mathbf{A}(\mathbf{A}^T\mathbf{V}\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}. \tag{3.8}$$

When $\mathbf{X}$ does not have full rank the estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y},$$

where $\mathbf{G}^-$ is the generalized inverse of $\mathbf{G}$.

Proof of theorem 3.5:

In order to shorten notation we define $\mathbf{B}^- = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^-$ and $\mathbf{M} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$, hence $\hat{\boldsymbol{\beta}} = \mathbf{B}^-\mathbf{M}$.

We will first look at the exponential parts of the likelihoods:

$$\begin{aligned}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= \mathbf{y}^T\mathbf{V}^{-1}\mathbf{y} - \mathbf{y}^T\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} - (\mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{V}^{-1}\mathbf{y} + (\mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y}^T\mathbf{V}^{-1}\mathbf{y} - \mathbf{M}^T\mathbf{B}^-\mathbf{M} - \mathbf{M}^T\mathbf{B}^-\mathbf{M} + \mathbf{M}^T\mathbf{B}^-\mathbf{B}\mathbf{B}^-\mathbf{M} \\ &= \mathbf{y}^T\mathbf{V}^{-1}\mathbf{y} - \mathbf{M}^T\mathbf{B}^-\mathbf{M}, \qquad \text{note } \mathbf{B}^-\mathbf{B}\mathbf{B}^- = \mathbf{B}^- \\ &= \mathbf{y}^T\mathbf{V}^{-1}\mathbf{y} - \mathbf{y}^T\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \\ &= \mathbf{y}^T\left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}^T\mathbf{V}^{-1}\right)\mathbf{y} \\ &\overset{(3.8)}{=} \mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{V}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}\end{aligned}$$

hence the exponential part of the likelihoods are the same. As for the non-exponential part, we let $\mathbf{W} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}$ and $\mathbf{V}_i = \frac{\partial \mathbf{V}}{\partial \psi_i}$. We then get

$$\begin{aligned}\frac{\partial \log|\mathbf{A}^T\mathbf{V}\mathbf{A}|}{\partial \psi_i} &= |\mathbf{A}^T\mathbf{V}\mathbf{A}|^{-1}|\mathbf{A}^T\mathbf{V}\mathbf{A}|\mathrm{tr}((\mathbf{A}^T\mathbf{V}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}_i\mathbf{A}) \\ &= \mathrm{tr}(\mathbf{A}(\mathbf{A}^T\mathbf{V}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}_i) = \mathrm{tr}(\mathbf{W}\mathbf{V}_i)\end{aligned}$$

and

$$\begin{aligned}&\frac{\partial}{\partial \psi_i}\left(\log|\mathbf{V}| + \log\left(|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|\right)\right) \\ &\quad= |\mathbf{V}|^{-1}|\mathbf{V}|\mathrm{tr}(\mathbf{V}^{-1}\mathbf{V}_i) + |\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-1}|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|\mathrm{tr}\left((\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}(-1)\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{X}\right) \\ &\quad= \mathrm{tr}(\mathbf{V}^{-1}\mathbf{V}_i) - \mathrm{tr}((\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{X}) \\ &\quad= \mathrm{tr}(\mathbf{V}^{-1}\mathbf{V}_i) - \mathrm{tr}(\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}_i) \\ &\quad= \mathrm{tr}((\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1})\mathbf{V}_i) = \mathrm{tr}(\mathbf{W}\mathbf{V}_i),\end{aligned}$$

where we have used the result for derivative of determinants:

$$\frac{\partial|\mathbf{Y}|}{\partial x} = |\mathbf{Y}|\mathrm{tr}\left(\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right)$$

Since $\frac{\partial}{\partial \psi_i}\log|\mathbf{A}^T\mathbf{V}\mathbf{A}|$ and $\frac{\partial}{\partial \psi_i}\left(\log|\mathbf{V}| + \log\left(|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|\right)\right)$ are the same up to an additive constant we get that

$$|\mathbf{A}^T\mathbf{V}\mathbf{A}| = (\text{Const.})|\mathbf{V}||\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|.$$

Hence the non-exponential parts are also the same, and we have now shown that equation (3.5) and equation (3.6) are the same.

■

We will now continue with estimation of the random effects.

3.2.3 Random effects estimate

The random effects, $\mathbf{U}$, are not parameters, hence the usual likelihood approach does not make sense to use in order to "estimate" these random effects. But it is still of interest to asses the latent variables. We can formulate the likelihood for all the parameters, by a so called hierarchical likelihood, which is a likelihood based on the joint density of $(\mathbf{Y}, \mathbf{U})$, i.e. the observable and the unobservable random quantities.

$$f(\mathbf{y}, \mathbf{u}; \boldsymbol{\beta}, \boldsymbol{\psi}) = f_{Y|u}(\mathbf{y}; \boldsymbol{\beta}) f_U(\mathbf{u}; \boldsymbol{\psi}),$$

where

$$\begin{aligned}\mathbf{U} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}) \\ \mathbf{Y}|\mathbf{U} = \mathbf{u} &\sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \boldsymbol{\Sigma})\end{aligned}$$

with probability density functions

$$f_U(\mathbf{u};\boldsymbol{\psi}) = \frac{1}{(\sqrt{2\pi})^q|\boldsymbol{\Psi}|}\exp\left(-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Psi}^{-1}\mathbf{u}\right)$$

and

$$f_{Y|u}(\mathbf{y};\boldsymbol{\beta}) = \frac{1}{(\sqrt{2\pi})^N|\boldsymbol{\Sigma}|}\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})\right).$$

Hence the hierarchical log-likelihood, when ignoring the constant terms, is

$$\begin{aligned}\ell(\boldsymbol{\beta},\boldsymbol{\psi},\mathbf{u}) = &-\frac{1}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})\\ &-\frac{1}{2}\log(|\boldsymbol{\Psi}|) - \frac{1}{2}\mathbf{u}^T\boldsymbol{\Psi}^{-1}\mathbf{u}.\end{aligned} \tag{3.9}$$

We can now take the derivative with respect to $\mathbf{u}$ to get an estimate of $\mathbf{u}$

$$\frac{\partial}{\partial\mathbf{u}}\ell(\boldsymbol{\beta},\boldsymbol{\Psi},\mathbf{u}) = \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}) - \boldsymbol{\Psi}^{-1}\mathbf{u}.$$

Setting this equal to zero and solve gives us the estimate

$$\begin{aligned}(\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\boldsymbol{\Psi}^{-1})\mathbf{u} &= \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\\ \hat{\mathbf{u}} &= (\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\boldsymbol{\Psi}^{-1})^{-1}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\end{aligned} \tag{3.10}$$

where $\boldsymbol{\beta}$ is replaced with the estimate $\hat{\boldsymbol{\beta}}$ found in equation (3.3). The estimate $\hat{\mathbf{u}}$ is called *the best linear unbiased predictor* (BLUP), if the variance parameter, $\boldsymbol{\psi}$, has been estimated we call $\hat{\mathbf{u}}$ the empirical BULP (EBLUP). We can take the second derivative of the log-likelihood with respect to $\mathbf{u}$ to asses the uncertainty of the estimate $\hat{\mathbf{u}}$

$$\frac{\partial^2}{\partial\mathbf{u}\partial\mathbf{u}^T}\ell(\boldsymbol{\beta},\boldsymbol{\psi},\mathbf{u}) = -\mathbf{Z}\boldsymbol{\Sigma}^{-1}\mathbf{Z} - \boldsymbol{\Psi}^{-1}.$$

From this we see the observed Fisher information for $\mathbf{u}$ is

$$\boldsymbol{I}(\hat{\mathbf{u}}) = \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \boldsymbol{\Psi}^{-1}$$

we can use this to asses the uncertainty of the estimate $\hat{\mathbf{u}}$. In order to find the values that maximises $\ell(\boldsymbol{\beta},\boldsymbol{\psi},\mathbf{u})$ from equation (3.9), we need to estimate $\boldsymbol{\beta}$ and $\mathbf{u}$ simultaneously for a fixed $\boldsymbol{\psi}$.

3.2.4 Estimation of $\boldsymbol{\beta}$ and u simultaneously

As mentioned above we need to estimate the values of $\boldsymbol{\beta}$ and $\mathbf{u}$ which simultaneously maximises the log-likelihood in equation (3.9) for a fixed $\boldsymbol{\psi}$. Hence from equation (3.9) we have

$$\frac{\partial}{\partial\boldsymbol{\beta}}\ell(\boldsymbol{\beta},\boldsymbol{\psi},\mathbf{u}) = \mathbf{X}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}),$$

equating this to zero we get the following equation:

$$\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}\mathbf{u} = \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}.$$

Combining this with equation (3.10) we get the so-called *mixed model equations*

$$\begin{bmatrix} \mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \boldsymbol{\Psi}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \\ \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \end{bmatrix}$$

We note that these equations estimate $\boldsymbol{\beta}$ and $\mathbf{u}$ without the need to calculate the marginal variance, $\mathbf{V}$, or its inverse. The estimation is done iterative, e.g. by a back-fitting algorithm which will be described in algorithm 3.7

Algoritm 3.7 Back-fitting algorithm

i) Initialize $\boldsymbol{\beta}$ by e.g. OLS

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

ii) Compute an adjusted observation

$$\mathbf{y}^{adj} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

and estimate $\mathbf{u}$ from $\mathbf{y}^{adj} = \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$ the same way as before:

$$\hat{\mathbf{u}} = (\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \boldsymbol{\Psi}^{-1})^{-1}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}^{adj}$$

iii) Update the adjusted observation by

$$\mathbf{y}^{adj} = \mathbf{y} - \mathbf{Z}\mathbf{u}$$

and estimate $\boldsymbol{\beta}$ from $\mathbf{y}^{adj} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ the same way as before

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}^{adj}.$$

We iterate between step ii) and iii) until we have reached convergence.

However if $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ depend on the variance parameter $\boldsymbol{\psi}$ we need to estimate $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Psi}}$ by e.g. REML, and then solve the mixed models equations

$$\begin{bmatrix} \mathbf{X}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X} & \mathbf{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X} & \mathbf{Z}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z} + \hat{\boldsymbol{\Psi}}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y} \\ \mathbf{Z}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y} \end{bmatrix}.$$

And then reestimate $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Psi}}$. We have now seen how to estimate the parameters in a mixed model. In the next section we will see how to compare two models, where one is a sub model of the other, in order determine whether the excluded variable is significant or not.

3.3 Model comparison

When comparing mixed models in order to determine whether the variable is significant or not, we first need to determine whether the variable is fixed or random. There are different ways to test significance of a variable depending on whether is it fixed or random. We will in the next subsections firstly see how to test for significance when the variable is fixed and secondly when it is random.

3.3.1 Fixed effects

When the variable is fixed there are a lot of ways to test whether a variable is significant, we could use a likelihood ratio test statistic which asymptotically follows a χ^2-distribution. However this approach can give misleading results if the sample size is small. Hence we need another way to approximate the distribution of the test statistic in this case. The approximation we will use is the one suggested by Kenward and Roger. The Kenward-Roger approximation builds on a **F** test for reduction of the mean structure. Before going into the Kenward-Roger approximation we will need some notation. This subsection is based on [Halekoh & Højsgaard, 2014]

Assuming we have the model from definition 3.4, where we assume $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$. We then get $\text{Var}[\mathbf{Y}] = \mathbf{V} = \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T + \sigma^2\mathbf{I}$. We want to test whether there is a significant difference between the model in equation (3.1) and the reduced model:

$$\mathbf{Y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{Z}\mathbf{U} + \boldsymbol{\epsilon},$$

where $\mathcal{C}(\mathbf{X}_0) \subset \mathcal{C}(\mathbf{X})$ and $\mathcal{C}$ denotes the column space of $\mathbf{X}$. Further let $d = \dim(\mathcal{C}(\mathbf{X})) - \dim(\mathcal{C}(\mathbf{X}_0))$.

To test reduction of the mean structure from $\text{E}[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$ to $\text{E}[\mathbf{Y}] = \mathbf{X}_0\boldsymbol{\beta}_0$ we can use e.g. the likelihood-ratio test

$$T = 2(\log L - \log L_0),$$

where T asymptotically will follow a χ^2_d distribution under the hypothesis, $\mathcal{H}_0 : \text{E}[\mathbf{Y}] = \mathbf{X}_0\boldsymbol{\beta}_0$. Another way to express the reduction from the larger model to the smaller model is by $\mathbf{L}\boldsymbol{\beta} = 0$, where $\mathbf{L}$ is a $d \times p$ restriction matrix of full rank. If we want to test the more general hypothesis $\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = 0$ we can use the Wald test statistic

$$W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T\mathbf{L}^T(\mathbf{L}\hat{\mathbf{V}}\mathbf{L}^T)^{-1}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{3.11}$$

where $\hat{\boldsymbol{\beta}}$ is an estimated value of $\boldsymbol{\beta}$ and $\hat{\mathbf{V}}$ is the covariance matrix of $\hat{\boldsymbol{\beta}}$. Further W will asymptotically follow a χ^2_d distribution under the hypothesis.

As mentioned in the beginning of this subsection the LR test can be poor for small samples. And since the Wald test, like the LR test, depends on asymptotic results it can also produce poor results for small samples. However Kenward and Roger proposed a modification of the scaled Wald test statistic:

$$F = \frac{1}{d}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T\mathbf{L}^T(\mathbf{L}\hat{\mathbf{V}}\mathbf{L}^T)^{-1}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

in order to improve small sample properties, they approximates the distribution of F by an $\mathcal{F}_{d,m}$ distribution. To estimate m they proposed to calculate the mean and the variance of the statistic, and then match the moments with those of a $\mathcal{F}$ distribution. This approximation is implemented in R in the *pbkrtest* package, the function is called *KRmodcomp*, and will be used in the data analysis.

3.3.2 Random effects

If the variable is a random effect there are no standard way to test whether this is significant or not. Therefore an approach is implemented in R, the implementation can be seen in Appendix B. The approach described in this subsection is based on the lecture notes [Waagepetersen, 2018(2)] on slides 29-30.

First we generalize definition 3.4.

Definition 3.8 A more general formulation of the model in equation (3.1) is:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^{K} \mathbf{Z}_i \mathbf{U}_i + \boldsymbol{\epsilon},$$

where $\mathbf{X}$ is a $N \times p$ design matrix, $\mathbf{Z}_i$ is a $N \times k$ design matrix, $\mathbf{U}_i \sim \mathcal{N}_{d_i}(\mathbf{0}, \sigma_i^2 \mathbf{I})$, $\boldsymbol{\epsilon} \sim \mathcal{N}_N(\mathbf{0}, \sigma^2 \mathbf{I})$, $\mathbf{U}_i$ and $\boldsymbol{\epsilon}$ are independent and K is the number of random effects.

To test the significance of a random effect we define $L = \text{span}\{\mathbf{X}, \mathbf{Z}_1, \ldots, \mathbf{Z}_K\}$ and $L_{-1} = \text{span}\{\mathbf{X}, \mathbf{Z}_2, \ldots, \mathbf{Z}_K\}$, where we assume that $L \neq L_{-1}$. Then we get

$$\mathbb{R}^N = L_{-1} \oplus V_1 \oplus V_{\mathbf{I}},$$

where $V_1 = L \ominus L_{-1}$ and $V_{\mathbf{I}} = \mathbb{R}^N \ominus L$. Let $\mathbf{Q}_1$ and $\mathbf{Q}_{\mathbf{I}}$ be the orthogonal projections on V_1 and $V_{\mathbf{I}}$ respectively. Hence $\mathbf{Q}_1 = \mathbf{P}_L - \mathbf{P}_{L_{-1}}$ and $\mathbf{Q}_{\mathbf{I}} = \mathbf{I} - \mathbf{P}_L$. Since $L \ominus L_{-1}$ is all the vectors in L which are orthogonal on the vectors in L_{-1} we get that the projection $\mathbf{Q}_1$ times any vector in L_{-1} is equal to 0. Hence $\mathrm{E}[\mathbf{Q}_1\mathbf{Y}] = \mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} = 0$, further since

$$\mathrm{Cov}[\mathbf{Y}] = \sum_{i=1}^{K} \sigma_i^2 \mathbf{Z}_i \mathbf{Z}_i^T + \sigma^2 \mathbf{I}$$

we get that

$$\mathrm{Cov}[\mathbf{Q}_1\mathbf{Y}] = \mathbf{Q}_1 \mathrm{Cov}[\mathbf{Y}] \mathbf{Q}_1^T = \sigma^2 \mathbf{Q}_1 + \sigma_1^2 \mathbf{Q}_1 \mathbf{Z}_1 \mathbf{Z}_1^T \mathbf{Q}_1.$$

Thus

$$\mathbf{Q}_1\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_1 + \sigma_1^2 \mathbf{Q}_1 \mathbf{Z}_1 \mathbf{Z}_1^T \mathbf{Q}_1).$$

We can use the same argumentation for $\mathbf{Q_I Y}$, i.e. $\mathbf{Q_I}$ is all the vectors in $\mathbb{R}^N$ which are orthogonal on the vectors in L hence $\mathbf{Q_I}$ times any vector in L is equal to $\mathbf{0}$. Again we get that $\mathrm{E}[\mathbf{Q_I Y}] = \mathbf{Q_I X}\boldsymbol{\beta} = \mathbf{0}$ and

$$\mathrm{Cov}[\mathbf{Q_I Y}] = \mathbf{Q_I}\mathrm{Cov}[\mathbf{Y}]\mathbf{Q_I}^T = \sigma^2\mathbf{Q_I},$$

Thus

$$\mathbf{Q_I Y} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{Q_I})$$

Under the hypothesis $\mathcal{H}_0 : \sigma_1^2 = 0$ we get that $\|\mathbf{Q}_1\mathbf{Y}\|^2$ and $\|\mathbf{Q_I Y}\|^2$ are independent scaled χ^2 distributed, hence we test that

$$\frac{\|\mathbf{Q}_1\mathbf{Y}\|^2/d_1}{\|\mathbf{Q_I Y}\|^2/d_\mathbf{I}} \sim \mathcal{F}(d_1, d_\mathbf{I}),$$

where $d_1 = \mathcal{C}(\mathbf{Q}_1)$ and $d_\mathbf{I} = \mathcal{C}(\mathbf{Q_I})$.

3.4 Residuals

We use residuals to validate our model assumptions. In this section residuals for the linear mixed models will be described. This section is based on [Nobre & Singer, 2007].

In this section we will look at the model on the subject level i.e.:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{U}_i + \boldsymbol{\epsilon}_i, \qquad i = 1, \ldots, m,$$

where $\mathbf{Y}_i$ is a $n \times 1$ vector for subject i, $\boldsymbol{\beta}$ is a $p \times 1$ vector, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are $n \times p$ and $n \times q$ design matrices respectively and of full rank, $\mathbf{U}_i$ is a $q \times 1$ random vector and $\boldsymbol{\epsilon}_i$ is a $n \times 1$ random vector. As earlier we assume

$$\mathbf{U}_1, \ldots \mathbf{U}_m \overset{iid}{\sim} \mathcal{N}_1(\mathbf{0}, \mathbf{G}) \quad \text{and} \quad \boldsymbol{\epsilon}_i \overset{ind}{\sim} \mathcal{N}_n(\mathbf{0}, \mathbf{R}_i), \quad i = 1, \ldots, m,$$

where $\mathbf{G}$ and $\mathbf{R}_i$ are $q \times q$ and $n \times n$ positive definite matrices respectively, and that $\mathbf{U}_i$ and $\boldsymbol{\epsilon}_i$ are independent. Remember that $\mathbf{G}$ and $\mathbf{R}_i$ can depend on some underlying parameter $\boldsymbol{\psi}$. Notice that $\mathbf{Y} = (\mathbf{Y}_1^T, \ldots, \mathbf{Y}_m^T)^T$, $\mathbf{X} = (\mathbf{X}_1^T, \ldots \mathbf{X}_m^T)^T$, $\mathbf{Z} = \oplus_{i=1}^m \mathbf{Z}_i$, where $\oplus$ is the direct sum, $\mathbf{U} = (\mathbf{U}_1^T, \ldots, \mathbf{U}_m^T)^T$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T, \ldots \boldsymbol{\epsilon}_m^T)^T$, $\boldsymbol{\Psi} = \mathbf{I}_m \otimes \mathbf{G}$ and $\boldsymbol{\Sigma} = \oplus_{i=1}^m \mathbf{R}_i$, where $\otimes$ is the Kronecker product.
In linear mixed models there are three kind of residuals, namely

- Marginal residuals, $\hat{\boldsymbol{\xi}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, that predicts the errors $\boldsymbol{\xi} = \mathbf{y} - \mathrm{E}[\mathbf{Y}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$,
- Conditional residuals, $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}}$, that predicts the errors $\boldsymbol{\epsilon} = \mathbf{y} - \mathrm{E}[\mathbf{Y}|\mathbf{u}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}$,
- The BLUP, $\mathbf{Z}\hat{\mathbf{u}}$, that predicts the random effects, $\mathbf{Z}\mathbf{u} = \mathrm{E}[\mathbf{Y}|\mathbf{u}] - \mathrm{E}[\mathbf{Y}]$.

These residuals can be used to check some of the model assumptions in the model in definition 3.4.

3.4.1 Marginal residuals

The marginal residuals can be used to check the linearity of $\mathbf{y}$ with respect to the explanatory variables. We can do this by plotting $\hat{\boldsymbol{\xi}}$ against the explanatory variables in $\mathbf{X}$. They can also be used to check the covariance structure $\mathbf{V}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^T + \mathbf{R}_i$. Here $\|\mathbf{I}_n - \mathcal{R}_i\mathcal{R}_i^T\|^2$, where $\mathcal{R}_i = \hat{\mathbf{V}}_i^{-1/2}\hat{\boldsymbol{\xi}}_i$ and $\|\mathbf{A}\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$ is the Frobenius norm of the matrix $\mathbf{A}$, can be used to check this covariance matrix. Since assuming the mean is modeled correctly by $\mathbf{X}\boldsymbol{\beta}$, and assuming that $\text{Var}[\mathbf{Y}_i]$ can be estimated by $\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i^T$. Then $\|\mathbf{I}_n - \mathcal{R}_i\mathcal{R}_i^T\|^2$ is expected to be near zero, hence plotting against subject indices can help to detect cases where the covariance structure does not fit well.

3.4.2 Conditional residuals

The conditional residuals can be used to check homoscedasticity and normality of the conditional residuals, $\boldsymbol{\epsilon}$. To check homoscedasticity we plot $\hat{\boldsymbol{\epsilon}}/\hat{\sigma}^2$ against the fitted values, $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$. To check normality we simply make Q-Q plots of $\hat{\boldsymbol{\epsilon}}/\hat{\sigma}^2$. We often studentize the residuals, such that they are on the same scale, in order to be able to compare them. The studentized residuals is found by dividing the estimate by the standard deviation of the estimate, i.e.

$$\hat{\epsilon}_j^* = \hat{\epsilon}_j(\text{Var}[\hat{\epsilon}_j])^{-1/2}, \quad j = 1, \dots N.$$

The studentized residuals can be used to check for outlying observations. To find the variance of $\hat{\epsilon}_j$ we need Woodbury's identity:

Theorem 3.9 (Woodbury's matrix identity) Let $\mathbf{A}$ be a $n \times n$ matrix, $\mathbf{B}$ be a $n \times k$ matrix, $\mathbf{C}$ be a $k \times k$ matrix and $\mathbf{D}$ be a $k \times n$ matirx then the following identity holds:

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}. \tag{3.12}$$

Remember the estimate of the conditional residuals is given by

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} - \mathbf{Z}(\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \boldsymbol{\Psi}^{-1})^{-1}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Before finding the variance of $\hat{\boldsymbol{\epsilon}}$ we will rewrite the estimate of $\hat{\mathbf{u}}$ and simplify the expression for $\hat{\boldsymbol{\epsilon}}$.

$$\begin{aligned}
\hat{\mathbf{u}} &= (\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \boldsymbol{\Psi}^{-1})\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \overset{(3.12)}{=} (\boldsymbol{\Psi} - \boldsymbol{\Psi}\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{Z}\boldsymbol{\Psi})\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
&= (\boldsymbol{\Psi}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Psi}\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
&= (\boldsymbol{\Psi}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Psi}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{V} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
&= (\boldsymbol{\Psi}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Psi}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Psi}\mathbf{Z}^T\mathbf{V}^{-1})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{\Psi}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}),
\end{aligned}$$

where we have used the fact that $\mathbf{V}-\mathbf{\Sigma}=\mathbf{Z}\mathbf{\Psi}\mathbf{Z}^T$. From this we get $\hat{\boldsymbol{\epsilon}}=\mathbf{y}-\mathbf{T}\mathbf{y}-\mathbf{Z}\mathbf{\Psi}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y}-\mathbf{T}\mathbf{y})$, where $\mathbf{T}=\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})\mathbf{X}^T\mathbf{V}^{-1}$. We can then simplify the expression for $\hat{\boldsymbol{\epsilon}}$ as follows:

$$\begin{aligned}\hat{\boldsymbol{\epsilon}}&=\mathbf{y}-\mathbf{T}\mathbf{y}-\mathbf{Z}\mathbf{\Psi}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y}-\mathbf{T}\mathbf{y})=\mathbf{y}-\mathbf{T}\mathbf{y}-\mathbf{Z}\mathbf{\Psi}\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{y}+\mathbf{Z}\mathbf{\Psi}\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{T}\mathbf{y}\\&=(\mathbf{I}-\mathbf{T}-(\mathbf{V}-\mathbf{\Sigma})\mathbf{V}^{-1}+(\mathbf{V}-\mathbf{\Sigma})\mathbf{V}^{-1}\mathbf{T})\mathbf{y}=(\mathbf{I}-\mathbf{T}-\mathbf{I}+\mathbf{\Sigma}\mathbf{V}^{-1}+\mathbf{T}-\mathbf{\Sigma}\mathbf{V}^{-1}\mathbf{T})\mathbf{y}\\&=\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T})\mathbf{y}.\end{aligned}$$

We can now find the variance of $\hat{\boldsymbol{\epsilon}}$.

$$\begin{aligned}\text{Var}[\hat{\boldsymbol{\epsilon}}]&=\text{Var}[\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T})\mathbf{y}]=\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T})\text{Var}[\mathbf{y}](\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T}))^T\\&=\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T})\mathbf{V}(\mathbf{I}-\mathbf{T})^T\mathbf{V}^{-1}\mathbf{\Sigma}=\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T})(\mathbf{V}\mathbf{V}^{-1}-\mathbf{V}\mathbf{T}^T\mathbf{V}^{-1})\mathbf{\Sigma}\\&=\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T})(\mathbf{I}-\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1})\mathbf{\Sigma}=\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T})(\mathbf{I}-\mathbf{T})\mathbf{\Sigma}\\&=\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T}-\mathbf{T}+\mathbf{T}\mathbf{T})\mathbf{\Sigma}\\&=\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T}-\mathbf{T}-\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1})\mathbf{\Sigma}\\&=\mathbf{\Sigma}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{T})\mathbf{\Sigma}.\end{aligned}$$

To get $\text{Var}[\hat{\epsilon}_j]$ we just take the j'th diagonal element in $\text{Var}[\hat{\boldsymbol{\epsilon}}]$.

3.4.3 BLUPs

The EBLUP $\mathbf{Z}_i\hat{\mathbf{u}}_i$ is the difference between the predicted value for the i'th subject and the population average. Hence we can check for outlying subjects by plotting the $\hat{\mathbf{u}}_i$s against subject indices. From such a plot we can see which subject(s) are outliers. We could also plot $\hat{\mathbf{u}}_i$ against the subject indices to identify outlying subjects. The EBLUPs can also be used check the normality assumption of the random effects, $\mathbf{U}_i$, we do this simply by making Q-Q plots of the estimated random effects, $\hat{\mathbf{u}}_i$.

In Table 1 we can see which residual to use depending on which of the model assumptions we are interested in checking.

Diagnostic for	Type of Residual	Plot
Linearity of effects	Marginal	$\hat{\boldsymbol{\xi}}$ vs. explanatory covariates.
Within subjects covariance structure	Marginal	$\|\mathbf{I}_n-\mathcal{R}_i\mathcal{R}_i^T\|^2$ vs. subject indices.
Outlying observations	Conditional	$\hat{\boldsymbol{\epsilon}}_j^*$ vs. observation indices.
Homoscedasticity of conditional errors	Conditional	$\hat{\boldsymbol{\epsilon}}_j^*$ vs. fitted values.
Normality of conditional errors	Conditional	Q-Q plots of $\hat{\boldsymbol{\epsilon}}/\hat{\sigma}^2$.
Outlying subjects	EBLUP	$\hat{\mathbf{u}}_i$ vs. subject indices.
Normality of the random effects	EBLUP	Q-Q plot of $\hat{\mathbf{u}}_i$.

Table 1: Residuals to use to check model assumptions.

4 Simulation Study

In this chapter we will make a simulation study in order to get an idea about how the residuals should look like when the model is specified correct. The simulated data have been simulated such that it resembles the given dataset, when we used the rotations as response, i.e. there are three groups with 30 subjects in each group, totalling to 90 subjects, and three measurements pr. subject, totalling to 270 observations. The covariates in the simulated dataset is:

- obs_ID: The subject ID,
- measure_NR: A time stamp of how long after e.g. an operation the measurement is taken,
- group: Indicates which group the observation belongs to,
- age: A positive discrete variable for each subject which could be something like the age of the subject when entering the study.

From theses covariates two responses are simulated $\mathbf{y}1$ and $\mathbf{y}2$. The first response, $\mathbf{y}1$, is simulated by the model

$$\mathbf{y}1 = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\epsilon},$$

and the second response $\mathbf{y}2$ is simulated by the model

$$\mathbf{y}2 = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathbf{X}$ is the design matrix containing measure_NR, group and age, $\mathbf{Z}$ is the design matrix for the random effects, obs_ID, $\mathbf{U} \sim \mathcal{N}(0, 0.5)$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, 1)$.

In Figure 4 we see the residual plots for the the first model. In the first plot we see, as expected, no pattern when plotting the age variables against the marginal residuals. In the second plot we see the within covariance structure plot which should lie around zero. However as we can see, despite the model being specified correct some of the points are still very large, so this plot needs to be taken with a grain of salt. In the third plot we see the studentized conditional residual vs. observation indices, we see that these residuals lie around zero and vary a bit with no outliers. In the forth plot we see the studentized conditional residual vs. fitted values, and it seems like the residuals lie around zero and vary a bit. In the fifth plot we see the Q-Q plot of the conditional residuals. We see that the points follow the line closely, indicating the conditional residuals are normally distributed. In the sixth plot we see the estimated random effect plotted against subject indices, here it does not seem like there are outlying subjects. And in the seventh and last plot we see the Q-Q plot of the estimated random effects, which indicates they are normally distributed as we know they are.

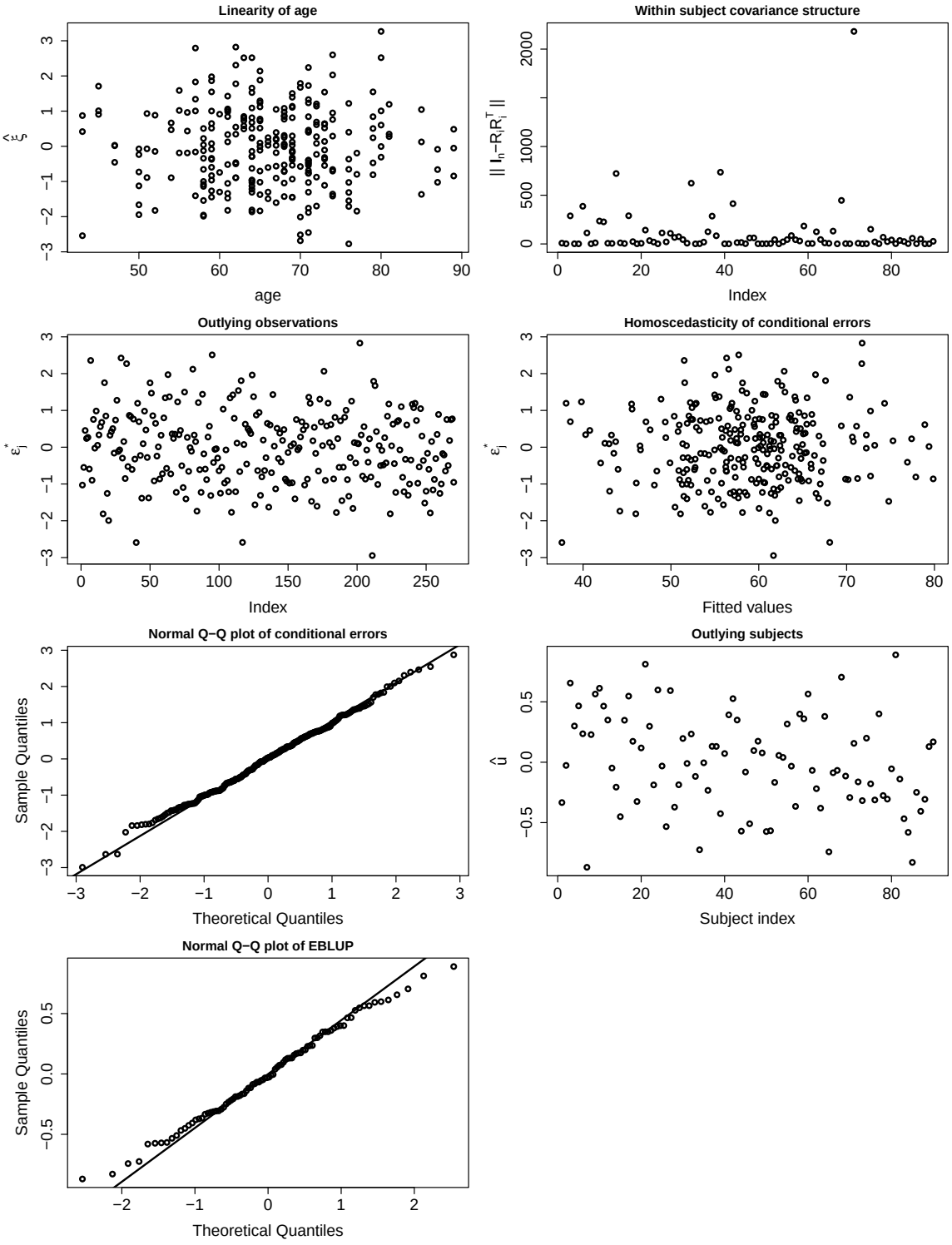

Figure 4: Plot of the residuals described in section 3.4.

We can further test the implemented approach to test whether the random effects are significant or not. We will fit the model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{ZU} + \boldsymbol{\epsilon}$ to both $\mathbf{y}1$ and $\mathbf{y}2$. When we test the significance of the random effect in the two models we get the p-values $2.11 \cdot 10^{-8}$ and 0.95 for the first and second model respectively. Hence as expected we get that the random effects are significant in the first model, and not significant in the second model.

5 Preliminary analysis

In this section we will take a closer look at the dataset. We will do this in order to better understand the data, and to get an idea about what results we can expect in the data analysis. As mentioned in section 2 we have the covariates:

- Group: Which treatment group the patient is in.
  - Group 1: 3 × 600mg Ibuprofen each day for seven days,
  - Group 2: 3 × 600mg Ibuprofen each day for three days and three placebo pills each days for four days,
  - Group 3: 3 × placebo pills each day for seven days.
- Age: The age of the patient when entering the study.
- Sex: The sex of the patient.
- Tradolan: The number of Tradolan pills a patient took in the 14 days after the operation.
- Time: Vary depending on the response. When rotation is response time is measured in weeks, and when pain is the response time is measured in days.
  - Week: An indicator which indicate how long after the operation the given measurement is taken. The measurements was taken 6 weeks, 3 months and 1 year after the operation. This is converted to 6 weeks, 13 weeks and 52 weeks.
  - Days: An indicator which indicate how many days after the operation the measurement is taken. The measurement are taken in the 14 days after the operation.

And the response variables:

- Supination and pronation rotation.
- Dorsal and volar flexion rotation.
- Radial and ulnar deviation rotation.
- Pain: How much pain, on a scale from 1 to 10, a patient felt in the 14 days after the operation.

5.1 Rotation response

We will start by looking at the rotations for each of the groups, in order to see whether one of the groups is obvious better or worse than the others. Below we can see the flexibility of each patient in each of the three pairs of movements. In figure 5 we see the improvement in the supination and pronation direction of each patient in each of the groups. It is not easy to see the improvement of each individual patient, but this is not what we are interested in. What we are interested in is the overall improvement in each

group. It might seem like group 3 is a little better than the other two groups, here all patients have gained around 90% or more (compared to the baseline) of their moveability back. Whereas the two other groups have a few patients whose moveability is around 50 - 70%.

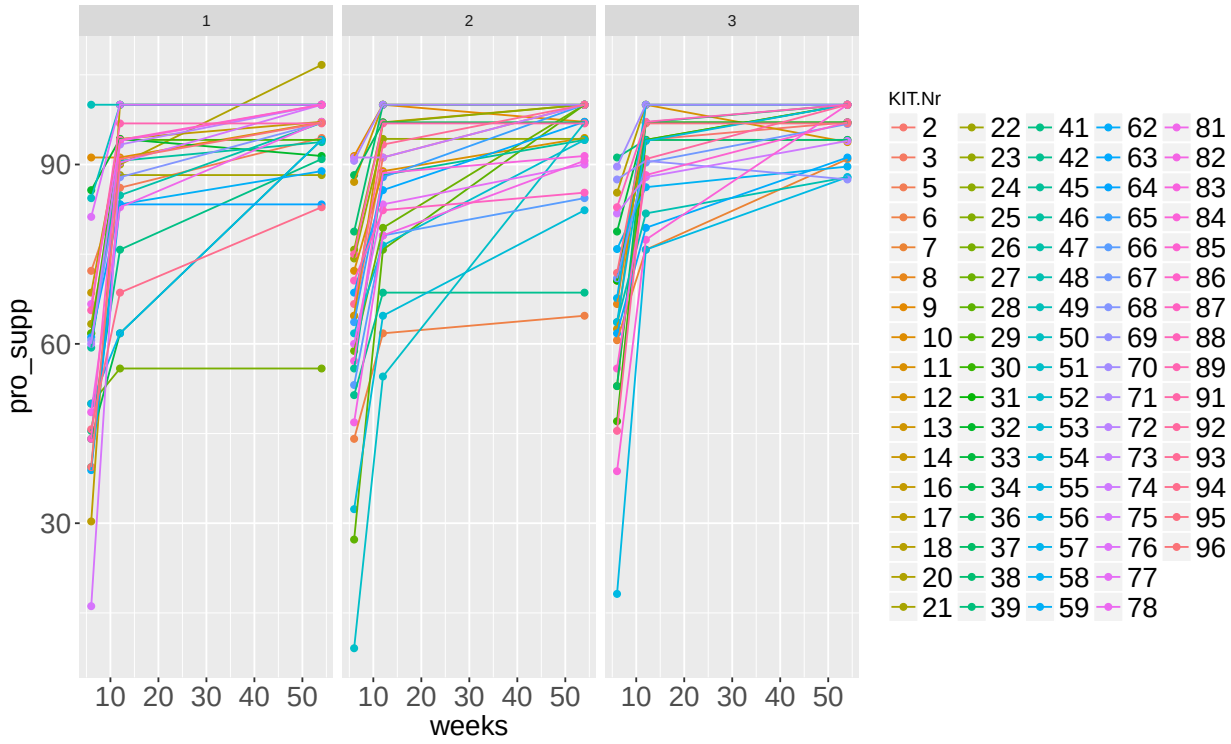

Figure 5: The improvements in the supination and pronation direction for each group in % (compared to the baseline) at each measurement.

In figure 6 we see the improvement in the dorsal and volar direction of each patient in the three groups. Again it's hard to see each patients improvement. From this figure it's hard to say one group should be better than the other two since they all vary quite a bit, however it seems like group 1 vary the least. But group 2 and 3 both have a few patients who exceeds 100% moveability quite a bit, further group 3 has a couple of patients whose moveability worsens from the second to the third measurement.

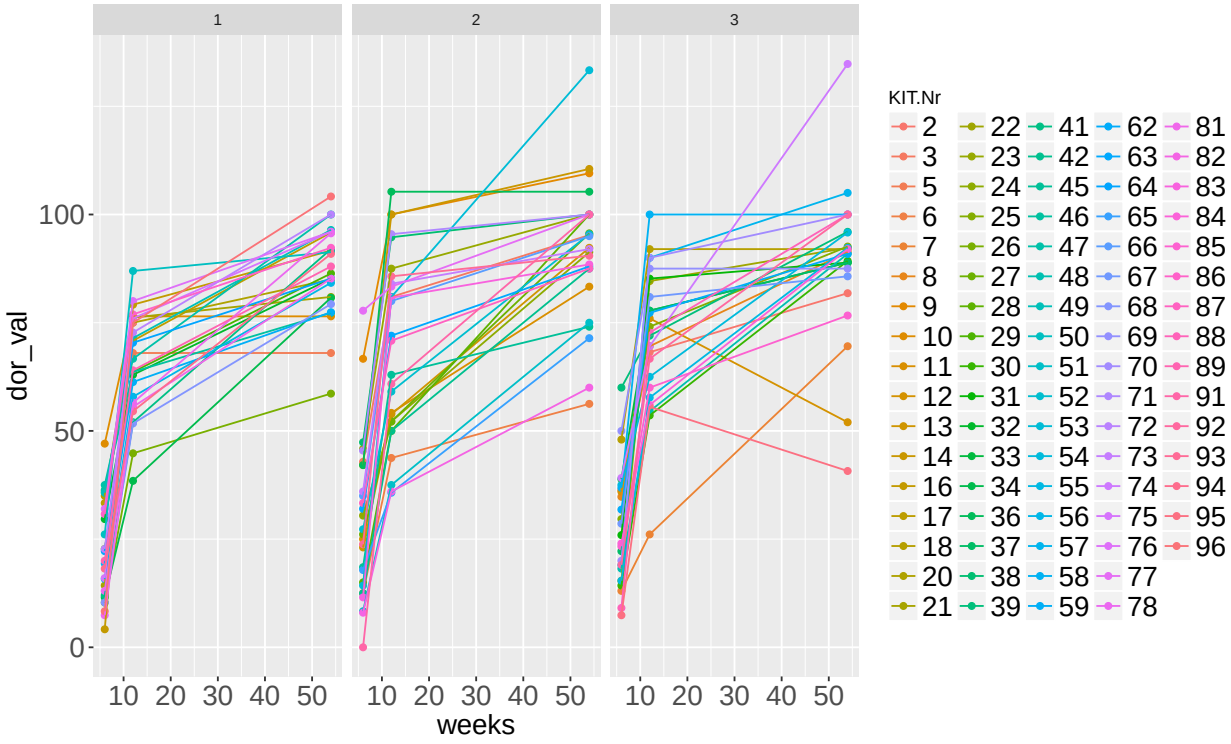

Figure 6: The improvements in the dorsal and volar direction for each group in % (compared to the baseline) at each measurement.

And lastly in figure 7 we see the improvements of each patients moveability in the radial and ulnar direction. And again it's hard to say one group should be better or worse than the other two, however there are multiple patients in group 1 and 3 who exceeds 100% moveability compared to the baseline but they also have one patient each whose moveability worsens from the second to the third measurement. Further they all vary quite a bit so again we can not say one group should be better than the other two.

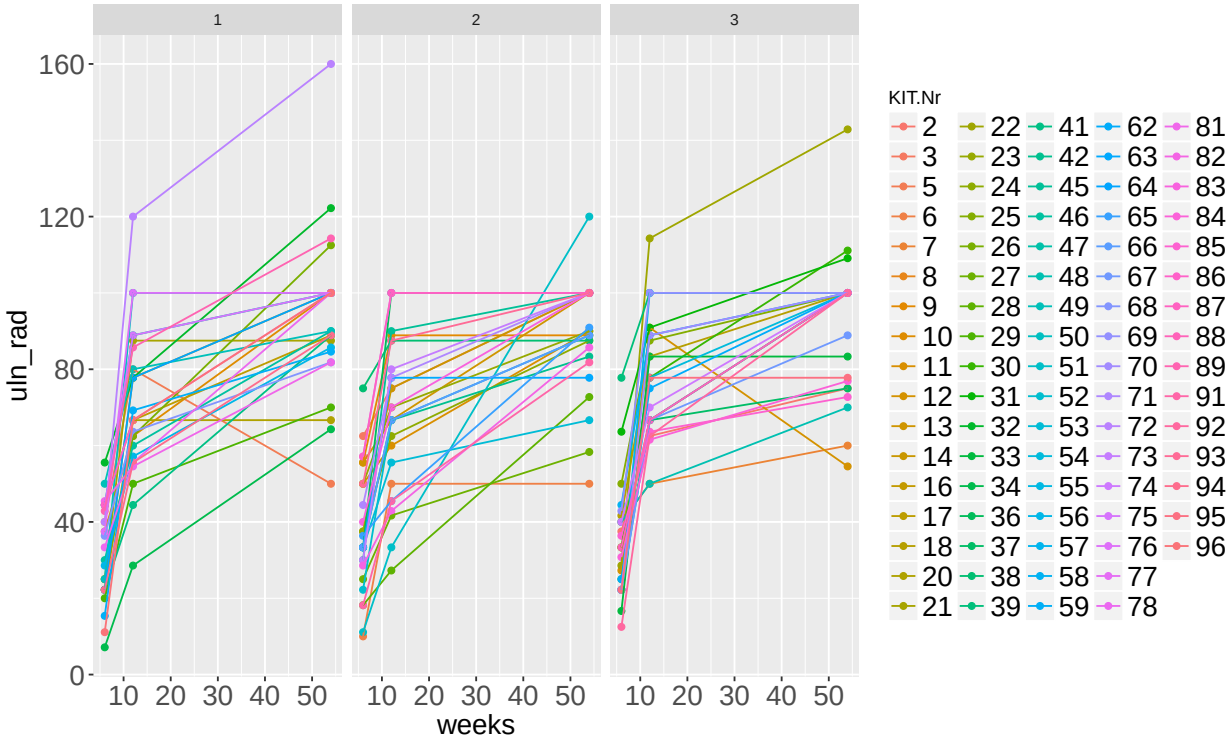

Figure 7: The improvements in the radial and ulnar direction for each group in % (compared to the baseline) at each measurement.

We have now examined the improvement in each of the groups moveability. And none of the groups were a clear "winner". I.e., we were not able to say one of the groups were better than the others, when looking at the moveability of the patients. We will now try to look at the pain of each group to see, whether any of the groups experience more pain than the other groups.

5.2 Pain response

As mentioned above we will now see whether any of the groups experience more pain than the others. In figure 8 we can see the average pain in each of the three groups, each day for 14 days. Group 1 and 3 lines almost on top of each other. And in group 2 we see an increase in pain at day four, this jump could be caused by the fact that the patients in group 2 switched from Ibuprofen to placebo after three days. From this figure we might get impression, that group 2s treatment is slightly worse when looking at the pain experienced. But the reason group 1 and 3 experience less pain might be because they took more Tradolan. In figure 9 we see the average number of Tradolan pills used by each group each day. Group 1 take quite a bit more Tradolan on day two than the other two groups, but all other days the groups use the same amount of Tradolan.

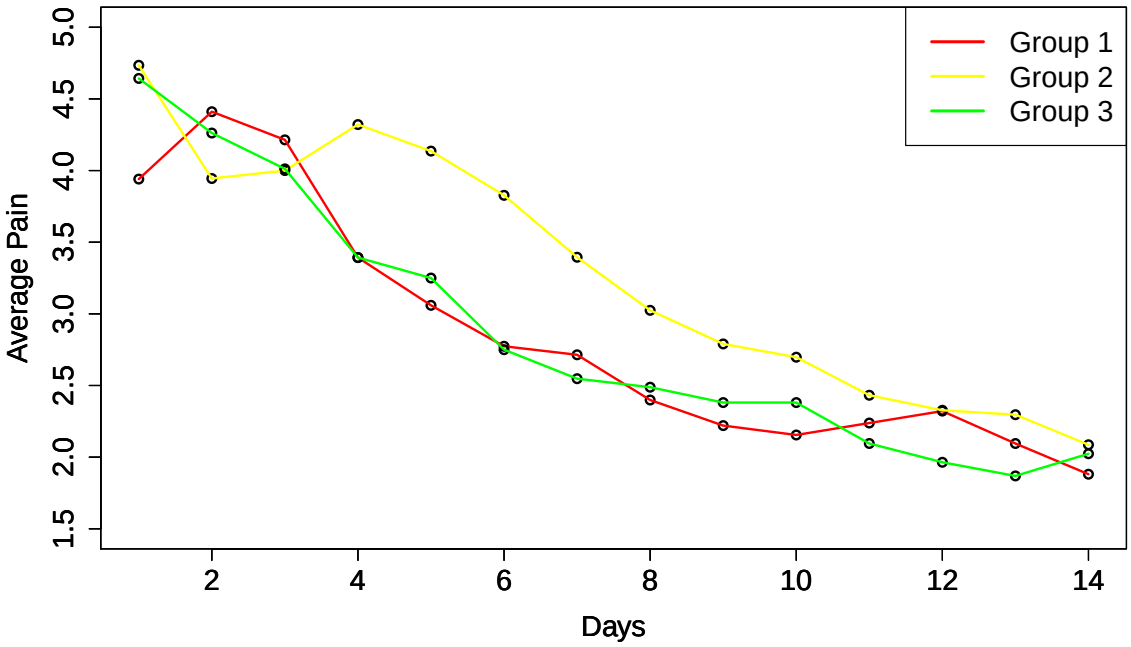

Figure 8: The average pain for each group each day for 14 days.

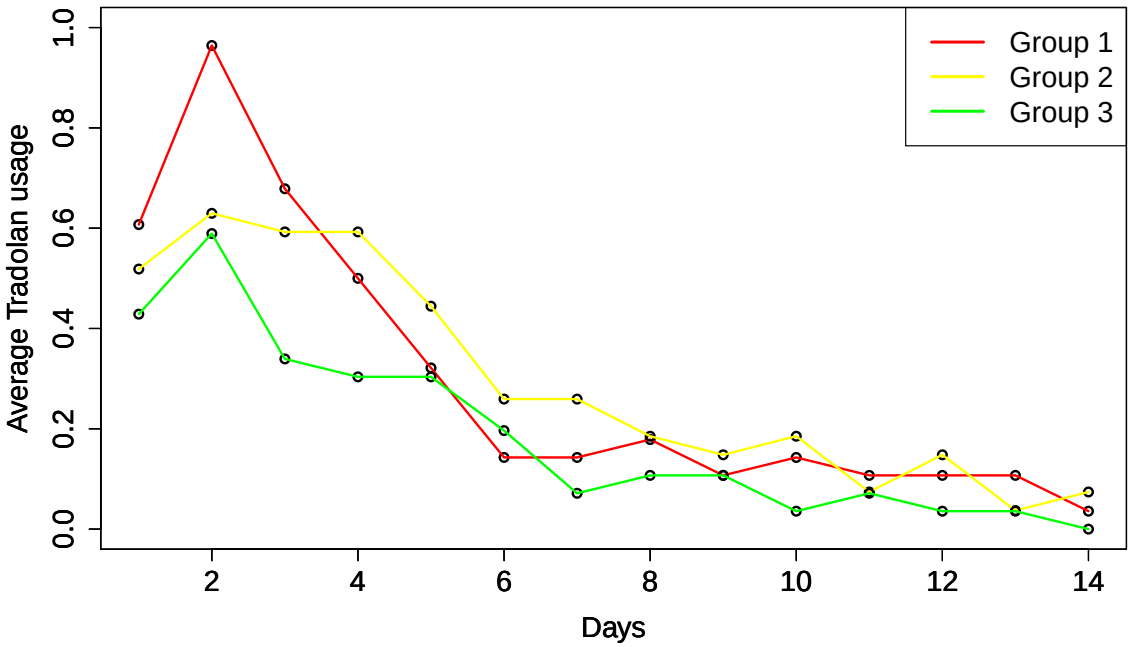

Figure 9: The average number of Tradolan used in each group each day

We now have an idea about what to expect and it does not seem like there are any difference between the groups. In the next chapter, chapter 6, we will see whether this is true or not, by using the mixed models introduced in chapter 3.

6 Data analysis

In this chapter we will analyse the data using the mixed model described in chapter 3. As mentioned in chapter 2 we have two responses, moveability and pain. We will look at these one at the time to see whether there's a difference between the groups or not.

We set up two models for each response both with the patient ID, KIT.Nr, as the random effect. The difference between the models lies in the fixed effects. In the first model, we include all the covariates and in the second we include all the covariates except Group. We then compare the two models to see whether Group is significant for the model. The data analysis is done in the statistic program R. To compare the two models we'll sure the KRmodcomp()-function from the package pbkrtest, which uses the Kenward-Roger approach. The KRmodcomp()-function test the null hypothesis $\mathcal{H}_0 : \mathbf{L}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = 0$.

Remember that group 1 is 600mg $\times$ 3 Ibuprofen each day for seven day, group 2 is 600 mg $\times$ 3 each day the first three days and three placebo pills each day the last four days and group 3 is three placebo pills each day for seven day.

6.1 Moveability

We will start by looking at the moveability of the patients in the groups. In chapter 5 it did not look like there were any difference in the moveability for the groups, we will check this here.

We will start by looking at the pair of movements individually. In table 2 we can see the p-values for the significance of Group in each rotation direction. With a significance level of 0.05, we can not reject $\mathcal{H}_0 : \mathrm{E}[\mathbf{Y}] = \mathbf{X}_0\boldsymbol{\beta}_0$. In other words we can not reject that the groups have no effect on the moveability when looking at the rotations individually. We will in the following shorten Pronation and supination to "Pro & Sup", dorsal and volar to "Dor & Vol" and radial and ulnar to "Rad & Uln".

	Pro & Sup	Dor & Vol	Rad & Uln
P-value	0.2417	0.4304	0.7935

Table 2: P-values for significance of group in each rotation direction.

Since Group is not significant for the model it does not matter which treatment we give the patients. It is not surprising that Group is not significant for the model, since the patients only receive treatment in the first week, and the rotations are measured up to a year after the operation. A more direct way to see whether the current used treatment, is significant different from the others is to look at the treatment difference i.e. β, the standard error and the confidence interval, of the groups when group 1 is set as reference. In Table 3 and 4 we can see the estimated treatment difference β, the corresponding standard error and the confidence interval for group 2 and group 3 compared to group 1. We can see that for both groups zero lies in the confidence interval, hence there are no significant difference between group 1 and the two other groups.

	Pro & Sup	Dor & Vol	Rad & Uln
Estimated group difference	−0.1706	4.0718	−0.9638
Standard error	2.8591	3.3633	3.5517
Lower conf. int.	−5.7744	−2.502	−7.9251
Upper conf. int.	5.4332	10.6639	5.9975

Table 3: The estimate of β, the standard error of β and lower and upper confidence interval for group 2 compared to group 1.

	Pro & Sup	Dor & Vol	Rad & Uln
Estimated group difference	4.0523	3.5261	1.5030
Standard error	2.8054	3.3001	3.4851
Lower conf. int.	−1.4463	−2.9420	−5.3278
Upper conf. int.	9.5509	9.9942	8.3338

Table 4: The estimate of β, the standard error of β and lower and upper confidence interval for group 3 compared to group 1.

We are also going to test whether the random effects are significant, i.e. whether there is a significant subject variation. We are testing this for the model without the Group covariate. In table 5 we can see the p-values for the significance of the random effects. With a significance level of 0.05 we can not reject $\mathcal{H}_0 : \sigma_u^2 = 0$. But we can not remove the random effects from the model, because then our data would not be modeled correct.

	Pro & Sup	Dor & Vol	Rad & Uln
P-value	0.9113	1	1

Table 5: P-values for the significance of the random effects in each rotation direction.

There might be some correlation between the rotations, we have not accounted for. In table 6 we can see the correlation between the three movement directions, and we see there is a strong correlation between the them. Hence we will make a model where we take this correlation into account, this model just looks at the moveability of the wrist as a whole. To account for the correlation between the rotations, we have to make the three rotation vectors into one vector simply by putting them at the end of each other. We also extent the other variables accordingly so that the dimensions corresponds to the new response. Further we add a new variable which indicates which of the rotations the measurements corresponds to. This approach also avoid the problem of multiple testing.

	Pro & Sup	Dor & Vol	Uln & Rad
Pro & Sup	1	0.78	0.72
Dor & Vol	0.78	1	0.86
Uln & Rad	0.72	0.86	1

Table 6: Correlation between the three movement directions.

When testing the model where we have taken the correlation between the rotations into account, we get a p-value, for the group significance, of 0.5031. Again we can not reject $\mathcal{H}_0$. Like before we will test more directly whether group 1 is significant different from the

other two groups. In Table 7 we can see the estimate group difference β, the corresponding standard error and the confidence interval for group 2 and 3 compared to group 1. We see that zero lies in the confidence interval, hence group 1s moveability improvement is not significant different from the other two groups.

	Group 2	Group 3
Estimated group difference	0.8590	3.0264
Standard error	2.6965	2.6522
Lower conf. int.	-4.4261	-2.1719
Upper conf. int.	6.1441	8.2247

Table 7: The estimate of β, the standard error of β and lower and upper confidence interval for group 2 and 3.

Since there is no significant difference between group 1 and the other two groups when looking at the improvement in moveability of the wrist, we can conclude that Ibuprofen is not better than the other two options. However again note that moveability is measured up to a year after the operation, and the treatment is only given for a week after the operation, hence it is not surprising that Group is not significant.

We will now check the model assumptions using the residuals described in section 3.4. We are mostly interested in the model where we have taken into account the correlation between the rotations, we will call this model *All_rot* model. The residual plots for the first three models, i.e. the models where we look at the rotations individually, will not be shown here but can be seen in Appendix A.

The residual plots for the All_rot model can be seen in Figure 10 below. In the first two plots we check the linearity of the covariates age and tradolan. We can see that there does not seem to be a pattern, hence the linearity assumption is kept. In the third plot we test the within subject covariance structure. We can see some of the observations are rather large, but as we saw in the simulation study this plot needs to be taken with a grain of salt. Hence we will accept that the within subject covariance structure is modeled correctly. In the fourth plot we test for outlying observations. There does not appear to be any outlying observations, the studentized residuals are large but none of the observations seem to be outliers. In the fifth plot we test the homoscedasticity of the conditional residuals. The residuals seems to have the same variance, but some of the residuals form a clear line, but we will still accept that the residuals have the same variance. In the sixth plot we test the normality assumption of the conditional errors, this assumption seems to be kept quite nicely with only a small tail. In the seventh plot we test for outlying subjects. And there does not appear to be any outlying subject. And in the eight and last plot we test the normality assumption of the BLUPs. Again the points follow the line quite nicely with a small tail, hence we will also accept that the BLUPs are normally distributed.

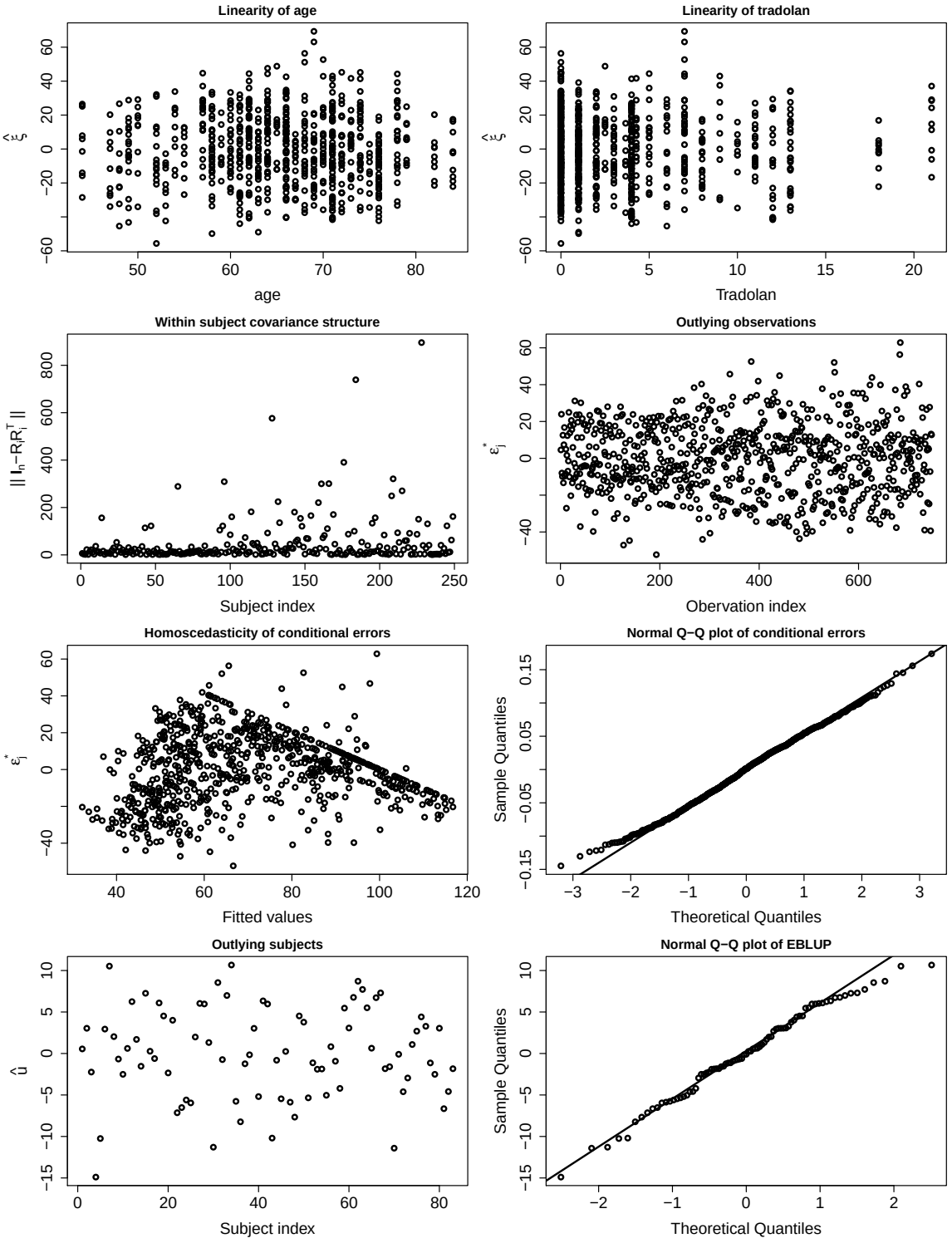

Figure 10: Residual plot of the model where the correlation between the rotations have been taken into account.

From Figure 10 we can conclude that the model fits data well, and thus the model is useful. In the next subsection we will see whether the groups experience a significant difference in pain.

6.2 Pain

As mentioned above we will now see whether the groups experience a significant difference in pain, in the 14 days after the operation.

Like before we set up two models one with all covariates and one where we have left out the covariate Group. We then compare the models to see whether Group is significant for the model when we use Pain as response. Note in this model we have 14 measurements on each subject, and instead of the total number of Tradolan each patient took, we will use the number of Tradolan pills each patient took each day. We get a p-value for the significance of the Group covariate of 0.4370. Again we can not reject $\mathcal{H}_0$, hence it does not matter which painkiller treatment the patients get, which indirectly tells us that the groups does not experience a significant difference in pain. Like with moveability we will check more directly whether group 1 experience a significant difference in pain compared to the other two groups. In Table 8 we can see the estimate of β, the corresponding standard error and the confidence interval for group 2 and 3. We see that zero lies in both confidence intervals, hence there is no significant difference between the pain group 1 experience compared to the other two groups.

	Group 2	Group 3
Estimated group difference	0.4363	0.0815
Standard error	0.3574	0.3541
Lower conf. int.	−0.2642	−0.6125
Upper conf. int.	1.368	0.7755

Table 8: The estimate of β, the standard error of β and lower and upper confidence interval for group 2 and 3, with pain as response.

When testing the significance of the random effects we get a p-value of 0.9102, hence the random effects are significant. This is not surprising since pain is a very subjective thing.

We will again check model assumptions using the residuals from section 3.4. We are testing the model without the Group covariate. In Figure 11 below we can see the residual plots for the model with pain as response. In the first plot there does not appear to be a pattern, hence the linearity assumption is kept for this covariate. In the second plot we see that the points become less and less dense as tradolan usage increases, hence the linearity assumption might not be kept in this case. In the third plot we again see that the values become quite large, but remember this plot needs to be taken with a grain of salt. In the fourth plot there appears to be a few outliers, but these are not so extreme what we will regard them as outliers. Hence there are no outlying observations. In the fifth plot we see there is a pattern, but the residuals seems to have the same variation. In the sixth plot we see that the conditional errors have some heavy tails, hence the normality assumption might not be kept. In the seventh plot we see that there does not appear to be any outlying subjects. And in the eight and last plot we see that the EBLUPs have a heavy tail, hence the normality assumption might not be kept here as well. However we assume that we have enough observations to assume normality.

Assuming the model is useful we can conclude that treatment group 1 does not experience a significant difference in pain compared to the other groups, hence we can conclude that there are no advantages to use treatment 1 over the other two.

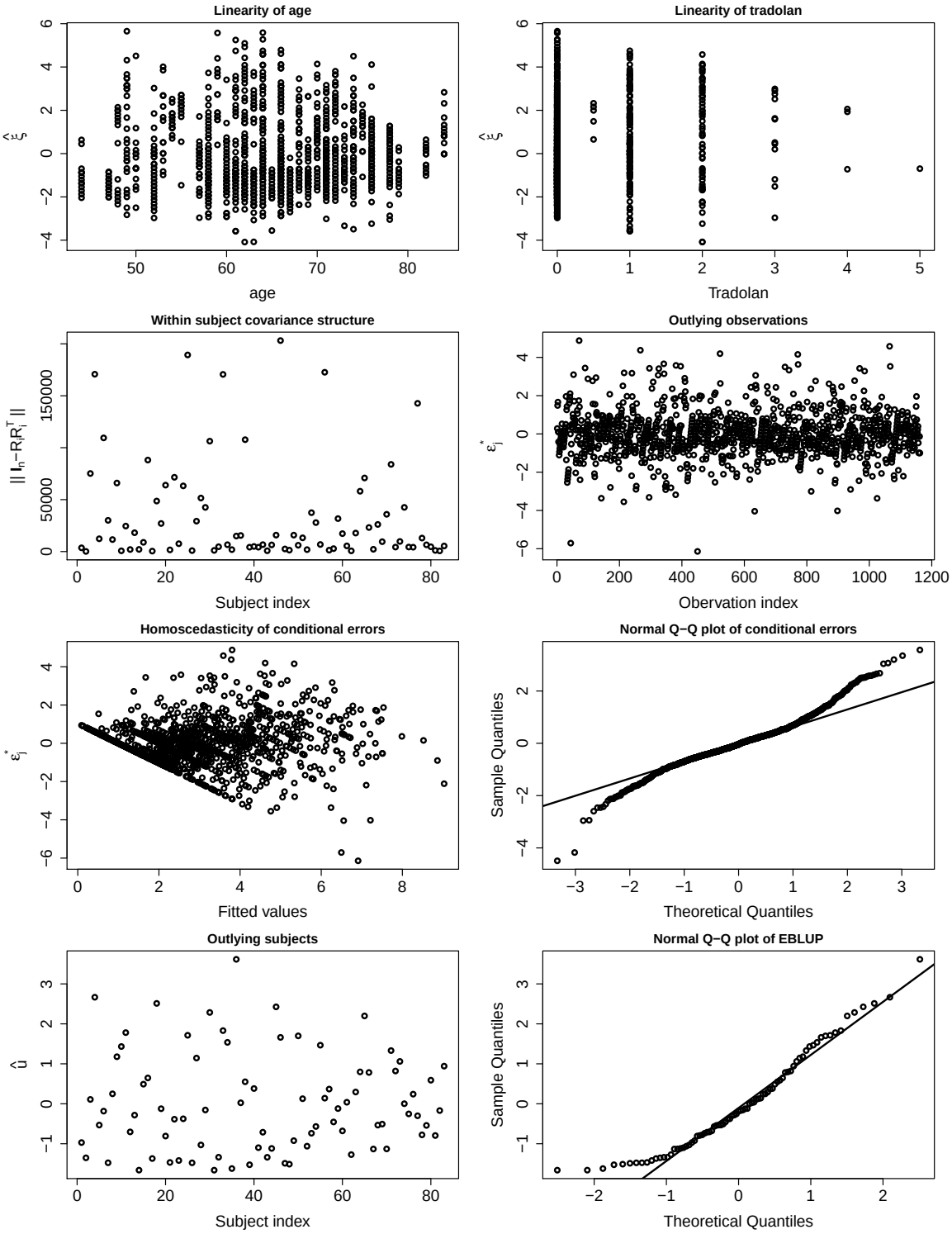

Figure 11: Residual plots of the model with pain as response.

7 Discussion

In this chapter the approaches used and choices made throughout this report will be discussed.

Data

As mentioned in the beginning of the report there are a few missing values in the dataset. These missing values are imputed in a very simple way since there are so few. The results might have change slightly if another imputation method had been used, but the conclusion would probably still be the same. The chosen imputation might not have been the best especially when imputing the pain felt on day 14, this value is probably larger than what would be expected. We could instead have used the pain felt the day before, this way we would also account for the subjects pain tolerance.

Simulation study

In the simulation study we simulated data to look like the given dataset. We did this to get an idea about, how we could expect the residuals to look like, and to test the implemented random effect hypothesis test. However the simulated data only resemble one of the datasets, and there are a lot more observations and repeated measures in the other dataset. Hence the simulated data might not be a good representation of what to expect from the second dataset.

Residuals

We saw that most of the residual were reliable, but when checking the within subject covariance structure we saw that close to zero might not be that close to zero, hence we needed to take the residual plot with a grain of salt. But all the other residual plots behaved as one would expect.

Data analysis

In the data analysis we tested to see whether group 1 had a significant different effect on moveability and pain compared to the other two groups, and saw that was not the case. We did however not test to see whether there was a significant difference between group 2 and group 3 because it was not directly of interest.

In the moveability model the Group covariate was not significant, this could be because the treatment is only given in one week, hence the effect hereof will most likely be small after 6 weeks, and completely gone after 3 months and a year. In the pain model we also saw that the Group covariate was not significant, and that there were no significant difference between group 1 and the other two groups. However it could be that if the usage of Tradolan explained the difference between the group 1 and the other two groups. Further we saw that the random effects were significant, which is not surprising since pain is very subjective.

8 Conclusion

The purpose of this project was to see whether there was a significant difference between the current used painkiller treatment, and two other alternatives after an operation on a broken wrist. We saw that there were no significant difference between treatment group 1 and the other two treatment groups, when look at the improvement in moveability of the wrist and at the pain felt in the 14 days after the operation. And from the residual plots we saw that the models fitted the moveability models well. However due to the large number of observations in the model with pain as response, we will assume that the data is normally distributed. And hence that we can use the pain model as well.

Appendices

A Extra figures

In this appendix the extra plots which were omitted in the rapport will be shown. In Figure 12 we can see the residual plots for the Pro & Sup model.

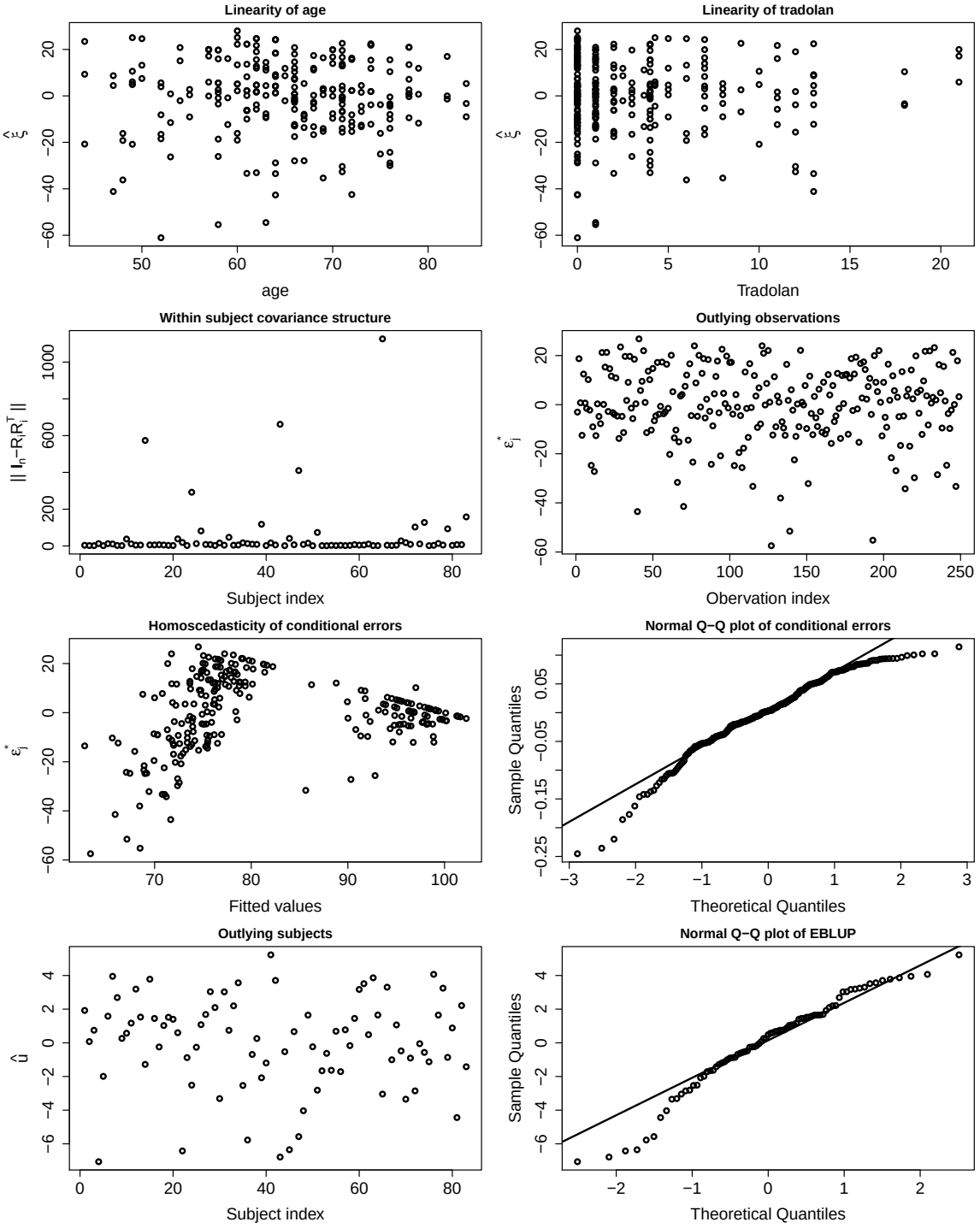

Figure 12: Residual plots of the Pro & Sup model.

In Figure 13 we can see the residual plots for the Dor & Vol model.

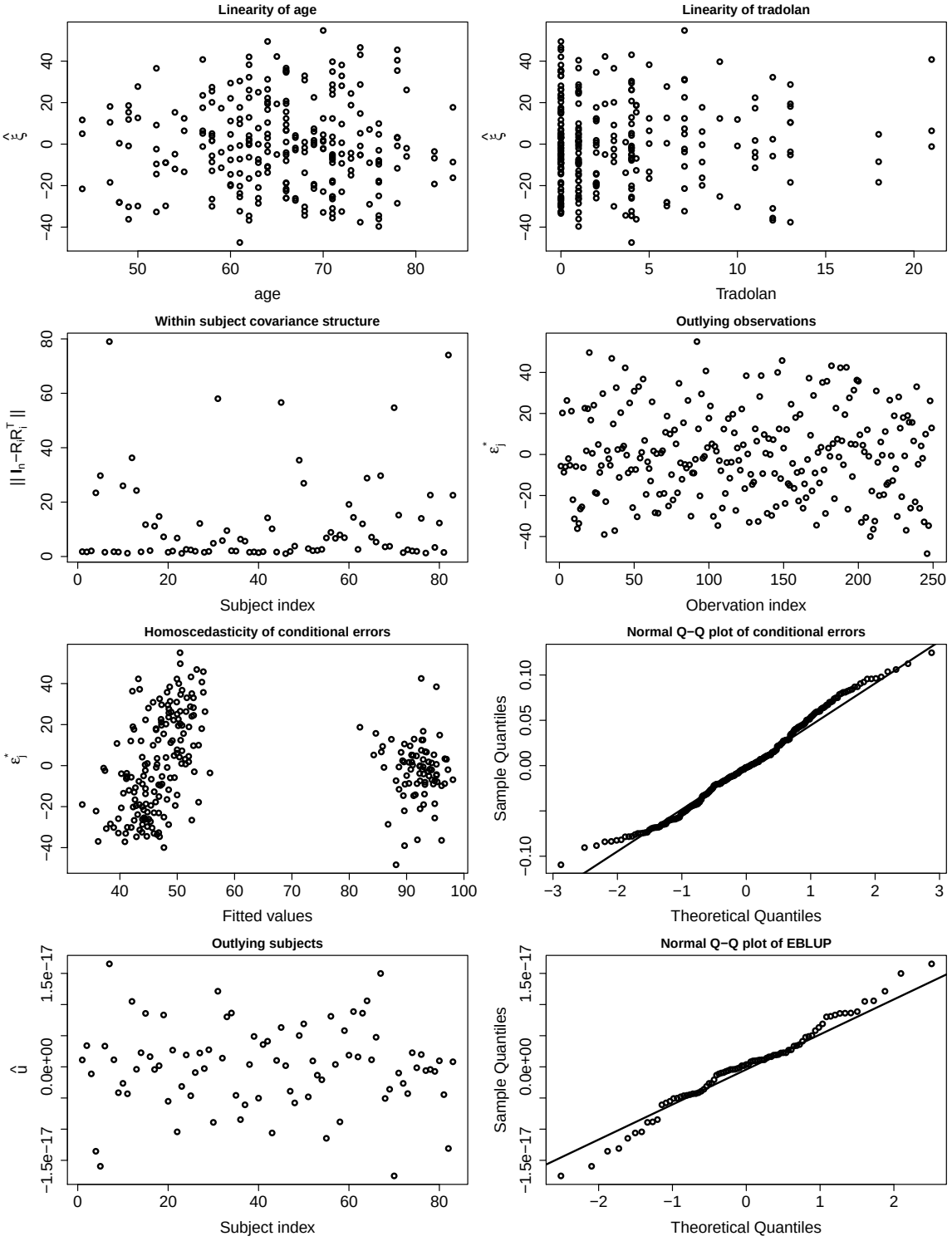

Figure 13: Residual plots of the Dor & Vol model.

In Figure 14 we can see the residual plots for the Rad & Uln model.

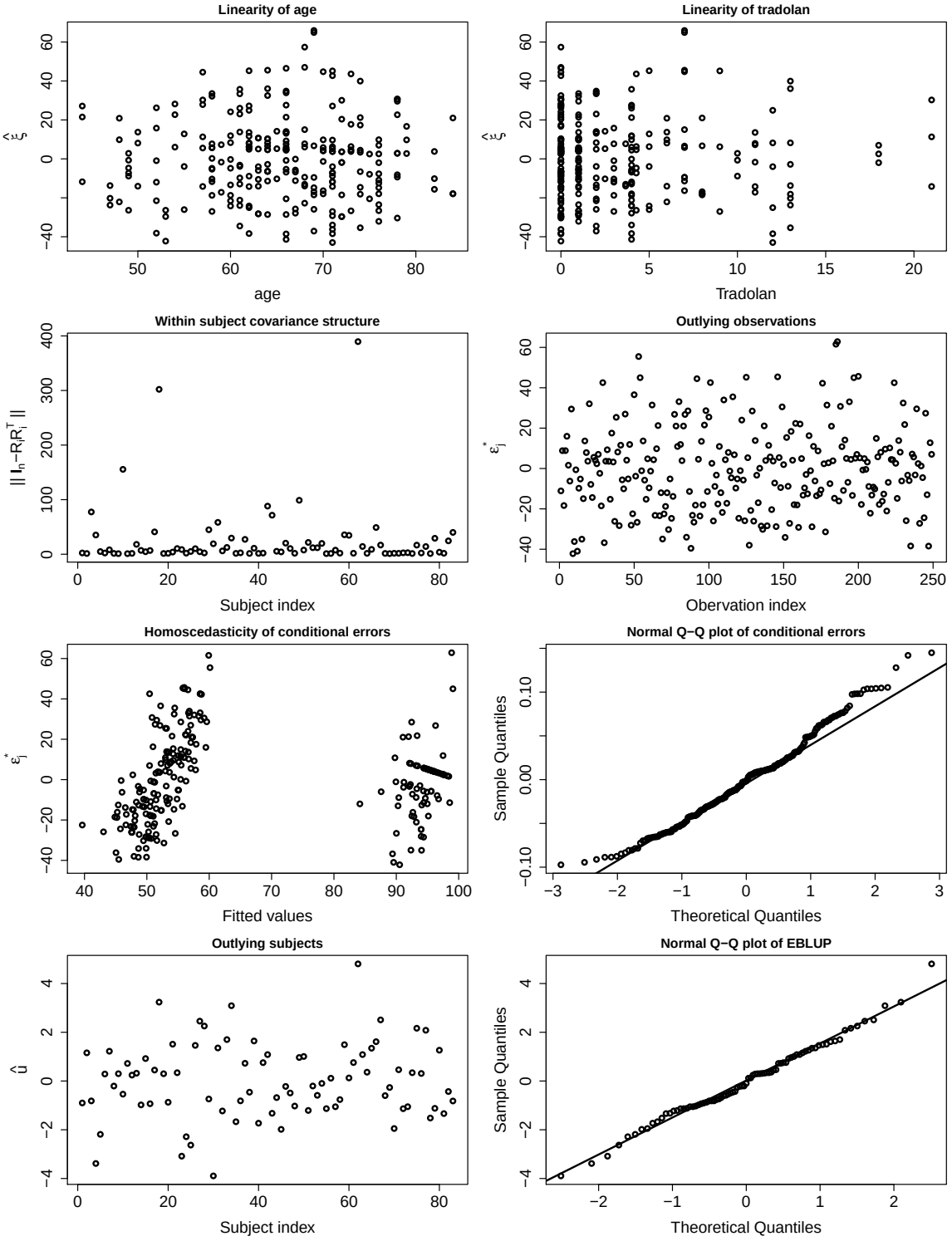

Figure 14: Residual plots of the Rad & Uln model.

B R code

In this appendix is some of the implmented R code used in this report.

Below is the implementation of the random effect test described in Subsection 3.3.2

```
1  var_comp_test = function(lmer_object)
2  {
3    model = lmer_object
4    y <- getME(model, "y") # the response vector
5    X = getME(model, "X") # The design matrix X
6    Z = as.matrix(getME(model, "Z")) # The design matrix
7    L = cbind(X, Z) # The matrix L
8
9    #checking to see whether the columns in L are independt, and
       if not we remove the non independt columns
10   reff_mat = rref(L)
11   col_rem = c()
12   for(i in 1:ncol(reff_mat))
13   {
14     if(sum(abs(reff_mat[,i])) != 1)
15     {
16       col_rem = c(col_rem, i)
17     }
18   }
19   L = L[,-col_rem]
20
21   # Calculating the projection matrices and the orthogonal
       matrices
22   P_L = L%*%solve(t(L)%*%L)%*%t(L)
23   P_NL = X%*%solve(t(X)%*%X)%*%t(X) # P_L-1
24   Q_I = diag(1, nrow = nrow(Z)) - Z%*%t(Z)/sum(Z[,1]) # assuming
       balanced data.
25   # Could also be calculated by Q_I = diag(1,nrow(Z))-P_L
26   Q_1 = P_L - P_NL
27
28   # Finding d_1 and d_I
29   d_1 = rankMatrix(Q_1)
30   d_I = rankMatrix(Q_I)
31
32   crit_value = (norm(Q_1%*%y)^2/d_1)/(norm(Q_I%*%y)^2/d_I)
33   p_value = 1 - pf(crit_value, df1 = d_1, df2 = d_I)
34
35   out <- list(
36     L = L,
37     P_L = P_L,
```

```
38      P_NL = P_NL,
39      Q_I = Q_I,
40      Q_1 = Q_1,
41      crit_value = crit_value,
42      p_value = p_value,
43      d_1 = d_1,
44      d_2 = d_I
45    )
46 }
```

Below is the implemented R code to calculate the marginal residuals, the conditional residuals and the BLUPs.

```
1  my_residual <- function(lmer_object)
2  {
3    model = lmer_object
4    # Calculate the marginal, and the conditional residuals and
       the BLUPs
5    Xbeta = model.matrix(model)%*%coef(summary(model))[,"Estimate"
       ]
6    marg_resi = model.response(model.frame(model)) - Xbeta
7
8    BLUP = as.matrix(getME(model, "Z")) %*% unlist(random.effects(
       model), use.names = F)
9    cond_resi = unlist(marg_resi, use.names = F) - BLUP
10
11   out <- list(
12     marg_resi = marg_resi, # overall residualer
13     cond_resi = cond_resi, # epsilon
14     BLUP = BLUP # BLUP
15   )
16 }
```

Below is an example of how the residuals from section 3.4 are calculated and plotted.

```
1  sim_lmer1 = lmer(y1 ~ measure_NR + group + age + (1|obs_ID),
       data = sim_dat)
2  sim_resi1 = my_residual(sim_lmer1)
3
4  # linearity of effects
5  plot(sim_dat$age, sim_resi1$marg_resi)
6
7  # Calculating the variances Sigma, Psi and V.
8  sigma_u2 = diag(attr(VarCorr(sim_lmer1)$obs_ID, "stddev")^2,
       nrow = 90) # sigma^2
9  Z = getME(sim_lmer1, "Z")
10 sigma2 = sigma(sim_lmer1)^2
11 var.u = (Z %*% sigma_u2 %*% t(Z)) # Psi
```

```
sI = sigma2 * diag(1, nrow = nrow(sim_dat), ncol = nrow(sim_dat)) # Sigma
var.y = var.b + sI # V

# within subject cov structure
sub_cov_str = c()
for(i in seq(1,270,3))
{
  V_i = var.y[i:(i+2),i:(i+2)]
  R_i = solve(sqrt(V_i))%*%sim_resi1$marg_resi[i:(i+2)]
  I_n = matrix(diag(1, nrow = length(R_i), ncol = length(R_i)), nrow = length(R_i), ncol = length(R_i))
  norm = norm(I_n - R_i%*% t(R_i))^2
  sub_cov_str = c(sub_cov_str, norm)
}
plot(sub_cov_str)

#outlying observations
X = getME(sim_lmer1, "X")
T_mat = X%*%solve(t(X)%*%solve(var.y)%*%X)%*%t(X)%*%solve(var.y) # The matrix T
var_eps = sI %*% solve(var.y)%*%(diag(1, nrow = 270, ncol = 270) - T_mat) # Variance of estimate epsilon

# Calculating the studentized epsilons
var_eps_j = c()
for(i in 1:nrow(sim_dat))
{
  var_eps_j = c(var_eps_j, sim_resi1$cond_resi[i]/sqrt(var_eps[i,i]))
}
plot(var_eps_j)

# Homoscedasticity of conditional errors
plot(var_eps_k~ fitted(sim_lmer1))

# normality of cond. err.
qqnorm(var_eps_k%*%solve(sI))
qqline(var_eps_k%*%solve(sI))

# outlying subjects
plot(getME(sim_lmer1, "b"))

# Normailtet af random effects
qqnorm(getME(sim_lmer1, "b"))
qqline(getME(sim_lmer1, "b"))
```

References

[Aliuskevicius, 2017] M. Aliuskevicius, *IBUPROFENS INFLUENCE ON CLINICAL AND FUNCTIONAL OUTCOMES OF COLLES FRACTURE: PAIN, RADIOLOGICAL MIGRATION AND WRIST FUNCTION*, 2017.

[Courrieu, 2009] P. Courrieu, *Fast solving of weighted pairing least-squares systems*, Journal of Computational and Applied Mathematics, 2009, pp. 39-48.

[Halekoh & Højsgaard, 2014] U. Halekoh & S. Højsgaard, *A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models {The R Package pbkrtest}*, Journal of statistical software, 2014.

[LaMotte, 2007] L. R. LaMotte, *A direct derivation of the REML likelihood function*, Springer, 2007.

[Madsen & Thyregod, 2011] H. Madsen & P. Thyregod, *Introduction to General and Generalized Linear Models*, CRC Press, 2011.

[Nobre & Singer, 2007] J. Santos Nobre & J. da Motta Singer, *Residual Analysis for Linear Mixed Models*, Biometrical Journal, 2007, pp. 863 - 875.

[Pawitan, 2013] Y. Pawitan, *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, OUP Oxford, 2013.

[Waagepetersen, 2018(1)] R. Waagepetersen, `http://people.math.aau.dk/~rw/Undervisning/MM_BI/`, 2018.

[Waagepetersen, 2018(2)] R. Waagepetersen, `http://people.math.aau.dk/~rw/Undervisning/MM_BI/Handouts/handout6.22x2.pdf`, 2018.