

Technische Universität München

ZENTRUM MATHEMATIK

**Influences on the loss of bone
density in perimenopausal women -
Applying linear mixed models**

Diplomarbeit

von

Elisabeth Hobmaier

Themenstellerin: Prof. Claudia Czado

Betreuer: Dr. Stephan Haug

Abgabetermin: 03.06.2011

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Garching, den 03.06.2011

Acknowledgments

First of all, I want to thank Prof. Claudia Czado for her excellent and intensive supervision and her belief in me during the whole process. I am also deeply grateful to my adviser Dr. Stephan Haug for his unlimited encouragement and the valuable support I received from him.

Furthermore, I want to express my gratitude to PD Dr. med. Vanadin Seifert-Klauss for the fruitful discussions and her expert knowledge concerning the practical part of this thesis. In this regard, I also want to thank Teresa Wimmer and Dr. Anita Ehle for their helpful and patient explanation of the medical background knowledge.

Finally, my biggest and sincerest thanks deserve my beloved boyfriend Thorsten and the best parents of the world, Josefine and Reinmund.

Contents

List of Figures	vi
List of Tables	vii
List of Notations	viii
1. Introduction	1
1. Linear mixed models for longitudinal data	2
2. The basic concept of linear mixed models	4
3. Model specification	6
4. The distribution of the response	13
4.1. The two-level hierarchical model	13
4.2. The marginal model	14
4.3. Coherence	15
4.4. Example: Simple mixed linear regression	15
5. Estimation of the linear mixed model parameters	17
5.1. Estimation of the fixed effects	17
5.1.1. Maximum likelihood estimation	18
5.1.2. Generalized least squares estimation	20
5.1.3. Properties of the estimate	21
5.2. Estimation of the random effects - The best predictor	21
5.3. Simultaneous estimation of the fixed and random effects	24
5.3.1. The best linear unbiased predictor for a linear function of the fixed and random effects	25
5.3.2. Henderson's mixed model equations	27
5.4. Estimation of the variance components	30
5.4.1. The profile log-likelihood	31
5.4.2. The restricted log-likelihood	32
5.4.3. The marginal log-likelihood	35
5.4.4. Computation of the estimates	36
5.4.5. Empirical estimates for the fixed and random effects	36

6. Hypothesis tests in linear mixed models	38
6.1. Likelihood ratio tests	39
6.1.1. Likelihood ratio test for the fixed effects	41
6.1.2. Likelihood ratio tests for the covariance parameters	42
6.1.3. Likelihood ratio tests for the random effects	44
6.1.4. Exact likelihood ratio tests for simple linear mixed models	47
6.2. Alternative tests for the fixed effects: Approximate Wald tests	49
7. Model building and model diagnostics	51
7.1. A general model building strategy	51
7.2. Information criteria	52
7.3. Examining a fitted model	53
8. Fitting and analyzing linear mixed models using R	56
8.1. Fitting a linear mixed model: The lme function	56
8.2. Analyzing the fitted model	59
8.2.1. Fitted values and residuals of a lme object	61
8.2.2. Estimated coefficients and prediction	63
8.3. Hypothesis tests in R for the model parameters in a linear mixed model . .	63
8.3.1. Likelihood ratio tests	63
8.3.2. Approximate marginal t-tests and approximate F-tests for the fixed effects	68
 II. Analyzing the PeKnO-Study	 71
9. Study and data description	72
10. Standard linear models to explain the loss of bone density	87
10.1. Exploratory data analysis	87
10.2. Modeling the change of bone density (X.BMD)	95
11. Linear mixed models to explain the bone markers (Longitudinal analysis)	101
11.1. Exploratory data analysis	101
11.2. Model building	111
11.3. Modeling the bone resorption marker CTX	114
11.3.1. Reduced model with random slope of logE2.st	121
11.3.2. Reduced model with random slope of FSH.st	127
11.3.3. Model comparisons	129
11.3.4. The final model for CTX	130
11.4. Modeling the bone formation marker osteocalcin	134
11.5. Modeling the bone formation marker BAP	139
11.6. Correlation between the bone markers	143
12. Conclusions and outlook	145

Appendix	147
A. Mathematical tools	148
A.1. Matrix Algebra	148
A.1.1. Properties of the determinant	148
A.1.2. Properties of the trace of a matrix	148
A.1.3. Differentiation of vector valued functions	149
A.1.4. Overview of properties of special matrices in a linear mixed model .	150
A.2. Statistical tools	150
A.2.1. The multivariate normal distribution	150
A.2.2. The mean squared error	151
A.2.3. Asymptotic normality of the MLE	151
A.3. Proofs of some used results	152
B. Additional information on the PeKnO-Study	154
B.1. Summary of variables	154
B.2. Summary of models	158
B.3. Additional graphics	159
Bibliography	164

List of Figures

2.1. Example of individual and population-specific effects	5
6.1. Densities of χ^2 distributions with different degrees of freedom	46
8.1. Example of plotting a <code>simulate.lme</code> object	67
9.1. Basic overview of the study for one single participant: <i>QCT</i> indicates the measurements of bone density, <i>BS</i> the number of the blood test	72
9.2. Barplot of the age of the participants at the begin of the study	73
9.3. Barplot of the BMI of each participant at 25 years of age, at the begin of the study and at the end of the study. BMI values within the dashed lines indicate that the corresponding woman is normalweight.	74
9.4. Barplot of the BMD of each participant at the begin and the end of the study. Values below the dashed line indicate that the corresponding woman has osteopenia.	75
9.5. Barplot of the absolute change of BMD (<code>X.BMD</code>) and the percental change of BMD with respect to the initial bone density (<code>X.BMD.PR</code>) for each participant. Negative values indicate the loss of bone density.	76
9.6. Barplot of <code>BMD.GROUP</code>	77
9.7. Barplot of the mean length of all cycles (<code>CYL.m</code>) for each participant	79
9.8. Barplot of the percental number of ovulatory (yellow) and anovulatory (orange) cycles with respect to all evaluable cycles in the whole study for each participant. The upper, dark part of the orange bars indicates the percentage of long anovulatory cycles (<code>anov.long.s.PR</code>)	79
9.9. Histograms of individual mean values of the variables determined by the luteal phase blood tests	82
9.10. Pairs plots of the individual mean values of the female sex hormones (a) and the bone turnover markers (b) determined by luteal phase blood samples	83
9.11. Scatter plot of the percentage of ovulatory cycles during the whole study and the mean values of the luteal phase serum concentrations of the female sex hormones	83
9.12. Barplot of participants with increasing or decreasing BMD and bone turnover markers during the course of the study	84
9.13. Scatter plot of the change of BMD and the change of the respective bone marker during the course of the study	85
9.14. Boxplots of the change of the respective bone markers classified by women who lost BMD (<code>X.BMD<0</code>) and women whose BMD increased during the course of the study	85

9.15. Boxplots of the percental change of the respective bone markers in women who lost BMD during the course of the study	86
10.1. Histogram and boxplot of the change of bone density	88
10.2. Scatter plots of the change of bone density with respect to age (left) and the bone density at the begin of the study (right)	89
10.3. Scatter plots of the change of bone density and the BMI values at different points of time as well as their changes over time	89
10.4. Scatter plots of the change of bone density and the percentage of ovulatory and (total, long, normal and short) anovulatory cycles during the study . .	90
10.5. Scatter plots of the change of bone density with respect to the mean length of cycles, it's transformation $\frac{1}{CYL.m^2}$ and the mean length of luteal phases . .	91
10.6. Scatter plots of the mean female sex hormones and the change of bone density with linear regression line and LOESS lines	92
10.7. Scatter plots of <code>X.BMD</code> with respect to <code>LH.m</code> and <code>FSH.m</code> as well as their logarithm (first row) and standardized residuals corresponding to simple standard linear models of <code>X.BMD</code> and the respective covariate (second row)	92
10.8. Optimal transformation of progesterone	93
10.9. Scatter plots of cortisol and the bone turnover markers and the change of bone density with linear regression line and LOESS lines	94
10.10.Boxplot of dummies	94
10.11.Q-Q plot of the standardized residuals (left) and scatter plot of the fitted values and the standardized residuals (right) of <code>mod.fin</code> (10.1)	100
11.1. Individual trajectories of the bone markers for the 13 participants with 5 complete measurements	102
11.2. Individual boxplots of the bone markers for the 13 participants with 5 complete measurements	103
11.3. Scatter plots of the bone markers and <code>AGE</code> with regression and LOESS lines	104
11.4. Scatter plots of the bone resorption marker <code>CTX</code> with respect to <code>FSH</code> , <code>ov.tot.PR</code> and <code>CORT</code> with regression and LOESS lines	104
11.5. Individual scatter plots and regression lines of the bone resorption marker <code>CTX</code> with respect to <code>FSH</code> , <code>log(E2)</code> , <code>log(PROG)</code> and <code>CORT</code> for the 13 participants with 5 complete measurements	106
11.6. 95% confidence interval for intercept and slope in individual simple standard linear models with response <code>CTX</code> and covariates <code>FSH</code> and <code>log(PROG)</code> . .	107
11.7. Boxplots of bone markers per group	107
11.8. Scatter plot of <code>CTX</code> and <code>FSH</code> partitioned by the level of <code>E2</code> . The red values in the upper graphic indicate <code>FSH</code> concentrations greater than 40 IU/l. In the lower graphic, the outlying value of high <code>FSH</code> and high <code>E2</code> was removed.	109
11.9. Scatter plots of <code>CTX</code> and <code>FSH</code> with respect to different groups	110
11.10.Effect of time	111
11.11. Scatter plots of the bone markers and <code>CYL.BS</code> (first row) and its transformed version (second row), respectively	112

11.12. Simulated distribution of the RLRT statistic compared to \mathcal{X}_1^2 , $0.5\mathcal{X}_2^2 + 0.5\mathcal{X}_1^2$ and \mathcal{X}_2^2 distributions. The dashed lines indicate the bisecting lines. .	118
11.13. Scatter plot of CTX and the fitted values of <code>mod.CTX.logE2</code> (11.3)	126
11.14. Diagnostics of the conditional Pearson residuals of <code>mod.CTX.logE2</code> (11.3)	131
11.15. Assessing assumptions of normality for <code>mod.CTX.logE2</code> (11.3)	132
11.16. Scatter plot of the estimated random intercept $\hat{\gamma}_{i0}$ and the estimated random slope $\hat{\gamma}_{i1}$ of <code>mod.CTX.logE2</code> (11.3)	132
11.17. Model diagnostics of model (11.3), (11.4) and (11.5)	133
11.18. Model diagnostics for <code>mod.OST.anovlong</code>	137
11.19. Model diagnostics for <code>mod.BAP.CYL</code>	142
11.20. Pairs plots of the bone markers and their conditional Pearson residuals corresponding to models (11.3), (11.7) and (11.9) with linear regression and LOESS lines	144
11.21. Contour plots of the bivariate empirical distributions of the standardized residuals of the bone markers CTX, OST and BAP corresponding to models (11.3), (11.7) and (11.9)	144
B.1. Scatter plot of the percentage of long anovulatory cycles during the whole study and the mean values of the luteal phase serum concentrations of the female sex hormones	159
B.2. Barplots of individual mean values of the luteal phase serum concentrations of the female sex hormones	159
B.3. Scatter plot of the percentage of ovulatory cycles during the whole study and the mean length of luteal phases <code>LPL.m</code>	160
B.4. Scatter plot of the transformed mean length of cycles and the mean values of the luteal phase serum concentrations of the female sex hormones	160
B.5. Histograms of summarized data exemplarily for FSH and CTX to compare different approaches (<i>mean</i> , <i>AUC</i> , <i>min</i> , <i>max</i>) for summarizing longitudinal data	160
B.6. Model diagnostics for an alternative linear mixed model for modeling BAP with random slope <code>QCT.1.st</code> . The model can briefly formulated as: $BAP_{it} = \beta_0 + \beta_1 \text{PROGhigh6}_{it} + \beta_2 \text{BMI0high}_i + \beta_3 \text{AGE.st}_i + \gamma_{i0} + \gamma_{i1} \text{QCT.1.st}_{it} + \epsilon_{it}$	161

List of Tables

8.1. Overview of possible outputs of the R functions <code>fitted</code> and <code>residuals</code> of the <code>nlme</code> library. The default calculations are labeled.	62
10.1. Empirical correlation coefficients (<code>emp.cor</code>) between the covariate \mathbf{x} and <code>X.BMD</code> and characteristics of corresponding simple standard linear models. The <code>pvalue</code> refers to a t-test to verify the significance of the slope.	96
11.1. Overview of covariates in the saturated fixed effects model	113
11.2. Properties of a linear mixed model of <code>CTX</code> where a random slope of the respective covariate was included to the random intercept model (11.2). <code>pvalue</code> and <code>pvalue.cor</code> correspond to asymptotic likelihood ratio tests of $H_0 : d_1^2 = d_{01} = 0$ and $H_0 : d_{01} = 0$, respectively. The values less than 0.1 are marked.	120
11.3. Table of $\hat{\beta}_1$ in the linear mixed model $\text{CTX}_{it} = \beta_0 + \beta_1 \cdot x_{i(t)} + \gamma_{i0} + \gamma_{i1} \cdot \log\text{E2.st}_{it} + \epsilon_{it}$ and the calculated p-value corresponding to an asymptotic F-test of $H_0 : \beta_1 = 0$. $x_{i(t)}$ represents the respective covariate given in the different rows. The p-values less than 0.1 are marked.	122
11.4. Properties of a linear mixed model of <code>OST</code> where a random slope of the respective covariate was included to the random intercept model (11.6). <code>pvalue</code> and <code>pvalue.cor</code> correspond to asymptotic likelihood ratio tests of $H_0 : d_1^2 = d_{01} = 0$ and $H_0 : d_{01} = 0$, respectively. The values less than 0.1 are marked. For details see Section 11.3.	135
11.5. Properties of a linear mixed model of <code>BAP</code> where a random slope of the respective covariate was included to the random intercept model (11.8). <code>pvalue</code> and <code>pvalue.cor</code> correspond to asymptotic likelihood ratio tests of $H_0 : d_1^2 = d_{01} = 0$ and $H_0 : d_{01} = 0$, respectively. The values less than 0.1 are marked. For details see Section 11.3.	140
B.1. Overview of variables	158
B.2. Overview of models	158

List of Notations

y	<i>small letter</i>	single value
$\mathbf{y}, \mathbf{Y}, \boldsymbol{\beta}$	<i>bold letter</i>	vector
X, Σ	<i>capital letter</i>	matrix or random variable (context-dependent)

$$\mathbf{1}_n := \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$$

$$I_n := \begin{pmatrix} 1 & \cdots & \cdots & 0 \\ \vdots & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}$$

$|M|$ denotes the determinant of a quadratic matrix M .

$\|\cdot\|$ denotes the Euclidian norm.

1. Introduction

Around eighty percent of the patients who suffer from osteoporosis are postmenopausal women. The postmenopause is the phase of a woman's life in which the ovaries become inactive. It is defined as the time beginning 12 months after menopause, i.e. one year after the last menstrual cycle. The loss of the ovulatory hormone production is seen as the determinant factor for the loss of bone density during this phase of life.

Recent findings have suggested that accelerated bone loss already starts in the perimenopause (Seifert-Klauss et al. (2006)). The perimenopause begins with the first cycle irregularities and hormonal changes and ends one year after menopause. The factors determining the bone resorption in this time period are subject to current research.

To investigate perimenopausal women with respect to their menstrual cycles, hormonal values and changes in bone density, a prospective longitudinal study was performed at the *Frauenklinik der Technischen Universität München* (the *PeKnO-Study*). One purpose of this work is the statistical analysis of this study.

A detailed description of the study and the most important results of our investigations are presented in the second part of the thesis. We propose a standard linear model to explain the change of bone density which was measured for each woman at the begin and the end of the study (Chapter 10).

In Chapter 11 we present *linear mixed models* to explain biochemical markers for the bone metabolism which were measured semiannually for each woman. The big advantage of linear mixed models is that they allow to incorporate correlations occurring due to repeated measurements on the same individual in a longitudinal study.

Whereas the theory of standard linear models is widely known and well documented (we refer to Fahrmeir et al. (2007) and Myers (1986)), there is much less literature on the theory of linear mixed models. It is a good deal more complex and, in particular, the theory of statistical hypothesis tests is still in a developing phase. Therefore, the first part of this thesis provides a basic overview of the theory of linear mixed models.

Part I.

Linear mixed models for longitudinal data

The major parts of the theory of linear mixed models have originally been developed in biostatistics, mainly in the domains of plant and animal breeding (see Searle (1988) and West et al. (2007) for a historical overview). Besides constant regression parameters (*fixed effects*) as in standard linear models, linear mixed models also include regression parameters which vary randomly with respect to a certain random distribution (the so-called *random effects*). These models provide methods of fitting multiple correlated data, in particular grouped data sets, in an adequate way. The most important representatives of grouped data are clustered and longitudinal data. In a clustered data set, data can be divided into different groups. Examples might be students of the same university or individuals born in the same litter.

The main characteristic of a longitudinal data set is that, in contrast to cross-sectional data, several measurements were made on the same individual. The number of repeated observations of the same subject or individual, however, is relatively small compared to time series data sets. It's obvious that measurements made on the same variable are likely to be correlated, implying that the response variables are no longer independent, which is crucial in the theory of standard linear models. A further feature of longitudinal studies is that in most of the cases measurements were taken at different time points or that we do not have the same number of observations for each subject. In these cases the data are called to be *unbalanced*. We recommend Fitzmaurice et al. (2004) which discusses the characteristics of longitudinal data, especially how to model the occurring correlation in a longitudinal data set.

The big advantage of linear mixed models is not only that such unbalanced data may be analyzed properly, where other methods fail, but also that they manage to capture the occurring correlation among the repeated measurements. Modelling longitudinal data with a linear mixed model commonly yields a very high accuracy of estimates and an appropriate explanation of the variance in the data. The standard and most commonly used formulation of the model was first suggested by Laird and Ware (1982).

The theory of linear mixed models for longitudinal data shall be introduced in the following chapters. Formulations concerning longitudinal data, however, can easily be transferred to any kind of grouped data set. Instead of specifying different measurements belonging to the same individual, for instance, different subjects belonging to the same cluster could be considered.

A good overview with many examples and applications of statistical software is contained in Pinheiro and Bates (2000), and West et al. (2007). To get a more detailed and advanced illustration of the theory of linear mixed models the reader should consult Verbeke and Molenberghs (2009). An elaborate presentation of estimating the model parameters in a linear mixed model is given in Searle et al. (1992), as well as in McCulloch and Searle (2001).

2. The basic concept of linear mixed models

As already mentioned, linear mixed models are designed for modelling longitudinal data, because they allow us to model the correlation occurring due to repeated measurements for each individual by incorporating so-called individual-specific effects in the analysis. In the theory of linear mixed models we assume that there is a certain global linear relationship between the response and the explanatory variables which underly all individuals. This population-averaged effect of the covariates on the response vector is called the *fixed* effect in the data, because it describes a fixed relationship holding for the entire population. This assumption about a fixed or population-specific effect in the data we know already from the theory of standard linear models.

Apart from this, we assume that the relationship between the explanatory variables and the response in the measurements of a single individual can vary from this global population-specific relationship. Fitting a standard linear regression model for each individual may show that the individual regression parameters differ from the fixed population regression parameters. As we require our sample of individuals in the data set to be randomly chosen from the overall considered population, we consider this individual-specific deviations as random component in the model and call them *random effects*. Thus, the individual effect of the covariates is assumed to deviate randomly from the entire population (fixed) effect, albeit we expect an average behavior as in the whole population.

In a linear mixed model we describe this mathematically by including apart from constant regression parameters, indicating the fixed effect, regression parameters which are represented by random variables, denoting the individual-specific effect. That way, the name linear *mixed* model can be explained due to the fact of incorporating *fixed* and *random* effects in the model. But before looking at the precise formulation of such a linear mixed model, we consider an introductory example to understand clearly the basic idea of the theory.

Example: simple mixed linear regression

The fictive data in Figure 2.1 may represent observations made on three different individuals. Each individual was measured five times over time in a variable X and a variable denoted by Y . There is a strong heterogeneity in the data which can be traced back to the fact of correlation occurring in the observations of the same individual. A standard linear model will fail to fit this variability in the data in a adequate way. Fitting a linear regression line for each individual (illustrated by the dashed lines in the plot) exposes, that these individual trajectories differ a lot in intercept and slope. If we think of a medical

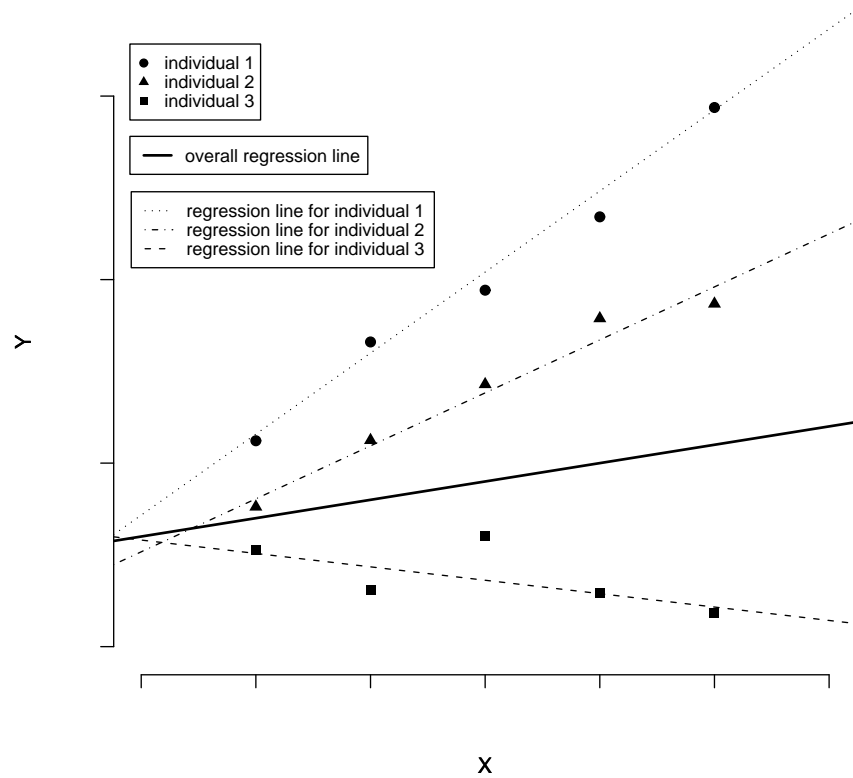


Figure 2.1.: Example of individual and population-specific effects

study then individual-specific genetic or environmental factors could only be one reason for an individual relationship between X and Y .

What we assume now is, that there is a systematic, population-averaged effect of X on Y which can be characterized as an overall population-specific regression line (represented by the continuous black line). The individual effects of X on Y , i.e. the intercept and slope of each individual regression line, are then assumed to vary randomly around the population intercept and the population slope. Thus, the individual trajectories can be explained as random deviations from the population regression line. Individual number one, for instance, has a similar intercept to the population intercept, but its individual slope value is much higher than the one of the entire population.

After this general description of the idea of linear mixed models, its mathematical formulation and formal assumptions are presented in the following chapter.

3. Model specification

So far we saw that the basic concept of linear mixed models is to take account of individual-specific random effects in the model. How to formulate this in a mathematical sense shall be described now, referring to the fundamental article of Laird and Ware (1982).

We consider a longitudinal data set where m individuals were observed at n_i ($i = 1, \dots, m$) points in time so that the data set consists of $n := \sum_{i=1}^m n_i$ measurements for the response and each covariate.

Let y_{it} denote the observed value of the continuous response variable for individual i at time t . The response vector for individual i , containing all its n_i observations, can therefore be expressed as $\mathbf{y}_i = (y_{i1}, \dots, y_{it}, \dots, y_{in_i})^T$.

To distinguish clearly between the observed response vector \mathbf{y}_i which is needed for calculating the estimated model parameters (see Chapter 5), and the random vector which shall be explained by the covariates, we denote the latter one by a capital letter, \mathbf{Y}_i .

Further, $\mathbf{x}_{it} = (x_{it}^{(1)}, \dots, x_{it}^{(p)})^T$ denotes the p -dimensional covariate vector for individual i at time t , which has a fixed effect on the response variable. That is, between these explanatory variables and the response there exists an overall population-specific relationship, which underly all individuals.

The matrix $X_i = (\mathbf{x}_{i1}^T, \dots, \mathbf{x}_{in_i}^T)^T \in \mathbb{R}^{n_i \times p}$ denotes the design matrix of the p known covariates for individual i , whose rows form the n_i covariate vectors observed at different times. Mostly, the model includes a fixed intercept term, so that the first column of X_i equals the $\mathbf{1}_{n_i}$ -vector, i.e. $x_{i1}^{(1)} = \dots = x_{it}^{(1)} = \dots = x_{in_i}^{(1)} = 1$. We assume X_i to be of full rank meaning nothing else as the linear independence of its columns, to avoid later identifiability problems for the fixed regression parameter.

The regression parameter vector indicating the fixed effects is denoted by $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$. Its values describe the entire population-averaged relationship between \mathbf{y}_i and the covariates in X_i for all individuals $i = 1, \dots, m$. If we include a population-specific intercept in the model, we often name the single components of $\boldsymbol{\beta}$ as $\beta_0, \dots, \beta_{p-1}$.

Likewise, let $\mathbf{z}_{it} = (z_{it}^{(1)}, \dots, z_{it}^{(q)})^T$ be the q -dimensional covariate vector for individual i at time t , containing the covariates which have an individual-specific effect. Naturally, most of the times, the components of \mathbf{z}_{it} are a subset of the values of \mathbf{x}_{it} , which implies $q \leq p$. Only in the case that the fixed effect of a covariate is equal to zero (i.e. the covariate has no effect on the population), there could be, however, an individual-specific effect, which means the variable is listed in \mathbf{z}_{it} but not in \mathbf{x}_{it} .

As before, $Z_i = (\mathbf{z}_{i1}^T, \dots, \mathbf{z}_{in_i}^T)^T$ indicates the $(q \times n_i)$ -dimensional design matrix for individual i associated with individual-specific random effects. Also in this case, the first

column of Z_i often is the $\mathbf{1}_{n_i}$ -vector, meaning that the model includes an intercept, which is randomly varying around the population intercept. If $Z_i = \mathbf{1}_{n_i}$, then we call our model *random intercept model*. To recapitulate, Z_i ($i = 1, \dots, m$) contains exactly these explanatory variables, whose effect varies randomly around the overall fixed population effect from individual to individual.

Note that the covariates given in X_i and Z_i may be either time-invariant, i.e. their values do not change within the measurements of one individual (e.g. gender or skin colour), or time-varying if their values vary over time.

The random effect for individual i is expressed by the specific regression parameter vector $\gamma_i = (\gamma_{i1}, \dots, \gamma_{iq})^T$. It describes the individual deviation from the overall population behavior expressed by β . As before, we denote the values in γ_i by $\gamma_{i0}, \dots, \gamma_{i(q-1)}$ if we incorporate a random intercept in the model.

Finally, let ϵ_{it} indicate the residual for individual i at time t , i.e. the deviation of the observed from the modeled response, like in standard linear models.

With $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{it}, \dots, \epsilon_{n_i})^T$ we denote the residual or error term vector for the n_i observations of individual i .

Using these notations, we formulate the linear mixed model as:

POSTULATE 1 (*Individual specification of a linear mixed model*)

For $i = 1, \dots, m$:

$$\mathbf{Y}_i = X_i \beta + Z_i \gamma_i + \epsilon_i ,$$

with $\mathbf{Y}_i \in \mathbb{R}^{n_i}$, $X_i \in \mathbb{R}^{n_i \times p}$, $\beta \in \mathbb{R}^p$, $Z_i \in \mathbb{R}^{n_i \times q}$, $\gamma_i \in \mathbb{R}^q$, $\epsilon_i \in \mathbb{R}^{n_i}$

$\text{rank}(X_i) = p$ and $\text{rank}(Z_i) = q$.

Consequently, the single observation of the response of individual i at time t satisfies

$$Y_{it} = \mathbf{x}_{it}^T \beta + \mathbf{z}_{it}^T \gamma_i + \epsilon_{it} .$$

Describing the population-averaged trend, the fixed effect β is an unknown but constant vector of regression parameters. The individual-specific effect γ_i , however, is meant to be a random vector. Assuming that the average population effect is already described by β , we expect γ_i to be the zero vector. Hence, we suppose γ_i to be multivariate normal distributed with mean vector zero and covariance matrix D :

POSTULATE 2 (*Random effect distribution*)

For $i = 1, \dots, m$:

$$\gamma_i \sim N_q(\mathbf{0}, D) \quad \text{with} \quad D = \begin{pmatrix} d_1^2 & d_{12} & \cdots & d_{1q} \\ d_{12} & d_2^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ d_{1q} & \cdots & \cdots & d_q^2 \end{pmatrix},$$

where D is a symmetric and positive definite covariance matrix.

The variance parameters in D indicate how strong the individual-specific effects can vary from the fixed effects and the covariance parameters show the dependence structure between the individual random effects.

Very often one assumes for D a so-called *variance components structure*, which implies the random effects of an individual to be uncorrelated. In this case, D can be written as a diagonal matrix with variance parameters d_1^2, \dots, d_q^2 :

$$D_{vcs} = \begin{pmatrix} d_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_q^2 \end{pmatrix}.$$

Note that we always suppose the random effects of different individuals $\gamma_1, \dots, \gamma_m$ to be independent.

As in standard linear models, we also assume in the theory of linear mixed models that the residuals are acting randomly. By them, we allow the response variable to vary randomly from the modeled, individual-specific behavior at any time. Hence, we make the following assumption:

POSTULATE 3 (*Error distribution*)

For $i = 1, \dots, m$:

$$\epsilon_i \sim N_{n_i}(\mathbf{0}, \Sigma_i),$$

where Σ_i is a symmetric and positive definite covariance matrix.

We see, that in contrast to standard linear models, we may allow here the residual components of the different observations in an individual to be correlated. In general one assumes in this case however that Σ_i depends on i only through its dimension n_i , but its

single elements are independent of i . For example one may describe an autocorrelated process of the residuals in Σ_i by assuming

$$\Sigma_i = \begin{pmatrix} \sigma^2 & \sigma^2 \rho & \dots & \sigma^2 \rho^{n_i-1} \\ \sigma^2 \rho & \sigma^2 & \dots & \sigma^2 \rho^{n_i-2} \\ \vdots & & \ddots & \vdots \\ \sigma^2 \rho^{n_i-1} & \sigma^2 \rho^{n_i-2} & \dots & \sigma^2 \end{pmatrix},$$

with a certain variance parameter $\sigma^2 > 0$ and a correlation parameter $\rho \in [-1; 1]$ (see West et al. (2007, Section 2.2)).

Nevertheless, it is often assumed, that the correlation between repeated measurements of an individual i is only explained by the covariance structure of γ_i and residuals within an individual are uncorrelated and have equal variance coefficients. In this case, we assume a diagonal structure for Σ_i ,

$$\Sigma_i = \sigma^2 \mathbf{I}_{n_i} \quad (i = 1, \dots, m),$$

with a certain constant variance factor σ^2 . We call this specific structure of the residual covariance matrices the *conditional independence structure*. Note that this is also the default structure used in the statistical software R which is applied in this thesis.

We always assume the residuals of different individuals $\epsilon_1, \dots, \epsilon_m$ to be independent.

Moreover, residuals and individual-specific random effects shall be independent of each other:

POSTULATE 4 (*Independence assumption*)

$\gamma_1, \dots, \gamma_m, \epsilon_1, \dots, \epsilon_m$ are independent.

Matrix specification of the model

Using the so far introduced individual model declaration, we can formulate an overall matrix specification of a linear mixed model, including all studied individuals. Remember, that we have an overall number of $n := \sum_{i=1}^m n_i$ measurements per variable, composed of the number of all individual measurements n_i ($i = 1, \dots, m$). By stacking above the individual regression components, we get:

POSTULATE 5 (*Matrix specification of a linear mixed model*)

$$\underbrace{\begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix}}_{=: \mathbf{Y} \in \mathbb{R}^n} = \underbrace{\begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}}_{=: \mathbf{X} \in \mathbb{R}^{n \times p}} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{=: \boldsymbol{\beta} \in \mathbb{R}^p} + \underbrace{\begin{pmatrix} Z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z_m \end{pmatrix}}_{=: \mathbf{Z} \in \mathbb{R}^{n \times mq}} \underbrace{\begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_m \end{pmatrix}}_{=: \boldsymbol{\gamma} \in \mathbb{R}^{mq}} + \underbrace{\begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix}}_{=: \boldsymbol{\epsilon} \in \mathbb{R}^n},$$

or more compactly:

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

with full-rank design matrices \mathbf{X} and \mathbf{Z} and symmetric, positive definite covariance matrices \mathbf{G} and \mathbf{R} of the random effects $\boldsymbol{\gamma}$ and the residuals $\boldsymbol{\epsilon}$, respectively,

$$\mathbf{G} = \begin{pmatrix} D & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D \end{pmatrix} \in \mathbb{R}^{mq \times mq} \quad \text{and} \quad \mathbf{R} = \begin{pmatrix} \Sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_m \end{pmatrix} \in \mathbb{R}^{n \times n},$$

so that

$$\begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim N_{mq+n} \left(0, \begin{pmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{pmatrix} \right).$$

Example: simple mixed linear regression

Let us look again at the example of simple mixed linear regression, where we incorporate only one covariate in the model by considering the fictive X and Y data shown in Figure 2.1. We formulate a linear mixed model, including a random intercept and a random slope coefficient. By this way we are able to indicate the individual deviation from the overall population regression line:

For $i = 1, 2, 3$ and $t = 1, \dots, 5$:

$$\begin{aligned}
 y_{it} &= \underbrace{\beta_0 + \beta_1 x_{it}}_{\text{population-specific influence, fixed}} + \underbrace{\gamma_{i0} + \gamma_{i1} x_{it}}_{\text{individual-specific influence, random}} + \epsilon_{it} \\
 &= \underbrace{(\beta_0 + \gamma_{i0})}_{\text{intercept of individual } i} + \underbrace{(\beta_1 + \gamma_{i1})}_{\text{slope of individual } i} x_{it} + \epsilon_{it} \\
 &=: \alpha_{i0} + \alpha_{i1} x_{it} + \epsilon_{it} , \\
 &\text{with } \begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{pmatrix} \right) \text{ and } \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{i5} \end{pmatrix} \sim N_5(\mathbf{0}, \Sigma_i) .
 \end{aligned}$$

The last line in the specification of y_{it} motivates already a model formulation presented in the following.

The two-stage approach

The introduced specification of a linear mixed model given by Laird and Ware (1982) can be motivated by a two-stage model formulation (see Fitzmaurice et al. (2004, Section 8.4) and Verbeke and Molenberghs (2009, Section 3.2)). In Chapter 2 we presented heuristically the main ideas underlying a linear mixed model. Those basic concepts may be illustrated more clearly through a two-stage model formulation of a linear mixed model:

Stage 1:

In this stage we fit a linear regression model to the data of each single individual. That is, for each individual we get individual regression coefficients. This provides to describe the individual relationship between the covariates and the explanatory variables. Hence, for $i = 1, \dots, m$:

$$\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i ,$$

where, as before, \mathbf{y}_i and $\boldsymbol{\epsilon}_i$ denote the individual response and residual vector, respectively, with $\boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \Sigma_i)$. $\boldsymbol{\alpha}_i \in \mathbb{R}^q$ denotes the individual regression coefficient. \mathbf{Z}_i is the design matrix of known, *time-varying* covariates of individual i explaining \mathbf{y}_i . Note that the effect of time-invariant covariates on \mathbf{y}_i would be expressed in an intercept term.

Stage 2:

The second stage explains the variability between the individual regression coefficients by fitting a linear regression model to $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m$. That way, the 2.-Stage regression coefficient describes the overall population effect which underly all individuals and the 2.-Stage residual vector expresses the individual-specific deviation of this entire population trend: For $i = 1, \dots, m$:

$$\boldsymbol{\alpha}_i = \mathbf{A}_i \boldsymbol{\beta} + \boldsymbol{\gamma}_i ,$$

where, as before, $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\boldsymbol{\gamma}_i \in \mathbb{R}^q$ denote the population-specific (fix) and the individual-specific (random) effects, respectively, with $\boldsymbol{\gamma}_i \sim N_q(\mathbf{0}, D)$. \mathbf{A}_i is the design matrix of known, *time-invariant* covariates of individual i explaining $\boldsymbol{\alpha}_i$.

Combining the two stages:

Combining Stage 1 and Stage 2 which are connected through a hierarchical structure provides, for $i = 1, \dots, m$:

$$\mathbf{y}_i = \underbrace{Z_i A_i}_{=: X_i} \boldsymbol{\beta} + Z_i \boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i .$$

By defining $Z_i A_i =: X_i$ we note that the in Postulate 1 to 5 given linear mixed model results from the two-stage specification. However, it is more general, i.e. not every linear mixed model can be expressed in two stages as presented. This is why the two-stage formulation requires that the design matrix X_i can be decomposed in $Z_i A_i$ where Z_i contains only time-varying covariates and A_i only time-invariant covariates. This implies that every time-varying covariate included in the model has a random effect on the response which makes the model fairly complex. Furthermore, random effects for time-invariant covariates cannot be incorporated in the two-stage model.

One may argue with respect to the latter point that it is not senseful to incorporate a random effect for a time-invariant covariate anyway, because this effects might be summed up with the random intercept. However, it is reasonable because random effects for time-invariant covariates may be important to explain the correlation structure between the measurements of the same individual.

The first drawback of modeling a random effect of every included time-varying covariate can be avoided by generalizing the two-stage approach: For including time-varying covariates without random effect as well, the first stage may be formulated as

$$\mathbf{y}_i = Z_i \boldsymbol{\alpha}_i + Z_i^{(F)} \boldsymbol{\alpha}^* + \boldsymbol{\epsilon}_i .$$

$Z_i^{(F)}$ denotes the design matrix of time-varying covariates which do not have an individual-specific effect on \mathbf{y}_i . The corresponding population behavior between these explanatory variables and the response is expressed in $\boldsymbol{\alpha}^*$. This model formulation, however, makes interpretation more complex.

4. The distribution of the response

Using the model specification introduced in Postulate 1 to 5, we concentrate now on considering the distribution of the response variable \mathbf{Y} .

4.1. The two-level hierarchical model

At first, we want to study the behavior of the response vector \mathbf{Y}_i of individual i , when the random effect γ_i , describing the individual-specific deviation from the overall population-averaged evolution, is given. For this, we have to consider the conditional distribution of \mathbf{Y}_i given γ_i . Remember that \mathbf{Y}_i was specified as

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + Z_i\gamma_i + \boldsymbol{\epsilon}_i \quad \text{with} \quad \gamma_i \sim N_q(\mathbf{0}, D) \quad \text{and} \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \Sigma_i) .$$

Hence, we can follow for $i = 1, \dots, m$:

$$\mathbf{Y}_i | \gamma_i \sim N_{n_i}(X_i\boldsymbol{\beta} + Z_i\gamma_i, \Sigma_i) \quad \text{with} \quad \gamma_i \sim N_q(\mathbf{0}, D) .$$

Specified in a overall matrix notation, including all individuals, we get:

$$\boxed{\mathbf{Y} | \boldsymbol{\gamma} \sim N_n(X\boldsymbol{\beta} + Z\boldsymbol{\gamma}, R) \quad \text{with} \quad \boldsymbol{\gamma} \sim N_{mq}(\mathbf{0}, G) .} \quad (4.1)$$

This formulation of the conditional distribution of \mathbf{Y} with the marginal distribution of $\boldsymbol{\gamma}$ we call the *two-level hierarchical model*. The resulting mean response profile for individual i , $X_i\boldsymbol{\beta} + Z_i\gamma_i$ indicates the values we can expect the observations of individual i at the different time points to be, when we know its individual effect γ_i .

The occurring covariance among the different response measurements can be totally explained by the residual covariance matrix R . Note that we assumed R to be a block-diagonal matrix with the individual residual covariance matrices Σ_i as blocks on the diagonal and zero blocks otherwise. This shows us, that the response vectors \mathbf{Y}_i of different individuals are uncorrelated. Due to the equivalence in a multivariate normal distribution of uncorrelated and independent random variables, the response vectors of different individuals are independent.

The individual residual covariance matrices Σ_i are in general no diagonal matrices, which indicates that the response measurements within an individual are correlated. If we assume, however, a diagonal structure for Σ_i , the response values of an individual are independent given its individual-specific effect. Therefore the special diagonal structure for Σ_i ($i = 1, \dots, m$), introduced in Chapter 3, is called *conditional independence structure*.

4.2. The marginal model

Let us look in the second step at the marginal distribution of \mathbf{Y} , i.e. we want to make conclusions about the behavior of the response variable with regard to the entire population. For this purpose we consider again the definition of \mathbf{Y}_i and define a new random variable ϵ_i^* , which includes the whole random part of the model:

For $i = 1, \dots, m$:

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + \underbrace{Z_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i}_{=:\boldsymbol{\epsilon}_i^*} .$$

Then, $E[\boldsymbol{\epsilon}_i^*] = \mathbf{0}$, since $E[\boldsymbol{\gamma}_i] = E[\boldsymbol{\epsilon}_i] = \mathbf{0}$, and $\text{Var}(\boldsymbol{\epsilon}_i^*) = \text{Var}(Z_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i) = Z_i D Z_i^T + \Sigma_i$.

Hence, we may deduce the marginal distribution of \mathbf{Y}_i :

$$\mathbf{Y}_i \sim N_{n_i}(X_i\boldsymbol{\beta}, \underbrace{Z_i D Z_i^T + \Sigma_i}_{=: V_i}) .$$

Likewise, for the overall matrix formulation:

$$\boxed{\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, \underbrace{ZGZ^T + R}_{=: V})} . \quad (4.2)$$

Thus, $X\boldsymbol{\beta}$ indicates the marginal mean of \mathbf{Y} , the population-averaged mean which is expected for all individuals and from which the individual-specific, conditional mean varies randomly. The marginal mean is an average over the distribution of the random effects $\boldsymbol{\gamma}_i$ ($i = 1, \dots, m$).

Note, that similar to R , V is a block-diagonal matrix, with blocks V_i ($i = 1, \dots, m$):

$$V = \begin{pmatrix} V_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_m \end{pmatrix} .$$

This shows, that also in the marginal case, the response variables of different individuals are independent. V_i , however, in fact is not diagonal, which yields correlation among the response variables within an individual. This correlation is explained not only by the residual covariance structure Σ_i , like in the case above, but also by the covariance structure of the random effects D . Even if R_i is a diagonal matrix, V_i is in general not diagonal. Then all of the occurring covariance in the data is induced by D . Hence, incorporating individual-specific effects in the model provokes covariance among the response values of an individual at different points of time.

4.3. Coherence

From the hierarchical model formulation we easily can get the marginal distribution of \mathbf{Y} . Let $f(\mathbf{y}|\boldsymbol{\gamma})$ denote the conditional density of \mathbf{Y} given $\boldsymbol{\gamma}$, and $f(\boldsymbol{\gamma})$ denote the marginal density of $\boldsymbol{\gamma}$. Both densities are normal with mean and covariance specified above. Then we can calculate the joint distribution $f(\mathbf{y}, \boldsymbol{\gamma})$ and out of it the marginal density of \mathbf{Y} using Bayesian theory:

$$f(\mathbf{y}, \boldsymbol{\gamma}) = f(\mathbf{y}|\boldsymbol{\gamma}) f(\boldsymbol{\gamma}) \Rightarrow f(\mathbf{y}) = \int f(\mathbf{y}, \boldsymbol{\gamma}) d\boldsymbol{\gamma} .$$

As seen before, $f(\mathbf{y})$ is the density of a n-dimensional normal distribution with mean $X\boldsymbol{\beta}$ and covariance matrix $(ZGZ^T + R)$. The marginal (population-specific) mean we get also directly from the conditional (individual-specific) mean in a very easy way, using the theory about conditional expectation:

$$E[\mathbf{Y}] = E[E[\mathbf{Y}|\boldsymbol{\gamma}]] = E[X\boldsymbol{\beta} + Z\boldsymbol{\gamma}] = X\boldsymbol{\beta} .$$

In contrast, from the marginal distribution of \mathbf{Y} we can not get its conditional distribution given $\boldsymbol{\gamma}$ and the distribution of $\boldsymbol{\gamma}$.

$\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, V)$ does not imply that variability in the data occurs due to random effects, neither we need in this specification D and Σ_i to be covariance matrices. Only V has to be positive definite and symmetric.

If we only want to estimate $\boldsymbol{\beta}$, we can use the marginal formulation, but for inferences about $\boldsymbol{\gamma}$ we have to consult the hierarchical model (see Chapter 5).

4.4. Example: Simple mixed linear regression

We consider again the example of fitting a linear mixed model including only one covariate. In Chapter 3 we specified the model formulation. In using that notation, we can formulate the marginal and conditional mean response for individual i at time t ($i = 1, 2, 3$ and $t = 1, \dots, 5$), respectively:

$$E[Y_{it}] = \beta_0 + \beta_1 x_{it} \quad \text{and} \quad E[Y_{it}|\boldsymbol{\gamma}_i] = \beta_0 + \beta_1 x_{it} + \gamma_{i0} + \gamma_{i1} x_{it} .$$

The expected value for the t^{th} response variable of individual i , if $\boldsymbol{\gamma}_i$ is known, is the value which is expected for the entire population plus or minus a certain individual-specific amount, indicated by the random effect of individual i , $\boldsymbol{\gamma}_i$.

Beyond that, we now want to investigate the (marginal) variance structure among the five response values of each of the three studied individuals.

Under the assumption that

$$\boldsymbol{\gamma}_i = \begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N_2(\mathbf{0}, \underbrace{\begin{pmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{pmatrix}}_D) \quad \text{and} \quad \boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{i5} \end{pmatrix} \sim N_5(\mathbf{0}, \Sigma_i) \quad \text{with} \quad \Sigma_i = \sigma^2 I_5 ,$$

we may calculate the variance of the i^{th} individual response at time t :
For $i = 1, 2, 3$ and $t = 1, \dots, 5$:

$$\begin{aligned}\text{Var}(Y_{it}) &= \text{Var}(\beta_0 + \beta_1 x_{it} + \gamma_{i0} + \gamma_{i1} x_{it} + \epsilon_{it}) \\ &= \text{Var}(\gamma_{i0}) + 2x_{it} \text{Cov}(\gamma_{i0}, \gamma_{i1}) + x_{it}^2 \text{Var}(\gamma_{i1}) + \text{Var}(\epsilon_{it}) \\ &= d_1^2 + 2x_{it}d_{12} + x_{it}^2 d_2^2 + \sigma^2 .\end{aligned}$$

We used here that $(\beta_0 + \beta_1 x_{it})$ is not random and that γ_i is independent of ϵ_i .

Furthermore, let us consider the covariance coefficient between the t^{th} and k^{th} response measurement of individual i :

For $i = 1, 2, 3$ and $t, k = 1, \dots, 5$ with $t \neq k$:

$$\begin{aligned}\text{Cov}(Y_{it}, Y_{ik}) &= \text{Cov}(\beta_0 + \beta_1 x_{it} + \gamma_{i0} + \gamma_{i1} x_{it} + \epsilon_{it}, \beta_0 + \beta_1 x_{ik} + \gamma_{i0} + \gamma_{i1} x_{ik} + \epsilon_{ik}) \\ &= d_1^2 + x_{ik}d_{12} + x_{it}d_{12} + x_{it}x_{ik}d_2^2 .\end{aligned}$$

We can see here clearly a dependence structure in the repeated response measurements of an individual i , induced by D . This means, the y-observations within an individual are not independent which explains, why including individual-specific effects allows us to model the occurring correlation among longitudinal data. Note here the difference from standard linear models, where we assume the response variables to be independent.

In addition, for $i, j = 1, 2, 3$ and $t, k = 1, \dots, 5$ with $i \neq j$:

$$\text{Cov}(y_{it}, y_{jk}) = \text{Cov}(\beta_0 + \beta_1 x_{it} + \gamma_{i0} + \gamma_{i1} x_{it} + \epsilon_{it}, \beta_0 + \beta_1 x_{jk} + \gamma_{j0} + \gamma_{j1} x_{jk} + \epsilon_{jk}) = 0 ,$$

because random effects of different individuals are assumed to be independent, as well as the residuals of different individuals. Moreover we made the assumption of γ and ϵ to be independent. This shows, that the response vectors of different individuals are independent.

5. Estimation of the linear mixed model parameters

After introducing the general model formulation of a linear mixed model and considering the conditional and marginal distribution of the response variable \mathbf{Y} in the previous chapter, we now want to deduce appropriate estimates for the unknown model parameters. The classical approaches for that purpose shall be presented in detail in the following sections.

In the first part of this chapter (Sections 5.1, 5.2 and 5.3) we concentrate on estimating the fixed and random effects of the model under the assumption that all variance and covariance parameters, the so-called *variance components* of the model, are given. That is we suppose the covariance matrices of the random effect $\boldsymbol{\gamma}$ and the residual vector $\boldsymbol{\epsilon}$, G and R respectively, to be known. Although this situation usually does not occur in practice, it is interesting and helpful to consider because it provides closed-form results for the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ which we may use to determine estimates also in the standard case of unknown variance components.

In the second part (Section 5.4), we drop this assumption and consider on the one hand how to achieve appropriate estimates for the model's variance components. On the other hand we aim to derive again estimates for the model effects under the lack of knowledge of the variance components.

5.1. Estimation of the fixed effects

We start our inferences by estimating the vector of fixed effects $\boldsymbol{\beta}$ when the model's variance components are known.

For this purpose we may refer to the marginal distribution of \mathbf{Y} given in (4.2) which turned out to be a n -dimensional multivariate normal distribution with mean vector $X\boldsymbol{\beta}$ and covariance matrix V . Since $V = ZGZ^T + R$ is a term of both covariance matrices G and R , which we assume here to be given, we consequently know also the exact form of V .

$$f(\mathbf{y}, \boldsymbol{\beta}) = (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - X\boldsymbol{\beta})^T V^{-1} (\mathbf{y} - X\boldsymbol{\beta}) \right\} \quad (5.1)$$

is the corresponding multivariate normal density function.

The common way to obtain an appropriate estimate for $\boldsymbol{\beta}$ is applying maximum likelihood (ML) or generalized least squares (GLS) estimation. We will introduce both methods in this section.

5.1.1. Maximum likelihood estimation

We assume here the reader to be familiar with the theory of maximum likelihood estimation and describe therefore this method in a short and summarizing manner. To get a detailed introduction to the ML concept the reader may consult Bickel and Doksum (2001, p. 114-121).

For the observed response variables \mathbf{y} , the value of $f(\mathbf{y}, \boldsymbol{\beta})$ depends only on the unknown fixed effect $\boldsymbol{\beta}$, a vector of the underlying parameter space $B = \mathbb{R}^p$. The basic principle of the ML theory is choosing this value as estimate for the unknown model parameter $\boldsymbol{\beta}$ which makes the observed data \mathbf{y} most likely. For that purpose we consider the marginal density function (5.1) as function of the unknown $\boldsymbol{\beta}$ while regarding the observed response data \mathbf{y} as fixed values. This function is called the *likelihood* and we denote it by $L(\boldsymbol{\beta})$:

$$\begin{aligned} L(\boldsymbol{\beta}) &= f(\mathbf{y}, \boldsymbol{\beta}) \\ &= (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - X\boldsymbol{\beta})^T V^{-1} (\mathbf{y} - X\boldsymbol{\beta}) \right\}. \end{aligned} \quad (5.2)$$

The concept of the maximum likelihood method is to try to maximize this function with respect to $\boldsymbol{\beta}$, i.e. we seek $\hat{\boldsymbol{\beta}}_{ML}$ which satisfies $L(\hat{\boldsymbol{\beta}}_{ML}) = \max_{\boldsymbol{\beta} \in B} L(\boldsymbol{\beta})$. In general, the ML estimate of $\boldsymbol{\beta}$ is defined as the value $\hat{\boldsymbol{\beta}}_{ML} \in B$ such that

$$L(\hat{\boldsymbol{\beta}}_{ML}) = \sup_{\boldsymbol{\beta} \in B} L(\boldsymbol{\beta})$$

If the existence of $\hat{\boldsymbol{\beta}}_{ML}$ is guaranteed (which will be shown in the following theorem), then

$$\hat{\boldsymbol{\beta}}_{ML} := \arg \max_{\boldsymbol{\beta} \in B} L(\boldsymbol{\beta})$$

and we estimate $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_{ML}$.

For reasons of technical facilitation, however, we consider the *log-likelihood*, the logarithm of the likelihood and deduce the ML estimate for $\boldsymbol{\beta}$ by maximizing this function. Since the logarithm is strictly monotone increasing, the maximum won't be affected in doing so.

$$\begin{aligned} l(\boldsymbol{\beta}) &= \ln(f(\mathbf{y}, \boldsymbol{\beta})) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(|V|) - \frac{1}{2} (\mathbf{y} - X\boldsymbol{\beta})^T V^{-1} (\mathbf{y} - X\boldsymbol{\beta}) \end{aligned} \quad (5.3)$$

In case of existence,

$$\hat{\boldsymbol{\beta}}_{ML} := \arg \max_{\boldsymbol{\beta} \in B} l(\boldsymbol{\beta}).$$

Theorem 5.1.1

The maximum likelihood estimate (MLE) of the fixed effect $\boldsymbol{\beta}$ in a linear mixed

model with log-likelihood (5.3) is

$$\begin{aligned}\hat{\beta}_{ML} &= (X^T V^{-1} X)^{-1} X^T V^{-1} \mathbf{y} \\ &= \left(\sum_{i=1}^m X_i^T V_i^{-1} X_i \right)^{-1} \sum_{i=1}^m X_i^T V_i^{-1} \mathbf{y}_i .\end{aligned}$$

Proof: Note that only the last summand of $l(\beta)$ depends on β and due to the negative sign the maximization problem is equivalent to minimizing a quadratic form:

$$\begin{aligned}\max_{\text{w.r.t. } \beta} l(\beta) &\iff \min_{\text{w.r.t. } \beta} (\mathbf{y} - X\beta)^T V^{-1} (\mathbf{y} - X\beta) \\ &\iff \min_{\text{w.r.t. } \beta} (\mathbf{y}^T V^{-1} \mathbf{y} - 2 \mathbf{y}^T V^{-1} X\beta + \beta^T X^T V^{-1} X\beta) .\end{aligned}\tag{5.4}$$

This function is convex, since V^{-1} is positive definite and has therefore a unique global minimum. Differentiating provides

$$-2X^T V^{-1} \mathbf{y} + 2X^T V^{-1} X\beta \stackrel{!}{=} \mathbf{0} \iff \beta = (X^T V^{-1} X)^{-1} X^T V^{-1} \mathbf{y} .$$

Basic properties of derivation with respect to a vector are listed in A.1.3. The second derivative of the function (5.4) w.r.t. β , i.e. the negative score $-\frac{\partial^2}{\partial \beta \partial \beta^T} l(\beta)$, is $2X^T V^{-1} X$.

According to Lemma A.3.1, this matrix is positive definite. Hence $\hat{\beta}_{ML}$ minimizes (5.4) and is therefore the unique maximizer of $l(\beta)$.

Note that

$$\begin{aligned}X^T V^{-1} X &= (X_1^T, \dots, X_m^T) \begin{pmatrix} V_1^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & V_m^{-1} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} \\ &= \sum_{i=1}^m X_i^T V_i^{-1} X_i\end{aligned}$$

and

$$\begin{aligned}X^T V^{-1} \mathbf{y} &= (X_1^T, \dots, X_m^T) \begin{pmatrix} V_1^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & V_m^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{pmatrix} \\ &= \sum_{i=1}^m X_i^T V_i^{-1} \mathbf{y}_i ,\end{aligned}$$

which shows the second statement. ■

5.1.2. Generalized least squares estimation

A further standard approach to derive an appropriate estimate for β is applying the method of generalized least squares to the marginal model. Remember that we may denote the marginal model (4.2) in form of a standard linear model under the constraint that the assumption of homogeneous and uncorrelated error variance may be violated, i.e. the covariance structure of the error terms is given by the general covariance matrix V :

$$\mathbf{Y} = X\beta + \epsilon^*, \quad \epsilon^* = Z\gamma + \epsilon \quad \text{and} \quad \epsilon^* \sim N_n(\mathbf{0}, V). \quad (5.5)$$

Since V is symmetric there exists a decomposition $V = V^{\frac{1}{2}}(V^{\frac{1}{2}})^T$. This representation can be achieved for example by spectral decomposition, i.e. matrix diagonalization in the form $V = ADA^T$ where D is real and diagonal and A is an orthogonal matrix. Then one may define $V^{\frac{1}{2}} := AD^{\frac{1}{2}}$. Since V is in addition positive definite, the Cholesky decomposition is an alternative for deriving $V^{\frac{1}{2}}$.

Transforming (5.5) by multiplying $V^{-\frac{1}{2}}$ provides a model in form of a standard linear model with homoscedastic and uncorrelated residual values:

$$\underbrace{V^{-\frac{1}{2}}\mathbf{Y}}_{\tilde{\mathbf{Y}}} = \underbrace{V^{-\frac{1}{2}}X}_{\tilde{X}} \beta + \underbrace{V^{-\frac{1}{2}}\epsilon^*}_{\tilde{\epsilon}} \quad \text{with} \quad \tilde{\epsilon} \sim N_n(\mathbf{0}, \underbrace{V^{-\frac{1}{2}}V(V^{-\frac{1}{2}})^T}_{\mathbf{I}_n}).$$

From the theory of standard linear models we know, that the corresponding ordinary least squares estimator turns out to be

$$\begin{aligned} \hat{\beta}_{LS} &= (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{\mathbf{y}} \\ &= ((V^{-\frac{1}{2}}X)^T V^{-\frac{1}{2}}X)^{-1} (V^{-\frac{1}{2}}X)^T V^{-\frac{1}{2}}\mathbf{y}. \end{aligned}$$

Thus, the generalized least squares estimator of β is

$$\boxed{\hat{\beta}_{LS} = (X^T V^{-1} X)^{-1} X^T V^{-1} \mathbf{y} .}$$

Hence, $\hat{\beta}_{LS}$ is the maximum likelihood estimator $\hat{\beta}_{ML}$ of Theorem 5.1.1.

From (5.4) we obtain that $\hat{\beta}_{LS}$ minimizes the so-called *weighted residual sum of squares*

$$(\mathbf{y} - X\beta)^T V^{-1} (\mathbf{y} - X\beta) .$$

Referring to Myers (1986, Section 7.1), it turns out that this estimator has better properties than the ordinary least squares estimator for the marginal model (5.5), which is obtained by minimizing the corresponding residual sum of squares, i.e.

$$\min_{\text{w.r.t. } \beta} (\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta) .$$

Therefore, in case of heterogeneous variance, always the generalized least squares estimator, i.e. the ordinary least squares estimator for the transformed model, is preferred.

5.1.3. Properties of the estimate

Since $\hat{\beta}_{ML}$ and $\hat{\beta}_{LS}$ turned out to be equal we omit the subindex and denote by

$$\boxed{\hat{\beta} := (X^T V^{-1} X)^{-1} X^T V^{-1} \mathbf{y}} \quad (5.6)$$

the estimate for β obtained by maximum likelihood or generalized least squares estimation.

It is straightforward to derive that $\hat{\beta}$ is unbiased:

$$\begin{aligned} E[\hat{\beta}] &= (X^T V^{-1} X)^{-1} X^T V^{-1} E[\mathbf{Y}] \\ &= (X^T V^{-1} X)^{-1} X^T V^{-1} X \beta \\ &= \beta . \end{aligned}$$

The variance of $\hat{\beta}$ may be calculated also fairly easy:

$$\begin{aligned} \text{Var}(\hat{\beta}) &= (X^T V^{-1} X)^{-1} X^T V^{-1} \text{Var}(\mathbf{Y}) V^{-1} X (X^T V^{-1} X)^{-1} \\ &= (X^T V^{-1} X)^{-1} X^T V^{-1} V V^{-1} X (X^T V^{-1} X)^{-1} \\ &= (X^T V^{-1} X)^{-1} . \end{aligned}$$

Furthermore, $\hat{\beta}$ is the best linear unbiased estimator (BLUE) of β . This was shown by Aitken (1934) who extended the Gauß-Markov theorem for the generalized least squares estimator. That is, under all linear, unbiased estimators $\hat{\beta}$ has minimal variance.

5.2. Estimation of the random effects - The best predictor

The fixed effect β is an unknown but constant parameter vector, the random effect γ , however, is assumed to be a random vector underlying a certain distribution. Therefore we can't apply exactly the same estimation methods for γ as for the fixed effects β . The particular individuals in the data set are required to be randomly chosen from the overall underlying population, so we speak about *prediction* of their individual influence on the response variable rather than *estimation*.

Unlike estimating β , we have to consider for the prediction procedure of γ the hierarchical model (4.1),

$$\mathbf{Y} \mid \gamma \sim N_n(X\beta + Z\gamma, R) \quad \text{with} \quad \gamma \sim N_{mq}(\mathbf{0}, G) ,$$

since the marginal distribution of \mathbf{Y} does not imply any information about the random effects.

The standard approach to derive a *best predictor* for γ is presented in this section. The term *best estimate* of a single model parameter is commonly known as this estimate which

has minimal variance or minimal mean squared error (MSE). Best prediction refers to the mean squared prediction error (MSPE). Thus, for $\gamma_i \in \mathbb{R}$ be a component of the vector of random effects $\boldsymbol{\gamma}$, the best predictor for γ_i minimizes

$$\text{MSPE}(\hat{\gamma}_i) = E[(\hat{\gamma}_i - \gamma_i)^2]$$

with respect to $\hat{\gamma}_i$, a measurable function of the response \mathbf{Y} . Definitions of the MSE and the MSPE are given in A.2.2. Referring to McCulloch and Searle (2001, Chapter 9), the best predictor in the multivariate case is defined as follows:

Definition 5.2.1

Let \mathbf{y} be an observed data vector of a statistical model with a random model parameter vector $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_q)^T$. Then the **best predictor** (BP) $\hat{\boldsymbol{\vartheta}}$ of $\boldsymbol{\vartheta}$ is the measurable function of \mathbf{Y} satisfying

$$\hat{\boldsymbol{\vartheta}} := \arg \min_{\tilde{\boldsymbol{\vartheta}} \text{ measurable function of } \mathbf{Y}} E[(\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})^T A (\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})] , \quad (5.7)$$

where A denotes any symmetric and positive definite matrix.

Note that for $\boldsymbol{\vartheta}$ being a single value ($q=1$) and $A = 1$, (5.7) is the MSPE.

Theorem 5.2.1 McCulloch and Searle (2001, Chapter 9)

Let \mathbf{y} be an observed data vector of a statistical model with a random model parameter vector $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_q)^T$. Then the best predictor does not depend on the symmetric and positive definite matrix A and is given by

$$\hat{\boldsymbol{\vartheta}} = E[\boldsymbol{\vartheta} | \mathbf{Y}] .$$

Proof: Let $\hat{\boldsymbol{\vartheta}}$ be some measurable function of \mathbf{Y} , then

$$\begin{aligned} E[(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})^T A (\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})] &= E[(\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}] + E[\boldsymbol{\vartheta} | \mathbf{Y}] - \boldsymbol{\vartheta})^T A (\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}] + E[\boldsymbol{\vartheta} | \mathbf{Y}] - \boldsymbol{\vartheta})] \\ &= E[(\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}])^T A (\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}])] + 2 E[(\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}])^T A (E[\boldsymbol{\vartheta} | \mathbf{Y}] - \boldsymbol{\vartheta})] \\ &\quad + E[(E[\boldsymbol{\vartheta} | \mathbf{Y}] - \boldsymbol{\vartheta})^T A (E[\boldsymbol{\vartheta} | \mathbf{Y}] - \boldsymbol{\vartheta})] . \end{aligned}$$

The last summand is independent of $\hat{\boldsymbol{\vartheta}}$ and the second summand turns out to be zero:

$$\begin{aligned} &E[(\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}])^T A (E[\boldsymbol{\vartheta} | \mathbf{Y}] - \boldsymbol{\vartheta})] \\ &= E[E[(\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}])^T A (E[\boldsymbol{\vartheta} | \mathbf{Y}] - \boldsymbol{\vartheta}) | \mathbf{Y}]] \\ &= E[(\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}])^T A E[E[\boldsymbol{\vartheta} | \mathbf{Y}] - \boldsymbol{\vartheta} | \mathbf{Y}]] , \text{ since } (\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}])^T A \text{ is measurable w.r.t. } \mathbf{Y} \\ &= E[(\hat{\boldsymbol{\vartheta}} - E[\boldsymbol{\vartheta} | \mathbf{Y}])^T A \underbrace{(E[\boldsymbol{\vartheta} | \mathbf{Y}] - E[\boldsymbol{\vartheta} | \mathbf{Y}])}_{=0}] \\ &= 0 . \end{aligned}$$

A was assumed to be positive definite, thus $(\hat{\boldsymbol{\vartheta}} - \mathbb{E}[\boldsymbol{\vartheta}|\mathbf{Y}])^T A (\hat{\boldsymbol{\vartheta}} - \mathbb{E}[\boldsymbol{\vartheta}|\mathbf{Y}]) \geq 0$ and $= 0 \Leftrightarrow \hat{\boldsymbol{\vartheta}} - \mathbb{E}[\boldsymbol{\vartheta}|\mathbf{Y}] = 0$. This proves that the conditional mean $\mathbb{E}[\boldsymbol{\vartheta}|\mathbf{Y}]$ of $\boldsymbol{\vartheta}$ is the best predictor for $\boldsymbol{\vartheta}$, independent of the choice of A . ■

From this result follows that for deriving the BP of $\boldsymbol{\gamma}$ we need to calculate its conditional mean given the observed data vector.

Corollary 5.2.1 McCulloch and Searle (2001, Chapter 9)

Given a linear mixed model specified by Postulate 1 to 5 of Chapter 3. Then

$$\hat{\boldsymbol{\gamma}}_{BP} = GZ^T V^{-1}(\mathbf{Y} - X\boldsymbol{\beta}) \quad (5.8)$$

is the best predictor for the vector of random effects $\boldsymbol{\gamma}$.

Proof: The BP was given in Theorem 5.2.1 to be $\mathbb{E}[\boldsymbol{\gamma}|\mathbf{Y}]$. Hence, we have to derive the conditional distribution of $\boldsymbol{\gamma}|\mathbf{Y}$. Consider the covariance of \mathbf{Y} and $\boldsymbol{\gamma}$

$$\text{Cov}(\mathbf{Y}, \boldsymbol{\gamma}) = \text{Cov}(X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \boldsymbol{\gamma}) = Z \text{Cov}(\boldsymbol{\gamma}, \boldsymbol{\gamma}) + \underbrace{\text{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\gamma})}_{=0} = ZG \quad (5.9)$$

and remember that $\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, V)$ and $\boldsymbol{\gamma} \sim N_{mq}(\mathbf{0}, G)$. The joint distribution of \mathbf{Y} and $\boldsymbol{\gamma}$ is also normal Fahrmeir et al. (2007):

$$\begin{pmatrix} \mathbf{Y} \\ \boldsymbol{\gamma} \end{pmatrix} \sim N_{n+mq} \left(\begin{pmatrix} X\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} V & ZG \\ GZ^T & G \end{pmatrix} \right). \quad (5.10)$$

Using A.2.1 shows

$$\boldsymbol{\gamma} | \mathbf{Y} \sim N_{mq}(GZ^T V^{-1}(\mathbf{Y} - X\boldsymbol{\beta}), G - GZ^T V^{-1} ZG), \quad (5.11)$$

which implies $\hat{\boldsymbol{\gamma}}_{BP} = \mathbb{E}[\boldsymbol{\gamma}|\mathbf{Y}] = GZ^T V^{-1}(\mathbf{Y} - X\boldsymbol{\beta})$. ■

Note that $\hat{\boldsymbol{\gamma}}_{BP}$ may also be derived using Bayesian techniques. Since $\boldsymbol{\gamma}$ is a random variable, we may consider its distribution, $\boldsymbol{\gamma} \sim N_{mq}(\mathbf{0}, G)$, as prior distribution which does not depend on the data. We denote the corresponding density by $\pi(\boldsymbol{\gamma})$, the prior density. Let $f(\mathbf{y}, \boldsymbol{\gamma})$ denote the density of (5.10). By using the observed data we then may deduce the posterior density of $\boldsymbol{\gamma}$ as

$$\pi(\boldsymbol{\gamma}|\mathbf{Y}) = \frac{f(\mathbf{y}, \boldsymbol{\gamma})\pi(\boldsymbol{\gamma})}{\int f(\mathbf{y}, \boldsymbol{\gamma})\pi(\boldsymbol{\gamma})d\boldsymbol{\gamma}}.$$

From (5.11) we know that this is the density of a multivariate normal distribution with mean vector $GZ^T V^{-1}(\mathbf{Y} - X\boldsymbol{\beta})$ and covariance matrix $G - GZ^T V^{-1} ZG$. Using the mean of the posterior distribution to predict $\boldsymbol{\gamma}$ is a commonly used method in the Bayesian framework. It yields exactly the best predictor derived in the previous theorem. For further reading we refer to Verbeke and Molenberghs (2009, Section 7.2) and Fahrmeir et al. (2007, Section 6.5).

Remark 5.2.1 Searle et al. (1992, Section 7.3.)

Without the assumption of normality, (5.8) may be derived as best linear predictor (BLP) for γ . That is, among all linear functions of \mathbf{Y} , $\mathbf{a} + B\mathbf{Y}$ with arbitrary \mathbf{a} and B , $\hat{\gamma}_{BLP} := GZ^TV^{-1}(\mathbf{Y} - X\beta)$ minimizes (5.7).

Properties

Considering distributional properties of $\hat{\gamma}_{BP}$ one first may notice that it is unbiased:

$$\begin{aligned} E[\hat{\gamma}_{BP} - \gamma] &= GZ^TV^{-1} E[\mathbf{Y} - X\beta] \\ &= GZ^TV^{-1}(X\beta - X\beta) \\ &= \mathbf{0} . \end{aligned}$$

Remember, since γ is random, unbiasedness of $\hat{\gamma}_{BP}$ refers to $E[\hat{\gamma}_{BP} - \gamma] = \mathbf{0}$. Asking for $E[\hat{\gamma}_{BP}] = \gamma$ is not meaningful (see A.2.2).

Furthermore,

$$\begin{aligned} \text{Var}(\hat{\gamma}_{BP}) &= \text{Var}(GZ^TV^{-1}\mathbf{Y} - GZ^TV^{-1}X\beta) \\ &= GZ^TV^{-1}VV^{-1}ZG \\ &= GZ^TV^{-1}ZG . \end{aligned}$$

The main drawback of the best predictor of γ is that $\hat{\gamma}_{BP}$ depends on β . This parameter vector, however, is as γ not known and has to be estimated. We therefore can't use $\hat{\gamma}_{BP}$ as predictor for γ in practice and have to search for a predictor which is independent of β . It turns out that the intuitive solution of this problem replacing β in $\hat{\gamma}_{BP}$ by $\hat{\beta}$ provides an appropriate predictor (see Subsection 5.3.1).

5.3. Simultaneous estimation of the fixed and random effects

After deriving estimates for each of the fixed and random effect separately, we consider in this section methods where estimates for both effects are derived simultaneously. One technique for that purpose is the derivation of the best linear unbiased predictor of a linear function of β and γ (Subsection 5.3.1). This way provides a predictor for γ not depending on β which makes it practically applicable (in contrast to the BP derived in the previous section). Another self-evident approach is the maximization of the joint log-likelihood of \mathbf{y} and γ with respect to β and γ (Subsection 5.3.2). It will turn out that using that method one gets the estimates by solving a linear equation system which (often) has computationally preferable properties.

5.3.1. The best linear unbiased predictor for a linear function of the fixed and random effects

To find an appropriate estimate of a linear function $C^T\boldsymbol{\beta} + \boldsymbol{\gamma}$ of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ poses initially the question whether we *estimate* or *predict* this function. As mentioned already in the beginning of Section 5.2, the term *estimation* related to random vectors is not very meaningful. Referring to Searle et al. (1992), we agree to use *predict* a function of both, fixed and random effects, rather than *estimate* it.

Definition 5.3.1

Let \mathbf{y} be an observed data vector of a statistical model with a random model parameter vector $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_q)^T$. Then the **best linear unbiased predictor** (BLUP) $\hat{\boldsymbol{\vartheta}}$ of $\boldsymbol{\vartheta}$

- is a linear function of \mathbf{Y} , i.e. $\hat{\boldsymbol{\vartheta}} = \mathbf{a} + B\mathbf{Y}$ with some vector \mathbf{a} and arbitrary matrix B ,
- is unbiased, i.e. $E[\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}] = \mathbf{0}$ and
- minimizes (5.7) .

Theorem 5.3.1 Henderson (1963)

Given a linear mixed model specified by Postulate 1 to 5 of Chapter 3. Let

$$\mathbf{c} = C^T\boldsymbol{\beta} + \boldsymbol{\gamma}$$

be any linear function of $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\boldsymbol{\gamma} \in \mathbb{R}^q$ where the known matrix C^T has a decomposition $C^T = MX$ for some M . Let $\hat{\boldsymbol{\beta}}$ denote the estimate of $\boldsymbol{\beta}$ given in (5.6). Then

$$\hat{\mathbf{c}} = C^T\hat{\boldsymbol{\beta}} + GZ^TV^{-1}(\mathbf{Y} - X\hat{\boldsymbol{\beta}})$$

is the best linear unbiased predictor of \mathbf{c} .

This result was first shown by Henderson (1963). In Theorem 5.3.1 his result is presented as formulated in Searle et al. (1992, pp. 269) and we prove it in detail also referring to the latter reference. Harville (1976) gave an extension of the Gauß-Markov theorem and proved it there in a very general way. In Searle et al. (1992, pp. 269-275) one may find, beyond the one given here, several different approaches to verify the statement.

Proof: Let $\hat{\mathbf{c}}$ be a linear function of \mathbf{Y} , $\hat{\mathbf{c}} = \mathbf{a} + B\mathbf{Y}$ with \mathbf{a} and B such that $\hat{\mathbf{c}}$ is unbiased and has minimal MSPE among all linear and unbiased predictors. Unbiasedness postulates

$$E[\hat{\mathbf{c}} - \mathbf{c}] \stackrel{!}{=} \mathbf{0} \quad \Rightarrow \quad \mathbf{a} + BX\boldsymbol{\beta} - C^T\boldsymbol{\beta} \stackrel{!}{=} \mathbf{0} \quad \text{for all } \boldsymbol{\beta} \text{ and } \mathbf{a}.$$

Since B is assumed to be independent of β ,

$$\mathbf{a} = \mathbf{0} \quad \text{and} \quad BX = C^T.$$

Hence, $\hat{\mathbf{c}} = B\mathbf{Y}$ with $BX = C^T$ such that $\hat{\mathbf{c}}$ has minimal MSPE. Note that we made explicitly the assumption that C^T can be represented in that way. Thus we have to minimize the following with respect to B where A denotes a positive definite and symmetric matrix (see 5.7):

$$\begin{aligned} & \mathbb{E}[(\hat{\mathbf{c}} - \mathbf{c})^T A (\hat{\mathbf{c}} - \mathbf{c})] \\ &= \mathbb{E}[(B\mathbf{Y} - C^T\beta - \gamma)^T A (B\mathbf{Y} - C^T\beta - \gamma)] \\ &\stackrel{(A.1)}{=} \text{tr}(A \text{Var}(B\mathbf{Y} - C^T\beta - \gamma)) + \underbrace{\mathbb{E}[B\mathbf{Y} - C^T\beta - \gamma]^T}_{=0} A \underbrace{\mathbb{E}[B\mathbf{Y} - C^T\beta - \gamma]}_{=0} \\ &= \text{tr}(A (\text{Var}(B\mathbf{Y}) + \text{Var}(\gamma) - \text{Cov}(B\mathbf{Y}, \gamma) - \text{Cov}(\gamma, B\mathbf{Y}))) \\ &\stackrel{(5.9)}{=} \text{tr}(A (BV B^T + G - BZG - GZ^T B^T)) \\ &= \text{tr}(A(BV B^T - BZG - GZ^T B^T)) + \text{tr}(AG), \end{aligned}$$

where the last summand is a constant independent of B .

By defining the matrices $T := X(X^T V^{-1} X)^{-1} X^T V^{-1}$ and $Q := (I - T)$, we express B as

$$\begin{aligned} B &= BT + B - BT = BT + (I - T)B = BT + BQ \\ &\quad \text{where } BT = C^T (X^T V^{-1} X)^{-1} X^T V^{-1} \text{ since } BX = C^T. \end{aligned} \quad (5.12)$$

For clear arrangement, basic properties of T and Q are summarized in A.1.4. Referring to this Subsection, it can be easily verified that

$$TVQ^T = 0 \quad \text{and} \quad QV = QVQ^T = VQ^T. \quad (5.13)$$

Hence, we get that

$$\begin{aligned} & BV B^T - BZG - GZ^T B^T \\ &= (BT + BQ)V(BT + BQ)^T - (BT + BQ)ZG - GZ^T(BT + BQ)^T \\ &\stackrel{(5.13)}{=} BTVT^T B^T + BQVQ^T B^T + \underbrace{2BT VQ^T}_{=0} B^T - BTZG - BQZG - GZ^T T^T B^T - GZ^T Q^T B^T \\ &= BTVT^T B^T + BQVQ^T B^T - BTZG - BQV V^{-1} ZG - GZ^T T^T B^T - GZ^T V^{-1} VQ^T B^T \\ &\stackrel{(5.13)}{=} BTVT^T B^T + \underline{BQVQ^T B^T} - BTZG - \underline{BQVQ^T V^{-1} ZG} - GZ^T T^T B^T - \underline{GZ^T V^{-1} QVQ^T B^T} \\ &= (BQ - GZ^T V^{-1} Q)V(BQ - GZ^T V^{-1} Q)^T \\ &\quad - GZ^T V^{-1} QVQ^T V^{-1} ZG + BTVT^T B^T - BTZG - GZ^T T^T B^T. \end{aligned}$$

BT was shown in (5.12) to be independent of B , the minimization problem remains therefore in

$$\min_{\text{w.r.t. } B} \text{tr}(A (BQ - GZ^T V^{-1} Q)V(BQ - GZ^T V^{-1} Q)^T).$$

A is positive definite and the quadratic form is positive semidefinite, therefore this function is greater or equal to zero (see A.1.2) and obtains its minimum if and only if the quadratic form is equal to zero, i.e.

$$\Leftrightarrow BQ = GZ^T V^{-1} Q = GZ^T V^{-1} (I - X(X^T V^{-1} X)^{-1} X^T V^{-1}) =: GZ^T P. \quad (5.14)$$

Note that this result does not depend on A . Substituting in (5.12) yields

$$B = C^T(X^T V^{-1} X)^{-1} X^T V^{-1} + GZ^T P .$$

Using (5.6) we get $PY = V^{-1}(Y - X\hat{\beta})$ and the BLUP of c :

$$\hat{c} = BY = C^T \hat{\beta} + GZ^T V^{-1}(Y - X\hat{\beta}) .$$

Basic properties of the defined auxiliary matrix P are also summarized in A.1.4. ■

Note that this result is not based on the assumption of normality.

The derived \hat{c} can be split up into two parts: the first part, $C^T \hat{\beta}$ is the BLUE of $C^T \beta$ (see Searle et al. (1992)). The second part, $GZ^T V^{-1}(Y - X\hat{\beta})$ is the best predictor $\hat{\gamma}_{BP}$ of γ with β replaced by its estimate $\hat{\beta}$. This provides the commonly used predictor for γ which we define as

$$\boxed{\hat{\gamma} := GZ^T V^{-1}(y - X\hat{\beta}) .} \quad (5.15)$$

It's advantage over $\hat{\gamma}_{BP}$ is, like already has been mentioned, that it does not involve the unknown vector of fixed effects and is therefore computable. Remember that we assume here the variance components describing G and V to be known. $\hat{\gamma}$ is often stated as the BLUP of γ . This might be confusing, since $\hat{\gamma}_{BP}$ is linear and unbiased as well. Correctly spoken, $\hat{\gamma}$ has minimal MSPE among all unbiased predictors for γ which are linear functions of Y and do not depend on β .

Properties

First we note that $\hat{\gamma}$ is unbiased since $\hat{\beta}$ was shown to be unbiased (see Subsection 5.1.3):

$$\begin{aligned} E[\hat{\gamma} - \gamma] &= GZ^T V^{-1} E[Y - X\hat{\beta}] \\ &= GZ^T V^{-1} (X\beta - X\beta) \\ &= \mathbf{0} . \end{aligned}$$

Using the auxiliary matrix P defined in (5.14) and referring to A.1.4 we may formulate $\hat{\gamma}$ briefly as $\hat{\gamma} = GZ^T P Y$. Thus, for the variance of $\hat{\gamma}$ one gets

$$\begin{aligned} \text{Var}(\hat{\gamma}) &= GZ^T \underbrace{PVP}_P ZG \quad (\text{see A.1.4}) \\ &= GZ^T (V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}) ZG \\ &= \text{Var}(\hat{\gamma}_{BP}) - GZ^T V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1} ZG . \end{aligned}$$

5.3.2. Henderson's mixed model equations

The approach of deriving the joint maximum likelihood estimates of β and γ was first suggested by Henderson (1950). The joint distribution of Y and γ can be derived by applying Bayes' theorem:

$$f(y, \beta, \gamma) = f(y|\gamma) f(\gamma) .$$

Referring to the hierarchical model (4.1) and inserting the corresponding multivariate densities yields:

$$f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = (2\pi)^{-\frac{n}{2}} |R|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - X\boldsymbol{\beta} - Z\boldsymbol{\gamma})^T R^{-1} (\mathbf{y} - X\boldsymbol{\beta} - Z\boldsymbol{\gamma}) \right\} \\ \cdot (2\pi)^{-\frac{mq}{2}} |G|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \boldsymbol{\gamma}^T G^{-1} \boldsymbol{\gamma} \right\} .$$

This is the density of a multivariate normal distribution with mean vector $(X\boldsymbol{\beta}, \mathbf{0})^T$ and covariance matrix $\begin{pmatrix} V & ZG \\ GZ^T & G \end{pmatrix}$, as we saw in (5.10). Hence the joint log-likelihood satisfies

$$l(\boldsymbol{\beta}, \boldsymbol{\gamma}) \propto -\frac{1}{2} (\mathbf{y} - X\boldsymbol{\beta} - Z\boldsymbol{\gamma})^T R^{-1} (\mathbf{y} - X\boldsymbol{\beta} - Z\boldsymbol{\gamma}) - \frac{1}{2} \boldsymbol{\gamma}^T G^{-1} \boldsymbol{\gamma} .$$

It can be easily seen that maximizing $l(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ is equivalent to the minimization of the following expression:

$$\max_{\text{w.r.t. } \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\gamma} \in \mathbb{R}^{mq}} l(\boldsymbol{\beta}, \boldsymbol{\gamma}) \iff \min_{\text{w.r.t. } \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\gamma} \in \mathbb{R}^{mq}} q(\boldsymbol{\beta}, \boldsymbol{\gamma})$$

with

$$q(\boldsymbol{\beta}, \boldsymbol{\gamma}) = (\mathbf{y} - X\boldsymbol{\beta} - Z\boldsymbol{\gamma})^T R^{-1} (\mathbf{y} - X\boldsymbol{\beta} - Z\boldsymbol{\gamma}) + \boldsymbol{\gamma}^T G^{-1} \boldsymbol{\gamma} \\ = \mathbf{y}^T R^{-1} \mathbf{y} - 2 \mathbf{y}^T R^{-1} X\boldsymbol{\beta} - 2 \mathbf{y}^T R^{-1} Z\boldsymbol{\gamma} + \boldsymbol{\beta}^T X^T R^{-1} X\boldsymbol{\beta} + 2 \boldsymbol{\beta}^T X^T R^{-1} Z\boldsymbol{\gamma} \\ + \boldsymbol{\gamma}^T Z^T R^{-1} Z\boldsymbol{\gamma} + \boldsymbol{\gamma}^T G^{-1} \boldsymbol{\gamma}$$

Taking derivatives (see A.1.3) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and equating to zero yields:

$$\frac{\partial q(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} = -2 X^T R^{-1} \mathbf{y} + 2 X^T R^{-1} X\boldsymbol{\beta} + 2 X^T R^{-1} Z\boldsymbol{\gamma} \stackrel{!}{=} \mathbf{0} \quad (5.16)$$

$$\frac{\partial q(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = -2 Z^T R^{-1} \mathbf{y} + 2 Z^T R^{-1} X\boldsymbol{\beta} + 2 Z^T R^{-1} Z\boldsymbol{\gamma} + 2 G^{-1} \boldsymbol{\gamma} \stackrel{!}{=} \mathbf{0} \quad (5.17)$$

Combining these equations in a joint matrix notation provides the so-called *mixed model equations* which were set up by Henderson (1950). The joint ML estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ consequently will solve the linear equation system

$$\boxed{\begin{pmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} X^T R^{-1} \mathbf{y} \\ Z^T R^{-1} \mathbf{y} \end{pmatrix}} \quad (5.18)$$

as necessary criterion to be maximizers of the likelihood.

Theorem 5.3.2 Henderson et al. (1959), Henderson (1963)

Given a linear mixed model specified by Postulate 1 to 5 of Chapter 3. Then $\hat{\boldsymbol{\beta}} =$

$(X^T V^{-1} X)^{-1} \mathbf{y}$ and $\hat{\boldsymbol{\gamma}} = G Z^T V^{-1} (\mathbf{y} - X \hat{\boldsymbol{\beta}})$ are the unique solutions of the mixed model equations (5.18).

Proof: Defining $K := (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1}$ we see from (5.17) that

$$\tilde{\boldsymbol{\gamma}} = K(\mathbf{y} - X \tilde{\boldsymbol{\beta}}). \quad (5.19)$$

Substituting this into 5.16 we get

$$\begin{aligned} X^T R^{-1} X \tilde{\boldsymbol{\beta}} - X^T R^{-1} Z K X \tilde{\boldsymbol{\beta}} &= X^T R^{-1} \mathbf{y} - X^T R^{-1} Z K \mathbf{y} \\ \Leftrightarrow X^T (R^{-1} - R^{-1} Z K) X \tilde{\boldsymbol{\beta}} &= X^T (R^{-1} - R^{-1} Z K) \mathbf{y} \\ \Leftrightarrow X^T W X \tilde{\boldsymbol{\beta}} &= X^T W \mathbf{y} \quad \text{with } W = R^{-1} - R^{-1} Z K. \end{aligned}$$

Hence, to proof that $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ we have to show that $W = V^{-1}$, therefore we calculate

$$\begin{aligned} VW &= (ZGZ^T + R)(R^{-1} - R^{-1} ZK) \\ &= I + ZGZ^T R^{-1} - ZK - ZGZ^T R^{-1} ZK \\ &= I + ZGZ^T R^{-1} - Z(I + GZ^T R^{-1} Z)K \\ &= I + ZGZ^T R^{-1} - ZG(G^{-1} + Z^T R^{-1} Z)(Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} \\ &= I. \end{aligned}$$

This shows the first part of the theorem.

Furthermore, it follows from (5.19):

$$\begin{aligned} \tilde{\boldsymbol{\gamma}} &= K(\mathbf{y} - X \tilde{\boldsymbol{\beta}}) \\ &= K V V^{-1} (\mathbf{y} - X \tilde{\boldsymbol{\beta}}) \\ &= (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} (ZGZ^T + R) V^{-1} (\mathbf{y} - X \tilde{\boldsymbol{\beta}}) \\ &= (Z^T R^{-1} Z + G^{-1})^{-1} (Z^T R^{-1} ZGZ^T + Z^T) V^{-1} (\mathbf{y} - X \tilde{\boldsymbol{\beta}}) \\ &= (Z^T R^{-1} Z + G^{-1})^{-1} (Z^T R^{-1} Z + G^{-1}) G Z^T V^{-1} (\mathbf{y} - X \tilde{\boldsymbol{\beta}}) \\ &= G Z^T V^{-1} (\mathbf{y} - X \tilde{\boldsymbol{\beta}}). \end{aligned}$$

Since $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ this shows that $\tilde{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}$. ■

Hence, solving the mixed model equations (5.18) provides the estimate of $\boldsymbol{\beta}$ (5.6) and the predictor of $\boldsymbol{\gamma}$ (5.15) which were derived in the previous sections separately. One big advantage of this procedure is that there is no need of inverting the $(n \times n)$ matrix V which is necessary in the direct calculation of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ as $(X^T V^{-1} X)^{-1} \mathbf{y}$ and $G Z^T V^{-1} (\mathbf{y} - X \hat{\boldsymbol{\beta}})$, respectively. In solving the mixed model equations we do only require the inversion of R and G , which is usually computationally less intensive since these covariance matrices often have diagonal form (see McCulloch and Searle (2001)).

One drawback, however, underlies all estimation approaches considered so far: we assumed that we know the exact form of the covariance matrices R , G and V but this is usually not the case. Therefore it remains the derivation of appropriate estimates for the variance components that we may use in the obtained formulas for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$.

5.4. Estimation of the variance components

Let us now concentrate on the standard case of unknown variance components. That means, that we have in addition to β and γ further unknown model parameters, namely the variance and covariance parameters of the model given by the covariance matrices D and Σ_i , $i = 1, \dots, m$.

The variance components vector

The convenient and commonly used way to represent these variables in a clear manner is to introduce a vector θ which is meant to consist of all unknown variance components. This vector is called the *variance components vector* of the model.

Hence, using that $D \in \mathbb{R}^{q \times q}$ is symmetric and in case that $\Sigma_i \in \mathbb{R}^{n_i \times n_i}$, $i = 1, \dots, m$, are any positive definite and symmetric covariance matrices, θ consists of the $q(q+1)/2$ elements of D and the $\sum_{i=1}^m n_i(n_i+1)/2$ elements of the individual residual covariance matrices. As mentioned in the following of Postulate 3, Σ_i , $i = 1, \dots, m$, commonly depend on i only through their dimension n_i , their respective elements, however, are independent of i . This is why the dimension of θ usually is smaller than $q(q+1)/2 + \sum_{i=1}^m n_i(n_i+1)/2$.

Assuming the conditional independence structure for the residual covariance matrices, $\Sigma_i = \sigma^2 I_{n_i}$, $i = 1, \dots, m$, θ can be written as

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_r \end{pmatrix} = (d_1^2, d_{12}, \dots, d_{1q}, d_2^2, \dots, d_{2q}, \dots, d_q^2, \sigma^2)^T \quad \text{with} \quad r := \frac{q(q+1)}{2} + 1$$

where $\theta \in (0, \infty) \times \mathbb{R} \times \dots \times \mathbb{R} \times (0, \infty) \times \dots \times \mathbb{R} \times \dots \times (0, \infty) \times (0, \infty)$
such that D is positive definite.

To make the unknown covariance structure clear, we denote the individual covariance matrices by $D(\theta)$ and $\Sigma_i(\theta)$, $i = 1, \dots, m$. Remember that the overall covariance matrices of the random effects vector γ and the residual vector ϵ , G and R respectively, are block diagonal matrices with diagonal composed out of the individual covariance matrices (see Postulate 5 of Chapter 3). In analogous manner we denote them therefore as $G(\theta)$ and $R(\theta)$.

The covariance matrix of the marginal distribution of \mathbf{Y} , $V = ZGZ^T + R$, is a term of both covariance matrices G and R , θ consequently specifies also V and we denote V as function of θ , $V(\theta)$.

The maximum likelihood approach

For estimating θ we present the standard approach of applying maximum likelihood methods, referring to the marginal model (4.2). To highlight the presence of additional unknown model components compared to (5.1) we restate the marginal distribution and the

corresponding density function as

$$\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, V(\boldsymbol{\theta})) \quad \text{and}$$

$$f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}) = (2\pi)^{-\frac{n}{2}} |V(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - X\boldsymbol{\beta})^T V(\boldsymbol{\theta})^{-1} (\mathbf{y} - X\boldsymbol{\beta}) \right\}. \quad (5.20)$$

This yields the likelihood

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}) &= f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}) \\ &= (2\pi)^{-\frac{n}{2}} |V(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - X\boldsymbol{\beta})^T V(\boldsymbol{\theta})^{-1} (\mathbf{y} - X\boldsymbol{\beta}) \right\}. \end{aligned} \quad (5.21)$$

Let B and Θ denote the parameter spaces for the fixed effects $\boldsymbol{\beta}$ and the variance components vector $\boldsymbol{\theta}$, respectively. Then

$B = \mathbb{R}^p$ and $\Theta = \{\boldsymbol{\theta} \mid D(\boldsymbol{\theta}) \text{ and } \Sigma_i(\boldsymbol{\theta}) \ (i = 1, \dots, m) \text{ are symmetric and positive definite}\}.$

If

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{ML} \\ \hat{\boldsymbol{\theta}}_{ML} \end{pmatrix} = \arg \max_{\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\theta} \end{pmatrix} \in B \times \Theta} L(\boldsymbol{\beta}, \boldsymbol{\theta}),$$

then $\hat{\boldsymbol{\beta}}_{ML}$ and $\hat{\boldsymbol{\theta}}_{ML}$ are the joint maximum likelihood estimates for the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Hence we have to find $\hat{\boldsymbol{\beta}}_{ML} \in B$ and $\hat{\boldsymbol{\theta}}_{ML} \in \Theta$ which satisfy

$$L(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\theta}}_{ML}) = \sup_{\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\theta} \end{pmatrix} \in B \times \Theta} L(\boldsymbol{\beta}, \boldsymbol{\theta}).$$

As mentioned in the beginning of the first section of this chapter, one usually considers the log-likelihood:

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \ln(f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta})) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(|V(\boldsymbol{\theta})|) - \frac{1}{2} (\mathbf{y} - X\boldsymbol{\beta})^T V(\boldsymbol{\theta})^{-1} (\mathbf{y} - X\boldsymbol{\beta}). \end{aligned} \quad (5.22)$$

ML estimation for $\boldsymbol{\theta}$ will turn out to be very complex. The existence of $\hat{\boldsymbol{\theta}}_{ML}$ cannot be guaranteed for all situations meaning that the (log-)likelihood may be unbounded with respect to $\boldsymbol{\theta}$. If $L(\boldsymbol{\beta}, \boldsymbol{\theta})$, however, assumes a maximum it is not possible to derive an analytical closed-form expressions for $\hat{\boldsymbol{\theta}}_{ML}$ and numerical techniques have to be applied.

5.4.1. The profile log-likelihood

From Theorem 5.1.1 we know that maximizing $l(\boldsymbol{\beta}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ and fixed $\boldsymbol{\theta}$ provides a unique, closed-form solution for every $\boldsymbol{\theta}$, which we denote now as function of $\boldsymbol{\theta}$:

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \arg \max_{\boldsymbol{\beta} \in B} l(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ &= (X^T V(\boldsymbol{\theta})^{-1} X)^{-1} X^T V(\boldsymbol{\theta})^{-1} \mathbf{y}. \end{aligned} \quad (5.23)$$

Therefore, instead of maximizing the log-likelihood simultaneously for β and θ , we may replace β in $l(\beta, \theta)$ by $\hat{\beta}(\theta)$ and maximize the so obtained function $l(\hat{\beta}(\theta), \theta)$ with respect to θ . This function we call the *profile log-likelihood*, since β was profiled out:

$$\begin{aligned} l_P(\theta) &:= l(\hat{\beta}(\theta), \theta) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(|V(\theta)|) - \frac{1}{2} (\mathbf{y} - X\hat{\beta}(\theta))^T V(\theta)^{-1} (\mathbf{y} - X\hat{\beta}(\theta)) . \end{aligned} \quad (5.24)$$

Then, $\hat{\theta}_{ML}$ is the maximum likelihood estimate of θ , if

$$\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} l_P(\theta) .$$

The MLE for β is obtained from there by inserting $\hat{\theta}_{ML}$ in (5.23). Then, $\hat{\beta}(\hat{\theta}_{ML})$ is commonly called the *empirical maximum likelihood estimate* (see Subsection 5.4.5).

5.4.2. The restricted log-likelihood

Deriving an estimate for θ by the ML method introduced in the previous subsection requires the estimation of β first. This, however, causes a loss of degrees of freedom which we don't incorporate when maximizing the profile log-likelihood. Due to Patterson and Thompson (1971) ML estimates for θ are therefore in general biased downwards. For this reason, the so-called *restricted* or *residual maximum likelihood* (REML) estimation is mostly preferred to estimate the variance components, because it provides less biased estimates. This method was introduced by Patterson and Thompson (1971), a general description was presented by Harville (1977).

The basic idea of the REML method is profiling out β from the likelihood by transforming the response variable \mathbf{Y} , so that the distribution of the transformed variable does not depend on the fixed effects any longer. REML estimates are then provided by maximizing the log-likelihood function of the transformed variable. This procedure allows us to take into account the degrees of freedom for the fixed effects in the model where the ML method introduced above fails.

Let $A \in \mathbb{R}^{n \times (n-p)}$ be any matrix whose columns are linearly independent and orthogonal to the columns of X , i.e. $\text{rank}(A) = n - p$ and $A^T X = 0$. We consider the linear transformation $A^T \mathbf{Y}$ of the response, the so-called *error contrasts*. Its distribution is given by a $(n-p)$ -dimensional normal distribution with mean vector zero, not containing the vector of fixed effects β anymore:

$$A^T \mathbf{Y} \sim N_{n-p}(\underbrace{A^T X \beta}_{= 0}, A^T V(\theta) A) .$$

For estimating the variance components we maximize the corresponding log-likelihood,

$$l_{REML}(\theta) = -\frac{n-p}{2} \ln(2\pi) - \frac{1}{2} \ln(|A^T V(\theta) A|) - \frac{1}{2} \mathbf{y}^T A (A^T V(\theta) A)^{-1} A^T \mathbf{y} , \quad (5.25)$$

the so-called *restricted log-likelihood* (since it is restricted to (n-p) error contrasts) with respect to $\boldsymbol{\theta}$.

Theorem 5.4.1 Harville (1974), Harville (1977)

Given a linear mixed model specified by Postulate 1 to 5 of Chapter 3. Let $A \in \mathbb{R}^{n \times (n-p)}$ with $\text{rank}(A) = n - p$ and $A^T X = 0$.

Then the log-likelihood (5.25) of $A^T \mathbf{Y} \sim N_{n-p}(\mathbf{0}, A^T V(\boldsymbol{\theta}) A)$ is equal to

$$l_{REML}(\boldsymbol{\theta}) = c - \frac{1}{2} \ln |X^T V(\boldsymbol{\theta})^{-1} X| + l_P(\boldsymbol{\theta}), \quad (5.26)$$

where c is denoting a constant independent of $\boldsymbol{\theta}$ and l_P the profile log-likelihood (5.24).

Proof: This result was first shown by Harville (1974) and Harville (1977). We give a detailed proof referring to LaMotte (2007)).

- (1) Since the matrix $(V(\boldsymbol{\theta})^{-1} X, A)$ is regular (Lemma A.3.2) we may consider the following expression by using that $A^T X = 0$:

$$\begin{aligned} (V(\boldsymbol{\theta})^{-1} X, A)^{-1} V(\boldsymbol{\theta})^{-1} \begin{pmatrix} X^T V(\boldsymbol{\theta})^{-1} \\ A^T \end{pmatrix}^{-1} &= \left(\begin{pmatrix} X^T V(\boldsymbol{\theta})^{-1} \\ A^T \end{pmatrix} V(\boldsymbol{\theta}) (V(\boldsymbol{\theta})^{-1} X, A) \right)^{-1} \\ &= \begin{pmatrix} X^T V(\boldsymbol{\theta})^{-1} X & 0 \\ 0 & A^T V(\boldsymbol{\theta}) A \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (X^T V(\boldsymbol{\theta})^{-1} X)^{-1} & 0 \\ 0 & (A^T V(\boldsymbol{\theta}) A)^{-1} \end{pmatrix} \\ \Rightarrow V(\boldsymbol{\theta})^{-1} &= (V(\boldsymbol{\theta})^{-1} X, A) \begin{pmatrix} (X^T V(\boldsymbol{\theta})^{-1} X)^{-1} & 0 \\ 0 & (A^T V(\boldsymbol{\theta}) A)^{-1} \end{pmatrix} \begin{pmatrix} X^T V(\boldsymbol{\theta})^{-1} \\ A^T \end{pmatrix} \\ &= V(\boldsymbol{\theta})^{-1} X (X^T V(\boldsymbol{\theta})^{-1} X)^{-1} X^T V(\boldsymbol{\theta})^{-1} + A (A^T V(\boldsymbol{\theta}) A)^{-1} A^T \\ \Rightarrow A (A^T V(\boldsymbol{\theta}) A)^{-1} A^T &= V(\boldsymbol{\theta})^{-1} - V(\boldsymbol{\theta})^{-1} X (X^T V(\boldsymbol{\theta})^{-1} X)^{-1} X^T V(\boldsymbol{\theta})^{-1}. \quad (5.27) \end{aligned}$$

Hence we may express the quadratic form in (5.25) in the following way:

$$\begin{aligned} &\mathbf{y}^T A (A^T V(\boldsymbol{\theta}) A)^{-1} A^T \mathbf{y} \\ &= \mathbf{y}^T (V(\boldsymbol{\theta})^{-1} - V(\boldsymbol{\theta})^{-1} X (X^T V(\boldsymbol{\theta})^{-1} X)^{-1} X^T V(\boldsymbol{\theta})^{-1}) \mathbf{y} \\ &=: \mathbf{y}^T P(\boldsymbol{\theta}) \mathbf{y} \\ &= (\mathbf{y} - X \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T V(\boldsymbol{\theta})^{-1} (\mathbf{y} - X \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) \quad (\text{see A.1.4 (3)}) \end{aligned}$$

- (2) $(X, A) \in \mathbb{R}^{n \times n}$ is square and regular, i.e. $|(X, A)| \neq 0$ (see Lemma A.3.2 for $V = I$). Using properties of the determinant listed in the Appendix A.1.1, we get

$$|(X, A)^T| |(X, A)| = \left| \begin{pmatrix} X^T \\ A^T \end{pmatrix} (X, A) \right| = \left| \begin{pmatrix} X^T X & 0 \\ 0 & A^T A \end{pmatrix} \right| = |A^T A| |X^T X|$$

$$\begin{aligned}
 |A^T A| |V(\boldsymbol{\theta})| |X^T X| &= |(X, A)^T| |V(\boldsymbol{\theta})| |(X, A)| \\
 &= \left| \begin{pmatrix} X^T \\ A^T \end{pmatrix} V(\boldsymbol{\theta}) (X, A) \right| \\
 &= \left| \begin{pmatrix} X^T V(\boldsymbol{\theta}) X & X^T V(\boldsymbol{\theta}) A \\ A^T V(\boldsymbol{\theta}) X & A^T V(\boldsymbol{\theta}) A \end{pmatrix} \right| \\
 &= |A^T V(\boldsymbol{\theta}) A| |X^T V(\boldsymbol{\theta}) X - X^T V(\boldsymbol{\theta}) A (A^T V(\boldsymbol{\theta}) A)^{-1} A^T V(\boldsymbol{\theta}) X| \\
 &= |A^T V(\boldsymbol{\theta}) A| |X^T \underbrace{(V(\boldsymbol{\theta}) - V(\boldsymbol{\theta}) A (A^T V(\boldsymbol{\theta}) A)^{-1} A^T V(\boldsymbol{\theta}))}_{=X(X^T V(\boldsymbol{\theta})^{-1} X)^{-1} X^T \text{ by (5.27)}} X| \\
 &= |A^T V(\boldsymbol{\theta}) A| |X^T X|^2 |X^T V(\boldsymbol{\theta})^{-1} X|^{-1} \\
 \Rightarrow |A^T V(\boldsymbol{\theta}) A| &= |A^T A| |X^T X|^{-1} |X^T V(\boldsymbol{\theta})^{-1} X| |V(\boldsymbol{\theta})|
 \end{aligned}$$

(3) Therefore, substituting the quadratic form and the determinant in (5.25) provides:

$$\begin{aligned}
 l_{REML}(\boldsymbol{\theta}) &= -\frac{n-p}{2} \ln(2\pi) - \frac{1}{2} \ln(|A^T V(\boldsymbol{\theta}) A|) - \frac{1}{2} \mathbf{y}^T A (A^T V(\boldsymbol{\theta}) A)^{-1} A^T \mathbf{y} \\
 &= \underbrace{\frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln |A^T A| + \frac{1}{2} \ln |X^T X|}_{=c} - \frac{1}{2} \ln |X^T V(\boldsymbol{\theta})^{-1} X| \\
 &\quad - \underbrace{\frac{1}{2} \ln |V(\boldsymbol{\theta})| - \frac{n}{2} \ln(2\pi) - \frac{1}{2} (\mathbf{y} - X \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T V(\boldsymbol{\theta})^{-1} (\mathbf{y} - X \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))}_{l_P(\boldsymbol{\theta})}
 \end{aligned}$$

which shows the theorem. ■

Hence, the maximum of the restricted log-likelihood with respect to the variance components vector $\boldsymbol{\theta}$ is independent of the choice of A and the estimates for $\boldsymbol{\theta}$ may be obtained by maximizing (5.26). Clearly, the constant c is negligible. Then, $\hat{\boldsymbol{\theta}}_{REML}$ is the so-called *restricted maximum likelihood estimate* of $\boldsymbol{\theta}$, if

$$\begin{aligned}
 \hat{\boldsymbol{\theta}}_{REML} &= \arg \max_{\boldsymbol{\theta} \in \Theta} l_{REML}(\boldsymbol{\theta}) \\
 &= \arg \max_{\boldsymbol{\theta} \in \Theta} \left(-\frac{1}{2} \ln |X^T V(\boldsymbol{\theta})^{-1} X| + l_P(\boldsymbol{\theta}) \right).
 \end{aligned}$$

As already has been mentioned, this procedure yields less biased estimates than maximizing the profile log-likelihood (5.24).

5.4.3. The marginal log-likelihood

Another method to eliminate β from the likelihood $L(\beta, \theta)$ so that it is defined only in terms of θ is to integrate it with respect to β . In this way we get a function depending only on θ , which we call the corresponding *marginal likelihood*, L_{MAR} . The logarithm of this marginal likelihood function is called the *marginal log-likelihood*

$$l_{MAR}(\theta) := \ln(L_{MAR})(\theta) = \ln\left(\int L(\beta, \theta) d\beta\right). \quad (5.28)$$

Theorem 5.4.2 Fahrmeir et al. (2007, Subsection 6.3.2)

Given a linear mixed model specified by Postulate 1 to 5 of Chapter 3. Let l_P denote the profile log-likelihood (5.24). Then the marginal log-likelihood (5.28) is given as

$$l_{MAR}(\theta) = c - \frac{1}{2} \ln |X^T V(\theta) X| + l_P(\theta),$$

where c is denoting a constant independent of θ .

Proof:

$$\int L(\beta, \theta) d\beta = (2\pi)^{-\frac{n}{2}} |V(\theta)|^{-\frac{1}{2}} \int \exp\left\{-\frac{1}{2} (\mathbf{y} - X\beta)^T V(\theta)^{-1} (\mathbf{y} - X\beta)\right\} d\beta$$

For calculating the integral we reformulate the quadratic form in the exponent such that we get the density of a multivariate normal distribution in β . Defining $K(\theta) := X^T V(\theta)^{-1} X$ and $\mu(\theta) := K(\theta)^{-1} X^T V(\theta)^{-1} \mathbf{y}$, and using that $\mu(\theta)^T K(\theta) = \mathbf{y}^T V(\theta)^{-1} X$, we get

$$\begin{aligned} & (\mathbf{y} - X\beta)^T V(\theta)^{-1} (\mathbf{y} - X\beta) \\ &= \mathbf{y}^T V(\theta)^{-1} \mathbf{y} - 2\mathbf{y}^T V(\theta)^{-1} X\beta + \beta^T K(\theta) \beta \\ &= [\beta^T K(\theta) \beta - 2\mu(\theta)^T K(\theta) \beta + \mu(\theta)^T K(\theta) \mu(\theta)] - \mu(\theta)^T K(\theta) \mu(\theta) + \mathbf{y}^T V(\theta)^{-1} \mathbf{y} \\ &= (\beta - \mu(\theta))^T K(\theta) (\beta - \mu(\theta)) - \mu(\theta)^T K(\theta) \mu(\theta) + \mathbf{y}^T V(\theta)^{-1} \mathbf{y}. \end{aligned}$$

Replacing $K(\theta)$ and $\mu(\theta)$ in the last term, using A.1.4, yields

$$\begin{aligned} & - \mu(\theta)^T K(\theta) \mu(\theta) + \mathbf{y}^T V(\theta)^{-1} \mathbf{y} \\ &= \mathbf{y}^T V(\theta)^{-1} X (X^T V(\theta)^{-1} X)^{-1} X^T V(\theta)^{-1} \mathbf{y} + \mathbf{y}^T V(\theta)^{-1} \mathbf{y} \\ &= \mathbf{y}^T P(\theta) \mathbf{y} \\ &= (\mathbf{y} - X\hat{\beta}(\theta))^T V(\theta)^{-1} (\mathbf{y} - X\hat{\beta}(\theta)). \end{aligned}$$

Hence, using that $K(\theta)$ is symmetric and positive definite (see A.3.1),

$$\begin{aligned} \int L(\beta, \theta) d\beta &= \underbrace{(2\pi)^{-\frac{n}{2}} \cdot |V(\theta)|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2} (\mathbf{y} - X\hat{\beta}(\theta))^T V(\theta)^{-1} (\mathbf{y} - X\hat{\beta}(\theta))\right\}}_{L(\hat{\beta}(\theta), \theta)} \\ &\quad \cdot \underbrace{\int \exp\left\{-\frac{1}{2} (\beta - \mu(\theta))^T K(\theta) (\beta - \mu)\right\} d\beta}_{=(2\pi)^{\frac{p}{2}} |K(\theta)^{-1}|^{\frac{1}{2}}} \end{aligned}$$

$$\Rightarrow l_{MAR}(\boldsymbol{\theta}) = \ln(L_{MAR}(\boldsymbol{\theta})) = l_P(\boldsymbol{\theta}) + \frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln |X^T V(\boldsymbol{\theta})^{-1} X|, \quad (5.29)$$

which proves the proposition. \blacksquare

Note that l_{MAR} and l_{REML} differ only in a constant summand, the part depending on $\boldsymbol{\theta}$ is equal and given by

$$- \frac{1}{2} \ln |X^T V(\boldsymbol{\theta}) X| + l_P(\boldsymbol{\theta}). \quad (5.30)$$

Consequently, provided existence, l_{MAR} and l_{REML} have the same maximum. Hence, if existence is guaranteed, the REML estimate of $\boldsymbol{\theta}$ satisfies

$$\hat{\boldsymbol{\theta}}_{REML} = \arg \max_{\boldsymbol{\theta} \in \Theta} l_{MAR}(\boldsymbol{\theta}).$$

5.4.4. Computation of the estimates

To recapitulate, deriving an estimator for the variance components vector $\boldsymbol{\theta}$ by applying the maximum likelihood approach implies maximizing l_P or (5.30) out of l_{REML} and l_{MAR} . This provides, in case of existence, the ML estimate $\hat{\boldsymbol{\theta}}_{ML}$ and the REML estimate $\hat{\boldsymbol{\theta}}_{REML}$, respectively.

$$\begin{array}{ccc} & \boldsymbol{\theta} & \\ \swarrow & & \searrow \\ \hat{\boldsymbol{\theta}}_{ML} & & \hat{\boldsymbol{\theta}}_{REML} \\ = \arg \max_{\boldsymbol{\theta} \in \Theta} l_P(\boldsymbol{\theta}) & & = \arg \max_{\boldsymbol{\theta} \in \Theta} \left(-\frac{1}{2} \ln |X^T V(\boldsymbol{\theta})^{-1} X| + l_P(\boldsymbol{\theta}) \right) \\ \text{maximizing } l_P(\boldsymbol{\theta}) \text{ (5.24)} & & \text{maximizing } l_{REML} \text{ (5.26) or } l_{MAR} \text{ (5.29)} \end{array}$$

In order to concretely determine the respective estimate, one has to apply numerical techniques. Analytical, closed-form expressions can in general not be derived. Maximization has to be performed within the parameter space Θ , i.e. we have a constrained maximization problem which makes the estimation of $\boldsymbol{\theta}$ computationally very complex and intensive. Basically two different iterative computation schemes established for that purpose, the *EM-algorithm* and the *Newton-Raphson method*.

Searle et al. (1992, Chapter 8) give a detailed presentation of these methods with its advantages and difficulties. Laird and Ware (1982) formulated the EM algorithm as well which was improved later with respect to convergence properties by Lindstrom and Bates (1988). The latter ones developed also an more efficient Newton-Raphson algorithm.

5.4.5. Empirical estimates for the fixed and random effects

In analogous manner as $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ in (5.23), $\hat{\boldsymbol{\gamma}}$ can be considered as function of $\boldsymbol{\theta}$. Plugging in $\hat{\boldsymbol{\theta}}$ ($\hat{\boldsymbol{\theta}}_{ML}$ or $\hat{\boldsymbol{\theta}}_{REML}$) in $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ provides an empirical estimate for $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}) = (X^T V(\hat{\boldsymbol{\theta}})^{-1} X)^{-1} X^T V(\hat{\boldsymbol{\theta}})^{-1} \mathbf{y} \quad (5.31)$$

and, accordingly, an empirical predictor for γ ,

$$\hat{\gamma}(\hat{\theta}) = G(\hat{\theta})Z^TV(\hat{\theta})^{-1}(\mathbf{y} - X\hat{\beta}(\hat{\theta})) . \quad (5.32)$$

$\hat{\gamma}(\hat{\theta})$ commonly is referred to as the *empirical best linear unbiased predictor* (EBLUP) of γ . Note that using the ML or REML estimate of θ in $\hat{\beta}(\hat{\theta})$ and $\hat{\gamma}(\hat{\theta})$ will yield different values.

The exact form of the variances of $\hat{\beta}(\hat{\theta})$ and $\hat{\gamma}(\hat{\theta})$ cannot be obtained analytically (see Fahrmeir et al. (2007)). The derived variances of $\hat{\beta}$ and $\hat{\gamma}$ in Subsections 5.1.3 and 5.3.1 where the covariance matrices are replaced by their estimates $V(\hat{\theta})$ and $G(\hat{\theta})$ can only be considered as rough approximations. That is, $\text{Var}(\hat{\beta})$ and $\text{Var}(\hat{\gamma})$ may be approximated, respectively, by

$$\begin{aligned} \hat{\text{Var}}(\hat{\beta}) &= (X^TV(\hat{\theta})^{-1}X)^{-1} \quad \text{and} \\ \hat{\text{Var}}(\hat{\gamma}) &= G(\hat{\theta})Z^T(V(\hat{\theta})^{-1} - V(\hat{\theta})^{-1}X(X^TV(\hat{\theta})^{-1}X)^{-1}X^TV(\hat{\theta})^{-1})ZG(\hat{\theta}) . \end{aligned}$$

Due to Dempster et al. (1981), these estimates underestimate the true variances $\text{Var}(\hat{\beta})$ and $\text{Var}(\hat{\gamma})$ since they do not take into account the variability of $\hat{\theta}$.

From now on, for simplicity, when there is no possibility of confusion, we denote (5.31) and (5.32) by $\hat{\beta}$ and $\hat{\gamma}$, respectively. In analogous manner, \hat{D} , $\hat{\Sigma}_i$, \hat{G} , \hat{R} and \hat{V} are meant to represent $D(\hat{\theta})$, $\Sigma_i(\hat{\theta})$, $G(\hat{\theta})$, $R(\hat{\theta})$ and $V(\hat{\theta})$, respectively.

6. Hypothesis tests in linear mixed models

After fitting a linear mixed model to a given data set one usually is interested if the model is correctly specified. This means basically to consider if additional or less model parameters should be included for modelling the response appropriately. Generally spoken we want to check if some of the included model parameters underly a certain restriction. For that purpose we introduce in this chapter the commonly used and convenient hypothesis tests. However, until now, there are besides some special cases only asymptotic results available. As has been briefly mentioned already in Chapter 1, theory of hypothesis tests in linear mixed models is fairly complex and still in a developing process.

Testing whether covariates have a significant random effect on the response is not as straightforward as testing fixed effects and has to be performed in an indirect manner: Since we assume the random effect γ to be a random variable, we apply tests to the components of its covariance matrix rather than to γ itself. In particular, to test the need of a certain random effect in the model we have to consider in the null hypothesis if its variance and the corresponding covariances are equal to zero. For verifying these hypotheses, there is still relatively little known.

Hypothesis formulation and nested models

Usually we are interested in testing hypotheses concerning only a subset of the model parameters. One standard example is to investigate if certain incorporated fixed or random effects are significant or if some of them might be eliminated from the model. Let ϑ denote the model parameter (in our context β or θ) and $\Lambda \subset \mathbb{R}^r$ its corresponding parameter space. Suppose that $\vartheta = (\vartheta^{(1)}, \vartheta^{(2)})^T$ with $\vartheta^{(1)} = (\vartheta_1, \dots, \vartheta_q)^T$ and $\vartheta^{(2)} = (\vartheta_{q+1}, \dots, \vartheta_r)^T$, $q \in \{0, \dots, r-1\}$, and we are interested if $\vartheta^{(2)}$ underlies a certain restriction. Thus we formulate the null hypothesis as

$$H_0 : \vartheta^{(2)} = \vartheta_0^{(2)}$$

with a certain specified vector (or value) $\vartheta_0^{(2)}$ determining the condition to be tested. However, if one wants to test a statement about the entire parameter vector ϑ , then we set $\vartheta^{(2)} = \vartheta$ ($q = 0$) and $\vartheta^{(1)}$ to be null.

This specification may be also interpreted as comparing two so-called *nested* models. One model is said to be nested within another model, called the full or reference model, if its parameter space is a subspace of that for the full model. Let us consider the nested

model which is obtained by imposing the constraint $\boldsymbol{\vartheta}^{(2)} = \boldsymbol{\vartheta}_0^{(2)}$ on the parameter $\boldsymbol{\vartheta}$ of the full model. Then

$$\Lambda_{\text{nested}} = \{ \boldsymbol{\vartheta} \in \Lambda \mid \boldsymbol{\vartheta}^{(2)} = \boldsymbol{\vartheta}_0^{(2)} \} \subset \Lambda$$

is the parameter space of the nested model. Note that the values in $\boldsymbol{\vartheta}^{(2)}$ are not considered as (unknown) model parameters in the nested model, because we know them to be $\boldsymbol{\vartheta}_0^{(2)}$. Hence the number of (free) model parameters in the nested model, which is commonly called the *dimension* of the model, is q .

Consequently, we may perceive $H_0 : \boldsymbol{\vartheta}^{(2)} = \boldsymbol{\vartheta}_0^{(2)}$ as the hypothesis whether the nested model is true,

$$H_0 : \boldsymbol{\vartheta} \in \Lambda_0 := \Lambda_{\text{nested}} \subset \Lambda .$$

We test this hypothesis against the alternative that the full model holds and is not of the specific form restricted by the nested model,

$$H_1 : \boldsymbol{\vartheta} \in \Lambda_1 := \Lambda \setminus \Lambda_0 .$$

6.1. Likelihood ratio tests

The likelihood ratio test is a standard test procedure in parametric statistical models. We want to give here first an overview about the general concept before considering the case of likelihood ratio tests about the model parameters in a linear mixed model.

Let $\mathbf{y} = (y_1, \dots, y_n)^T$ be observed data of a statistical study assumed to be realizations of independent and identical distributed random variables Y_1, \dots, Y_n . Let these random variables follow a distribution which depends on a unknown model parameter $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_r)^T$ lying in a certain parameter space $\Lambda \subset \mathbb{R}^r$. The corresponding likelihood function is denoted by $L(\boldsymbol{\vartheta})$, the corresponding log-likelihood by $l(\boldsymbol{\vartheta})$.

For testing the null hypothesis $H_0 : \boldsymbol{\vartheta} \in \Lambda_0 \subset \Lambda$ against the alternative hypothesis $H_1 : \boldsymbol{\vartheta} \in \Lambda_1 = \Lambda \setminus \Lambda_0$, one may consider the so-called *likelihood ratio* which compares the values of the likelihood functions corresponding to the respective hypothesis:

Definition 6.1.1

For testing $H_0 : \boldsymbol{\vartheta} \in \Lambda_0 \subset \Lambda$ against $H_1 : \boldsymbol{\vartheta} \in \Lambda_1 = \Lambda \setminus \Lambda_0$,

$$\mathcal{L}(\mathbf{y}) := \frac{\sup_{\boldsymbol{\vartheta} \in \Lambda_0} L(\boldsymbol{\vartheta})}{\sup_{\boldsymbol{\vartheta} \in \Lambda_1} L(\boldsymbol{\vartheta})}$$

is called the **likelihood ratio**.

In case the underlying maximum likelihood estimates restricted to the null hypothesis, $\hat{\boldsymbol{\vartheta}}_{ML}^{\Lambda_0}$, and the alternative, $\hat{\boldsymbol{\vartheta}}_{ML}^{\Lambda_1}$, do exist, $\mathcal{L}(\mathbf{y})$ obviously turns out to be

$$\mathcal{L}(\mathbf{y}) = \frac{L(\hat{\boldsymbol{\vartheta}}_{ML}^{\Lambda_0})}{L(\hat{\boldsymbol{\vartheta}}_{ML}^{\Lambda_1})} .$$

To determine a critical value at which the null hypothesis is rejected we need to know the (at least asymptotic) distribution of $\mathcal{L}(\mathbf{y})$ which, however, might be hard to find out.

If the null hypothesis represents a "special case" of the alternative, then results about the asymptotic behavior of the likelihood ratio are available. This case implies comparing nested models for which Wilks (1938) showed the following theorem. It is based upon the well known result, that under certain regularity conditions, maximum likelihood estimates are asymptotically normal distributed (Theorem A.2.1).

Theorem 6.1.1 Wilks (1938)

Let

$$\boldsymbol{\vartheta} = \begin{pmatrix} \boldsymbol{\vartheta}^{(1)} \\ \boldsymbol{\vartheta}^{(2)} \end{pmatrix} \in \Lambda \subset \mathbb{R}^r, \Lambda \text{ open}, \boldsymbol{\vartheta}^{(1)} = \begin{pmatrix} \vartheta_1 \\ \vdots \\ \vartheta_q \end{pmatrix} \text{ and } \boldsymbol{\vartheta}^{(2)} = \begin{pmatrix} \vartheta_{q+1} \\ \vdots \\ \vartheta_r \end{pmatrix}$$

and the null hypothesis be given by $H_0 : \boldsymbol{\vartheta}^{(2)} = \boldsymbol{\vartheta}_0^{(2)}$ with a certain specified vector $\boldsymbol{\vartheta}_0^{(2)}$. Suppose that the regularity conditions (1)-(6) stated in Theorem A.2.1 for the full model are satisfied and let (6) be satisfied for the nested model.

Let $\hat{\boldsymbol{\vartheta}}_{ML}$ be the MLE of $\boldsymbol{\vartheta}$ and $\hat{\boldsymbol{\vartheta}}_{ML}^{\Lambda_0}$ be the MLE of $\boldsymbol{\vartheta}$ under H_0 .

Then, under H_0 :

$$-2 \ln \mathcal{L}(\mathbf{y}) = -2(l(\hat{\boldsymbol{\vartheta}}_{ML}^{\Lambda_0}) - l(\hat{\boldsymbol{\vartheta}}_{ML})) \xrightarrow{d} \chi_{r-q}^2. \quad (6.1)$$

For the proof of this theorem we refer to Wilks (1938) and Bickel and Doksum (2001, Theorem 6.3.2.).

Definition 6.1.2

Using the notations of Theorem 6.1.1,

$$\text{LRT} := -2 \ln \mathcal{L}(\mathbf{y})$$

is called the **likelihood ratio test (LRT) statistic**.

Hence, H_0 may be rejected at level α if

$$\text{LRT} > \chi_{r-q, 1-\alpha}^2,$$

given that the number of observations is large enough to assume the asymptotic distribution to be accurate.

Theorem 6.1, known as *Wilk's Theorem*, provides asymptotic likelihood ratio tests for the fixed effects and the covariance parameters in linear mixed models. These tests are

presented in the Subsections 6.1.1 and 6.1.2.

6.1.1. Likelihood ratio test for the fixed effects

Applying Wilk's Theorem to our context, we may derive an (asymptotic) likelihood ratio test whether the vector of fixed effects, $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)})^T \in B = \mathbb{R}^p$, underlies a certain restriction. In practice, this problem is of primarily interest. We may formulate the null hypothesis as

$$H_0 : \boldsymbol{\beta}^{(2)} = \boldsymbol{\beta}_0^{(2)} ,$$

$$\text{or equivalently, } H_0 : \boldsymbol{\beta} \in B_0 = \{ \boldsymbol{\beta} \in B \mid \boldsymbol{\beta}^{(2)} = \boldsymbol{\beta}_0^{(2)} \} \subset B .$$

To find out for instance if certain explanatory variables have a significant influence on the response \mathbf{Y} or if some of them might be eliminated from the model, we test if their corresponding fixed effect in $\boldsymbol{\beta}$ is equal to zero (set $\boldsymbol{\beta}_0^{(2)} = \mathbf{0}$). As explained in the introductory part of this chapter, we may interpret this test as comparing nested models fitted to the same data,

$$(F) : \mathbf{Y} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{\epsilon} \quad \text{and} \quad (N) : \mathbf{Y} = X \begin{pmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}_0^{(2)} \end{pmatrix} + Z\boldsymbol{\gamma} + \boldsymbol{\epsilon} ,$$

where (F) represents the full, and (N) the nested model, and we test whether (N) is valid or (F) should be preferred. The parameter space B_0 of $\boldsymbol{\beta}$ in the nested model is a subspace of the parameter space B of the full model. Thus, the two models differ in the specification of their fixed effects but have identical covariance structure (particularly the same set of random effects).

Let $l(\boldsymbol{\beta}, \boldsymbol{\theta})$ be the log-likelihood (5.22) of the marginal full model, $\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, V(\boldsymbol{\theta}))$, which was derived as

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(|V(\boldsymbol{\theta})|) - \frac{1}{2} (\mathbf{y} - X\boldsymbol{\beta})^T V(\boldsymbol{\theta})^{-1} (\mathbf{y} - X\boldsymbol{\beta}) .$$

Let

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\beta} \in B} l(\boldsymbol{\beta}, \boldsymbol{\theta}) \quad \text{and} \quad \hat{\boldsymbol{\beta}}^{B_0}(\boldsymbol{\theta}) := \arg \max_{\boldsymbol{\beta} \in B_0} l(\boldsymbol{\beta}, \boldsymbol{\theta})$$

denote the MLEs of $\boldsymbol{\beta}$ in the full and the nested model, respectively. Further, let

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{ML} &= \arg \max_{\boldsymbol{\theta} \in \Theta} l_P(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\theta} \in \Theta} l(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) \quad \text{and} \\ \hat{\boldsymbol{\theta}}_{ML}^0 &= \arg \max_{\boldsymbol{\theta} \in \Theta} l_P^0(\boldsymbol{\theta}) := \arg \max_{\boldsymbol{\theta} \in \Theta} l(\hat{\boldsymbol{\beta}}^{B_0}(\boldsymbol{\theta}), \boldsymbol{\theta}) \end{aligned}$$

denote the MLEs of $\boldsymbol{\theta}$ in the full and the nested model, respectively, obtained by maximizing the corresponding profile log-likelihood. Then, $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_{ML})$ and $\hat{\boldsymbol{\beta}}^{B_0}(\hat{\boldsymbol{\theta}}_{ML}^0)$ denote the empirical estimates of $\boldsymbol{\beta}$ in the respective model. Thus the pair $(\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_{ML}), \hat{\boldsymbol{\theta}}_{ML})$ is obtained from maximizing (5.22) over the parameter space $B \times \Theta$ and $(\hat{\boldsymbol{\beta}}^{B_0}(\hat{\boldsymbol{\theta}}_{ML}^0), \hat{\boldsymbol{\theta}}_{ML}^0)$ is the maximum of (5.22) with respect to the parameter space $B_0 \times \Theta$.

Hence, the LRT statistic is given by

$$\text{LRT} = -2 \ln \mathcal{L}(\mathbf{y}) = -2(l(\hat{\boldsymbol{\beta}}^{B_0}(\hat{\boldsymbol{\theta}}_{ML}^0), \hat{\boldsymbol{\theta}}_{ML}^0) - l(\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_{ML}), \hat{\boldsymbol{\theta}}_{ML})) .$$

Then, according to Theorem 6.1.1, provided that the regularity conditions are satisfied,

$$\text{LRT} \xrightarrow{d} \chi_{df}^2 \quad \text{under } H_0,$$

where df is the difference between the number of model parameters in the full and the nested model.

Hence, an asymptotic test is to reject H_0 at level α , if

$$\text{LRT} > \chi_{df, 1-\alpha}^2 ,$$

where $\chi_{df, 1-\alpha}^2$ is the $(1 - \alpha)$ -quantile of the χ_{df}^2 distribution.

Pinheiro and Bates (2000, Subsection 2.4.2) do not recommend to use this test. They showed in their book that the sample often is too small that the test statistic follows approximately the stated asymptotic distribution. By simulating various data sets from the nested model and calculating the corresponding likelihood ratio test statistics they found out that the tests often tend to be "anticonservative", meaning that the empirical (the "true") p-values are greater than the theoretical (the "assumed") p-values of the asymptotic chi-square distribution. Hence H_0 would be rejected too often at trusting on the theoretical p-values. Therefore they suggest to use approximate t- or F-tests for the fixed effects rather than likelihood ratio tests (see Section 6.2).

6.1.2. Likelihood ratio tests for the covariance parameters

Another typical problem of interest is testing whether the model's covariance parameters underly certain restrictions. This applies to the elements d_{ij} , $i \neq j$, of the covariance matrix D of the random effects or covariance parameters in the residual covariance matrices Σ_i , $i = 1, \dots, m$. A standard example is testing whether the random effects within an individual are uncorrelated, i.e. whether $d_{ij} = 0$, $i < j$, $i = 1, \dots, q - 1$. In this case one is interested to check if a variance components structure for D is valid or if covariance between the random effects should be included.

Remember that the variance and covariance parameters of the model are stored in the variance components vector $\boldsymbol{\theta}$. Hence, tests concerning the elements of the covariance matrices D and Σ_i are equivalent to tests about the corresponding elements of $\boldsymbol{\theta}$. For example, consider the following linear mixed model:

For $i = 1, \dots, m$:

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + Z_i \begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} + \boldsymbol{\epsilon}_i, \quad \text{with} \quad \begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N_2 \left(\mathbf{0}, \begin{pmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{pmatrix} \right) \quad \text{and} \quad \boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma^2 I_{n_i}) .$$

Then the variance components vector is given by

$$\boldsymbol{\theta} = (d_1^2, d_{12}, d_2^2, \sigma^2) .$$

Hence, the null hypothesis $H_0 : d_{12} = 0$ for testing whether γ_{i0} and γ_{i1} are uncorrelated is the same as $H_0 : \theta_2 = 0$.

In an analogous manner as for the fixed effects and based on Theorem 6.1.1, we may derive an asymptotic likelihood ratio test for the model's covariance parameters. Restructuring the variance components vector $\boldsymbol{\theta}$ appropriately as $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})^T$, such that $\boldsymbol{\theta}^{(2)}$ contains the subset of covariance parameters to be tested, we formulate the null hypothesis as $H_0 : \boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}_0^{(2)}$. As before, this may be considered as comparing nested models with the same set of fixed and random effects but different covariance structure. The parameter space of $\boldsymbol{\theta}$ in the full model is defined as

$$\Theta = \{ \boldsymbol{\theta} \mid D(\boldsymbol{\theta}) \text{ and } \Sigma_i(\boldsymbol{\theta}) \ (i = 1, \dots, m) \text{ are symmetric and positive definite} \} ,$$

and the parameter space for $\boldsymbol{\theta}$ in the nested model can be described as

$$\Theta_0 = \{ \boldsymbol{\theta} \mid \boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}_0^{(2)} \text{ and } D(\boldsymbol{\theta}) \text{ and } \Sigma_i(\boldsymbol{\theta}) \ (i = 1, \dots, m) \text{ are symmetric and positive definite} \} .$$

Let $l(\boldsymbol{\theta})$ denote the profile log-likelihood (5.24),

$$l(\boldsymbol{\theta}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(|V(\boldsymbol{\theta})|) - \frac{1}{2} (\mathbf{y} - X\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T V(\boldsymbol{\theta})^{-1} (\mathbf{y} - X\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) ,$$

and let the MLEs $\hat{\boldsymbol{\theta}}_{ML}$ and $\hat{\boldsymbol{\theta}}_{ML}^{\Theta_0}$ of $\boldsymbol{\theta}$ for the full and the nested model exist. That is, $\hat{\boldsymbol{\theta}}_{ML}$ and $\hat{\boldsymbol{\theta}}_{ML}^{\Theta_0}$ are obtained by maximizing $l(\boldsymbol{\theta})$ over the parameter spaces Θ and Θ_0 , respectively. Then the LRT statistic is given by

$$\text{LRT} = -2 (l(\hat{\boldsymbol{\theta}}_{ML}^{\Theta_0}) - l(\hat{\boldsymbol{\theta}}_{ML})) . \quad (6.2)$$

According to Theorem 6.1.1, it holds under H_0 and the regularity conditions stated in Theorem A.2.1:

$$\text{LRT} \xrightarrow{d} \chi_{df}^2 ,$$

where df is the difference between the dimensions of the full and the nested model.

Hence, if the sample size is large, H_0 can be rejected at a level α if

$$\text{LRT} > \chi_{df, 1-\alpha}^2 .$$

According to Verbeke and Molenberghs (2009, Subsection 6.3.2), this test may also be based on the restricted log-likelihood $l_{REML}(\boldsymbol{\theta})$ (5.26). Then, instead of the MLEs, the REML estimates $\hat{\boldsymbol{\theta}}_{REML}$ and $\hat{\boldsymbol{\theta}}_{REML}^{\Theta_0}$ of $\boldsymbol{\theta}$ have to be used, obtained by maximizing $l_{REML}(\boldsymbol{\theta})$ over Θ and Θ_0 , respectively. In this case, the test statistic

$$\boxed{\text{RLRT} := -2 \ln \mathcal{L}_{REML}(\mathbf{y}) := -2 (l_{REML}(\hat{\boldsymbol{\theta}}_{REML}^{\Theta_0}) - l_{REML}(\hat{\boldsymbol{\theta}}_{REML})) } \quad (6.3)$$

is called *restricted likelihood ratio test (RLRT) statistic*. Then, under H_0 , the asymptotic result of Theorem 6.1.1 is still true and the RLRT statistic satisfies

$$\text{RLRT} \xrightarrow{d} \chi_{df}^2 .$$

In fact, it is recommended to apply this asymptotic likelihood ratio test based on REML estimation rather than the test based on ML estimation. This is due to the fact that REML estimates are less biased than ML estimates, as explained in the introductory part of Subsection 5.4.2.

6.1.3. Likelihood ratio tests for the random effects

As mentioned in the beginning of this chapter, for testing the significance of a certain random effect i we have to check if its corresponding variance d_i^2 in D is equal to zero. This implicates that also all corresponding covariance parameters d_{i1}, \dots, d_{iq} are equal to zero. However, under the null hypothesis of the form

$$H_0 : d_i^2 = 0, d_{ij} = 0 \ (j = 1, \dots, q) , \quad (6.4)$$

D would no longer be positive definite. Therefore, remembering that the elements of D are specified in the variance components vector $\boldsymbol{\theta}$, we redefine the parameter space Θ of $\boldsymbol{\theta}$ as

$$\Theta = \{ \boldsymbol{\theta} \mid D(\boldsymbol{\theta}) \text{ is symmetric and positive semidefinite,} \\ \Sigma_i(\boldsymbol{\theta}) \ (i = 1, \dots, m) \text{ are symmetric and positive definite} \} .$$

Similarly as in Section 5.4, when assuming $\Sigma_i = \sigma^2 I_{n_i}, i = 1, \dots, m$, $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta} = (d_1^2, d_{12}, \dots, d_{1q}, d_2^2, \dots, d_{2q}, \dots, d_q^2, \sigma^2)^T ,$$

and Θ can be formulated as

$$\Theta = [0, \infty) \times \mathbb{R} \times \dots \times \mathbb{R} \times [0, \infty) \times \dots \times \mathbb{R} \times \dots \times [0, \infty) \times [0, \infty) \\ \text{such that } D \text{ is positive definite} .$$

Note that the parameter space for $\boldsymbol{\theta}$ introduced in Section 5.4 and considered in the previous subsection was expanded in allowing the variances to be zero and therefore D to be semidefinite.

It is crucial to note that Θ no longer is open and that hypothesis of the form (6.4) place $\boldsymbol{\theta}$ on the boundary of the parameter space Θ . Therefore the profile log-likelihood (5.24) is not differentiable in $\boldsymbol{\theta}$ when H_0 is true, which is, however, an essential regularity assumption of Theorem A.2.1. Hence Theorem 6.1.1 can not be applied for performing a likelihood ratio test of hypothesis of the form (6.4).

For testing whether a random effect should be kept in a model or not we may apply the following result.

Theorem 6.1.2 Stram and Lee (1994)

Given a linear mixed model specified by Postulate 1 to 5 of Chapter 3 with residual covariance matrices $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$, $i = 1, \dots, m$, $\sigma^2 > 0$.

Then, under the null hypothesis

$$H_0 : D = \begin{pmatrix} D_1 & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \in \mathbb{R}^{q \times q}, \quad D_1 \in \mathbb{R}^{(q-1) \times (q-1)} \text{ positive definite},$$

the LRT statistic (6.2) satisfies

$$\text{LRT} \xrightarrow{d} 0.5\chi_q^2 + 0.5\chi_{q-1}^2. \quad (6.5)$$

This theorem was shown by Stram and Lee (1994) who used results of Self and Liang (1987) about the asymptotic distribution of MLEs and likelihood ratio test statistics in nonstandard testing situations. For another detailed proof we refer to Scheipl (2007, Section 4.3).

Giampaoli and Singer (2009) argue that results of Self and Liang (1987) are only applicable if the individual response variables $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ may be considered as independent and identically distributed, implying that the individual design matrices satisfy $X_i = X$ and $Z_i = Z$, for $i = 1, \dots, n$. Referring to Giampaoli and Singer (2009), this assumption does not hold for longitudinal studies with unbalanced data. However they show, by using results of Vu and Zhou (1997), that Theorem 6.1.2 is satisfied for the case $q \leq 2$.

Since we consider in this thesis (Part II) only linear mixed models which incorporate maximal 2 random effects, the result of this Theorem should be valid.

Note that the null hypothesis in Theorem 6.1.2 can be formulated as $H_0 : d_q^2 = 0$, $d_{qj} = 0$ ($j = 1, \dots, q-1$), and is therefore equivalent to (6.4).

Morrell (1998) showed that the result of the theorem is still true when, instead of the LRT statistic, the RLRT statistic (6.3) is used. In fact, he recommends to apply the RLRT statistic rather than the LRT statistic, according to an extensive Monte Carlo study where the first one performed slightly better.

In practice, boundary problems are often ignored and for testing the need of a certain random effect, the quantiles of the asymptotic distribution in the case without boundary problems are used. Hence, the likelihood ratio test statistic often is assumed to follow asymptotically a χ_q^2 distribution instead of a 50 : 50 mixture of a χ_q^2 and a χ_{q-1}^2 distribution. As illustrated in Figure 6.1 for some values of q , from a certain point the χ_q^2 -density is greater than the latter one and p-values consequently tend to be overestimated. H_0 therefore would be accepted too often which results in a too simple covariance structure of the model. Unfortunately this is the case in many statistical softwares, in R, for instance, the `anova` function (see Subsection 8.3.1). One should keep this problem in mind when testing the significance of the random effects in the model and, where appropriate, prefer manual calculations to software procedures.

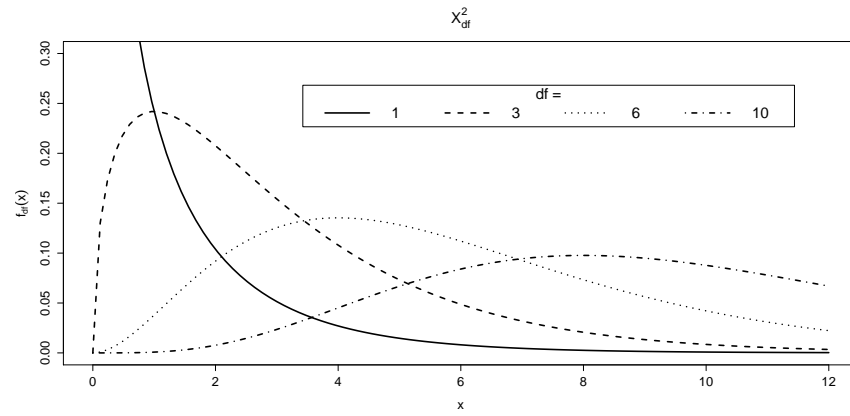


Figure 6.1.: Densities of χ^2 distributions with different degrees of freedom

6.1.4. Exact likelihood ratio tests for simple linear mixed models

Crainiceanu and Ruppert (2004) found out that the $0.5\mathcal{X}_q^2 + 0.5\mathcal{X}_{q-1}^2$ asymptotic approximations, presented in the previous subsection, can be very poor in some cases. They consider a linear mixed model with only one variance component (for random effects) and use a more general formulation as introduced in Postulate 5:

$$\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{b} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma_\epsilon^2 I_n), \quad \mathbf{b} \sim N_K(\mathbf{0}, \sigma_b^2 \Sigma), \quad \boldsymbol{\epsilon} \text{ and } \mathbf{b} \text{ independent} \quad (6.6)$$

where \mathbf{b} is a K -dimensional vector of random effects and Σ is a *known* symmetric and positive definite matrix. As before, \mathbf{Y} denotes the n -dimensional response vector, $\boldsymbol{\beta}$ a p -dimensional vector of fixed effects and X the corresponding $(n \times p)$ -dimensional design matrix. Note that in this specification it is not explicitly assumed that \mathbf{Y} can be partitioned into independent and identically distributed subvectors \mathbf{Y}_i indicating single models for different groups (individuals or cluster).

For this model they derived in their article the finite sample distribution of the LRT (and RLRT) statistic related to the null hypothesis that the variance component of the random effects is equal to zero and the fixed effects underly certain restrictions:

Theorem 6.1.3 Crainiceanu and Ruppert (2004, Theorem 1)

Given the linear mixed model (6.6).

Let μ_s be the eigenvalues of $\Sigma^{\frac{1}{2}} Z^T (I_n - X(X^T X)^{-1} X^T) Z \Sigma^{\frac{1}{2}} \in \mathbb{R}^{K \times K}$ and ξ_s the eigenvalues of $\Sigma^{\frac{1}{2}} Z^T Z \Sigma^{\frac{1}{2}} \in \mathbb{R}^{K \times K}$ and let $\lambda = \frac{\sigma_b^2}{\sigma_\epsilon^2}$. Further, let w_s and u_s denote independent standard normal random variables and

$$\begin{aligned} N(\lambda) &= \sum_{s=1}^K \frac{\lambda \mu_s}{1 + \lambda \mu_s} w_s^2, \\ D(\lambda) &= \sum_{s=1}^K \frac{w_s^2}{1 + \lambda \mu_s} + \sum_{s=K+1}^{n-p} w_s^2 \\ \text{and } f(\lambda) &= n \ln \left(1 + \frac{N(\lambda)}{D(\lambda)} \right) - \sum_{s=1}^K \ln(1 + \lambda \xi_s). \end{aligned}$$

For testing

$$H_0 : \beta_{p+1-q} = \beta_{p+1-q}^0, \dots, \beta_p = \beta_p^0 \text{ and } \sigma_b^2 = 0$$

versus

$$H_1 : \beta_{p+1-q} \neq \beta_{p+1-q}^0 \text{ or } \dots, \text{ or } \beta_p \neq \beta_p^0 \text{ or } \sigma_b^2 > 0$$

the corresponding likelihood ratio test statistic satisfies under H_0

$$\text{LRT} \stackrel{d}{=} n \ln \left(1 + \frac{\sum_{s=1}^q u_s^2}{\sum_{s=1}^{n-p} w_s^2} \right) + \sup_{\lambda \geq 0} f(\lambda), \quad (6.7)$$

where $\stackrel{d}{=}$ denotes equality in distribution.

If H_0 does not impose restrictions on the fixed effects, i.e. for testing only $H_0 : \sigma_b^2 = 0$ versus $H_1 : \sigma_b^2 > 0$, then the corresponding restricted likelihood ratio test statistic satisfies under H_0

$$\text{RLRT} \stackrel{d}{=} \sup_{\lambda \geq 0} \left[(n-p) \ln \left(1 + \frac{N(\lambda)}{D(\lambda)} \right) - \sum_{s=1}^K \ln(1 + \lambda \mu_s) \right]. \quad (6.8)$$

For the proof of this theorem we refer to Crainiceanu and Ruppert (2003).

We may apply this result to our model formulation of Postulate 5 when only one random effect is incorporated in the linear mixed model ($q=1$) and the conditional independence structure for the residual covariance matrix is assumed. The case of one single random effect included in a linear mixed model usually is called *simple linear mixed model*. The individual specific model formulation then turns out to be

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + \mathbf{z}_i \gamma_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma^2 I_{n_i}), \quad \gamma_i \sim N(0, d^2), \quad i = 1, \dots, m$$

and for the overall matrix specification one gets by stacking above the single group components

$$\mathbf{Y} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 I_n) \text{ and } \boldsymbol{\gamma} \sim N_m(\mathbf{0}, d^2 I_m). \quad (6.9)$$

Remember that $\boldsymbol{\epsilon}$ and $\boldsymbol{\gamma}$ are assumed to be independent as well as the random effects of different individuals i . It is easily seen that this model describes a special case of (6.6) with $\Sigma = I$.

Hence, Theorem 6.1.3 provides an exact likelihood ratio test for the need of the random effect (and the validity of some restrictions on the fixed effects) in a simple linear mixed model with independent and identically distributed residuals. That is, for testing

$$H_0 : \beta_{p+1-q} = \beta_{p+1-q}^0, \dots, \beta_p = \beta_p^0 \text{ and } d^2 = 0$$

versus

$$H_1 : \beta_{p+1-q} \neq \beta_{p+1-q}^0 \text{ or } \dots, \text{ or } \beta_p \neq \beta_p^0 \text{ or } d^2 > 0$$

in (6.9). (6.7) presents the exact distribution of the corresponding LRT statistic. Note that this test may be interpreted as comparing the nested model

$$\mathbf{Y} = X\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 I_n) \text{ with } \boldsymbol{\beta}_0 = (\beta_1, \dots, \beta_{p-q}, \beta_{p+1-q}^0, \dots, \beta_p^0)^T, \quad (6.10)$$

with the full model (6.9).

If the null hypothesis does not place any restrictions on the fixed effects, the likelihood ratio test should be based on the restricted likelihood functions, since REML estimates for the variance components are less biased than ML estimates (see Subsection 5.4.2). Thus, for testing

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

the corresponding RLRT statistic is distributed as (6.8).

Note that the distributions (6.7) and (6.8) depend only on the eigenvalues μ_s and ξ_s of $(K \times K)$ -matrices. Crainiceanu and Ruppert (2004) argue that once they have been calculated, simulations of this distribution are straightforward and can be performed much more rapidly than direct bootstrapping. In the statistical software R this is implemented by the function `LRLsim` of the package `RLRsim` (see Subsection 8.3.1).

6.2. Alternative tests for the fixed effects: Approximate Wald tests

The commonly used tests for the variance components (tests for the covariance parameters and the random effects) in a linear mixed model are the likelihood ratio tests described in the previous section. For the fixed effects, however, asymptotic forms of Wald tests are usually preferred. A short overview of these tests is presented in this section.

So far we considered only tests whether a certain subset $\beta^{(2)}$ of the vector of fixed effects $\beta = (\beta^{(1)}, \beta^{(2)})^T$ is restricted to be a certain specified vector $\beta_0^{(2)}$. Clearly, the standard example for such tests which is of primary interest is verifying whether certain fixed effects are needed in the model or if some of them might be eliminated. In this case these fixed effects of interest are tested to be equal to the zero vector. The null hypothesis is then given as

$$H_0 : \beta^{(2)} = \mathbf{0} . \quad (6.11)$$

Sometimes, however, one might be interested in testing hypothesis which are not of the form given in (6.11). For example, we would like to investigate if certain covariates have the same fixed effect on the response, i.e.

$$H_0 : \beta_j = \beta_k, \quad j \neq k, \quad j, k \in \{1, \dots, p\} . \quad (6.12)$$

Generally, we therefore are interested to test whether any linear combination of the coefficients in β is restricted to be zero. That is, we consider so-called *linear hypothesis* of the form

$$H_0 : L\beta = 0 ,$$

where L denotes a certain specified matrix determining the condition to be tested. Note that this null hypothesis includes the hypotheses of the form (6.11) and (6.12) where L is given, respectively, as

$$L = (0, I_q) \text{ where } q \text{ denotes the length of } \beta^{(2)}, \quad \text{and} \quad L = (0, \dots, 0, 1_j, \dots, -1_k, \dots, 0) .$$

According to Cnaan et al. (1997) and Verbeke and Molenberghs (2009, Subsection 6.2.1), an approximate Wald test for testing such linear hypotheses can be derived by approximating the distribution of $\hat{\beta}$ by a multivariate normal distribution with mean vector β and covariance matrix $(X^T \hat{V}^{-1} X)^{-1}$. Remember, however, that $(X^T \hat{V}^{-1} X)^{-1}$ is only a

rough approximation of $\text{Var}(\hat{\beta})$ (see Subsection 5.4.5). Further, the distribution of $L\hat{\beta}$ is approximated by

$$\begin{aligned} L\hat{\beta} &\approx N_{\text{rank}(L)}(L\beta, L(X^T\hat{V}^{-1}X)^{-1}L^T) \quad \text{and} \\ L(\hat{\beta} - \beta) &\approx N_{\text{rank}(L)}(\mathbf{0}, L(X^T\hat{V}^{-1}X)^{-1}L^T) . \end{aligned}$$

Then, the Wald statistic satisfies under the null hypothesis $H_0 : L\beta = \mathbf{0}$,

$$W := \hat{\beta}^T L^T [L(X^T\hat{V}^{-1}X)^{-1}L^T]^{-1} L\hat{\beta} \xrightarrow{d} \chi_{\text{rank}(L)}^2 .$$

Hence, H_0 may be rejected at level α if

$$W > \chi_{\text{rank}(L), 1-\alpha}^2 .$$

Since, as mentioned in Subsection 5.4.5, $(X^T\hat{V}^{-1}X)^{-1}$ underestimates the true variability in $\hat{\beta}$, obtained p-values might be imprecise. Therefore approximate F- or t-tests should be preferred because these tests take into account the inaccuracy of the presented Wald test.

Approximate F- and t-tests

An approximate F-test for testing the linear null hypothesis $H_0 : L\beta = 0$ can be performed by dividing the Wald statistic by the rank of the matrix L . The distribution of this F-statistic can be approximated by an F-distribution with numerator degrees of freedom $\text{rank}(L)$ and denominator degrees of freedom DF estimated from the data (see Cnaan et al. (1997), Verbeke and Molenberghs (2009, Subsection 6.2.2)):

$$F := \frac{\hat{\beta}^T L^T [L(X^T\hat{V}^{-1}X)^{-1}L^T]^{-1} L\hat{\beta}}{\text{rank}(L)} = \frac{W}{\text{rank}(L)} \xrightarrow{d} F_{\text{rank}(L), DF} .$$

For testing whether one single covariate has a significant fixed effect on the response \mathbf{y} or not, the null hypothesis is given as $H_0 : \beta_j = 0, j \in \{1, \dots, p\}$. Then the corresponding t-statistic satisfies

$$T := \frac{\hat{\beta}_j}{\sqrt{(X^T\hat{V}^{-1}X)^{-1}_{jj}}} \xrightarrow{d} t_{DF} ,$$

where the degrees of freedom of the asymptotic t-distribution are again estimated from the data.

Approximate methods for estimating the degrees of freedom for the F- and t-test are described for example by Satterthwaite (1941) and Kenward and Roger (1997). One method which is used in the statistical software R is presented in Subsection 8.3.2.

7. Model building and model diagnostics

What is still outstanding is how to fit an appropriate linear mixed model to a given data set in practice. This question implies procedural methods for model building as well as further tools (in addition to hypothesis tests) for model selection to compare any (also non-nested) models. Finally we consider in this chapter how to examine if a certain fitted model explains the data in an adequate way.

7.1. A general model building strategy

The question how to proceed in fitting an appropriate linear mixed model to a given data set is sophisticated. Neither there is a single strategy for model building that may be applied in every case, nor there exists a strategy which provides automatically the most appropriate model. However, there are several guidelines proposed by West et al. (2007, Section 2.7), and Verbeke and Molenberghs (2009, Chapter 9) which may help to find an adequate model. These guidelines indicate a general procedural method for model selection which is referred to as the *Top-Down Strategy*:

- (1) **Start with a saturated fixed effects model:** Include the maximum number of fixed effects to explain the systematic, population-specific variation in the data, although the model might be overfitted. Overfitting is deliberately preferred at this stage and facilitates the choice of an appropriate covariance structure.
- (2) **Choice of random effects:** Include random effects in the model to describe the individual deviation of the overall population trend. Furthermore, consider which covariance structure of the random effects, i.e. the general form of D , should be chosen.
- (3) **Selection of the residual covariance structure:** Specify the form of the residual covariance matrices Σ_i to explain the remaining variation in the data.
- (4) **Model reduction:** Eliminate certain fixed effects which are not necessary to explain the response.

Note that Verbeke and Molenberghs (2009) emphasize that following the proposed strategy does not necessarily yield a linear mixed model where all distributional assumptions are automatically satisfied. It is essential to check if the model assumptions appear valid (see Section 7.3).

7.2. Information criteria

As seen in Chapter 6, hypothesis tests permit to compare nested models. To discriminate between any two models fitted to the same data set, Pinheiro and Bates (2000) suggest to apply so-called *information criteria*. Information criteria for a given model are functions of the maximized log-likelihood and a penalty term based on the number of model parameters. Basically two information criteria are used:

- The **Akaike information criterion** (suggested by Akaike (1974)),

$$AIC = -2l(\hat{\beta}, \hat{\theta}) + 2r ,$$

- and the **Bayesian or Schwarz's information criterion** (Schwarz (1978)),

$$BIC = -2l(\hat{\beta}, \hat{\theta}) + r \cdot \ln(n) ,$$

where $l(\hat{\beta}, \hat{\theta})$ denotes the log-likelihood with respect to ML or REML estimation and r the total number of model parameters, i.e.

$$r = \text{number of fixed effects} + \text{number of variance components} .$$

Note that the *BIC* penalizes the number of included model parameters more than the *AIC* and incorporates in addition the total number n of observations in the data set. Then, when comparing models, this model should be preferred which has the lowest *AIC* or *BIC* ("the smaller the better").

Information criteria based on REML estimates of θ and the corresponding value of the restricted log-likelihoods should only be used to compare models with the same set of fixed effects. For comparing models with different specifications of fixed effects REML estimation is not senseful and ML estimation is to be used. The reason for this is based upon the origin of the REML log-likelihood described in Subsection 5.4.2. Remember that the REML log-likelihood is the log-likelihood of error contrasts $A^T \mathbf{y}$ with $A \in \mathbb{R}^{n \times (n-p)}$, $\text{rank}(A) = n - p$ and $A^T X = 0$. Hence, different design matrices X lead to different matrices A and this is why for models with different sets of fixed effects the corresponding error contrasts $A^T \mathbf{y}$ differ as well. Thus, due to the latter point, the corresponding REML log-likelihoods are based on different observations which makes them no longer comparable (see Verbeke and Molenberghs (2009, Section 6.2)).

Although Pinheiro and Bates (2000) apply information criteria for discriminating between (non-nested) models, Verbeke and Molenberghs (2009) strongly emphasize that information criteria only provide an empirical rule for model selection. They should be applied with great caution and can not be interpreted as formal statistical tests of significance. Considering the *AIC* or *BIC* might be "rule of thumb" for comparing models and can be applied, for example, to confirm results of hypothesis tests.

7.3. Examining a fitted model

A very important point in deciding whether a fitted model is accurately specified is to check if the model assumptions appear valid. This basically concerns the assumptions about the distribution of the random effects and the residual vector made in Postulate 2 and 3 of Chapter 3. Furthermore one usually is interested to identify outliers, i.e. observations with large residuals, and to detect whether the model fails to fit this data well. These model diagnostics are commonly performed by considering graphics of the fitted model components. A detailed overview with many examples how such graphics can be easily designed with R is given in Pinheiro and Bates (2000, Section 4.3) and shown later in the second, applicatory part of this thesis.

The distributional assumption of normality can be verified by so-called *quantile-quantile (Q-Q) plots*. These graphics provide to compare the empirical distribution with the standard normal distribution by plotting the empirical quantiles and the standard normal quantiles against each other. If the points of the Q-Q plot lie approximately on a line, the normality assumption can be considered as valid.

A general model diagnostic can be performed by drawing the observed against the fitted response values. If the points of such a plot form approximately a line, the linear mixed model can be seen as successful in explaining the data.

It has to be verified for the error term vector ϵ , besides the assumption of normality, the assumption of the zero-mean. If the conditional independence structure for the residual covariance matrices, $\Sigma_i = \sigma^2 I_{n_i}$, is assumed, then, additionally it has to be checked if the variability arising in the error terms is constant.

Referring to the article of Schabenberger (2004), an adequate approximation of the residual or error term vector of individual i at time t , ϵ_{it} , is provided by the so-called *raw residual* of the fitted model. That is the difference between the observed response y_{it} and the fitted one. The fitted response may either refer to the marginal model (4.2). In this case it is the fitted population mean,

$$\hat{y}_{it}^M := \mathbf{x}_{it}^T \hat{\beta} \quad \text{and} \quad r_{it}^M := y_{it} - \hat{y}_{it}^M.$$

Or it may refer to the conditional distribution (4.1) of the response vector, i.e. the predicted, individual-specific mean,

$$\hat{y}_{it}^C := \mathbf{x}_{it}^T \hat{\beta} + \mathbf{z}_{it}^T \hat{\gamma}_i \quad \text{and} \quad r_{it}^C := y_{it} - \hat{y}_{it}^C.$$

However, raw residuals are not suitable to check distributional assumptions about the residual vector ϵ . Even if the individual residual vectors underly a conditional independence covariance structure, i.e. $\Sigma_i \sim \sigma^2 I_{n_i}$, which includes homoscedastic variances, and the assumption that the residual vectors $\epsilon_1, \dots, \epsilon_m$ are independent is true, the raw residuals could be correlated and have unequal variances. Therefore it is preferable for model diagnostics to consider standardized raw residuals. However, the exact standard deviation of the raw residuals is not available in practice. Therefore they are usually scaled by

the estimated standard deviation of the response. These residuals generally are called *Pearson residuals*.

Pinheiro and Bates (2000, Section 4.3) use the conditional Pearson residuals for checking distributional assumptions of the error term vector in a linear mixed model. These are the conditional raw residuals $y_{it} - \hat{y}_{it}^C$ divided by the standard deviation of $Y_{it}|\gamma_i$. According to Section 4.1, the latter quantity is given by

$$\sqrt{[\hat{\Sigma}_i]_{tt}} .$$

Clearly, under the assumption that $\Sigma_i = \sigma^2 I_{n_i}, i = 1, \dots, m$, the conditional Pearson residuals are simplified to

$$\frac{y_{it} - \mathbf{x}_{it}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_{it}^T \hat{\boldsymbol{\gamma}}_i}{\hat{\sigma}} .$$

Fitzmaurice et al. (2004, Chapter 9) recommend to apply the so-called *Scaled residuals* for model diagnostics. These residuals are calculated by scaling the vector of marginal raw residuals

$$\mathbf{y} - X\hat{\boldsymbol{\beta}} := \mathbf{y} - \hat{\mathbf{y}}^M ,$$

accounting that way also for occurring correlations. They argue that, since the covariance matrix of $\mathbf{Y} - X\boldsymbol{\beta}$ is given by V (see (4.2)), the covariance matrix of $\mathbf{y} - \hat{\mathbf{y}}^M$ can be approximated by the same matrix when the variability of $\hat{\boldsymbol{\beta}}$ is ignored. Thus, the scaled residuals are defined by

$$\hat{V}^{-\frac{1}{2}} (\mathbf{y} - \hat{\mathbf{y}}^M) ,$$

where the scaling matrix is given by a Cholesky decomposition of the estimated covariance matrix \hat{V} ,

$$\hat{V} = \hat{V}^{\frac{1}{2}} (\hat{V}^{\frac{1}{2}})^T .$$

Using the previous definitions, the subsequent table provides an overview of the different types of residuals, following the presentation of Schabenberger (2004).

	marginal residuals	conditional residuals
Raw residuals	$r_{it}^M = y_{it} - \mathbf{x}_{it}^T \hat{\boldsymbol{\beta}}$	$r_{it}^C = y_{it} - \mathbf{x}_{it}^T \hat{\boldsymbol{\beta}} - \mathbf{z}_{it}^T \hat{\boldsymbol{\gamma}}_i$
Pearson residuals	$\frac{r_{it}^M}{\sqrt{\hat{\text{Var}}(Y_{it})}}$	$\frac{r_{it}^C}{\sqrt{\hat{\text{Var}}(Y_{it} \gamma_i)}}$
Scaled residuals	$\hat{V}^{-\frac{1}{2}} (\mathbf{y} - X\hat{\boldsymbol{\beta}})$	

Since the scaled residuals are difficult to access in R, we refer to Pinheiro and Bates (2000) and focus on the conditional Pearson residuals for verifying the distributional assumption about the residual vector $\boldsymbol{\epsilon}$. In Subsection 8.2.1 is described how residuals are defined and can be calculated in the `nlme` library of R.

To check the assumptions made on ϵ in Postulate 3 of Chapter 3, usually a scatter plot is considered, where the residuals are drawn versus the fitted values or the values of a certain covariate. If the points of this plot vary around zero the assumption $E[\epsilon] = \mathbf{0}$ probably is valid. If the points deviate constantly around zero without forming a specific pattern, a conditional independence covariance structure of the residuals can be considered as valid. In linear mixed models it is important that such scatter plots are additionally considered separately for each individual to check if variances are constant between the individuals.

8. Fitting and analyzing linear mixed models using R

In order to analyze the *PeKnO-Study* (see Part II), the statistical software R (version 2.8.1) was used. Therefore we complete the first part of this thesis by briefly outlining how linear mixed models can be adjusted and analyzed using R. We basically describe those R functions, which were frequently used in our statistical analysis in Chapter 11. Thus, the reader is referred to that chapter for explicit examples. We focus especially on how to perform the introduced hypothesis tests in R, as it turns out that some inaccuracies arise which have to be kept in mind.

We primarily apply tools provided by the `nlme` library (used version: 3.1-89). This is the commonly used package for fitting and analyzing linear mixed models in R. It was mainly developed by Jose Pinheiro and Douglas Bates and is broadly known and widely used. As an alternative, we recommend the recently developed R package `lme4`.

For an exhaustive description of the functions presented in the following we suggest to consult the respective R manuals. A detailed presentation of the `nlme` library with many applicative examples is given in Pinheiro and Bates (2000). Further examples can be found in West et al. (2007) and Fox (2002).

8.1. Fitting a linear mixed model: The `lme` function

To fit linear mixed models with R, the `lme` function of the `nlme` library is most commonly used. We applied this function to adjust the respective models in Chapter 11 as well. Therefore, this initial section is devoted to the description of the `lme` function. It allows to specify several arguments in order to adjust the linear mixed model as desired.

```
> library(nlme)

> lme( fixed, random, data=..., method="REML", na.action="na.fail",
      correlation=..., weights=..., control=..., ....)
```

Specification of fixed and random effects

The argument `fixed` of the `lme` function stands for a two sided formula of the form

$$y \sim x_1 + x_2 + \dots + x_{(p-1)}$$

specifying the "fix" part of the linear mixed model, $\mathbf{Y} = X\boldsymbol{\beta}$. y represents the response variable and $x_1, x_2, \dots, x_{(p-1)}$ the covariates forming the columns of X , i.e. the covariates which have a fixed, population-specific effect on \mathbf{Y} .

The argument `random` specifies the covariates which have a random effect on y as well as the grouping variable(s). It is generally a one-sided formula of the form

```
~ z1 + z2 + ... + z(q-1) | group
```

where $z_1, \dots, z_{(q-1)}$ represent the columns of the random effects' design matrix Z . The `group` variable gives the grouping structure of the linear mixed model specifying by which variable the data should be grouped. In our case this is the variable indicating which observations belong to the same individual i for $i = 1, \dots, m$.

The form of the covariance matrix D of the random effects $\boldsymbol{\gamma}$ can be also specified by the `random` argument. The default structure is a general symmetric and positive definite matrix (`pdSymm`) but this can be changed to any matrix of the class `pdMat`. For example,

```
> lme(..., random = list ( group = pdDiag( ~ z1 + ... + z(q-1) ) ), ...)
```

specifies D to be a diagonal matrix, i.e. the random effects $\gamma_{i0}, \dots, \gamma_{i(q-1)}$, $i = 1, \dots, m$, to be uncorrelated (variance components structure).

Note that if not explicitly stated "- 1" in the formulas defining the fixed and random structure of the model, an intercept is fitted automatically. Hence we get by using these specifications a p -dimensional vector of fixed effects $\boldsymbol{\beta}$ and a q -dimensional vector of random effects $\boldsymbol{\gamma}$ as defined in Chapter 3. For further informations about the formula specifications we refer to the R manual, for further reading about the `pdMat` classes we recommend to consult Pinheiro and Bates (2000, Subsection 4.2.2).

Grouping Levels

R provides examining data sets with several grouping levels. That is, the `lme` function may incorporate nested random effects. For example, in the fairly well known clustered data sets `MathAchieve` and `MathAchSchool` of the `nlme` library in R, student's scores on a math-achievement test, their socioeconomic status, their High School and the High School Sector (Public or Catholic) were reported. A detailed analysis of these data is given for example by Fox (2002, Chapter 3). In these data sets groups are formed by considering students of one school of one sector. This could even be extended to schools within one county within one state and so forth. Then the state variable represents the *outermost* grouping level and the particular students form the *innermost* grouping level. The grouping structure can be adjusted by defining a list in the `random` argument of the `lme` function, giving the formulas of random effects on the respective grouping level. For example

```
> lme(..., random = list( school ~ mathach , county ~ 1 )...)
```

includes a random intercept and a random slope with respect to `mathach` on the school level and a random intercept on the county level. To repeat the specified random effects

formula $z_1 + z_2 + \dots + z_{(q-1)}$ for all levels of grouping, the grouping structure may be declared by `| group1 / group2 /`. For example

```
> lme( ..., random = ~1 | county / school , ... )
```

includes a random intercept at both levels, county and school. For more details we refer to Pinheiro and Bates (2000, Chapter 3 and 4).

The different grouping levels are specified in R with numbers ascending from the outermost to the innermost level. The outermost level, i.e. the level concerning the whole population, is numbered by 0. In our context we have 2 levels of grouping: the *population level* (level 0) represented through the marginal model. This level (or model) indicates the global relationship between the covariates in X and the response which underly all individuals. At this level only the fixed effects are taken into account. The second level is the *individual level* (level 1) represented through the conditional model where the response is explained for each individual and therefore the individual deviation of the overall population trend is incorporated.

The "data", "method" and "na.action" arguments

The (optional) `data` argument determines the data set from which the variables arise, i.e. the name of the data frame containing the variables.

By the argument `method`, the estimation method for the variance components θ may be specified. The default adjustment is "REML" meaning that $\hat{\theta}$ is derived by maximizing (5.30). The second option is "ML" where the profile log-likelihood (5.24) is maximized.

The default specification `na.action=na.fail` causes to print an error message if there are incomplete observations (NA entries). By changing this setting to `na.action=na.omit`, incomplete observations, i.e. the data rows with NA entries, are removed and a linear mixed model is fitted without these observations. Note that in this case the number of observations given in the summary table (see Section 8.2) is less than the number of rows in the data frame.

The residual covariance structure

By default, it is assumed that the residual covariance matrices underly the conditional independence structure, $\Sigma_i = \sigma^2 I_{n_i}$, $i = 1, \dots, m$, i.e. the single residuals are supposed to be uncorrelated and have equal variances σ^2 . Hence, the occurring correlation between repeated measurements of an individual i is only explained by the covariance structure D of the random effects vector γ_i .

This default structure of the residual covariance matrix can be changed by the arguments `weights` and `correlation`. These arguments allow for modelling heteroscedasticity by setting `weights` equal to a certain `VarFunc` object and provide modelling dependences among the residuals by setting `correlation` equal to a `CorStruct` object (see Pinheiro and Bates (2000, Sections 5.2 and 5.3) for details).

Controlling the estimation algorithm

The computational algorithm for maximizing the respective log-likelihood function to derive the estimate of θ is based on a modification of the EM algorithm and the Newton-Raphson algorithm West et al. (2007). It follows the general framework presented in Lindstrom and Bates (1988). For a detailed description we also refer to Scheipl (2007). Default settings of the estimation algorithm can be changed by the `control` argument. Sometimes the algorithm does not converge, then it might help to increase the number of iterations or change the default optimizer from `nlminb` to `optim`, for example:

```
> lme(..., control = list( maxIter=500, msMaxIter=500, niterEM=100,
                           msMaxEval=1000, opt="optim" ), ...)
```

Note, that not all components in the list have to be necessarily setted and that this adjustment can not be considered as a panacea to solve all kinds of convergency problems.

8.2. Analyzing the fitted model

The first step when analyzing the fitted model usually is to consider the `summary` table of the model. This table summarizes basic model characteristics, returns calculated estimates of the model parameters and performs asymptotic t-Tests for the fixed effects. In the following we explain the single components of a summary table. "... " stands for individual characterizations or calculated values of the respective fitted model and `[mod]` represents a `lme`-object.

```
> summary( [mod] )
```

```
Linear mixed-effects model fit by ...
```

```
Data: ...
      AIC   BIC logLik
...     ...     ...
```

The estimation method for the model is given in the first line of the summary table. This could be either `REML` or `maximum likelihood` estimation, depending on which method was specified when adjusting the model by the `lme` function. In the second line the name of the underlying data set is printed. Subsequently some calculated model characteristics are returned, the information criteria *AIC* and *BIC* as defined in Section 7.2 as well as the value of the model's log-likelihood evaluated at the estimate $\hat{\theta}$. Note that when ML estimation is applied, the `logLik` is the maximized value of the profile log-likelihood (5.24). However, referring to Pinheiro and Bates (2000), under REML estimation the optimal value of the marginal log-likelihood (5.29) is reported, which in contrast to the restricted log-likelihood does not depend on a specific matrix *A* defining the error contrasts, but provides the same estimate of θ .

```
Random effects:
```

```
Formula: ~ [z1] + [z2] + ... + [z(q-1)] | [group]
```

```

Structure: ...
              StdDev      Corr
(Intercept)   ...      (Intr)  [z1]   ...  [z(q-2)]
[z1]          ...      ...      ...
[z2]          ...      ...      ...
...
[z(q-1)]      ...      ...      ...      ...
Residual      ...

```

This paragraph of the summary table presents in **Formula** the in **lme** adjusted random structure of the linear mixed model. **[z1], ..., [z(q-1)]** are the covariates for which a random effect on the response was fitted and **[group]** stands for the grouping variable. **Structure** gives the specified structure of the covariance matrix D of the random effects, by default this is "General positive-definite". The **StdDev** column indicates the estimated standard deviations of the random effects, i.e. $\hat{d}_1, \dots, \hat{d}_q$, as well as the standard deviation of the residuals, σ , in the last line. Remember that the **lme** function fits by default $\Sigma_i = \sigma^2 I_{n_i}$, $i = 1, \dots, m$. Aside, the correlations between the single random effects are returned, i.e.

$$\text{Corr}_{ij} = \frac{\hat{d}_{ij}}{\hat{d}_i \hat{d}_j}$$

where Corr_{ij} denotes the correlation between the i^{th} and the j^{th} random effect given in row i and column j for $i > j$.

```

Fixed effects: [y1] ~ [x1] + [x2] + ... + [x(p-1)]
              Value Std.Error   DF  t-value  p-value
(Intercept)   ...      ...    ...    ...    ...
[x1]          ...      ...    ...    ...    ...
[x2]          ...      ...    ...    ...    ...
...
[x(p-1)]      ...      ...    ...    ...    ...

```

Furthermore, the specification of the "fix" part of the linear mixed model is printed. The column **Value** reports the estimated value for each fixed effect according to (5.31) and the column **Std.Error** the estimated standard deviation of $\hat{\beta}_0, \dots, \hat{\beta}_{p-1}$ calculated by square root the diagonal elemens of the matrix $\hat{\text{Var}}(\hat{\beta}) = (X^T \hat{V}^{-1} X)^{-1}$. Remember that this is only a rough approximation of the covariance matrix of $\hat{\beta}$ (see Subsection 5.4.5).

For each single fixed effect an approximate t-Test as described in Section 6.2 is automatically performed. By these tests the marginal significance of each fixed effect when all other fixed effects are present in the model is investigated. That is, the null hypothesis $H_0 : \beta_k = 0$ is tested for each fixed effect $k = 0, \dots, p - 1$. Aside the columns **Value** and **Std.Error** in the above part of the summary table, the calculated components of these t-tests are presented, i.e. estimated degrees of freedom (DF), the value of the t-statistic (**t-value**) and the corresponding p-value. It holds that

$$\text{t-value} = \frac{\text{Value}}{\text{Std.Error}} \quad \text{and} \quad \text{p-value} = \mathbb{P}(T_{\text{DF}} > \text{t-value})$$

where T_{DF} denotes a random variable following a t-distribution with DF degrees of freedom. For details about the determination of DF we refer to Section 8.3.

Correlation:

```
(Intr)   [x1]   [x2]   .....
[x1]      ...
[x2]      ...   ...
....
```

These lines indicate the approximate correlation among the single fixed effects estimates $\hat{\beta}_0, \dots, \hat{\beta}_{p-1}$ deduced from the approximate covariance matrix $\text{Var}(\hat{\beta}) = (X^T \hat{V}^{-1} X)^{-1}$.

Standardized Within-Group Residuals:

```
Min      Q1      Med      Q3      Max
...      ...      ...      ...      ...
```

This paragraph indicates successively the minimal value, the empirical 25% quantile, the median, the empirical 75% quantile and the maximal value of the conditional Pearson residuals (see below).

Number of Observations: ...

Number of Groups: ...

Finally the number of (complete, without NA entries) observations is given as well as the number of different groups. In our case, the latter value is the number of individuals m .

8.2.1. Fitted values and residuals of a lme object

Fitted values of a linear mixed model can be extracted by

```
> fitted( [mod], level = ., ...)
```

where `[mod]` represents a `lme`-object. The argument `level` specifies whether the response values should be fitted with respect to the marginal model (`level=0`), i.e. $\hat{\mathbf{y}}^M := X\hat{\beta}$, or, the default specification, with respect to the conditional model (`level=1`), i.e. $\hat{\mathbf{y}}^C := X\hat{\beta} + Z\hat{\gamma}$. `level=0:1` returns both.

The residuals of a fitted model can be retrieved by the function `residuals`:

```
> residuals( [mod], type=..., level=.,...)
```

with a `lme`-object `[mod]` and with `level` indicating the grouping level as in the function `fitted`.

By specifying `type="r"` one can ask for the marginal raw residuals (when `level=0`) or the conditional raw residuals (when `level=1`) as defined in Subsection 7.3.

By `type="p"`, the conditional as well as the marginal raw residuals are scaled by the estimated residual standard error $\hat{\sigma}$. These values are called Pearson or standardized

residuals in R. Clearly, under the assumption of the conditional independence structure, $\Sigma_i = \sigma^2 I_{n_i}, i = 1, \dots, m$, the conditional Pearson residuals thus are defined as in Subsection 7.3, because $\text{Var}(Y_{it}|\gamma_i) = \sigma^2$. But the calculation of the marginal Pearson residuals differs from the introduced definition ($\sigma^2 \neq \text{Var}(Y_{it})$). As mentioned in Subsection 7.3, Pinheiro and Bates (2000) focus on these conditional Pearson residuals for verifying distributional assumptions about the error term vector ϵ .

!!! Note that the conditional *raw* residuals are returned by default although the corresponding R manual currently says, that conditional *Pearson* residuals are returned by default.

A third option is `type="n"`, where so-called *normalized residuals* are calculated. This is only relevant in the case when correlation within the individual error term vectors is fitted (provided by the function `CorStruct`). Then the individual residual covariance matrix is assumed to be of the form

$$\Sigma_i = \sigma^2 \Lambda_i, \quad i = 1, \dots, m,$$

where σ^2 denotes the (homoscedastic) variance component and

$$\Lambda_i = (\Lambda_i^{\frac{1}{2}})^T \Lambda_i^{\frac{1}{2}}$$

describes the error correlation matrix. The normalized residuals are then calculated by scaling the Pearson residuals by

$$(\hat{\Lambda}_i^{-\frac{1}{2}})^T.$$

Table 8.1 provides an overview of the calculated quantities. The default values are marked.

		marginal level level=0	conditional level level=1
fitted		$\hat{\mathbf{y}}^M := X\hat{\beta}$	$\hat{\mathbf{y}}^C := X\hat{\beta} + Z\hat{\gamma}$
residuals	type="r"	$\mathbf{y} - \hat{\mathbf{y}}^M$	$\mathbf{y} - \hat{\mathbf{y}}^C$
	type="p"	$\frac{1}{\hat{\sigma}} (\mathbf{y} - \hat{\mathbf{y}}^M)$	$\frac{1}{\hat{\sigma}} (\mathbf{y} - \hat{\mathbf{y}}^C)$
	type="n"	$\frac{1}{\hat{\sigma}} (\hat{\Lambda}_i^{-\frac{1}{2}})^T (\mathbf{y} - \hat{\mathbf{y}}^M)$	$\frac{1}{\hat{\sigma}} (\hat{\Lambda}_i^{-\frac{1}{2}})^T (\mathbf{y} - \hat{\mathbf{y}}^C)$

Table 8.1.: Overview of possible outputs of the R functions `fitted` and `residuals` of the `nlme` library. The default calculations are labeled.

For the sake of completeness it shall be mentioned that calculating the scaled residuals, which were also defined in Subsection 7.3 and are recommended for model diagnostics

by Fitzmaurice et al. (2004), turns out to be complex. However, under the assumption of the conditional independence structure, $\Sigma_i = \sigma^2 I_{n_i}, i = 1, \dots, m$, the estimated covariance matrix $\hat{V} = Z\hat{G}Z^T + \hat{\sigma}^2 I$ of \mathbf{Y} can be calculated fairly easy by the function `extract.lmeDesign` of the `RLRsim` library. This function provides to extract the design matrices X and Z as well as the estimated variance of the residuals $\hat{\sigma}$. Additionally the estimated covariance matrix \hat{G} of the random effects divided by $\hat{\sigma}$ can be accessed. Then \hat{V} can be calculated by using the following code.

```
> Z = extract.lmeDesign([mod])$Z
> G = extract.lmeDesign([mod])$Vr * extract.lmeDesign([mod])$sigmasq
> si.sq = extract.lmeDesign([mod])$sigmasq
> V = Z %*% G %*% t(Z) + si.sq
```

Hereupon, a Cholesky decomposition of this matrix has to be performed and the scaled residuals can be calculated as defined in Subsection 7.3.

8.2.2. Estimated coefficients and prediction

Further useful functions with respect to a `lme` object which should be mentioned here are `fixed.effects` which returns the estimated fixed effects $\hat{\beta}$ and `random.effects` which gives the predicted random effects for each group, $\hat{\gamma}_1, \dots, \hat{\gamma}_m$ calculated as described in Subsection 5.4.5. By the function `coefficients` the effects (fix and random) of each covariate for each single group can be retrieved.

The response value for new data can be predicted by applying the function `predict` to a fitted linear mixed model `[mod]` (a `lme` object) and a data frame `[newdata]` defining the new data.

```
> predict ( [mod], [newdata], level= . ,...)
```

As in `fitted` or `residuals`, the argument `level` specifies at which level prediction should be performed. That is, when `level=0`, then the response for new data of group i is predicted by $X_{\text{new}}\hat{\beta}$, when `level=1`, then prediction is calculated by $X_{\text{new}}\hat{\beta} + Z_{\text{new}}\hat{\gamma}_i$. Note that prediction for a new group is only possible at level 0. Furthermore it is important to know that predicting a single value currently fails.

8.3. Hypothesis tests in R for the model parameters in a linear mixed model

8.3.1. Likelihood ratio tests

In Section 6.1 we presented (asymptotic) likelihood ratio tests for the fixed effects, the covariance parameters and the random effects (i.e. the variance and covariance parameters in the random effects' covariance matrix D). In practice it is of primarily interest to test whether certain model parameters are significant or if some of them might be eliminated from the model. That is, the null hypothesis is considered whether a subset of coefficients in β or θ is equal to zero. As explained in the introductory part of Chapter 6, testing if the

respective model parameters underly certain restrictions can be considered as comparing nested models. The nested model then arised out of the full model without containing the model parameters of interest. Likelihood ratio tests in R about the need of certain model parameters in β or θ are implemented also in the context of comparing nested models and provided by

```
> anova( [mod1], [mod0], ... ) .
```

The argument `[mod1]` is a `lme` object representing the fitted full model, `[mod0]`, a `lme` or `lm` object, stands for the fitted nested model.

For example, for testing the need of certain fixed effects, i.e. $H_0 : \beta^{(2)} = \mathbf{0}$, we have to compare models which differ in their set of fixed effects but have same covariance structure, i.e. the full model is given by $\mathbf{y} = X(\beta^{(1)}, \beta^{(2)})^T + Z\gamma + \epsilon$, and the nested model by $\mathbf{y} = X(\beta^{(1)}, \mathbf{0})^T + Z\gamma + \epsilon$ (see Subsection 6.1.1). Hence, for specifying the nested model simply the covariates of interest corresponding to $\beta^{(2)}$ are omitted. Therefore `[mod1]` and `[mod0]` are of the form

```
> mod1 = lme( y ~ x1 + x2 + ... + xq + x(q+1) + ... + x(p-1),
              random = ..., method="ML", ... )
> mod0 = update(mod1, fixed = ~ x1 + x2 + ... + xq )
```

when the covariates $x(q+1), \dots, x(p-1)$ shall be tested to be significant.

Note that both models necessarily have to be fitted using ML estimation. As already mentioned in Section 7.2, REML estimation does not make sense because models with different fixed effects structure provide different error contrasts $A^T \mathbf{y}$. Hence, the corresponding REML log-likelihoods are based on different observations and therefore not comparable. If a linear mixed model `[mod.REML]` was fitted by `lme` using REML estimation, it can be easily refitted under ML estimation applying

```
> [mod.ML] = update([mod.REML], method="ML") .
```

However, for testing the significance of certain covariance parameters (see Subsection 6.1.2) or testing whether a certain random effect of the model is needed (see Subsection 6.1.3), REML estimation should be used in fitting the models. As mentioned in the respective subsections, this is why REML estimates are less biased than ML estimates and in tests for a certain random effect the RLRT statistic performs generally better in converging against the asymptotic distribution (6.5) Morrell (1998).

Note, that `anova` provides also to compare `lme` and `lm` objects. This is used for performing an asymptotic likelihood ratio test for testing the need of the random effect in a simple linear mixed model. Then `[mod1]` represents the linear mixed model with one single random effect and `[mod0]` the corresponding standard linear model with the same fixed effects structure (a `lm` object).

The output of the `anova` function is given by the following table:

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
[mod1]	1			
[mod0]	2	1 vs 2

The values of the column `df` stand for the number of model parameters of the respective model. The columns `AIC` and `BIC` indicate the corresponding information criteria as defined in Section 7.2 for each of the two models. Further, `logLik` gives the values of the models' log-likelihood or restricted log-likelihood, depending on whether the models were fitted under ML or REML estimation, evaluated at the respective estimate $\hat{\theta}$. `L.Ratio` represents the LRT or RLRT statistic for testing if the nested model is valid. As defined in Definition 6.1.2 it is given by

$$\text{L.Ratio} = -2(\text{logLik}_{[\text{mod0}]} - \text{logLik}_{[\text{mod1}]}).$$

In this notation the index shall represent to which model the value refers, i.e. it specifies the row of the column `logLik`. Let DF denote the difference between the number of model parameters of the full and the nested model, i.e.

$$DF = \text{df}_{[\text{mod1}]} - \text{df}_{[\text{mod0}]}.$$

Then, the corresponding p-value given in the final column of the `anova` table is calculated by

$$\text{p-value} = \mathbb{P}(\mathcal{X}_{DF}^2 > \text{L.Ratio}),$$

where \mathcal{X}_{DF}^2 denotes a \mathcal{X}^2 distributed random variable with DF degrees of freedom.

Remember that in tests for the fixed effects and the covariance parameters the LRT statistic satisfies

$$\text{L.Ratio} \xrightarrow{d} \mathcal{X}_{DF}^2 \quad \text{under } H_0.$$

Hence, provided that the sample is large enough for approximating the distribution of the LRT statistic by the \mathcal{X}_{DF}^2 distribution, the calculated p-value may be valid to decide whether the null hypothesis should be rejected.

However, it is crucial to note that for testing the need of certain random effects, the corresponding RLRT statistic according to Theorem 6.1.2 satisfies

$$\text{L.Ratio} \xrightarrow{d} 0.5\mathcal{X}_{DF}^2 + 0.5\mathcal{X}_{DF-1}^2 \quad \text{under } H_0.$$

Hence, the calculated p-values are overestimated by `anova` in this case and H_0 therefore tends to be accepted too often. This is why manual calculations of the p-value should be preferred:

$$> 0.5 * (1 - \text{pchisq}([\text{L.Ratio}], [\text{DF}])) + 0.5 * (1 - \text{pchisq}([\text{L.Ratio}], [\text{DF}-1])) .$$

Simulations of the distribution

To verify whether the sample is large enough that approximating the distribution of the (R)LRT statistic by its asymptotic distribution is appropriate and whether the p-value given in the `anova` table is trustable, one may perform a simulation study: In a first step, generate data according to the fitted nested model. Hereupon, in a second step, fit both, the nested and the alternative model to each set of these simulated data and calculate the (R)LRT statistics. This may be performed in R by

```
> simulate.lme( [mod0], [mod1], nsim=..., method = ... , seed=..., ...) .
```

The two `lme` objects `[mod0]` and `[mod1]` stand for the nested and the full model, respectively. Note, since both models have to be necessarily fitted by `lme`, it is not possible to simulate the distribution of the (R)LRT statistic for testing the presence of one single random effect in a standard linear model. `nsim` indicates the number of simulations to perform and `method` specifies the estimation method to be used for fitting the simulated data sets. The default setting is `method=c("REML", "ML")` which indicates to fit the nested and the alternative model twice for each data set, one using ML estimation and one using REML estimation. For testing fixed effects, `method="ML"` should be specified since, as already mentioned, REML estimation is not sensible in that case.

Note that it is necessary to declare by `seed` a starting value for the pseudo random number generator which is used for generating data from the fitted nested model.

`simulate.lme` then returns the calculated values of the log-likelihood of each fitted model, i.e. `nsim` values of fitted nested models and `nsim` values of fitted full models. When both estimation methods are applied, then twice this number. The respective (R)LRT statistic can easily be calculated out of it as -2 times the difference between the value of the log-likelihood for the nested and the full model.

This procedural method provides to get an empirical distribution of the (R)LRT statistic under the null hypothesis. Particularly, for each of the generated test statistics, an empirical p-value can be calculated. Hence, to compare the empirical distribution with the asymptotic distribution one may plot these empirical p-values against the theoretical (nominal) ones, calculated for each generated test statistic with respect to the asymptotic distribution. To this end, one may apply

```
> library(lattice)
> plot( [simulate.lme object], df= ... , weights = ... , ...) .
```

The `plot` function yields so-called *trellis plots*, i.e. multiple plots per page. Therefore the library `lattice` is needed. By `df`, the degrees of freedom of the asymptotic χ^2 distribution have to be specified. This argument may also be a vector, which provides several plots where the empirical p-values are compared to each of the specified χ^2 distributions. The argument `weights` indicates whether the empirical p-values shall also be compared to a mixture of χ^2 distributions with degrees of freedom given in `df` and weights specified by a vector of corresponding length in `weights`. The default setting are equal weights. If `weights=F`, the graphic which shows the empirical p-values drawn against the nominal p-values of the χ^2 mixture distribution is omitted. Clearly, there can be additionally specified panel functions in the `plot` function to adjust the graphic.

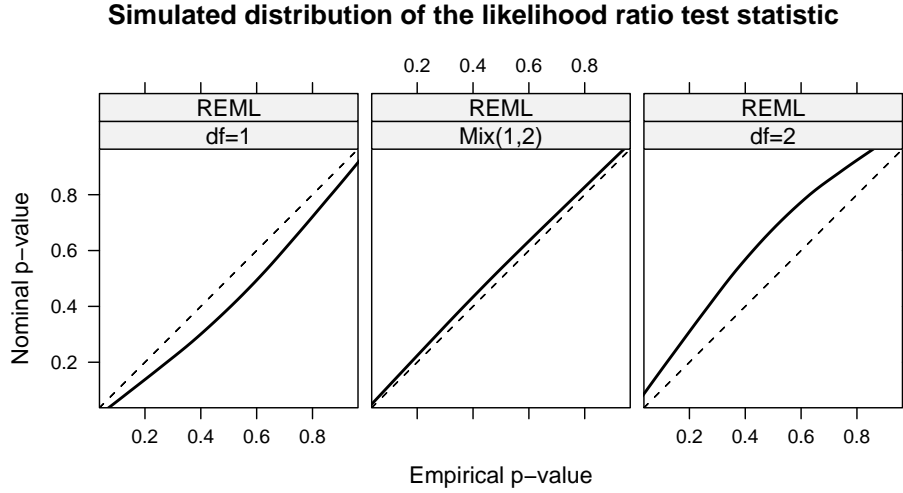

 Figure 8.1.: Example of plotting a `simulate.lme` object

Figure 8.1 shows an example for such a plot. In this example two models with different sets of random effects but same set of fixed effects are compared. In the full model there is adjusted a random intercept and a random slope, in the nested model, only the random intercept is included. Hence, the significance of the random slope shall be investigated. According to Theorem 6.1.2, the asymptotic distribution of the (R)LRT statistic is $0.5\chi_2^2 + 0.5\chi_1^2$.

Plotting the empirical p-values against the theoretical p-values of the $0.5\chi_2^2 + 0.5\chi_1^2$ distribution (indicated by the graphic in the middle) yields nearly a line. From this we may conclude that the asymptotic distribution is an appropriate approximation.

The points in the right plot lie above the diagonal. Hence, calculating the p-value with respect to a χ_2^2 distribution provides a "conservative" test since the true p-value would be overestimated and the null hypothesis might be accepted to lightly.

The nominal p-values of a χ_1^2 distribution are smaller than the empirical p-values (left graphic). A likelihood ratio test which assumes the test statistic to follow this distribution consequently is "anticonservative", i.e. the p-value would be underestimated and the null hypothesis might be rejected to readily.

Exact likelihood ratio tests in simple linear mixed models

For testing the need of the random effect and certain fixed effects in a simple linear mixed model, specified in (6.9) as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n) \text{ and } \boldsymbol{\gamma} \sim N_m(\mathbf{0}, d^2 \mathbf{I}_m),$$

the finite sample distribution of the LRT and RLRT statistics can be derived (see Theorem 6.1.3). Simulating these finite sample distributions enables the function `LRTSim` of the R package `RLRsim`. The functions `exactLRT` and `exactRLRT` of the same package provide exact likelihood ratio tests, based on the simulations of `LRTSim`.

An exact likelihood ratio test about the need of a certain set of fixed effects and the random effect, $H_0 : \beta_{p+1-q} = \dots = \beta_p = 0$ and $d^2 = 0$, can be performed by

```
> library(RLRsim)
> exactLRT( [mod1.ML], [mod0], ... ) .
```

[mod1.ML] represents the full model (6.9) (a `lme` object) fitted by applying ML estimation, and [mod0] the nested model as specified in (6.10) (a `lm` object).

A list is returned containing the value of the LRT statistic, LRT, and the corresponding p-value based on simulated values of the corresponding finite sample distribution (6.7).

```
      simulated finite sample distribution of RLRT.
      (p-value based on 10000 simulated values)
data:
LRT = ..., p-value < ...
```

Note that the default number of simulated values is 10000. This can be changed by the argument `nsim` in `exactLRT`.

As mentioned in Subsection 6.1.4, for testing only the presence of the random effect, $H_0 : d^2 = 0$, the likelihood ratio test should be based on the restricted log-likelihood. Then the null hypothesis $H_0 : d^2 = 0$ can be tested by

```
> exactRLRT( [mod1.REML], ... )
```

where the `lme` object [mod1.REML] represents the full model fitted under REML estimation. Note that in this case it is not necessary to specify the nested model. The output is a list, containing similarly as `exactLRT`, the RLRT statistic as well as the corresponding p-value based on a default number of 10000 simulations of the corresponding finite sample distribution (6.8).

8.3.2. Approximate marginal t-tests and approximate F-tests for the fixed effects

As already described in Section 8.2, an approximate marginal t-test, where the significance of each single fixed effect is tested, is automatically performed by `summary`. To print the results of these t-tests only,

```
> summary([mod])$tTable
```

may be applied. Thus, the estimated values $\hat{\beta}_k$, the estimated standard deviations $\sqrt{\hat{\text{Var}}(\hat{\beta}_k)} = \sqrt{(X^T \hat{V}^{-1} X)_{kk}^{-1}}$, the estimated degrees of freedom, the value of the t-statistics, $\hat{\beta}_k / \sqrt{\hat{\text{Var}}(\hat{\beta}_k)}$, which follows under H_0 asymptotically a t-distribution with DF degrees of freedom, and the corresponding p-values are presented for each fixed effect $k = 0, \dots, p - 1$.

The degrees of freedom are estimated in R by using the following definitions:

- A fixed effect is *estimated at level 2*, if the value of the corresponding covariate can change within the different measurements of one individual, i.e. if it is time-varying.

- A fixed effect is *estimated at level 1*, if the value of the corresponding covariate does not change within the measurements of one individual, but it can change between the individuals, i.e. if it is time-invariant.
- The intercept is *estimated at level 0*.

Furthermore, m_i , $i = 0, 1, 2$, denotes the total number of groups in level i , thus

$$m_0 = 1, m_1 = m \text{ and } m_2 = n$$

and p_i , $i = 1, 2$, the number of fixed effects at level i . Then the degrees of freedom for fixed effects estimated at level i , $i = 1, 2$, are estimated by

$$DF_i = m_i - m_{i-1} - p_i \quad (8.1)$$

and the degrees of freedom for the intercept are estimated by DF_2 .

If a linear mixed model includes a nested structure of random effects, i.e. several levels of grouping, the estimation of the degrees of freedom turns out to be slightly more complex. For this case we refer to Pinheiro and Bates (2000, Subsection 2.4.2).

An approximate F-test for testing the general null hypothesis $H_0 : L\beta = \mathbf{0}$ as described in Section 6.2 can be performed by

```
> anova( [mod], type = "sequential", Terms = NULL , L = NULL , ... ) .
```

It is returned a table indicating the numerator degrees of freedom, **numDF**, i.e. the rank of the matrix L , the denominator degrees of freedom, **denDF**, estimated by (8.1), the calculated value of the F-statistic, **F-value** and the corresponding p-value, **p-value**. Remember that it holds

$$\text{p-value} = \mathbb{P}(F_{\text{numDF}, \text{denDF}} > \text{F-value}),$$

where $F_{\text{numDF}, \text{denDF}}$ represents a random variable following a F-distribution with **numDF** and **denDF** degrees of freedom.

By default, an approximate sequential F-test for each fixed effect is accomplished, i.e. the fixed effects are tested sequentially in the order they appear in the model. The null hypothesis for each fixed effect β_k can be formulated as $H_0 : \beta_k = 0$ when $\beta_0, \dots, \beta_{k-1}$ are present in the model.

Changing to the second option **type="marginal"** provides that for each fixed effect an approximate marginal F-test is performed. That is, $H_0 : \beta_k = 0$ when all other fixed effects are present in the model for $k = 0, \dots, p - 1$.

If several fixed effects shall be jointly tested to be zero, the names of the corresponding covariates have to be specified in a character vector in the argument **Terms**. For example,

```
> anova( [mod], Terms = c("[x(p-k)]", ..., "[x(p-1)]") )
```


tests if $\beta_{p-k} = \dots = \beta_{p-1} = 0$. Note, that it is only possible for fixed effects with same estimated degrees of freedom. This implies that it is not possible to test jointly the significance of a time-varying and a time-invariant covariate.

For testing other linear hypothesis of the form $H_0 : L\boldsymbol{\beta} = \mathbf{0}$, the matrix L defining the linear combinations to be tested can be declared in the argument `L`. By naming the columns of the matrix with the corresponding covariates, it can be specified to which of the fixed effects the linear combinations correspond. For example, $H_0 : \beta_1 = \beta_2$, i.e. $(1, -1)(\beta_1, \beta_2)^T = 0$ can be tested by

```
> anova( [mod], L = c( "[x1]" = 1, "[x2]" = -1 ) ) .
```

Note that for both tests, approximate marginal t-tests and approximate F-tests, only the full model has to be fitted. Hence, in comparison to the likelihood ratio test, it is not necessary to fit both models, the full and the nested model, which is tested to be valid.

Part II.

Analyzing the PeKnO-Study

9. Study and data description

In the second part of this thesis a detailed statistical analysis of the *PeKnO-Study* (Perimenopausale Knochendichte [bone density] und Ovulation) is performed. As mentioned in the introduction, it is a prospective observational study over the course of two years investigating perimenopausal women to characterize the changes in bone metabolism during this phase of life. It was carried out at the *Frauenklinik der Technischen Universität München* under the direction of PD Dr. med. Vanadin Seifert-Klauss. The study is still ongoing, meaning that still further participants are included or participants are observed over more than one two-year period. Inclusion criteria for the women in this study are that they are older than 45 years or age, their cycle lengths are not greater than 42 days, no exogenous hormones were used during the 6 months prior to study onset, and that there are no medical reasons for low bone mass. In a nutshell, the study is based on the following observations:

- measurement of the lumbar spine trabecular bone mineral density, using quantitative computed tomography at the begin of the study and after two years (end of the study)
- documentations of the menstrual cycles for the whole two-year period of the study using a cycle monitor
- measurements of the serum concentration of hormonal values (FSH, LH, progesterone, estradiol and cortisol basal) and bone turnover markers (CTX, osteocalcin and BAP) by luteal phase blood tests performed approximately every 6 months from the begin of the study (i.e. 5 blood samples for each participant).

A basic illustration of the study, indicating when the respective measurements were carried out, is shown by Figure 9.1.

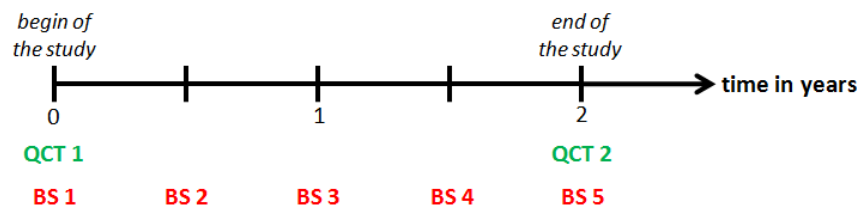


Figure 9.1.: Basic overview of the study for one single participant: *QCT* indicates the measurements of bone density, *BS* the number of the blood test

At the time this thesis was written, data were available for 49 different women. These women were coded by a number indicated by the acronym `PAT.ID`. However, two participants are represented twice in the study (`PAT.ID` 12 and 24), i.e. they were observed over all in all 4 years. In our initial investigations of this chapter we incorporate both two-year observations of these women and refer therefore to 51 participants.

A concrete description of the single variables available for our statistical analysis is presented in the following, referring to Seifert-Klauss (2007), Seifert-Klauss et al. (2006), Seifert-Klauss and Prior (2010) and Seifert-Klauss et al. (2011). For an introduction to the subject of osteoporosis we recommend Bartl (2009). Table B.1 (p. 158) provides an overview of all variables which are considered in this thesis.

Basic characterizations of the participants

At first, some general properties of the participants concerning their age and their body mass index (BMI) were recorded. At the beginning of the study participants were aged between 45 and 56 years with an average of 48.2 years. The age at study onset is expressed by the variable `AGE`. Figure 9.2 shows the number of participants with the respective age at the begin of the study.

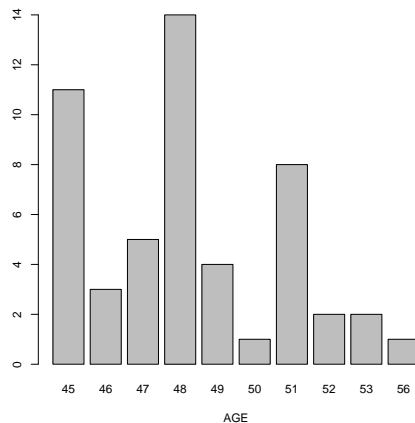


Figure 9.2.: Barplot of the age of the participants at the begin of the study

The BMI is a heuristic proxy for the percentage of body fat based on the individual's weight and height:

$$\text{BMI} = \frac{\text{weight}[\text{kg}]}{(\text{height}[\text{m}])^2}.$$

This index represents a measure for rating the weight. If it is greater than 25kg/m^2 the person can be classified as overweight, if it is less than 18.5kg/m^2 the person is underweight. Underweight is said to have a negative effect on the bone density, thus the BMI provides an appropriate measure to investigate this correlation. For each of the participants an approximate value for the BMI at 25 years of age (`BMI.25`), the BMI at

the begin of the study (BMI.0) and the BMI at the end of the study (BMI.1) was recorded. Figure 9.3 shows these values for every participant. Remember that participant number 12 and 24 are observed twice since they took part in the study over two two-year periods. Note that at 25 years of age 5 women were slightly underweight (PAT.ID 14, 30, 34, 45

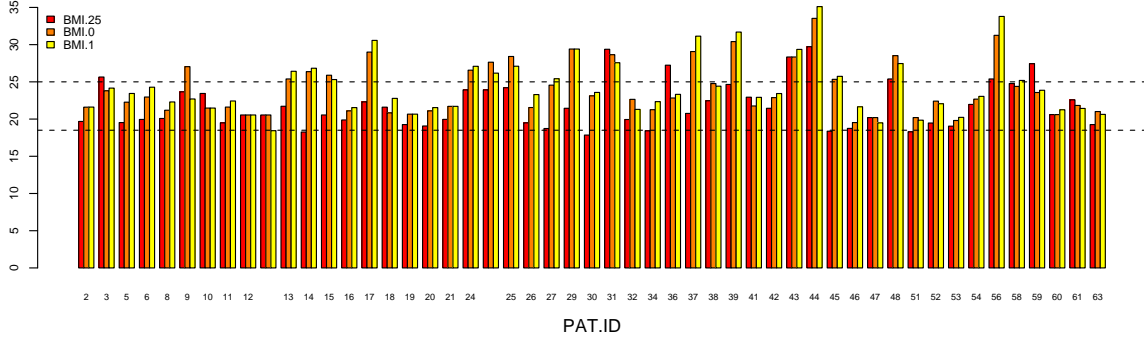


Figure 9.3.: Barplot of the BMI of each participant at 25 years of age, at the begin of the study and at the end of the study. BMI values within the dashed lines indicate that the corresponding woman is normalweight.

and 51). At study onset, however, none of the women was underweight. Furthermore, the changes in the BMIs over time are denoted by the variables

$$\begin{aligned} X.BMI.0.25 &:= BMI.0 - BMI.25, & X.BMI.1.25 &:= BMI.1 - BMI.25 \\ & & \text{and} & X.BMI.1.0 &:= BMI.1 - BMI.0. \end{aligned}$$

These changes refer to the differences of the later value and the respective value earlier in time, hence a positive value indicates the gain in weight over time and a negative value the loss in weight. From Figure 9.3 it can be seen that most of the participants gained weight from 25 years of age to the begin of the study. Between the begin and the end of the study there are no significant changes of the BMI values apart from participant 9, who did severely lose weight during the course of study.

Since participants were asked for this weights, rather than being weighted on site, these variables might be less trustable than others. Certain outliers might be explained by exaggerated statements.

Measurements of the bone density

At the begin and the end of the study measurements of the lumbar spine trabecular bone mineral density (BMD) were performed. To this end, quantitative computed tomography (QCT) was employed. The corresponding values are expressed in mg Calcium-Hydroxyapatite (Ca-Ha)/ml and denoted by QCT.1 for the study onset and QCT.2 for the end of the study.

Pursuant to the World Health Organization the values of the BMD are classified as follows:

120 mgCa-Ha/ml <	BMD	→	<i>normal BMD</i>
80 mgCa-Ha/ml ≤	BMD ≤ 120 mgCa-Ha/ml	→	<i>osteopenia</i>
	BMD < 80 mgCa-Ha/ml	→	<i>osteoporosis</i>

At the begin of the study, the mean level of the BMD values is given by 139.3 mgCa-Ha/ml. At this time 9 of the 51 participants had osteopenia. At the end of the study 16 participants had osteopenia and the average BMD value dropped to 134.0 mgCa-Ha/ml. Figure 9.4 shows the two BMD measurements for each participant. Note that for participants number 18 and 19 the first and the second BMD measurement, respectively, was not carried out. The bone density increased for only 9 participants, all others lost BMD.

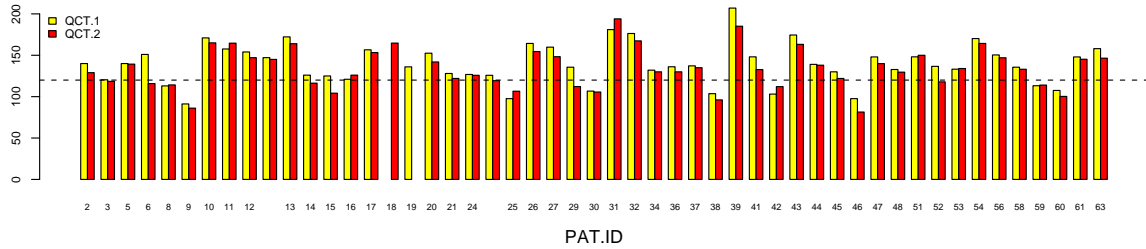


Figure 9.4.: Barplot of the BMD of each participant at the begin and the end of the study. Values below the dashed line indicate that the corresponding woman has osteopenia.

The difference of BMD over time is defined by the variable

$$X.BMD := QCT.2 - QCT.1 ,$$

taking negative values when BMD is lost and positive values when BMD increases. $X.BMD$ takes on values between -35.2 mgCa-Ha/ml (PAT.ID 6) and 13.0 mgCa-Ha/ml (PAT.ID 31) and has a mean of -6.1 mgCa-Ha/ml.

To make this change of BMD comparable, it might also be interesting to consider the percental change of BMD with respect to the initial bone density. This relative value of the change of BMD is expressed by

$$X.BMD.PR := \frac{X.BMD}{QCT.1} \cdot 100\%$$

and takes values from -23.30% to 9.22% with mean -4.22%. Figure 9.5 illustrates the absolute and percental change of BMD for each participant. It shows that the absolute loss of BMD of participant 39 is much higher than of participant 41, but since number 41 starts at a lower initial BMD level, the relative loss is equal. In total however, the two graphics do not differ a lot in their general form so that $X.BMD$ is meaningful for comparisons as well. On this variable we focus henceforth.

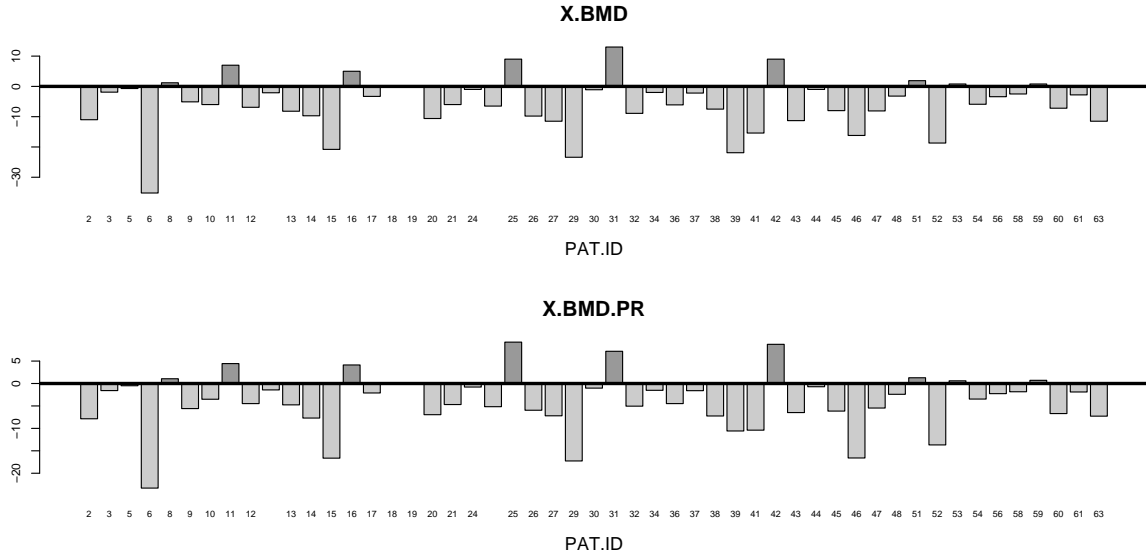


Figure 9.5.: Barplot of the absolute change of BMD ($X.BMD$) and the percent change of BMD with respect to the initial bone density ($X.BMD.PR$) for each participant. Negative values indicate the loss of bone density.

For further characterizations of the change of bone density, it is classified by the variable `BMD.GROUP` as follows:

- `BMD.GROUP = 11`, if $QCT.1 > 120$ and $QCT.2 > 120$, i.e. normal BMD at the begin and the end of the study
- `BMD.GROUP = 12`, if $QCT.1 > 120$ and $QCT.2 \leq 120$, i.e. normal BMD at the begin of the study and osteopenia at the end of the study
- `BMD.GROUP = 22`, if $QCT.1 < 120$ and $QCT.2 \leq 120$, i.e. osteopenia at the begin and the end of the study.

Note that none of the participants has osteopenia at the begin of the study and normal BMD at the end of the study. Figure 9.6 shows the number of women in the respective group.

Menstrual cycle documentation

During the whole time of the study, characterizations of the menstrual cycles of each participant were documented in detail. To this end, a fertility monitor (*Clearplan® Fertilitätsmonitor*) was used, which usually is applied by women who want to get pregnant. This monitor was allocated to each participant at the begin of the study. It allows to detect days on which there is a high probability of ovulation by measuring hormone concentrations on the basis of urine tests.

It's concrete manner of use is briefly described as follows: At the beginning of each

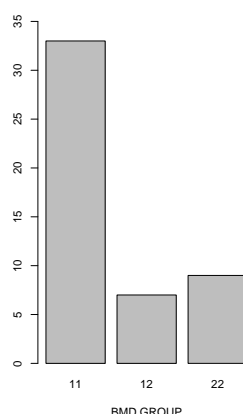


Figure 9.6.: Barplot of BMD.GROUP

cycle, i.e. at the first day of menstruation, the monitor has to be set up to start the documentation of a new cycle. From day five of each cycle up to day 14, participants had to insert urine test sticks from the first urine of the day into the corresponding slot of the monitor. This is requested by a test stick symbol on the display. When cycles are very long or irregular a urine test might be necessary for a total of 20 days. In that case, the first day of the urine test can vary between day 6 and 9, according to what the monitor has learned about earlier cycles. On the basis of these first morning urine samples, the cycle monitor measures the concentration of the luteinizing hormone (LH), which triggers ovulation, and estrogen (estrone-3-glucuronide) and calculates their ratio. A significant surge of this ratio makes ovulation 24-36h later highly probable. Such an increase is indicated by an ovulation symbol on the display. For a detailed description of the monitor we refer to Clearblue (2007).

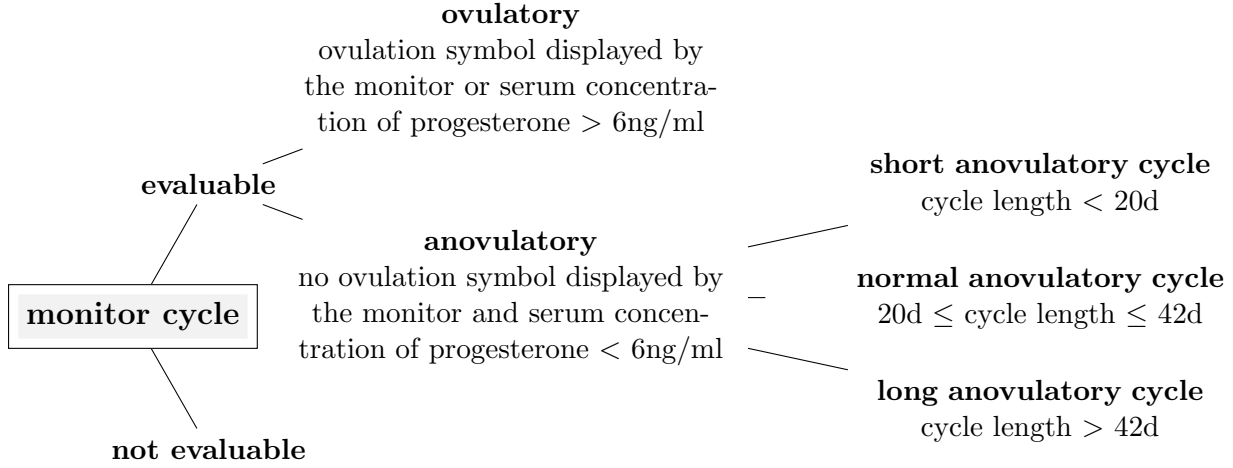
The monitor is able to store data of up to eight menstrual cycles. Approximately every six months, the monitor data were transmitted to the computers of the study group and analyzed. This was done when the women came to the clinic in order to take of the blood sample.

Cycles were categorized as *not evaluable* if urine tests were not performed at more than two consecutive days and, at the same time, no increase of the hormone ratio was reported by the monitor during the whole cycle.

According to the monitor information, evaluable cycles then were categorized as ovulatory or anovulatory. Anovulatory cycles were further classified by their cycle lengths.

However, it was not relied on the monitor declaration when progesterone levels in the blood sample were high. Progesterone is a female sex hormone which is produced in a great extent only in case ovulation occurred. It prepares the uterus for implantation of the fertilized egg, to maintain pregnancy, and to promote development of the mammary glands. The commonly considered cut-off value for the serum concentration of progesterone indicating ovulation is 6ng/ml. It is one of the female sex hormones which were measured at the semiannual blood tests (see the following subsection). In the case that the serum concentration of progesterone was greater than 6ng/ml, but the monitor did not

display an ovulation symbol during the cycle at which the blood test was performed, the respective cycle was categorized as ovulatory and a failure of the monitor was assumed. This classification of the menstrual cycles are shown by the following diagram:



The percentage of ovulatory cycles with respect to all evaluable cycles between the ca. semiannual time interval of two blood samples is recorded by the variable

$$\text{ov.tot.PR} := \frac{\text{number of evaluable } \textit{ovulatory} \text{ cycles since the last blood test}}{\text{number of evaluable cycles since the last blood test}} \cdot 100\% . \quad (9.1)$$

At the end of the study, data were aggregated to a variable ov.tot.s.PR , indicating the percentage of ovulatory cycles with respect to all evaluable cycles during the study,

$$\text{ov.tot.s.PR} := \frac{\text{total number of evaluable } \textit{ovulatory} \text{ cycles during the study}}{\text{total number of evaluable cycles during the study}} \cdot 100\% . \quad (9.2)$$

The percental number of the single anovulatory cycle types was documented analogously.

The mean length of cycles of the individual average cycle lengths (CYL.m) is 45.0 days. Figure 9.7 illustrates these individual average lengths of cycles for each participant. There are six women with extremely long average lengths of cycles (PAT.ID 6, 9, 15, 41, 47 and 48). These participants did not have a menstruation during a very long time of the study (up to one year). For participants 6, 9 and 17 even menopause occurred within the time of the study.

Figure 9.8 shows the percentage of ovulatory and anovulatory cycles of the total number of evaluable cycles during the study for each woman. Furthermore, the percentage of long anovulatory cycles is recorded. It illustrates that the portion of anovulatory cycles in total is very high. There are also quite many extremely long, anovulatory cycles. These observations are typical for perimenopausal women, since cycles become highly irregular and ovulatory hormone production is reduced in this phase of life. However, there are two women who only had ovulatory cycles during the study (PAT.ID 10 and 45). It is noticeable that participant 15, who did not have one single ovulatory cycle shows a relatively high loss in BMD (see Figure 9.5).

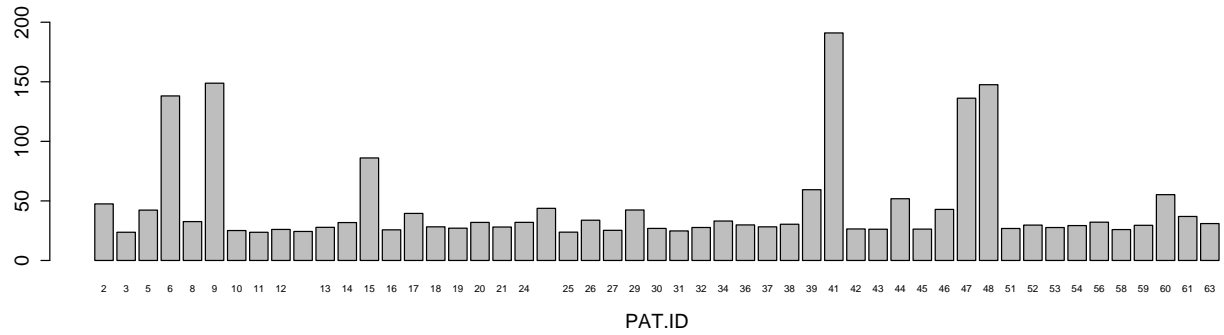


Figure 9.7.: Barplot of the mean length of all cycles (CYL.m) for each participant

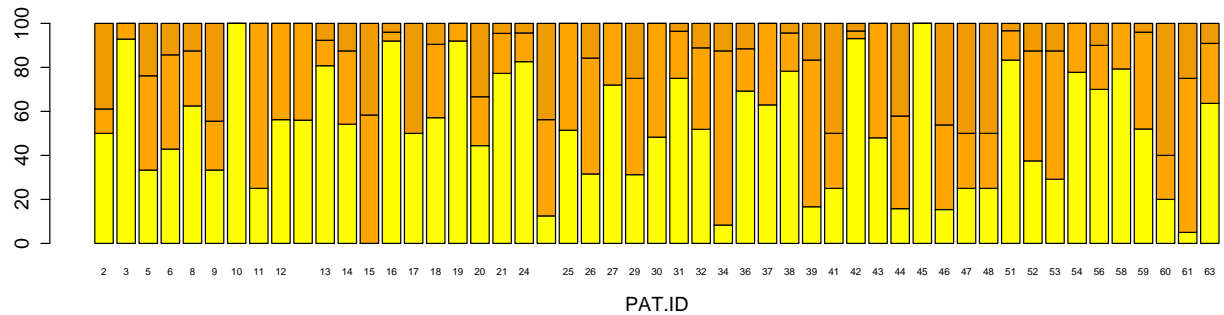
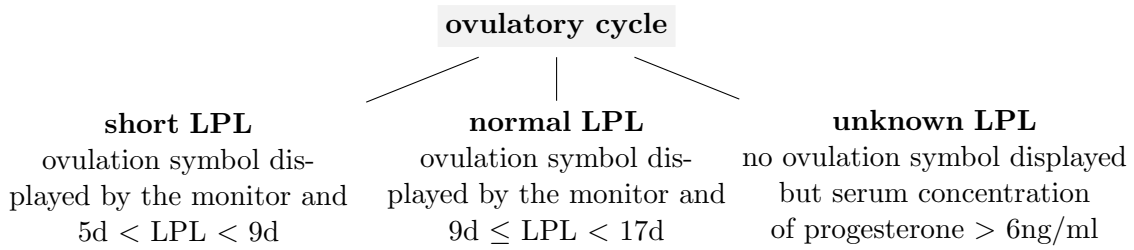


Figure 9.8.: Barplot of the percental number of ovulatory (yellow) and anovulatory (orange) cycles with respect to all evaluable cycles in the whole study for each participant. The upper, dark part of the orange bars indicates the percentage of long anovulatory cycles (anov.long.s.PR)

The big advantage of the cycle monitors is that not only the length of each cycle is recorded (as in former studies), but also the (likely) day of ovulation. This information allows to derive the (likely) length of the luteal phase (LPL) of each cycle. The luteal phase is defined as the second half of the cycle, i.e. the phase from ovulation to the end of the cycle. Typically, it is around 10-16 days in length. It is calculated, in days, from the data stored by the cycle monitor by

$$\text{LPL} = \text{length of the cycle in days} - \text{day of the ovulation symbol} - 2. \quad (9.3)$$

Anovulatory cycles (i.e. no ovulation occurs) are formally coded by a LPL of 0.1 days. The mean length of luteal phases of the individual average lengths of luteal phases (LPL.m) is 6.1 days. This low value indicates that there is a great portion of anovulatory cycles. Using this information, ovulatory cycles can be further classified by the length of their luteal phase as follows:



Since, however, our investigations did not show any necessity of this classification, we focus on discriminating the menstrual cycles referring to the previous diagram.

Semiannual blood samples

One purpose of the PeKnO study which is of primarily interest is to investigate whether women with higher rates of ovulatory cycles have a lower loss of bone density. Therefore, besides a detailed documentation of lengths and types of the menstrual cycles during the study, semiannual blood tests were performed to measure concentrations of the most important female sex hormones.

As already has been mentioned, one of these hormones is progesterone from which conclusions about the type of the cycle can be drawn, since its production notably increases only when ovulation occurs. Therefore, to allow for a feasible interpretation, blood tests have to be performed during the luteal phase. To recapitulate, in case of ovulation the serum concentration of progesterone usually is higher than 6ng/ml and vice versa, progesterone values greater than 6ng/ml indicate ovulation. Progesterone is suggested to have a bone-constructive effect.

Furthermore, the serum concentrations of the follicle-stimulating hormone (FSH) in IU/l, which is suggested to have a bone-destructive effect, and the luteinizing hormone (LH) in IU/l as well as (17β)-estradiol in pg/ml were recorded.

The production of these hormones is highly correlated. The basic relation among them and their main functions in the menstrual cycle can be briefly described as follows: Ini-

tially, significant amounts of FSH and LH are produced to stimulate egg follicles to grow. Hereupon, the maturing follicles release estradiol. As progesterone, estradiol prepares the uterus for implantation of the fertilized egg. When the estradiol level reaches a certain point, it causes an increased production of LH. This surge of LH triggers the most mature follicle to burst open (ovulation) usually happening in the middle of the cycle (day 12-14). Up from there the luteal phase begins: The egg is leaving the ovary ready to be fertilized the next 12 to 24 hours. At the same time the egg's follicle is transformed into the corpus luteum which produces significant amounts of progesterone whereas concentrations of FSH, LH and estradiol drop. If ovulation does not occur, the levels of FSH and LH do further increase. Thus, increased concentrations of these hormones during the luteal phase may indicate that the cycle is anovulatory. For further reading about this subject we recommend the book of Gruhn and Kazer (1989).

To recapitulate, the cycle monitor displays an ovulation symbol when ovulation is highly probable 24-36h later. Therefore, the begin of the luteal phase was defined as two days after the ovulation symbol was shown (see 9.3). Consistently to this, a blood test was classified as *luteal phase blood test*, i.e. as correct, if it was performed

- 6-9 days after the ovulation symbol was displayed by the cycle monitor or at day 19-22 of the cycle when there is no ovulation symbol displayed and
- at least 4 days before the next menstruation.

When the serum concentration of progesterone was measured to be greater than 6ng/ml, blood samples were classified as correct, regardless of the point of time the sample was taken. Failing these conditions, a blood sample was categorized as incorrect.

The 255 data ($\cong 51 \times \text{five blood samples}$) are split up as follows:

Classification of the blood test	correct	not correct	missing
Number of observations	182	58	15

In our statistical analysis we refer only to serum concentrations taken from correct (luteal phase) blood tests. The table indicates that unfortunately by far not for all participants there are 5 luteal phase blood sample available. It is especially regrettably that participants number 6 and 47 do not have any correct taken blood sample over the whole study so that we lose them for our investigations.

Additionally, the concentration of

- the stress hormone cortisol in $\mu\text{g}/\text{dl}$

as well as the bone turnover markers

- carboxyterminal cross-linking telopeptide of bone collagen (CTX) in ng/ml, a bone resorption marker,
- osteocalcin in ng/ml, a bone formation marker and

- bone-specific alkaline phosphatase (BAP) in $\mu\text{g/l}$ given on an ordinal scale with usual values from 4 to 21, another bone formation marker

were recorded from an analysis of the blood samples. The latter ones are biochemical markers characterizing the bone metabolism. However, bone formation and resorption always is linked, so that an increase of the bone formation markers is usually accompanied by an increase of CTX.

Figure 9.9 shows histograms of the individual means of the serum concentrations of the respective female sex hormones, cortisol and the bone markers measured in the luteal phase blood tests. This graphic illustrates how the single variables are distributed. Especially FSH and LH are concentrated to the left part of their range but there are also very high values observable corresponding to long anovulatory cycles.

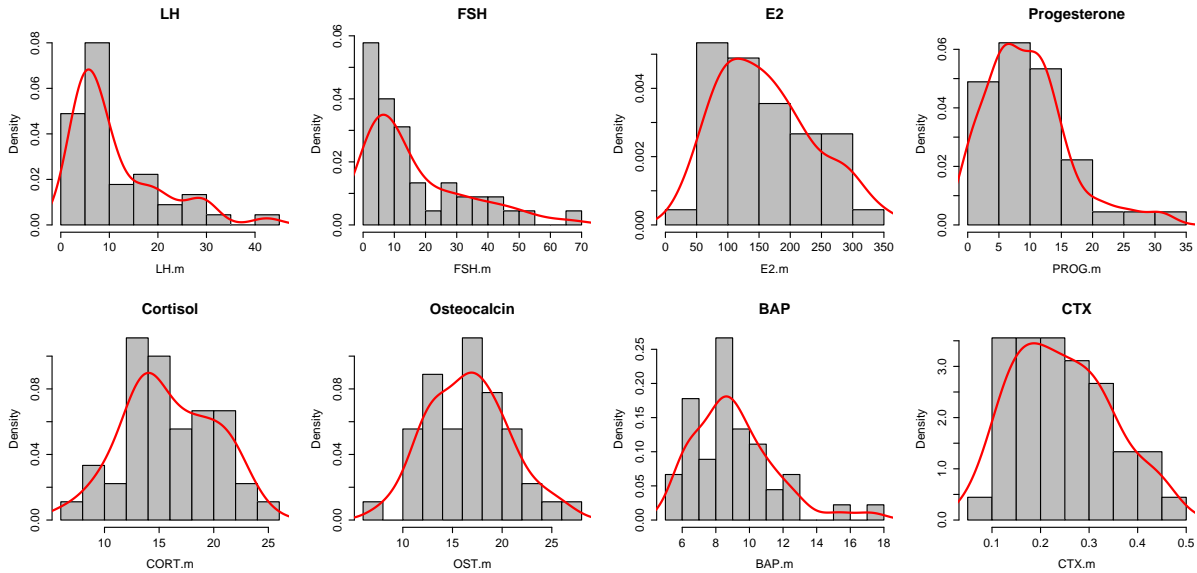


Figure 9.9.: Histograms of individual mean values of the variables determined by the luteal phase blood tests

Figure 9.10(a) illustrates the correlation among the sex hormones. Especially the high correlation between FSH and LH can be seen very well. As expected, these hormones are negative correlated with estradiol and progesterone. Figure 9.11 highlights the correlation of the sex hormones and ovulation. Clearly, it is not surprising that the ratio of ovulatory cycles is highly correlated with the average concentration of progesterone.

Figure 9.10(b) confirms that mechanism within the bone metabolism are linked. Especially the bone resorption marker CTX and the bone formation marker osteocalcin correlate strongly.

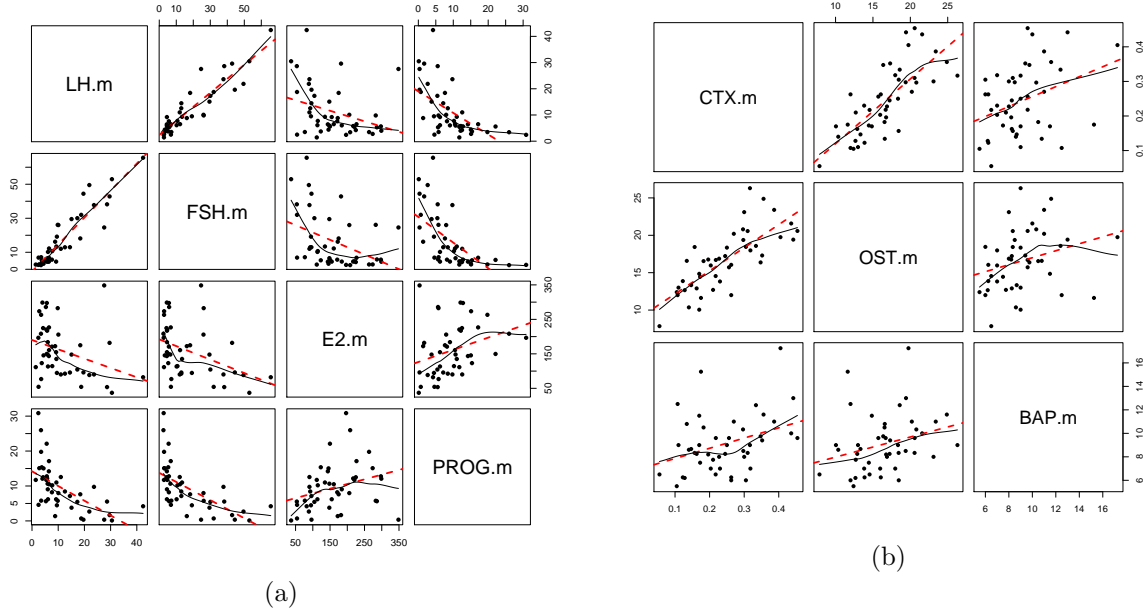


Figure 9.10.: Pairs plots of the individual mean values of the female sex hormones (a) and the bone turnover markers (b) determined by luteal phase blood samples

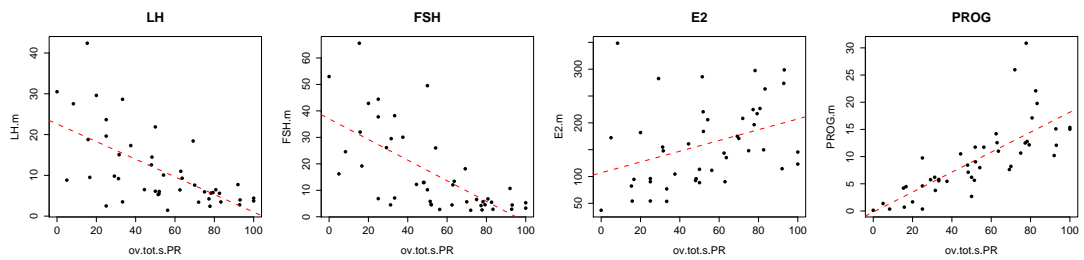


Figure 9.11.: Scatter plot of the percentage of ovulatory cycles during the whole study and the mean values of the luteal phase serum concentrations of the female sex hormones

Response variables and goals

In the following chapters we investigate which variables influence the bone metabolism and lead to the accelerated bone loss occurring in the perimenopause. We focus at first on explaining and predicting the change of bone density ($X.BMD$) by fitting a standard linear model (Chapter 10). We will only outline the model building process and focus on results.

Furthermore we verify which factors have an influence on the bone markers. Since several measurements of CTX, osteocalcin and BAP are available for each woman, we use linear mixed models (see Part I) to account for arising correlations among the repeated measurements. As already mentioned, we only incorporate measurements of blood tests which can be clearly classified as luteal phase blood tests.

The bone markers characterize the bone metabolism and are supposed to be very rough indicators for the bone density. But since bone formation and resorption are linked, changes of bone marker concentrations can be hardly interpreted (Seifert-Klauss et al. (2011)). Therefore we aim to compare here the change of the bone markers and the change of the bone density during the course of the study. We declare the change of CTX, osteocalcin and BAP by $X.CTX$, $X.OST$ and $X.BAP$, respectively. Similarly as $X.BMD$, we define them as the difference between the value at the end of the study (5th blood sample) and the value at the begin of the study (first blood sample).

Figure 9.12 shows the number of participants where the BMD and the bone markers increase or decrease during the study. As already illustrated by Figure 9.4 and Figure 9.5, it can be clearly seen that there are only few women whose bone density increases over the two years. Furthermore we note that for the greater part of participants the serum concentration of the bone markers is higher at the end of the study than at study onset.

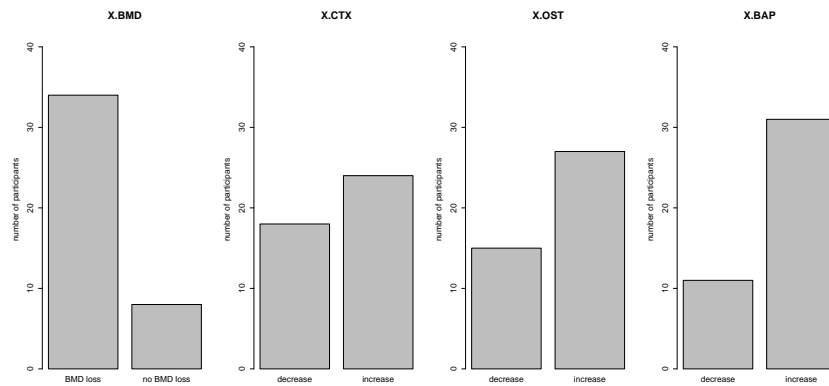


Figure 9.12.: Barplot of participants with increasing or decreasing BMD and bone turnover markers during the course of the study

Figure 9.13 shows the correlations among the change of BMD and the change of the respective bone marker. It illustrates that the loss of bone density is slightly accompanied

by an increase of the bone markers.

This observation can be even better clarified by Figure 9.14 which shows how the changes

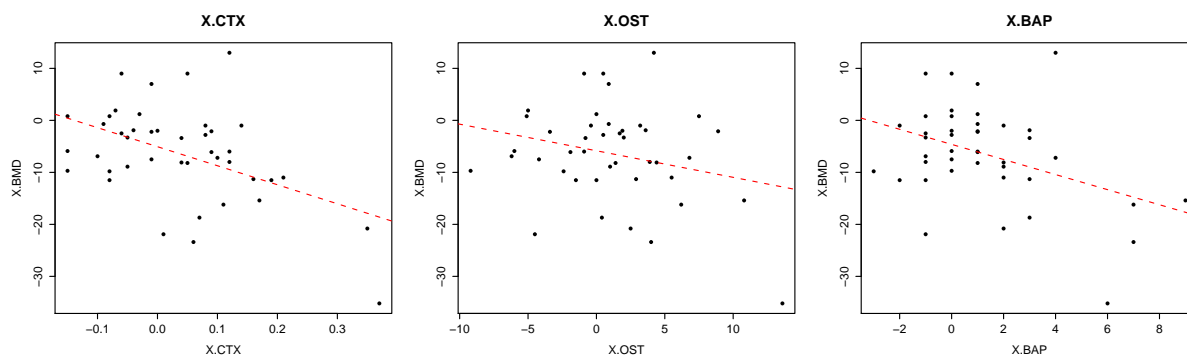


Figure 9.13.: Scatter plot of the change of BMD and the change of the respective bone marker during the course of the study

of bone markers are distributed in women who lose BMD and women who gain BMD. In the latter group, the bone formation markers osteocalcin and BAP do on average not change highly since the corresponding means are almost zero. The bone resorption marker CTX actually decreases in most of the participants. In women who lose BMD, bone marker concentrations are higher at the end than at the begin of the study.

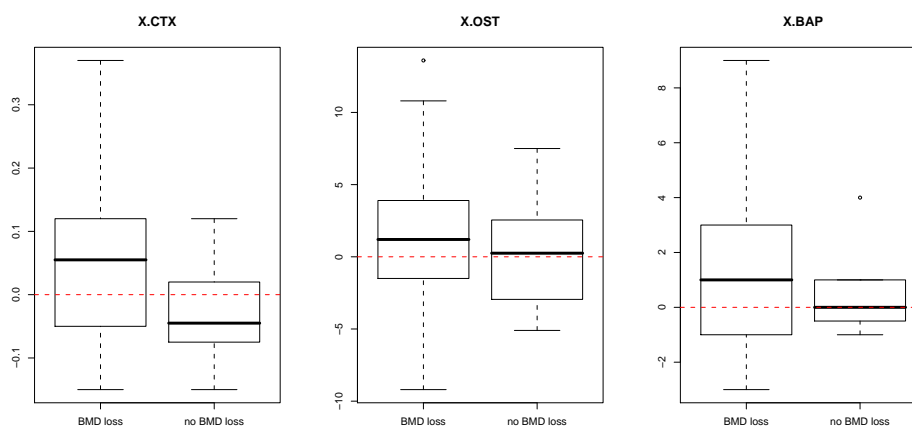


Figure 9.14.: Boxplots of the change of the respective bone markers classified by women who lost BMD ($X.BMD < 0$) and women whose BMD increased during the course of the study

Figure 9.15 shows the percental change of the bone markers for this group of participants, i.e. the change of the bone markers with respect to the value measured at the first blood test. It illustrates that for women who lose bone density during the course of the study, the

bone resorption marker CTX increases on average more than the bone formation markers.

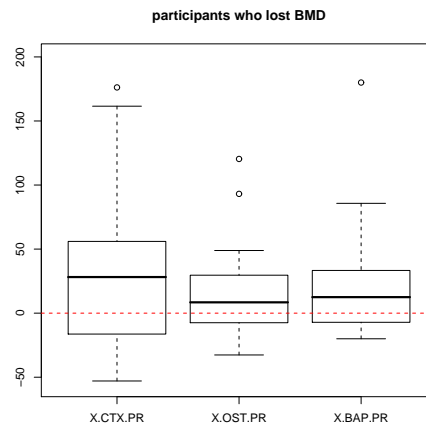


Figure 9.15.: Boxplots of the percental change of the respective bone markers in women who lost BMD during the course of the study

This result might motivate our further investigations of the bone turnover markers in Chapter 11. We especially focus on the factors influencing the bone resorption marker CTX. It will turn out that linear mixed models are adequate to explain the data. The model building process and verifying model assumptions will be explained in detail since these models are not as widespreadly known as standard linear models.

10. Standard linear models to explain the loss of bone density

In this chapter we investigate which factors determine the change of bone density ($X.BMD$) over the two years of the study. Clearly, for participants number 12 and 24, who are represented twice in the study, we only incorporate the observations corresponding to the first two-year period. The quantitative computed tomography to determine the bone density was performed only once for participants 18 and 19, thus $X.BMD$ is missing for these women. Furthermore, as already mentioned in the previous chapter, we do not incorporate participants number 6 and 47, since for these women there was not performed any luteal phase blood sample. Blood samples were either not performed or they were taken in the wrong time interval. Hence, there are in total data of **45** different perimenopausal women available for our statistical analysis.

For each participant measurements of the semiannual blood tests are summarized to one single value. To this end, several approaches for summarizing longitudinal data are possible. The most common and intuitive method is calculating the mean. But, for example, taking the maximal or the minimal value observed during the course of the study is also feasible in some cases. An approach recommended by Fitzmaurice et al. (2004, Section 3.6) is taking the area under the curve (AUC). We tried several ways to aggregate the repeated measurements to one value for each woman. At considering corresponding histograms, there were no many differences in the basic form of the distribution to see (see Figure B.5). Thus we apply the generally used summarizing method and take the mean of the serum concentrations.

In Section 10.2 we fit standard linear models to explain $X.BMD$. As has been said already in Chapter 1, standard linear models are matters of common knowledge and there is a plenty of literature on them. Therefore we focus on results and refer to Fahrmeir et al. (2007) and Myers (1986) to get a detailed introduction to the theory of standard linear models.

10.1. Exploratory data analysis

Before adjusting a concrete model for $X.BMD$ in the second section, we perform here a exhaustive graphical analysis of the data. To verify if the assumption of normality of the response variable $X.BMD$ can be justified, we consider a histogram and a boxplot of this variable to get an impression of its distribution (Figure 10.1). The distribution seems to be a little bit left-skewed but this might also be explained by the covariates. In total, there is no evidence that a standard linear model might be improper. Clearly, we have

to consider the standardized residuals after fitting the model to check whether concrete model assumptions are valid (see Section 10.2).

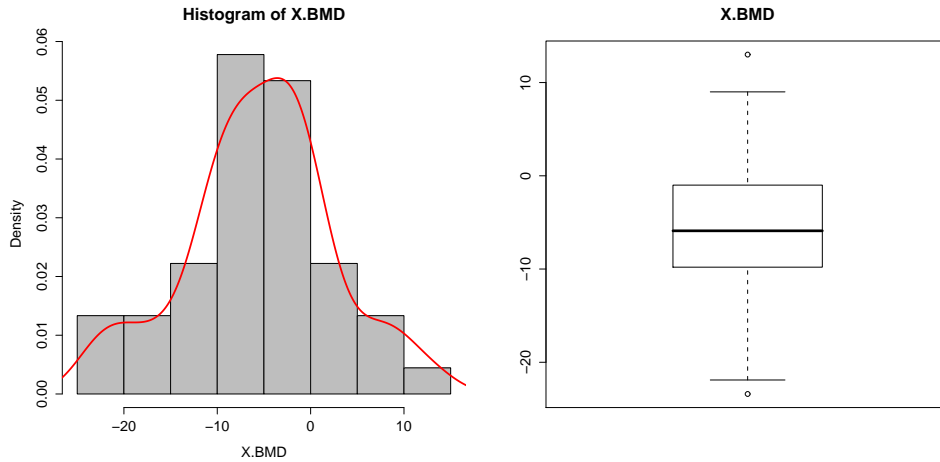


Figure 10.1.: Histogram and boxplot of the change of bone density

The basic graphical tool of the exploratory data analysis when fitting standard linear models are scatter plots of the response and the covariates of interest. These graphics suggest how strong variables are correlated and may indicate which of the covariates have a strong influence and which of them probably are not needed to explain the response. Therefore we consider in the following scatter plots of `X.BMD` and each single covariate. In order to better illustrate the arising correlations and to indicate possibly necessary transformations of variables, the respective linear regression lines (red and dashed) and LOESS lines are added in every single graphic.

Figure 10.2 shows scatter plots of `X.BMD` with respect to `AGE` and `QCT.1`. Both covariates seem to be negative correlated with `X.BMD` meaning that the older the women are and the higher their bone density at study onset, the more bone density they lose. The second fact is not surprisingly since we investigate here the absolute value of BMD loss and not the percental.

Figure 10.3 illustrates that there is a slightly positive effect of `BMI.25` on `X.BMD`. Furthermore it shows that the more the women gain in weight, the more they lose bone density. Data are concentrated to the right in the middle plot below due to the outlier we detected already in the previous chapter: participant number 9 reduced her BMI by -4 during the course of the study. Without this observation, `X.BMI.1.0` is ranged between -1.35 and 2 and does not seem to influence `X.BMD`.

Figure 10.4 suggests that there is a fairly high positive linear influence of the percentage of ovulatory cycles on `X.BMD` and vice versa for `anov.tot.s.PR`. It does not seem to be necessary to classify anovulatory cycles by normal and short length, only anovulatory long cycles might be investigated separately.

Figure 10.5 shows the relationship between `X.BMD` and the mean length of cycles and

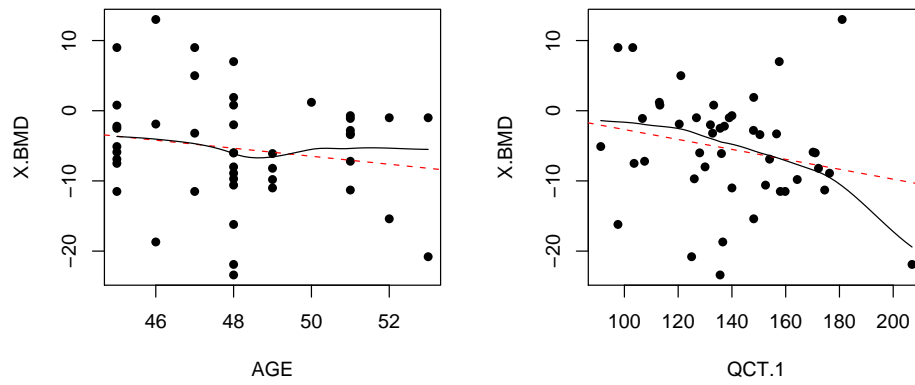


Figure 10.2.: Scatter plots of the change of bone density with respect to age (left) and the bone density at the begin of the study (right)

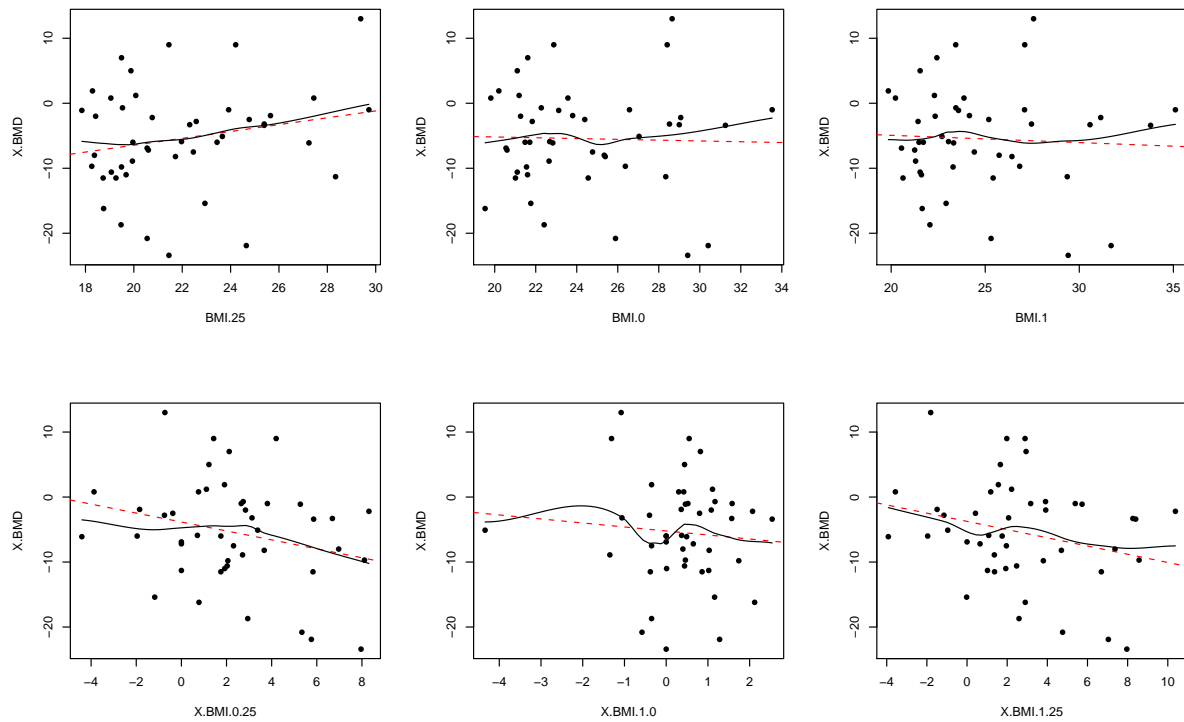


Figure 10.3.: Scatter plots of the change of bone density and the BMI values at different points of time as well as their changes over time

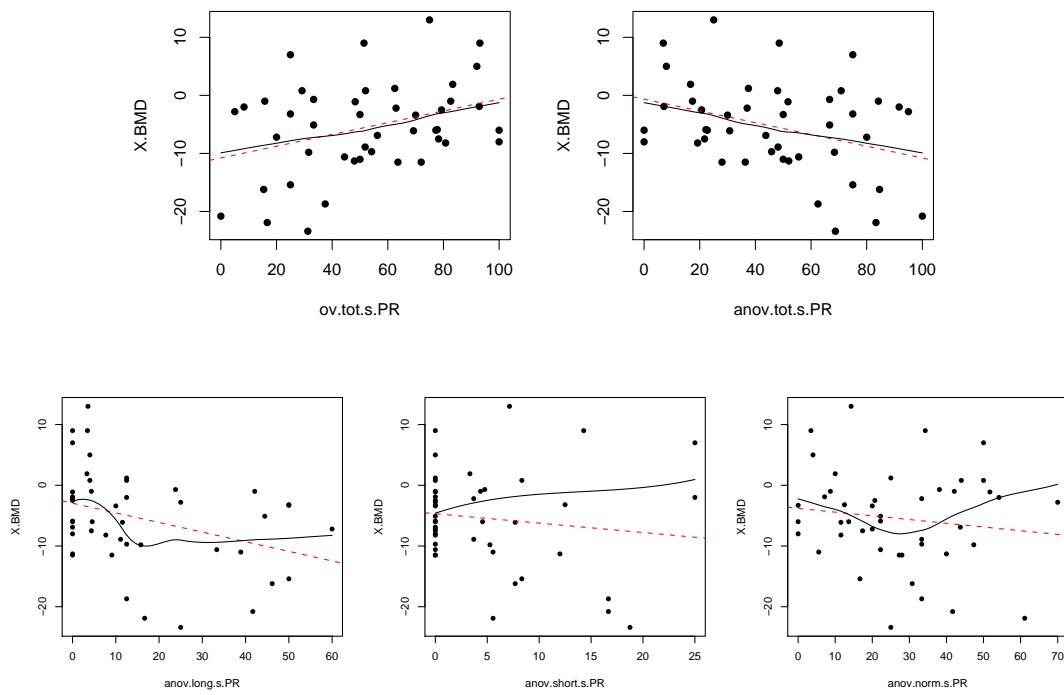


Figure 10.4.: Scatter plots of the change of bone density and the percentage of ovulatory and (total, long, normal and short) anovulatory cycles during the study

luteal phases. One may immediately note that `CYL.m` should be transformed since there are a few huge values which skew the distribution highly. This fact we noticed already at considering Figure 9.7. The quadratic form of the LOESS line suggests the following transforming of the mean length of cycles:

$$\text{CYL.m.tr} := \frac{1}{\text{CYL.m}^2} .$$

Indicated by the graphic in the middle, this transformation appears appropriate. It points out that there is a high, linear, positive correlation between `X.BMD` and `CYL.m.tr`. It has to be kept in mind that `CYL.m.tr` increases when `CYL.m` decreases. Thus the interpretation of this positive correlation is that longer cycles provoke a higher loss in bone density which is actually what we expected. At the same time, longer luteal phases cause a lower decrease of BMD. Figure 10.6 shows the correlation between the female sex hormones

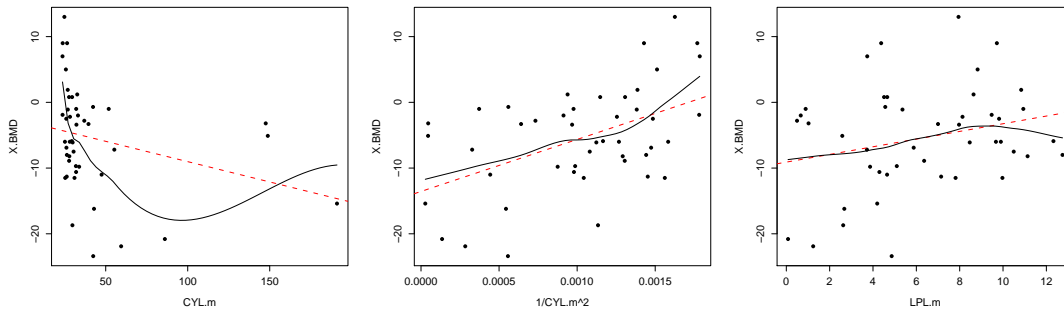


Figure 10.5.: Scatter plots of the change of bone density with respect to the mean length of cycles, it's transformation $\frac{1}{\text{CYL.m}^2}$ and the mean length of luteal phases

and `X.BMD`. `LH` and `FSH` are negative correlated with the change of BMD. As already seen in the corresponding histograms (Figure 9.9), these measurements are right-skewed, i.e. concentrated to the left part of their scale. Therefore we transform them by taking the logarithm. The corresponding scatter plots are shown by Figure 10.7. Additionally, scatter plots of the standardized residuals and the non-transformed and log-transformed covariates corresponding to a simple standard linear model to explain `X.BMD` by `LH.m`, $\log(\text{LH.m})$, `FSH.m` or $\log(\text{FSH.m})$, respectively, is shown by this graphic. It can be seen that, although the non-transformed data are not homogeneously distributed, variability does not increase. Thus, it is not absolutely necessary to consider their logarithm.

Furthermore, Figure 10.6 illustrates that the mean values of estradiol and progesterone have a positive effect on `X.BMD`. However, one may note that the variability within the scatter plot of `X.BMD` and `PROG.m` is not constant. It decreases for high mean values of progesterone. Therefore we should also investigate for progesterone, if a transformation might be preferable. To this end, Figure 10.8 shows a box-cox-transformation plot (see Myers (1986, Section 7.3)), generated by

```
> boxcox(PROG.m ~ X.BMD) .
```

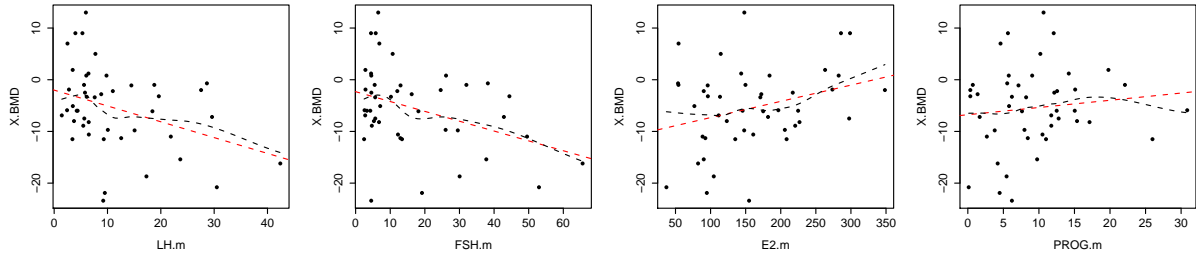


Figure 10.6.: Scatter plots of the mean female sex hormones and the change of bone density with linear regression line and LOESS lines

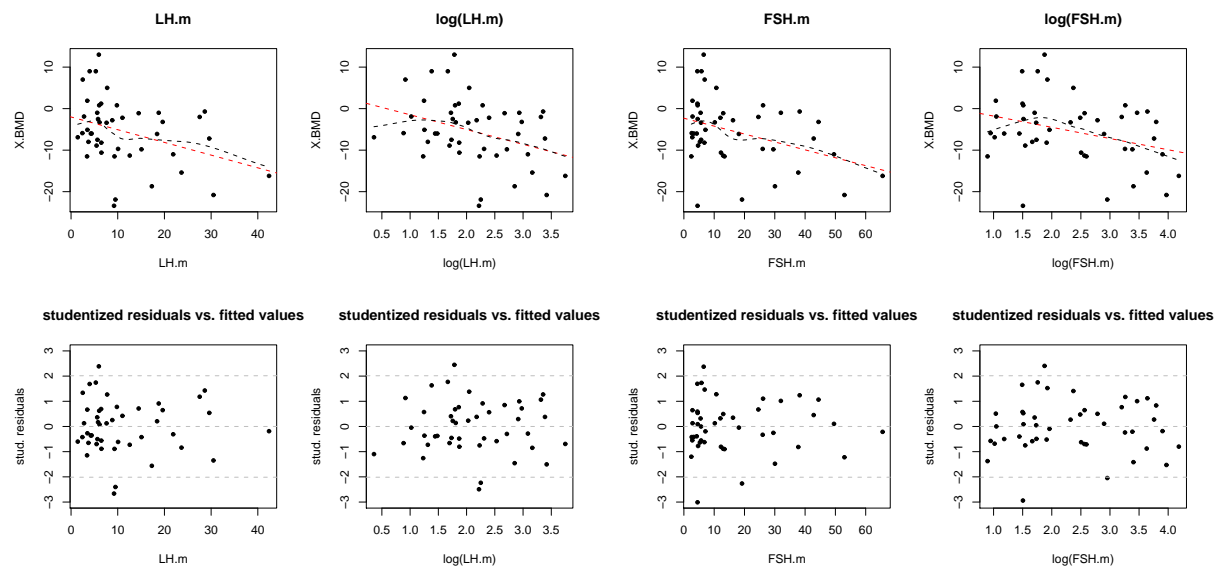


Figure 10.7.: Scatter plots of X.BMD with respect to LH.m and FSH.m as well as their logarithm (first row) and standardized residuals corresponding to simple standard linear models of X.BMD and the respective covariate (second row)

It indicates that the optimal transformation of progesterone is taking its square root (box-cox coefficient $\lambda = \frac{1}{2}$). Although variability is still heterogeneous, which is shown by the right plot of Figure 10.8, we will find out in the next chapter, that this transformation indeed is preferable (see Table 10.1).

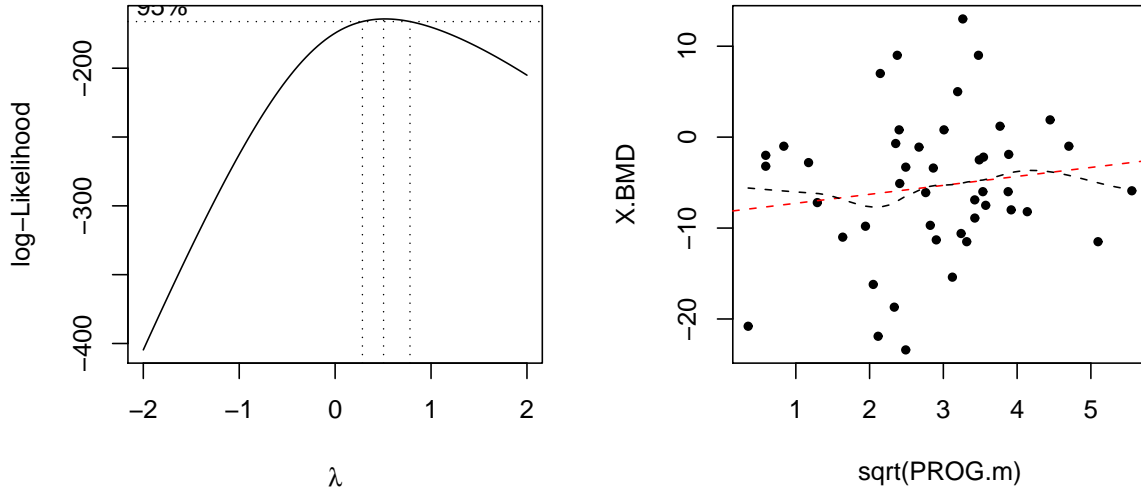


Figure 10.8.: Optimal transformation of progesterone

Finally, Figure 10.9 suggest no high influence of the mean values of the bone markers on the change of bone density. The LOESS line in the left plot might suggest a quadratic relationship, but by trying several transformations of `OST.m`, this trend is still be seen. Thus we let `OST.m` untransformed.

It is noticeable that comparing `X.BMD` to the change of the bone markers with respect to the end of the study and study onset, which was considered in the previous chapter (Figure 9.13), showed a stronger correlation than comparing `X.BMD` to the mean values. This can be explained by the fact that the bone markers are not monotonically decreasing or increasing during the course of the study, but can vary highly. Figure 11.1 of the next chapter illustrates this fact.

An investigation whether `X.BMD` should be transformed did not show any result. All box-cox values were given by 1 which does not suggest a transformation. The same was indicated by scatter plots of `X.BMD` and transformed versions, for example $\log(X.BMD)$, with respect to the single covariates. Thus, we do not transform the change of bone density.

Additionally, we investigate if some covariates have a significant influence on `X.BMD` in a manner that indicates if they are higher or lower than a certain so-called *cutoff* value. To this end we introduce dummy variables which assume only the values 1 ($\hat{=}$ "true", "yes") and 0 ($\hat{=}$ "false", "no"). For example, `QCT1low120` is coded by 1 if `QCT.1` is less than 120

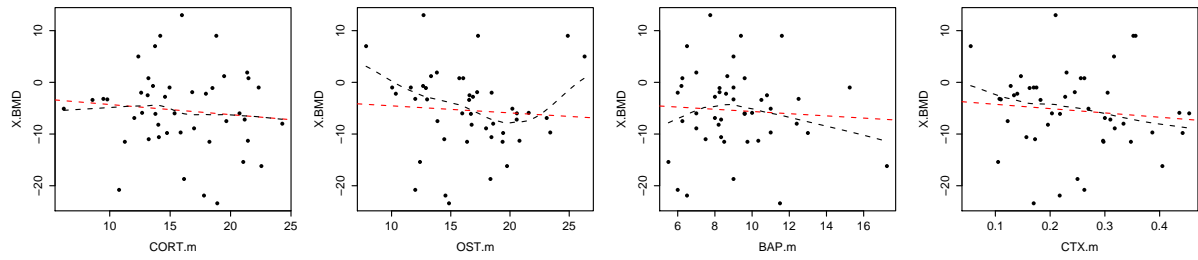


Figure 10.9.: Scatter plots of cortisol and the bone turnover markers and the change of bone density with linear regression line and LOESS lines

mgCa-Ha/ml and by 0 if `QCT.1` is greater than 120 mgCa-Ha/ml. `BMI25high25` is defined to be 1 if the respective participant is overweight (`BMI.25` greater than 25 kg/m^2) and 0 if the participant has normal weight. Figure 10.10 shows some boxplots of `X.BMD` per each of the two groups of the respective dummy variable. It is basically illustrated that bone density decreases on average less for participants which were overweight at 25 years of age, for women which had osteopenia at study onset and for women with low a low mean concentration of FSH.

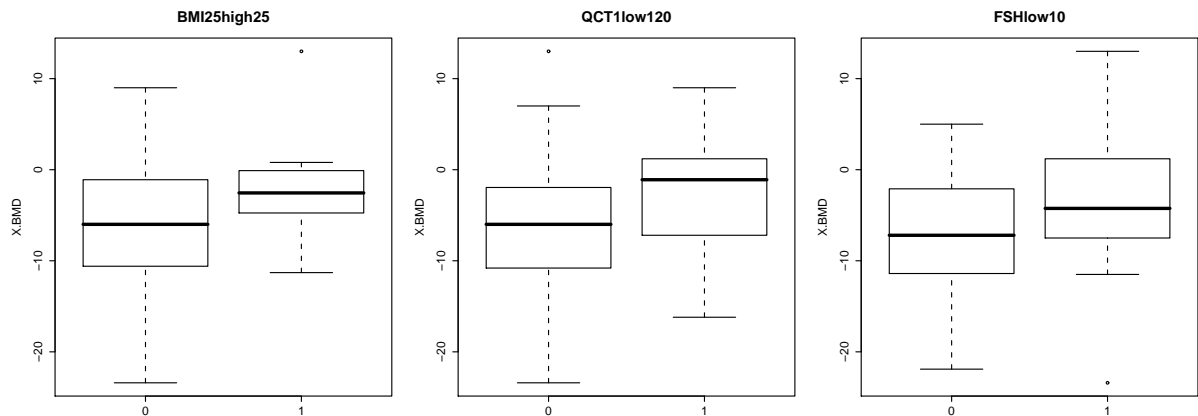


Figure 10.10.: Boxplot of dummies

10.2. Modeling the change of bone density (X.BMD)

Based on the scatter plots of the previous section, we got already an impression which covariates might have a high explanatory potential on the change of bone density. But since there are only little data available and covariates are highly correlated among each other, the model selection process turns out to be difficult. We suppose that models get fastly overfitted and by highly correlated covariates effects are reversed.

Therefore we started our inference by considering simple standard linear models. That is, by denoting each of the available covariates and also some dummy versions with \mathbf{x} we adjusted the following model:

$$\text{X.BMD} = \beta_0 + \beta_1 \mathbf{x} + \epsilon \quad \text{with} \quad \epsilon \sim N(\mathbf{0}, \sigma^2 I) .$$

Table 10.1 shows the coefficient of determination R^2 and the p-value corresponding to a t-test of $H_0 : \beta_1 = \mathbf{0}$ to verify the significance of the respective covariate for each of these simple standard linear models. Additionally, the empirical correlation coefficient between X.BMD and the respective variable is listed in the second column. P-values less than 0.05 indicating that the covariate is significant at level 5% are marked.

It is noticeable that the female sex hormones LH, FSH and estradiol are highly significant. LH and FSH have a strong negative effect on the change of bone density, i.e. if their mean value is high, the more bone density is lost. From the middle column we get that only incorporating an intercept and FSH.m, 15% of the variability in the data can be explained by the model. A similar statement can not be affirmed for progesterone nor for its transformed version. The simple standard linear model which has the greatest coefficient of determination is the model with slope of CYL.m.tr, i.e. the mean length of the cycles transformed by the squared inverse. This model is able to explain almost 25% of variation in the response data. Furthermore the percentage of ovulatory and anovulatory long cycles is highly significant as was suggested already by the scatter plots.

We continue our investigations by considering a saturated model for X.BMD, meaning to include the maximal number of covariates. To this end there are 17 covariates available. Obviously, we have to avoid complete linear dependences between the covariates. For example, BMI.0 is not included since it can be derived from BMI.25 and X.BMI.0.25. Among the "dependent" variables we select these variables with lowest p-values in Table 10.1. In this first multivariate linear model we do not incorporate the dummy versions either. The transformed mean cycle length CYL.m.tr, defined as $\frac{1}{\text{CYL.m}^2}$, assumes very low values which leads to a huge estimated regression parameter. Therefore we additionally standardize this covariate, i.e. subtract its mean and divide it by its standard deviation. The obtained standardized and transformed mean cycle length variable we declare by CYL.m.tr.st. We decide to take the square root transformation of progesterone based on the observations of Section 10 and due to its lower p-value in Table 10.1. This initial model we declared as mod.full:

```
> mod.full = lm(X.BMD ~ AGE + QCT.1 + BMI.25 + X.BMI.0.25 + X.BMI.1.25 +
+               LH.m + FSH.m + E2.m + sqrtPROG.m +
```

x	emp.cor	R^2	pvalue
AGE	-0.17	0.029	0.26
QCT.1	-0.22	0.050	0.14
BMI.25	0.21	0.045	0.16
BMI.0	-0.03	0.001	0.86
BMI.1	-0.05	0.003	0.73
X.BMI.0.25	-0.26	0.066	0.09
X.BMI.1.25	-0.26	0.069	0.08
X.BMI.1.0	-0.09	0.008	0.57
LH.m	-0.36	0.128	0.02
log(LH.m)	-0.35	0.124	0.02
FSH.m	-0.38	0.147	0.01
log(FSH.m)	-0.33	0.107	0.03
E2.m	0.30	0.091	0.04
PROG.m	0.12	0.014	0.44
sqrtPROG.m	0.14	0.020	0.35
CORT.m	-0.10	0.010	0.50
OST.m	-0.07	0.005	0.65
BAP.m	-0.07	0.004	0.67
CTX.m	-0.10	0.011	0.50
CYL.m	-0.27	0.073	0.07
CYL.m.tr	0.49	0.245	0.00
LPL.m	0.26	0.065	0.09
ov.tot.s.PR	0.35	0.122	0.02
anov.tot.s.PR	-0.35	0.122	0.02
anov.norm.s.PR	-0.14	0.018	0.38
anov.short.s.PR	-0.13	0.018	0.38
anov.long.s.PR	-0.35	0.126	0.02
Dummy versions			
BMI0high25	-0.08	0.006	0.61
BMI25high25	0.22	0.050	0.14
QCT1low150	0.10	0.011	0.49
QCT1low120	0.22	0.050	0.14
PROGhigh6	0.10	0.009	0.53
PROGhigh12	0.17	0.028	0.27
E2high200	0.21	0.043	0.17
E2high150	0.10	0.011	0.50
FSHlow10	0.31	0.098	0.04

Table 10.1.: Empirical correlation coefficients (emp.cor) between the covariate x and X.BMD and characteristics of corresponding simple standard linear models. The pvalue refers to a t-test to verify the significance of the slope.

```
+          CORT.m + CTX.m + OST.m + BAP.m +
+          ov.tot.s.PR + anov.long.s.PR +
+          CYL.m.tr.st + LPL.m, x=T, y=T)
```

It is able to explain almost 57% of the variability arising in the data, but it is highly overfitted. This is indicated by the fact that only two covariates are significant at level 10% with respect to a t-test. Furthermore, as initially mentioned, effects are reversed. For example, the effect of progesterone (`sqrtPROG.m`) is estimated to be negative, but the scatter plot (see Figure 10.8) and Table 10.1 indicate a positive effect.

```
> summary(mod.full)
```

```
Call: ...
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-10.949	-3.257	-0.538	3.845	11.951

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.0780	37.6173	0.37	0.711
AGE	0.1668	0.6372	0.26	0.795
QCT.1	-0.1015	0.0543	-1.87	0.072 .
BMI.25	0.2874	0.4789	0.60	0.553
X.BMI.0.25	-0.9529	1.3980	-0.68	0.501
X.BMI.1.25	0.5163	1.2780	0.40	0.689
LH.m	-0.2089	0.3470	-0.60	0.552
FSH.m	-0.0331	0.2228	-0.15	0.883
E2.m	0.0085	0.0187	0.45	0.654
sqrtPROG.m	-2.0825	2.4498	-0.85	0.403
CORT.m	-0.4005	0.2943	-1.36	0.185
CTX.m	11.5857	20.0296	0.58	0.568
OST.m	-0.7817	0.4556	-1.72	0.098 .
BAP.m	0.0385	0.6556	0.06	0.954
ov.tot.s.PR	0.0607	0.1421	0.43	0.672
anov.long.s.PR	0.1580	0.1454	1.09	0.287
CYL.m.tr.st	6.2521	2.5607	2.44	0.021 *
LPL.m	-0.1517	1.2939	-0.12	0.908

```
---
```

```
Residual standard error: 6.68 on 27 degrees of freedom
```

```
Multiple R-squared: 0.569, Adjusted R-squared: 0.298
```

```
F-statistic: 2.1 on 17 and 27 DF, p-value: 0.0412
```

We suppose that this occurs due to high correlations among the single covariates. Therefore we consider the corresponding variance inflation factors (see Myers (1986, Section 3.8)), which are indicators for arising multicollinearity.

```
> vif(mod.full)
```

AGE	QCT.1	BMI.25	X.BMI.0.25	X.BMI.1.25	LH.m	FSH.m	E2.m	sqrtPROG.m
2.26	1.86	2.31	17.18	17.63	10.30	12.60	2.02	7.92

CORT.m	CTX.m	OST.m	BAP.m	ov.tot.s.PR	anov.long.s.PR	CYL.m.tr.st	LPL.m
1.48	3.92	3.35	2.54	15.04	6.62	6.46	20.10

The *VIFs* of various covariates are very high (even greater than 10) which affirms our assumption of occurring multicollinearity.

Hence, we reduce this model in a further step and search for significant covariates which may explain **X.BMD** "best". Since we have 17 covariates, there are $2^{17} = 131.072$ sub-models. Furthermore, we also aim to investigate the dummy versions of covariates which have not be incorporated so far and interactions between the covariates. This makes the model selection process even more complex. Clearly, it can not be investigated all these possible models, therefore sequential procedures are usually applied for the model building process. To this end, several approaches are known, basically concerning forward, backward or stepwise selection with respect to t-tests, F-tests, the adjusted R^2_{ad} statistic, the Akaike information criterion or Mallows' C_p statistic. For a detailed description of variable selection procedures and the mentioned tests and statistics, the reader is referred to Myers (1986, Section 4.4).

We actually applied all the stated methods for our model building process. By using forward selection methods, the variable **CYL.m.tr.st** was always first included and by applying backward (or stepwise) selection procedures, the same variable were last eliminated from the model. Taking into account the results of Table 10.1, this fact is not surprising. It basically established one model, referred to as **mod.fin**:

$$\begin{aligned}
 \text{X.BMD} &= \beta_0 + \beta_1 \text{CYL.m.tr.st} + \beta_2 \text{QCT1low120} \\
 &\quad + \beta_3 \text{CTX.m} + \beta_4 \text{CORT.m} + \epsilon \\
 &\text{with } \epsilon \sim N(\mathbf{0}, \sigma^2 I) .
 \end{aligned} \tag{10.1}$$

Considering the corresponding summary table indicates that all covariates are significant with respect to a t-test at level 10%. This model is able to explain 42% of the variation arising in the data. The adjusted R^2 increases to 0.363 compared to **mod.full** which has a R^2_{ad} of 0.298. Since this statistic is appropriate for comparing models, we can conclude that **mod.fin** is preferred to **mod.full**. Indeed, this model has the highest R^2_{ad} of all the models we considered.

```
> summary(mod.fin)
```

Call:

```
lm(formula = X.BMD ~ CYL.m.tr.st + QCT1low120 + CTX.m + CORT.m)
```

Residuals:

Min	1Q	Median	3Q	Max
-12.809	-4.993	0.219	3.677	13.312

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.573	4.328	1.06	0.297
CYL.m.tr.st	4.854	1.003	4.84	2.0e-05 ***
QCT1low1201	6.336	2.424	2.61	0.013 *
CTX.m	-19.888	10.047	-1.98	0.055 .
CORT.m	-0.405	0.238	-1.71	0.096 .

Residual standard error: 6.37 on 40 degrees of freedom
Multiple R-squared: 0.421, Adjusted R-squared: 0.363
F-statistic: 7.26 on 4 and 40 DF, p-value: 0.000171

Again considering the *VIFs* corresponding to this model,

```
> vif(mod.fin)
CYL.m.tr.st    QCT1low120    CTX.m    CORT.m
      1.09          1.04        1.09      1.07
```

does not suggest tht among these variables multicollinearity arises.

To verify the model assumptions, Figure 10.11 shows a Q-Q plot of the standardized residuals as well as a scatter plot of the fitted values and the standardized residuals. Keeping in mind that there are only 46 measurements, the graphics do not suggestst a lack of fit. The standardized residuals are centered to zero and show constant variability. There are only four outliers, i.e. observations with standardized residuals greater in absolute value than the 97.5% quantile of the standard normal distribution.

Using this model, the change of bone density might be predicted as follows:

$$\widehat{X.BMD} = 4.57 + 4.85 \cdot \text{CYL.m.tr.st} + 6.34 \cdot \text{QCT1low120} - 19.89 \cdot \text{CTX.m} - 0.41 \cdot \text{CORT.m} \quad (10.2)$$

Interpreting the dummy variable *QCT1low120*, the formula is given by

- Participants with $BMD \leq 120$ mgCa-Ha/ml at study onset:

$$\widehat{X.BMD} = 10.91 + 4.85 \cdot \text{CYL.m.tr.st} - 19.89 \cdot \text{CTX.m} - 0.41 \cdot \text{CORT.m}$$

- Participants with $BMD > 120$ mgCa-Ha/ml at study onset:

$$\widehat{X.BMD} = 4.57 + 4.85 \cdot \text{CYL.m.tr.st} - 19.89 \cdot \text{CTX.m} - 0.41 \cdot \text{CORT.m}$$

To make the regression coefficients comparable, we standardize all continuous (non-dummy) covariates and refit *mod.fin* using these standardized covariables. The corresponding regression coefficients then are estimated as shown in the following Table:

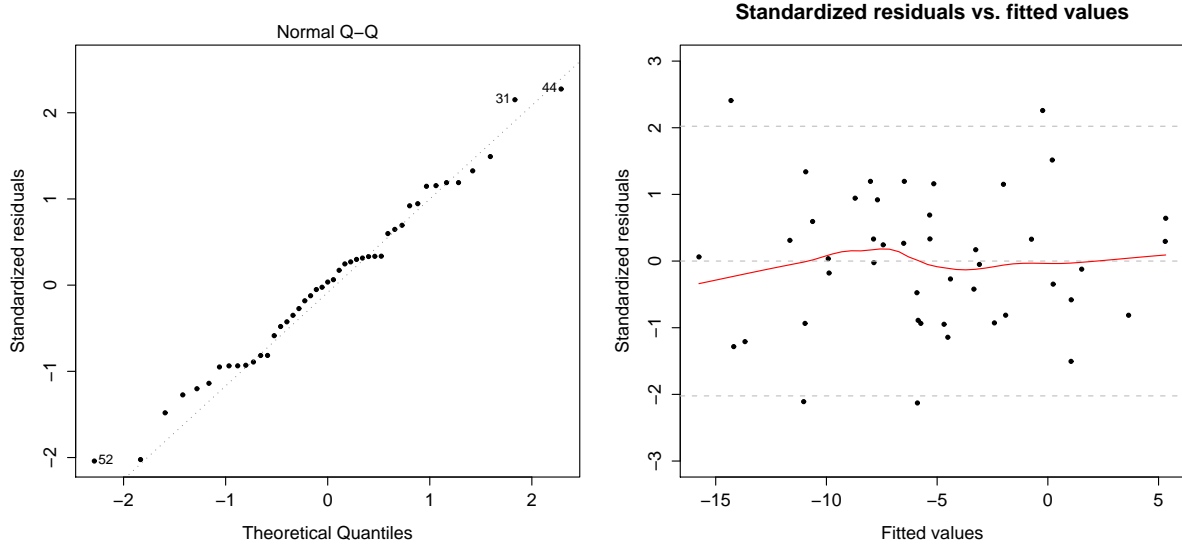


Figure 10.11.: Q-Q plot of the standardized residuals (left) and scatter plot of the fitted values and the standardized residuals (right) of `mod.fin` (10.1)

(Intercept)	CYL.m.tr.st	QCT1low1201	CTX.m.st	CORT.m.st
-6.71	4.85	6.34	-1.98	-1.69

Interpreting the suggested model, we first note that among the continuous covariates, the standardized and transformed mean length of cycles is most influential. This variable explains the majority of arising variability. Adjusting the model without including `CYL.m.tr.st` yields a R^2 of 0.08. The explanatory power of `QCT1low120`, `CTX.m.st` and `CORT.m.st` is therefore relatively low. The higher the mean length of cycles, the lower $\frac{1}{\text{CYL.mean}^2}$ and thus the higher the loss of bone density.

The bone resorption marker CTX and the stress hormone cortisol are have an relatively low influence on the change of bone density. An increase of these variables provokes a higher loss of BMD. Hence, stressed women are exposed to a higher risk for getting osteoporosis.

If women had already osteopenia at study onset, they loose relatively less BMD. This fact is not very surprising since we considered the absolute value of BMD change and not the percental.

11. Linear mixed models to explain the bone markers (Longitudinal analysis)

Goal in this chapter is to describe the bone turnover markers characterizing the bone metabolism by an appropriate model. The central question in this chapter is therefore which variables influence the serum concentration of CTX, osteocalcin (OST) and BAP. Since several measurements of the bone markers were made on each individual we fit linear mixed models which were described in the first part of this thesis. To recapitulate, these models account for occurring correlations among the repeated measurements and individual behavior by incorporating, besides constant, fixed effects, also random effects.

11.1. Exploratory data analysis

For the statistical analysis in this chapter there are **170** complete data of **46** different participants available. Remember that participants 12 and 24 are represented twice in the study (in total observed over 4 years). Nevertheless, for participant 12 there are all in all only 5 and for participant 24 only 4 complete data available. The other blood tests can not be classified as luteal phase blood tests (i.e. were taken in the wrong time interval) or some data are completely missing. The following table shows the number of participants per number of complete measurements.

Number of measurements	1	2	3	4	5
Number of participants	2	6	9	16	13

As has been already mentioned, the exact time over which the participants were observed differs from participant to participant. For most of them it is approximately two years, but for participant 12 and 24 it is 4 years. Therefore and for reasons to facilitate prediction we consider in the following only the time-invariant covariates which refer to study onset, i.e. AGE, BMI.25, BMI.0, X.BMI.0.25 and omit the time-invariant variables which refer to study end (for example BMI.1 and QCT.2).

As usual, we start our inferences with a detailed graphical analysis of the data to get to know them well and to suggest first hypotheses for concrete model building later on. We basically refer here to Pinheiro and Bates (2000) but techniques are also well described in Verbeke and Molenberghs (2009) and concrete examples for an exploratory data analysis in R are also presented in West et al. (2007) and Fox (2002). Clearly, we give here, however,

only an overview of our complete performed exploratory data analysis by presenting the most meaningful graphics. We basically focus on the bone resorption marker **CTX** on which we also concentrate by describing the concrete model building process (Section 11.3).

Since the data are longitudinal it suggests itself that we investigate at first if there is any time trend in the data. Therefore we consider the individual trajectories of the bone markers. To have a higher informative value we consider only those 13 participants with 5 available measurements. We define this subset of participants by the variable **obs**. Among them is participant 12 who has complete data corresponding to blood sample number 4 to 8. The corresponding graphic is shown by Figure 11.1. It can not be seen that the bone markers are monotonically decreasing or increasing during the course of the study. Instead, some of the data show a high variation between the single measurements. This indicates that there is probably no influence of time. It is interesting to note that our results of Chapter 9 show, that the change of the bone markers with respect to the end of the study and the study onset are correlated with the change of bone density (**X.BMD**) (see Figure 9.15). Moreover it can be seen that especially the bone markers **CTX** and **OST** show a similar behavior over time, which suggests that they might be correlated.

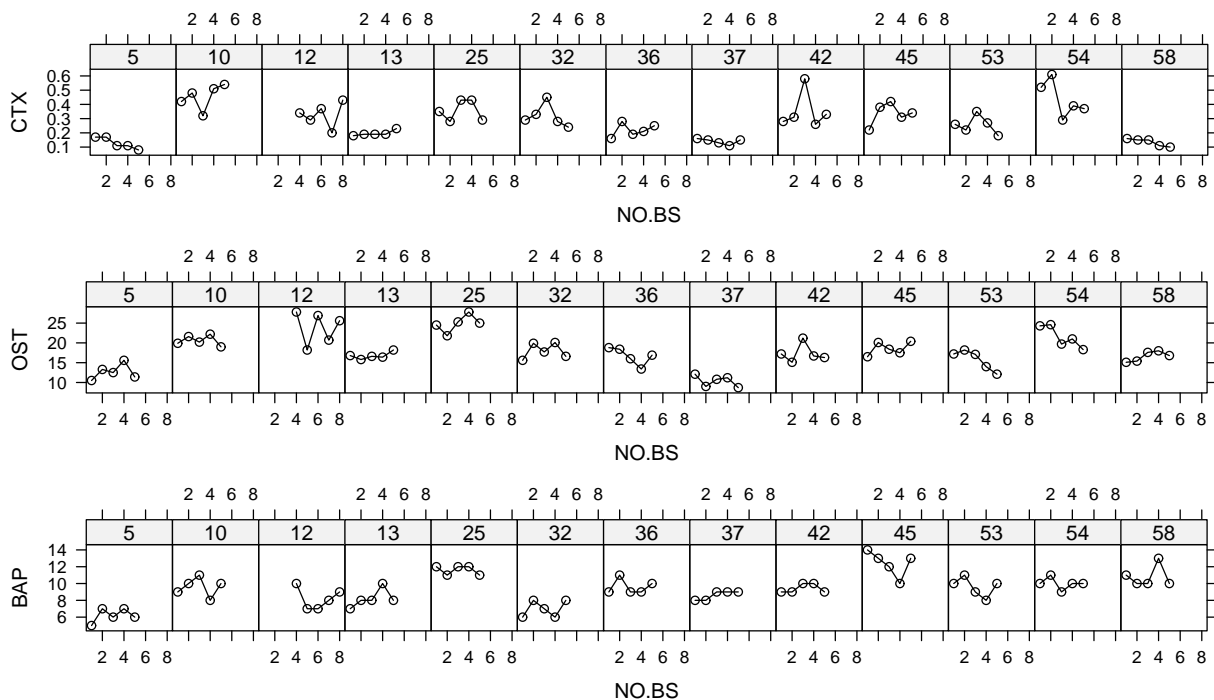


Figure 11.1.: Individual trajectories of the bone markers for the 13 participants with 5 complete measurements

The individual trajectories shown in Figure 11.1 can be generated in R by plotting a grouped data object which is provided by the package **nlme**. The first plot, for instance, was produced by the following code.

```
> group.CTX = groupedData( CTX ~ NO.BS | PAT.ID, order.groups=F )
> plot( group.CTX, layout=c(13,1), aspect=1, subset=obs, grid=F, as.table=T)
```

Additionally it can be observed in Figure 11.1 that the individual mean profiles of the markers vary from participant to participant. This impression may be reinforced by considering individual boxplots of the bone markers which is shown in Figure 11.2. Although the single boxes only consist out of 5 observations and may therefore be not very meaningful, it can be observed that the individual mean concentrations, indicated by the bold, black lines, highly differ between the participants. This might be a first indication that a random intercept is needed to explain the individual deviations of the overall population level. The latter one is represented by the red, dashed line. In order to confirm this assumption we have to show that these differences in the individual average levels can not be explained by different levels of certain covariates.

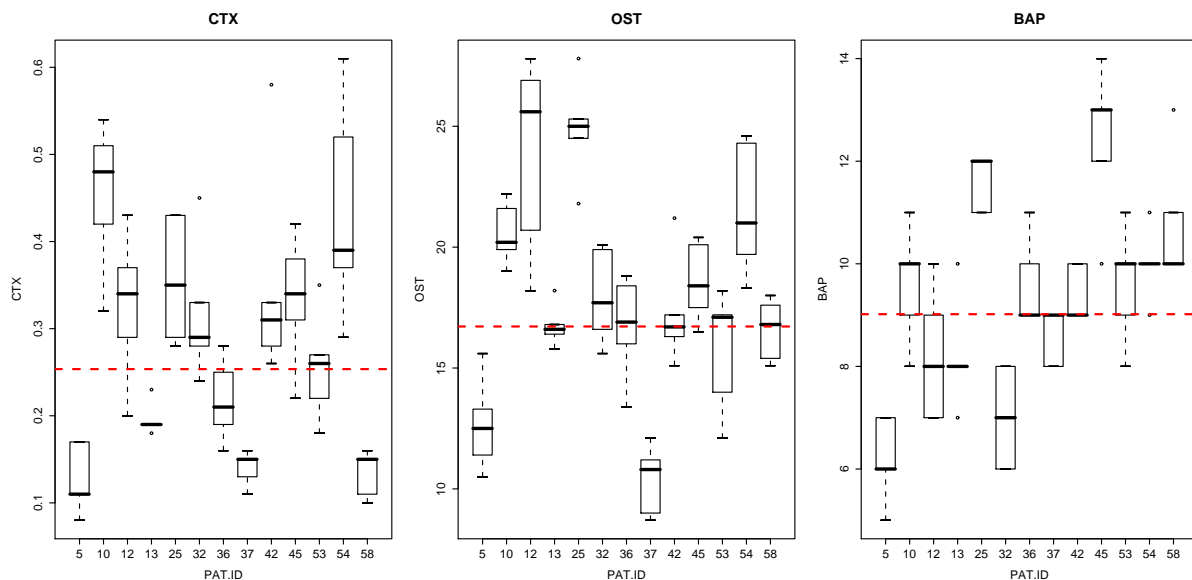


Figure 11.2.: Individual boxplots of the bone markers for the 13 participants with 5 complete measurements

The standard graphical tool of the exploratory data analysis when fitting standard linear models are scatter plots of the response and the covariates of interest (see Section 10.1). However, when analyzing longitudinal data, overall scatter plots mostly are not very meaningful. Although a certain explanatory variable has a significant influence on the response, this correlation has not necessarily to be seen in the overall scatter plot. The reason for this is that variability in the data can be explained by individual-specific deviations of a certain significant global population effect. Consider Figure 2.1 (p. 5) for an appropriate illustration of this fact.

Figure 11.3 shows an overall scatterplot of the bone markers and the time-invariant covariate **AGE**. Although we will detect later of this chapter, by applying statistical hypothesis

tests, that there is a significant effect of **AGE** on each of the bone markers, the scatterplots do not suggest a *high* correlation between the variables. The right graphic does not even indicate that there is any correlation between **BAP** and **AGE**.

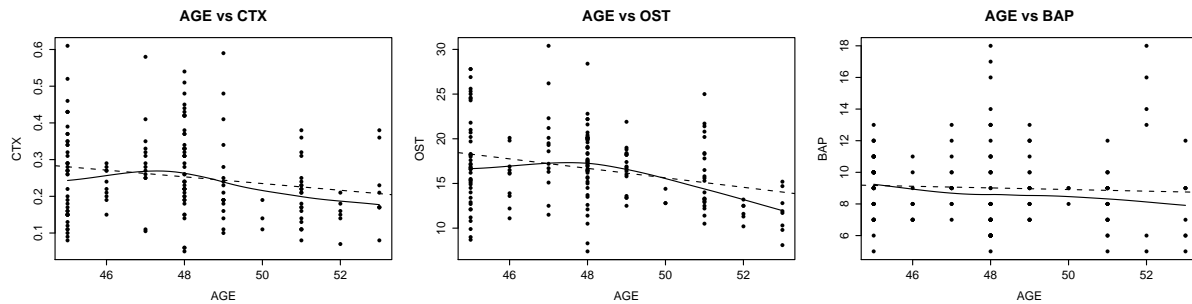


Figure 11.3.: Scatter plots of the bone markers and **AGE** with regression and LOESS lines

Figure 11.4 illustrates these observations as well by showing scatter plots of the bone resorption marker **CTX** with respect to the time-varying covariates **FSH**, **ov.tot.PR** and **CORT**. It will turn out in Section 11.3 that **FSH** and **ov.tot.PR** have a significant fixed effect on **CTX**, whereas the latter one not. One definitely might not suppose these results by only considering the corresponding scatter plots.

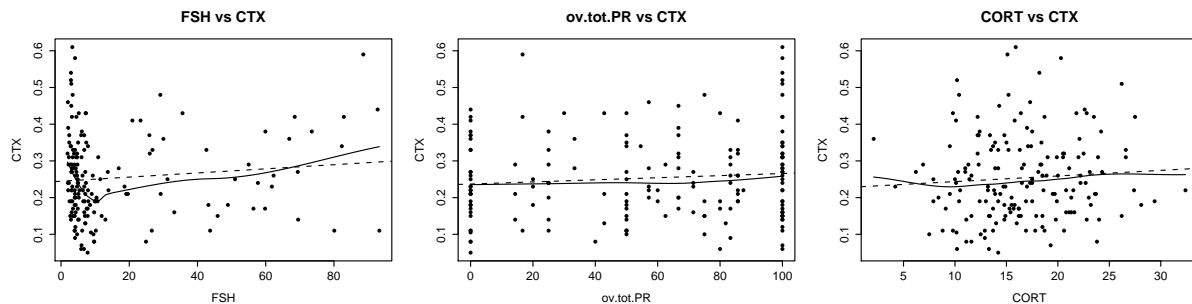


Figure 11.4.: Scatter plots of the bone resorption marker **CTX** with respect to **FSH**, **ov.tot.PR** and **CORT** with regression and LOESS lines

Therefore, *individual* scatter plots have to be considered in the exploratory longitudinal data analysis. Central questions in this case are if there is a certain global relationship between the explanatory variable and the response which can be seen in each single graphic or if there are differences between the individual intercept and slope values. If the form of the individual scatter plots differs from individual to individual, then this might indicate that there are individual-specific effects needed in the model.

Unfortunately, in our data set are even in the optimal case only 5 measurements per participant available. Consequently individual plots, as already seen at the individual

boxplots of Figure 11.2, are kind of poor and it turns out to be quite difficult to draw conclusions from the graphical analysis of the data. Looking at individual scatter plots of the bone markers and every single covariate for the participants with 5 data available shows that individual regression lines highly differ from participant to participant. This might indicate the need of a random slope of every covariate. Therefore we focus on statistical hypothesis testing (confirmatory data analysis) for the choice of random effects.

To illustrate this difficulty, individual scatter plots of the bone resorption marker **CTX** with respect to **FSH**, $\log(\text{E2})$, $\log(\text{PROG})$ and **CORT** are shown in Figure 11.5. **FSH** will turn out in Section 11.3 to have a significant fixed and random effect on **CTX**, $\log(\text{E2})$ will be detected to have a significant random effect, $\log(\text{PROG})$ has a significant fixed effect and the stress hormone **CORT** has neither a significant fixed nor a significant random effect on **CTX**. As described above, intercept and slope of the individual regression lines vary highly among the participants. The individual regression lines of $\log(\text{PROG})$ and **CORT** do not differ that much in their slopes than these of **FSH** and $\log(\text{E2})$, but in total there can not be seen *high* differences in the general nature of these individual scatter plots.

The single graphics of Figure 11.5 were generated by the function `xyplot` of the library `lattice`. For example, to generate the first graphic showing scatter plots of **CTX** and **FSH** for each participant,

```
> xyplot(CTX ~ FSH | as.factor(PAT.ID), data = DATA, subset = obs, as.table=T,
+        scales = list(ylim=range(CTX)), layout = c(7,2),
+        panel = function(x,y) {panel.xyplot(x,y,pch=16)
+                                panel.lmline(x,y,lty=2)
+                                panel.grid(h=3,v=3) } )
```

was applied.

Fitting an individual standard linear model for each participant can be performed by the function `lmList` of the `nlme` package. For example,

```
> lmlistfsh = lmList( CTX ~ FSH | PAT.ID, data=DATA, subset=obs)
> plot( intervals(lmlistfsh))
```

adjusts simple standard linear models to explain the bone resorption marker **CTX** by the hormone **FSH** for the participants with 5 available measurements (`subset=obs`). An approach for comparing these individual models which can be more illustrative than looking at the individual scatter plots is provided by plotting confidence intervals for the single model parameters. This is shown by the left graphic of Figure 11.6. The right graphic shows confidence intervals for intercept and slope for individual simple linear models with covariate $\log(\text{PROG})$. In both graphics we may note that some intervals for the intercept values highly differ which might indicate the need of a random intercept in the model. However, confidence intervals for the slope values overlap to a great extend which does not suggest the need of a random slope in the respective covariate. Nevertheless, by applying asymptotic likelihood ratio tests we will find out that a random slope of **FSH** is significant (see Section 11.3). Consequently, the explanatory power of these graphics again is not very high, simply due to the fact that there are not enough data available.

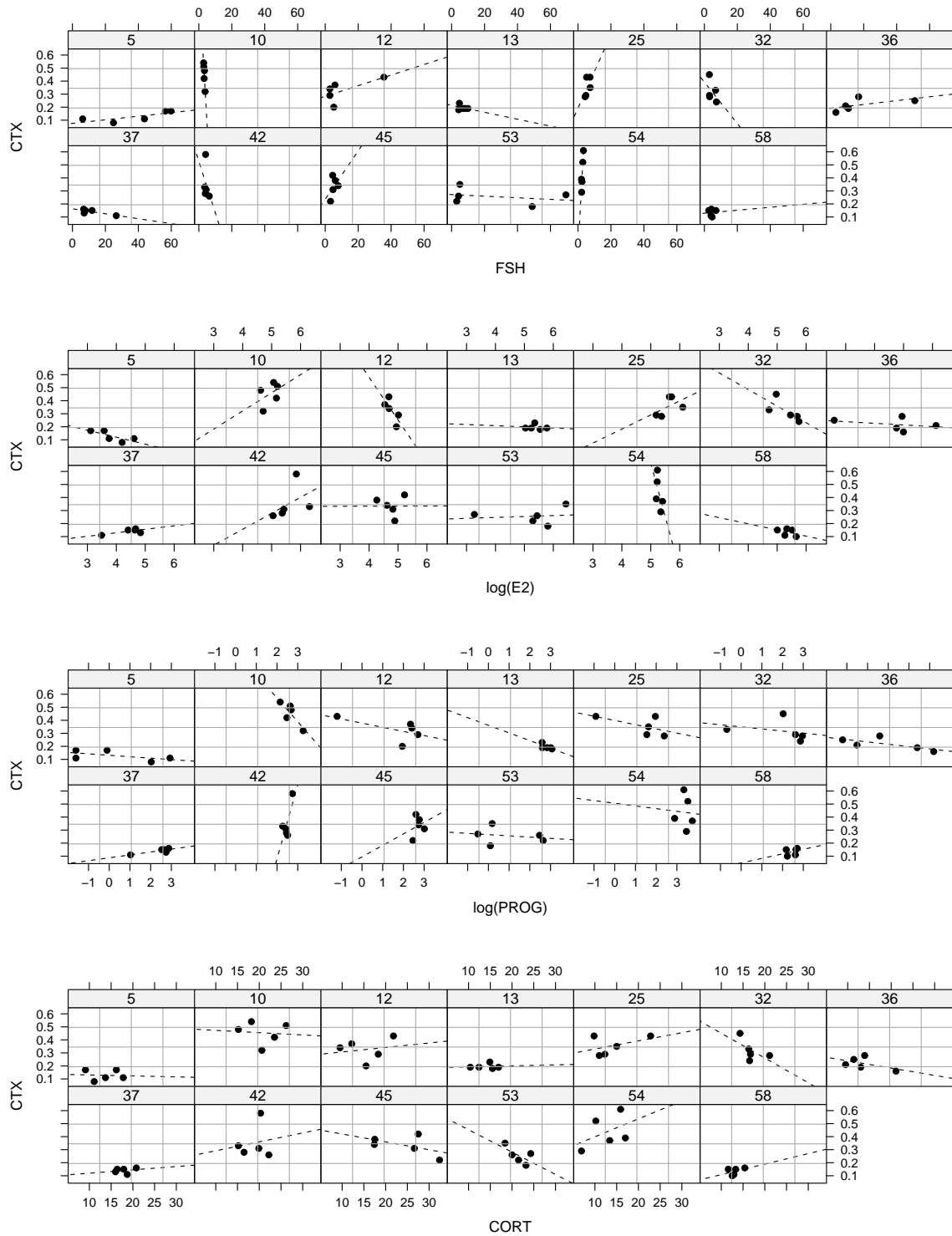


Figure 11.5.: Individual scatter plots and regression lines of the bone resorption marker CTX with respect to FSH, $\log(E2)$, $\log(PROG)$ and CORT for the 13 participants with 5 complete measurements

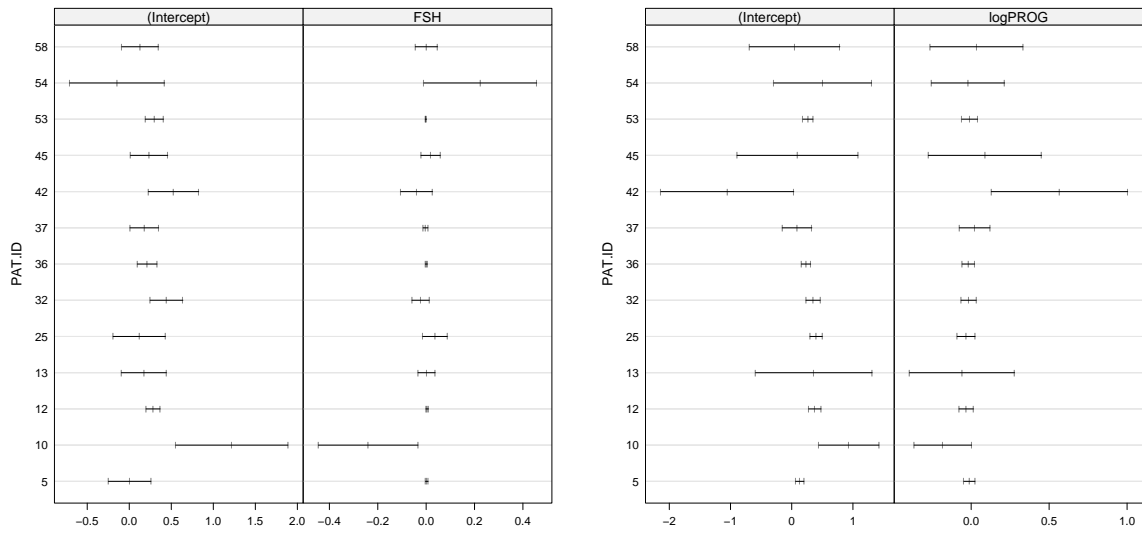


Figure 11.6.: 95% confidence interval for intercept and slope in individual simple standard linear models with response CTX and covariates FSH and log(PROG)

Similar as in the previous chapter, we furthermore investigate if some dummy versions of covariates might influence the bone markers. Figure 11.7 shows some boxplots of the bone marker concentrations per each of the two groups of the respective dummy variable. It can be seen that the mean profile of CTX is much higher when the the bone density measured at the beginning of the study is higher than 150 mgCa-Ha/ml. Serum concentrations of osteocalcin tend to be a little bit higher if the concentration of estradiol is higher than 200 pg/ml. BAP is somewhat higher when progesterone is lower than 6 ng/ml (indicating an anovulatory cycle) and shows significant higher values when the respective participant is overweight.

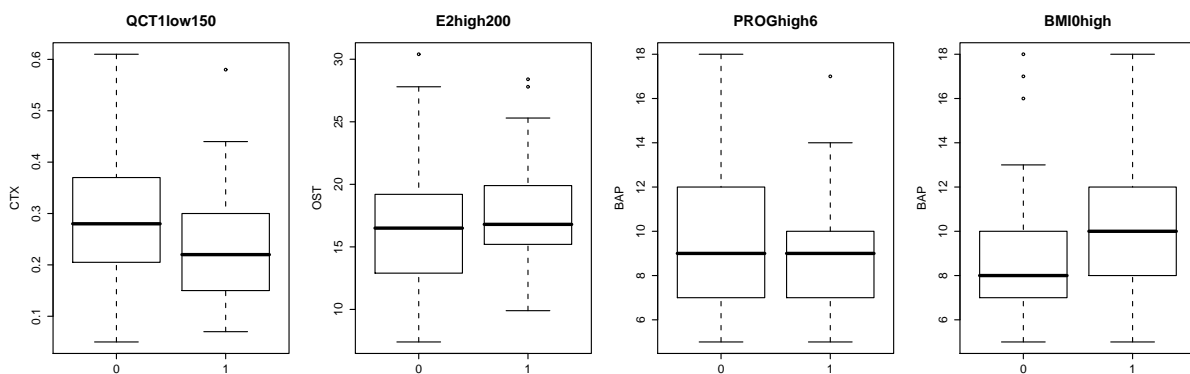


Figure 11.7.: Boxplots of bone markers per group

The following table shows the corresponding number of measurements in the respective

groups:

	0	1
QCT1low150	55	115
E2high200	123	47
PROGhigh6	60	110
BMI0high	117	53

Since there are all in all 170 measurements available, data mostly are grouped by the dummy variables approximately into 1/3 and 2/3 parts.

By introducing these dummy variables it can also be looked for any interactions among the covariates. Considering Figure 11.8 shows different effects of FSH on CTX when the serum concentration of E2 is higher or lower than a cutoff value of 150 pg/ml. Clearly, this might be due to the fact, that FSH and E2 are highly negative correlated which explains the high values of FSH (red) in the left graphic.

One also could suspect that the outlying observation (red) in the right graphic pulls down the regression line and that therefore an interaction might be not significant. However, omitting this measurement does not change the basic form and the single regression lines for different E2 levels do still differ. This is why we investigate this interaction further in Section 11.3.

Since only few observations for each participant are available, it is obvious to investigate if there can be formed groups of participants which show a similar behavior of the bone markers with respect to certain covariates. This means to check if random effects should be included rather for groups of participants than for every single participant. To this end, we may analyze for example the similarities of the participants with increasing, constant or decreasing regression lines in the individual scatter plots (Figure 11.5). Unfortunately these investigations did not show any results.

To illustrate an example, Figure 11.9 shows the relationship between the bone resorption marker CTX and FSH for different groups of participants. In the first scatter plot participants were aggregated by their age. Though the single regression lines differ in their slopes this is mostly due to one single outlier (for example at age 47) or because of only few observations in one class (for example at age 50). The general distribution of the measurements does not show high differences. Instead, all single plots suggest a slight positive correlation of FSH and CTX which might indicate a fixed effect of FSH. In the second to 5th graphic of Figure 11.9 participants are equally partitioned in groups of weight (BMI.0), initial level of the bone density (QCT.1), initial level of FSH (defined by FSH.start) and the average level of estradiol (defined by E2.m). Note that FSH values are higher when E2.m is low is due to the fact that FSH and E2 are negative correlated. Since neither these graphics nor concrete investigations by statistical hypotheses indicate the need of random effects on the level of grouped participants we concentrate in the following only on including individual-specific effects (random effects for each participant).

Since the study is longitudinal it may also be considered, besides the initial investigation of a present time trend, if there is any random effect of time. That is we investigate if

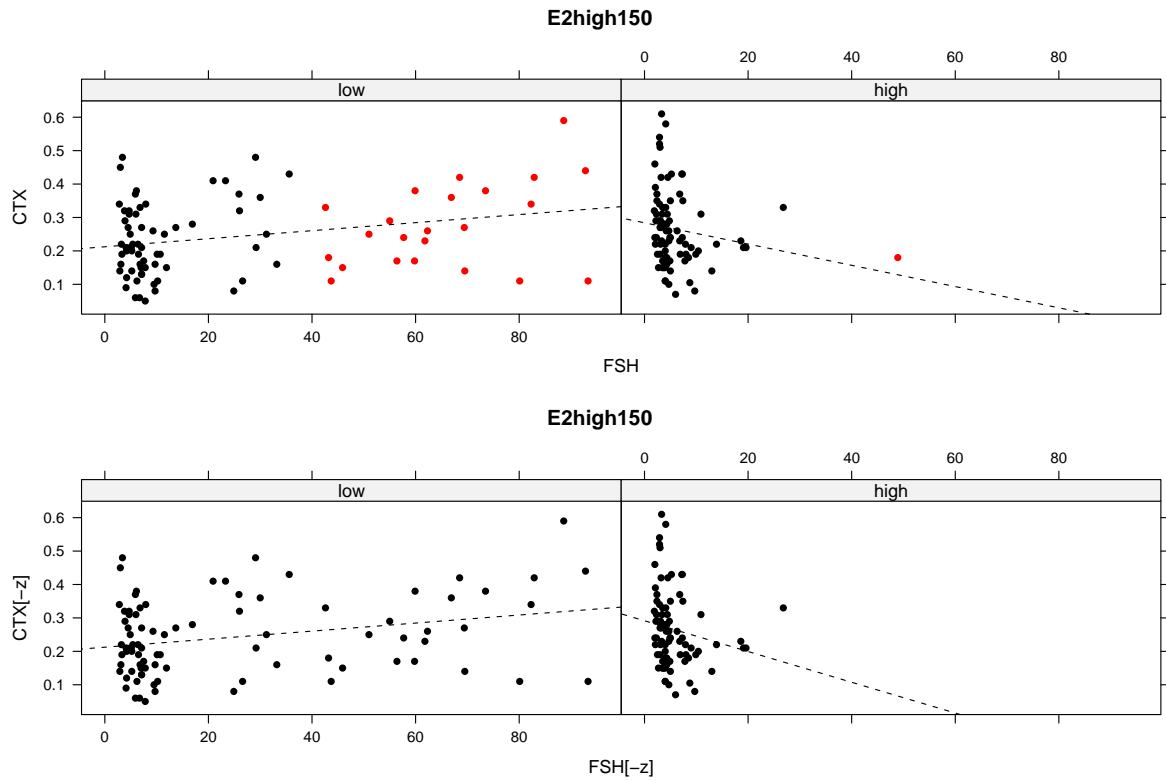


Figure 11.8.: Scatter plot of CTX and FSH partitioned by the level of E2. The red values in the upper graphic indicate FSH concentrations greater than 40 IU/l. In the lower graphic, the outlying value of high FSH and high E2 was removed.

measurements made at different points in time show different or similar behavior. To this end we generated scatter plots of the bone markers and the covariates grouped by the corresponding number of blood sample. Figure 11.10 shows one example for CTX and FSH. Neither there can be seen different behavior of the measurements taken at different times nor such an effect of time would be well interpretable. This is why we do not assume that there is any influence of time on the bone markers.

11. Linear mixed models to explain the bone markers (Longitudinal analysis)

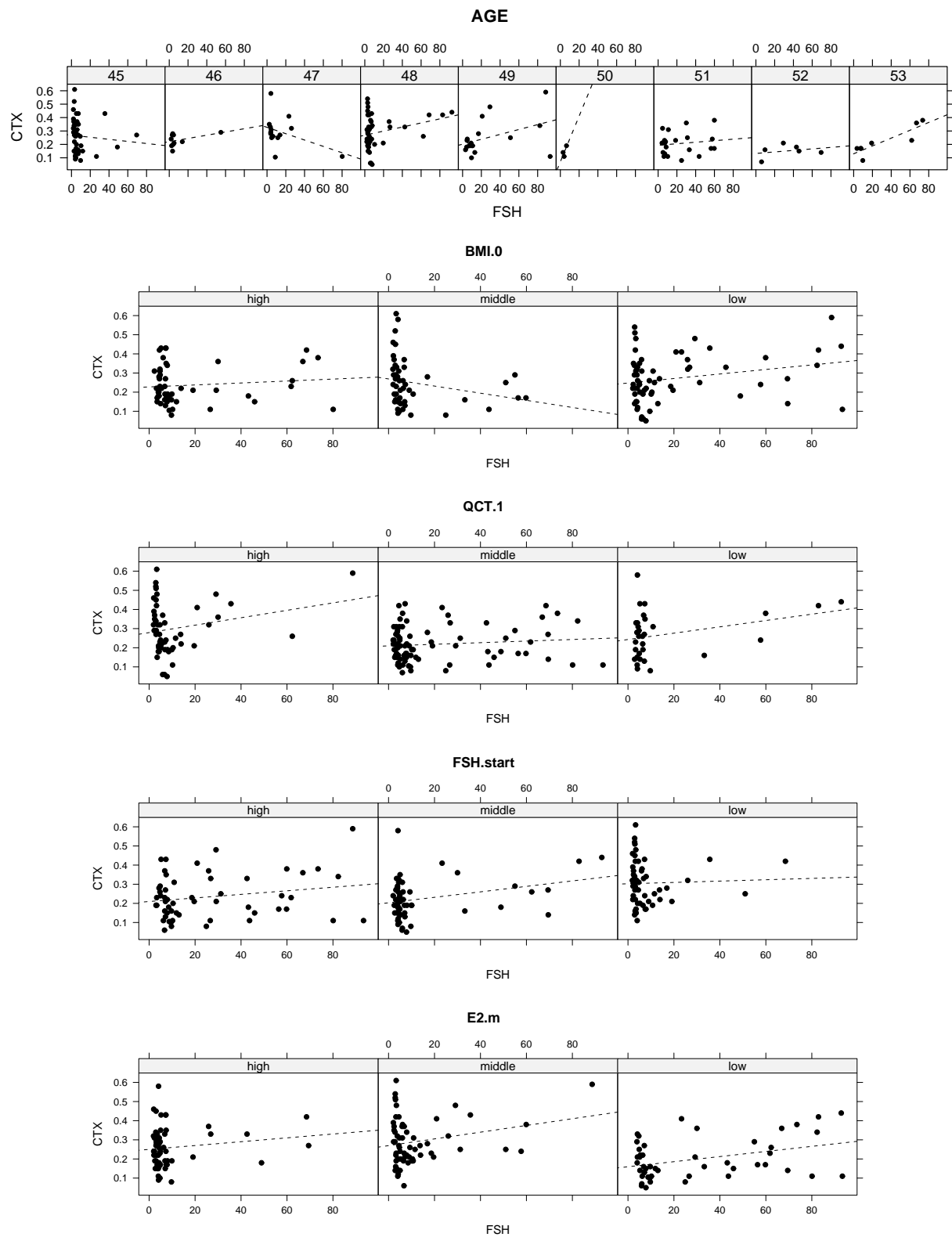


Figure 11.9.: Scatter plots of CTX and FSH with respect to different groups

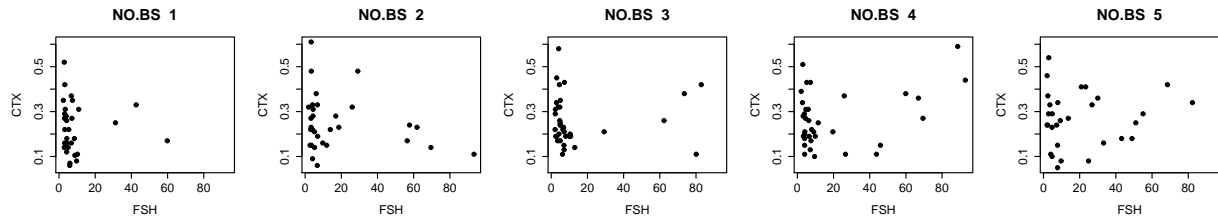


Figure 11.10.: Effect of time

11.2. Model building

To find an appropriate linear mixed model for each of the three bone turnover markers we proceed by applying the *Top-Down Strategy* introduced in Section 7.1. That is, we start with fitting a standard linear model which incorporates the maximum number of fixed effects, then we focus on the choice of an appropriate structure of the random effects before we finally eliminate non-significant fixed effects. Note that since there are at the most only 5 measurements available for each of the 46 women, it is not very meaningful to investigate if there is any specific structure of the residuals. Models would turn out to be too complex for this small data set.

As for modeling the change of bone density $X.BMD$ in Chapter 10, we consider for $CYL.BS$ and $CYL.mean$ their transformed versions

$$CYL.BS.tr := \frac{1}{CYL.BS^2} \quad \text{and} \quad CYL.mean.tr := \frac{1}{CYL.mean^2}.$$

To recapitulate, these variables indicate the length of the cycle at which the blood sample was taken and the mean length of cycles since the last blood test for each woman, respectively. Figure 11.11 shows exemplarily for $CYL.BS$ that data corresponding to the transformed variables are distributed more homogeneously.

Considering scatter plots of every single covariate does not clearly suggest that further transformations of variables might be necessary. However, by applying statistical hypotheses in the following sections, we indeed will see that taking the logarithm of some of the female sex hormones yields better results. A transformation of the response variables (the bone marks) did not turn out to be necessary.

Moreover, we consider in the following model building process the standardized version for every continuous, non-dummy covariate. That is, we subtract the mean and divide this difference by the variable's standard deviation. These standardized covariates are labeled by ".st". For example, instead of FSH we incorporate in the following $FSH.st$, defined as

$$FSH.st = \frac{FSH - \text{mean}(FSH)}{\text{sd}(FSH)}. \quad (11.1)$$

The reason for this is on the one hand that estimated fixed effects of the covariates are directly comparable, meaning that the higher the absolute value of the estimated slope of

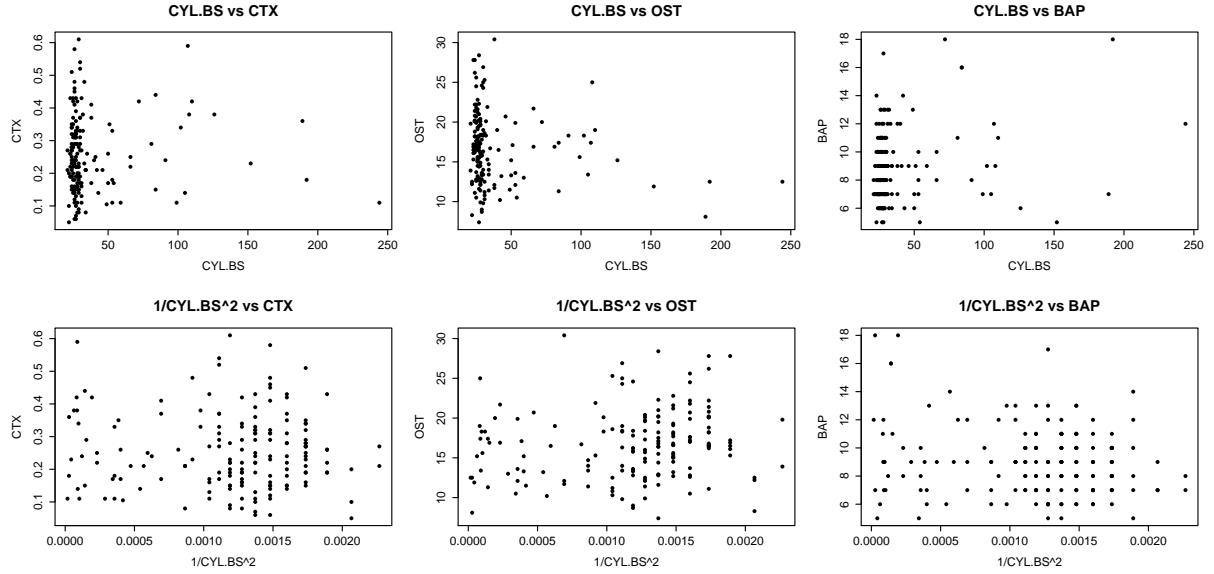


Figure 11.11.: Scatter plots of the bone markers and CYL.BS (first row) and its transformed version (second row), respectively

a covariate, the higher its influence on the response. On the other hand, calculations of estimates are often numerically more stable.

As already mentioned in the previous section, we incorporate only these time-invariant covariates which refer to the begin of the study. Furthermore, we do not incorporate the length of the luteal phase of the cycle at which the blood test was performed (LPL.BS). This is due to the fact that there are too many missing values (NA entries), since all cycles where progesterone was measured to be greater than 6ng/ml were categorized as ovulatory, even if the cycle monitor did not show a ovulation symbol. Clearly, for these cycles the corresponding length of the luteal phase is not available.

Hence, by avoiding complete dependencies among the variables, the pool of covariates in the respective saturated fixed effects model, from which we start our statistical analysis in the following sections, can be summarized as shown by Table 11.1.

time-invariant coavariates	time-varying covariates	
AGE.st BMI.25.st BMI.0.st QCT.1.st	LH.st FSH.st E2.st PROG.st CORT.st	CY.BS CYL.BS.tr.st CYL.mean.tr.st LPL.mean.st ov.tot.PR.st anov.long.PR.st

Table 11.1.: Overview of covariates in the saturated fixed effects model

11.3. Modeling the bone resorption marker CTX

We start our inferences by explaining the bone resorption marker CTX. As mentioned in the previous section, to find an appropriate model we proceed by applying the *Top-Down Strategy* introduced in Section 7.1. To this end we begin with fitting a standard linear model which incorporates the maximum number of fixed effects to explain adequately the population-specific variability in the data. This model is called the *saturated fixed effects model* or *model with loaded mean structure*. According to the notations which were introduced in Part I of this thesis, it can be briefly formulated as

$$\text{CTX}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \epsilon_{it} \quad \text{with} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I),$$

where CTX_{it} represents the t^{th} measurement of CTX of participant i , $i = 1, \dots, 46$. \mathbf{x}_{it}^T symbolizes the corresponding covariate vector consisting out of the 15 variables listed in Table 11.1 and a 1-entry corresponding to an intercept.

As seen in Chapter 10.2, fitting a standard linear model in R is provided by the function `lm`, thus we declare

```
> modfix = lm( CTX ~ AGE.st + ... + anov.long.PR.st, , x=T, y=T) .
```

”...” stands for the remaining covariates of Table 11.1. The corresponding coefficient of determination R^2 is calculated to 0.179, meaning that almost 18% of the variability in the data can be explained by the model. The adjusted R_{ad}^2 statistic is 0.099 and the residual standard error σ is estimated to 0.109.

The random intercept model

The fitted standard linear model `modfix`, however, probably is not appropriate for explaining the data. Due to the repeated measurements of the same individual, data are correlated. Therefore the assumption of independent response variables is not satisfied. This is why we now focus on including individual-specific, random effects to this initial model to incorporate the occurring correlations. First of all, we include a random intercept, γ_{i0} , to describe individual deviations of the mean population level:

For participant $i = 1, \dots, 46$ and blood sample number t :

$$\text{CTX}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \gamma_{i0} + \epsilon_{it} \quad \text{with} \quad \gamma_{i0} \sim N(0, d_0^2) \quad \text{and} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I). \quad (11.2)$$

This random intercept model can be adjusted in R by the function `lme` of the `nlme` package. The `lme` function as well as the other functions applied in the following are described in Chapter 8.

```
> library(nlme)
> mod.ri = lme( CTX ~ AGE.st + ... + anov.long.PR.st,
               random = ~1 | PAT.ID, data=DATA) .
```

To get a first impression of the fitted model we consider the corresponding summary table:

```
> summary(mod.ri)

Linear mixed-effects model fit by REML
Data: DATA
   AIC      BIC    logLik
  -204    -150     120

Random effects:
Formula: ~1 | PAT.ID
      (Intercept)      Residual
StdDev:    0.0903      0.0696

Fixed effects:
              Value Std.Error  DF t-value p-value
(Intercept)   0.260    0.022 113   11.85  0.000
AGE.st        -0.032    0.015  41    -2.06  0.046
BMI.25.st      0.000    0.019  41     0.01  0.992
BMI.0.st       -0.019    0.018  41    -1.07  0.289
QCT.1.st       0.012    0.015  41     0.82  0.419
LH.st          0.008    0.017 113     0.49  0.626
FSH.st         0.016    0.018 113     0.88  0.383
E2.st         -0.003    0.008 113    -0.42  0.672
PROG.st        -0.007    0.011 113    -0.65  0.517
CORT.st         0.002    0.007 113     0.28  0.780
CY.BS1         -0.017    0.025 113    -0.69  0.491
CYL.BS.tr.st   -0.007    0.012 113    -0.59  0.555
CYL.mean.tr.st -0.014    0.013 113    -1.08  0.281
LPL.mean.st    -0.012    0.030 113    -0.39  0.696
ov.tot.PR.st   0.025    0.032 113     0.80  0.427
anov.long.PR.st -0.017    0.012 113    -1.44  0.152
...
Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.8926 -0.4447 -0.0946  0.4786  3.5112

Number of Observations: 170
Number of Groups: 46
```

From the second paragraph we get that the standard deviation d_0 of the random intercept is estimated to 0.0903 and the estimated residual standard deviation $\hat{\sigma}$ is 0.0696. Hence, $\hat{\sigma}$ of the random intercept model is less than the the estimated residual standard deviation of modfix and variation also is explained by γ_{i0} . This might indicate already the significance of the random intecept.

Further we note that only a few of the incorporated fixed effects are significant at level 10% with respect to an asymptotic t-test (see Section 6.2). This indicates that we won't need all the corresponding covariates in the model which we will test concretely after choosing an appropriate structure of random effects.

To check whether the random intercept is significant ($H_0 : d_0^2 = 0$), we perform an asymptotic likelihood ratio test as presented in Subsection 6.1.3. From this subsection we

know that the corresponding RLRT statistic satisfies under H_0 :

$$\text{RLRT} \xrightarrow{d} 0.5\mathcal{X}_1^2 + 0.5\mathcal{X}_0^2.$$

Referring to Subsection 8.3.1 we may apply the `anova` function to calculate the corresponding RLRT statistic. The first argument of the function is the full model, i.e. the random intercept model `mod.ri`, the second argument is the nested model, that is the model given that H_0 is true. The latter one is the standard linear model `modfix` which we fitted initially. Note that the default setting in `lme` is that the variance components vector θ is estimated using REML estimation (see Section 8.1). Hence, the given test statistic `L.Ratio` indicates the value of the corresponding RLRT statistic.

```
> (anova.ri = anova(mod.ri, modfix))
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
mod.ri	1	18	-204	-149.6	120.1			
modfix	2	17	-141	-89.1	87.4	1 vs 2	65.5	<.0001

Remember that the returned p-value is calculated with respect to $\text{RLRT} \xrightarrow{d} \mathcal{X}_1^2$ and is therefore overestimated. Manual calculation of the p-value with respect to the correct asymptotic distribution, i.e.

$$0.5 \cdot \mathbb{P}(\mathcal{X}_1^2 > \text{L.Ratio}),$$

yields obviously an even smaller p-value:

```
> 0.5 * (1-pchisq( anova.ri$L.Ratio[2], 1))
[1] 0 .
```

Consequently H_0 is rejected which indicates that the random intercept is needed.

Since the random intercept model is a simple linear mixed model (only one single random effect is included) it can be applied also an exact likelihood ratio test as presented in Subsection 6.1.4. Under $H_0 : d_0^2 = 0$, the finite sample distribution of the RLRT statistic is given by (6.8). Simulations of this distribution are implemented in the R package `RLRsim` and a likelihood ratio test based on these simulations can be performed by the function `exactRLRT` (see Subsection 8.3.1).

```
> library(RLRsim)

> exactRLRT(mod.ri)
      simulated finite sample distribution of RLRT.
      (p-value based on 10000 simulated values)
data:
RLRT = 65.5, p-value < 2.2e-16
```

This test also provides a very small p-value, confirming that H_0 can be rejected and the random intercept is highly significant.

Inclusion of a random slope

Still concentrating on the ideal structure of random effects, we furthermore investigate if additionally a random slope is needed in the model. We check this by using again likelihood ratio tests as presented in Subsection 6.1.2 and 6.1.3. In the following we illustrate the applied procedure exemplarily to verify if a participant-specific slope of `logE2.st` is significant. Certainly, the presented approach was applied to various covariates. The corresponding results are summarized in Table 11.2.

Thus, we consider in the following a linear mixed model incorporating besides a random intercept, γ_{i0} , also a random slope, γ_{i1} , of `logE2.st`. Similar as the saturated fixed effects model and the random intercept model, it may be briefly formulated as follows:

For participant $i = 1, \dots, 46$ and blood sample number t :

$$\text{CTX}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \gamma_{i0} + \gamma_{i1} \text{logE2.st} + \epsilon_{it}$$

with $\begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N \left(\mathbf{0}, \begin{pmatrix} d_0^2 & d_{01} \\ d_{01} & d_1^2 \end{pmatrix} \right)$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I)$.

As before, \mathbf{x}_{it}^T symbolizes the corresponding covariate vector consisting out of the 15 variables listed in Table 11.1 and a 1-entry corresponding to an intercept. It can be fitted in R by an update of the random intercept model:

```
> mod.rirs.logE2 = update(mod.ri, random = ~ logE2.st | PAT.ID) .
```

From the respective part of the summary table we get the estimates of the corresponding standard deviations of γ_{i0} , γ_{i1} and ϵ_{it} as well as the estimated correlation $\frac{\hat{d}_{01}}{\hat{d}_0 \hat{d}_1}$ between the random effects:

```
> summary(mod.rirs.logE2)
...
Random effects:
Formula: ~logE2.st | PAT.ID
Structure: General positive-definite, Log-Cholesky parametrization
          StdDev   Corr
(Intercept) 0.0874 (Intr)
logE2.st    0.0217 -0.505
Residual    0.0673
...
```

Again, the estimated residual standard error $\hat{\sigma}$ was slightly reduced as well as the estimated standard deviation of the random intercept. In exchange, a small part of the variation is now explained by the random slope.

To investigate if the random slope of `logE2.st` is needed in the model we perform an asymptotic likelihood ratio test for testing random effects. From Subsection 6.1.3 we know that under

$$H_0 : d_1^2 = d_{01} = 0$$

the RLRT statistic satisfies

$$\text{RLRT} \xrightarrow{d} 0.5\chi_2^2 + 0.5\chi_1^2.$$

Since the `anova` function returns a p-value which is based on $\text{RLRT} \xrightarrow{d} \chi_2^2$, we only extract the calculated test statistic and calculate the corresponding p-value manually, similarly as before:

```
> LR.logE2 = anova(mod.rirs.logE2, mod.ri)$L.Ratio[2]
> 0.5 * (1 - pchisq(LR.logE2,2)) + 0.5 * (1 - pchisq(LR.logE2,1))
[1] 0.0734
```

However, this p-value refers to the asymptotic distribution of the test statistic. Thus it should be verified if the sample is large enough that approximating the distribution of the RLRT statistic by its asymptotic distribution is appropriate and the calculated p-value is trustable. To this end, the function `simulate.lme` (described in Subsection 8.3.1) provides to simulate the distribution of the RLRT statistic under H_0 .

```
> data.sim.E2 = simulate.lme( mod.ri, mod.rirs.logE2, nsim=1000, seed=2,
+                             method="REML" )
> plot(data.sim.E2, df=c(1,2), between=list(x=c(0.5,0.5)),
+ panel = function(x,y){panel.abline(0,1,lty=2); panel.loess(x,y,lwd=2)})
```

By plotting the generated `simulate.lme` object, the empirical distribution of the RLRT statistic is compared to a χ^2 distribution with "df" degrees of freedom (Figure 11.12). Concretely, in the left graphic, the empirical p-values are drawn against the p-values of a χ_1^2 distribution, in the middle plot against the asymptotic distribution, $0.5\chi_2^2 + 0.5\chi_1^2$, and in the right plot against a χ_2^2 distribution. In the middle plot, the empirical and nominal p-values do nearly lie on the bisecting line. This indicates that the asymptotic distribution is appropriate and the calculated p-value of 7.3% is trustable.

Hence we may reject H_0 at level 10% and conclude that the random slope of `logE2.st` is necessary.

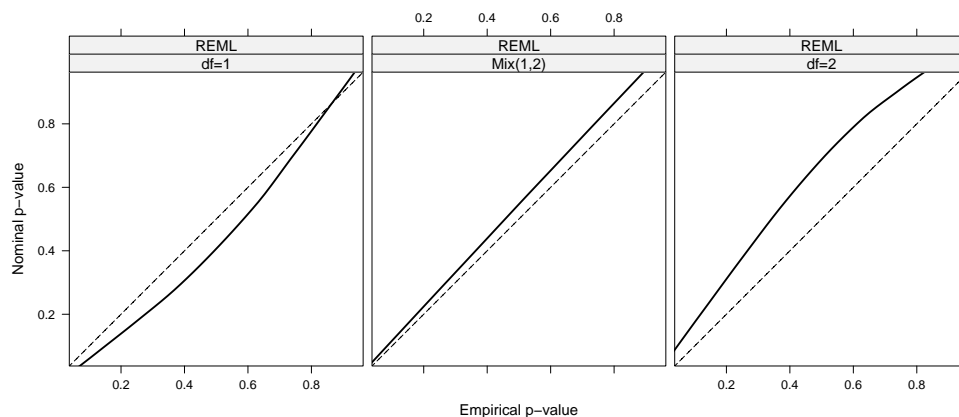


Figure 11.12.: Simulated distribution of the RLRT statistic compared to χ_1^2 , $0.5\chi_2^2 + 0.5\chi_1^2$ and χ_2^2 distributions. The dashed lines indicate the bisecting lines.

We may proceed by verifying if the fitted correlation between the random intercept and the random slope, which was estimated to be -50.5%, is significant or if the random effects are uncorrelated. Hence, the null hypothesis is $H_0 : d_{01} = 0$. For this test an asymptotic likelihood ratio test as presented in Subsection 6.1.2 can be applied. Under H_0 it holds that $\text{RLRT} \xrightarrow{d} \chi_1^2$. The 95% quantile of this distribution is 5.99. The test can be performed in R directly by the `anova` function. To this end, however, we first have to fit the nested model. Given that H_0 is true, the covariance matrix of the random effects is a diagonal matrix. This can be adjusted by the argument `pdDiag`.

```
> mod.rirs.logE2.diag = update(mod.ri,
+                               random = list(PAT.ID = pdDiag(~logE2.st)))

> anova(mod.rirs.logE2, mod.rirs.logE2.diag)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
mod.rirs.logE2	1	20	-205	-144	122			
mod.rirs.logE2.diag	2	19	-206	-148	122	1 vs 2	1.13	0.288

The corresponding p-value is calculated to 28.8%. Thus, the null hypothesis can not be rejected (at level 10%) and `mod.rirs.logE2.diag` is therefore preferred. This model can be briefly formulated as follows: For participant $i = 1, \dots, 46$ and blood sample number t :

$$\text{CTX}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \gamma_{i0} + \gamma_{i1} \log \text{E2.st} + \epsilon_{it}$$

with $\begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} d_0^2 & 0 \\ 0 & d_1^2 \end{pmatrix}\right)$ and $\epsilon \sim N(\mathbf{0}, \sigma^2 I)$

Usually, when fitting linear mixed models, the initial explorative data analysis gives an idea which covariates probably have a significant random effect. Unfortunately this has not been the case in our investigations. Therefore we applied this procedure presented for `logE2.st` to all covariates. Since we saw in our previous data descriptions (Chapter 9 and 10) that the distribution of the female sex hormones is not very homogeneous, we transformed them by their logarithm and incorporated also these transformed variables in our investigations (`logLH.st`, `logFSH.st`, `logE2.st` and `logPROG.st`).

More precisely, according to the approach which was just presented for `logE2.st`, we proceeded with all covariates in the following order:

- (1) inclusion of the respective covariate in the random intercept model (11.2),
- (2) asymptotic likelihood ratio test whether the respective random slope is significant,
- (3) asymptotic likelihood ratio test whether the random intercept and the respective random slope are correlated.

Table 11.2 provides an overview of results of this procedure applied to every covariate:

to (1) It shows for each of these models the AIC , the estimated residual standard error, $\hat{\sigma}$, the estimated standard deviation of the random intercept, \hat{d}_0 , the estimated standard deviation of the random slope, \hat{d}_1 , and the estimated correlation between the random effects ("cor.ri.rs").

to (2) The corresponding test statistic and the corresponding p-value are given in the columns "L.Ratio" and "pvalue".

to (3) The corresponding p-value is given in the column "pvalue.cor".

	AIC	$\hat{\sigma}$	\hat{d}_0	\hat{d}_1	cor.ri.rs	pvalue.cor	L.Ratio	pvalue
AGE.st	-202	0.0696	0.0896	0.0165	-0.9426	0.2455	1.3486	0.3775
QCT.1.st	-201	0.0696	0.0753	0.0590	0.4933	0.2848	1.2010	0.4108
BMI.25.st	-200	0.0696	0.0898	0.0080	-0.7267	0.7181	0.1303	0.8275
BMI.0.st	-200	0.0696	0.0902	0.0038	-0.0776	0.9590	0.0026	0.9788
X.BMI.0.25.st	-200	0.0696	0.0890	0.0164	-0.1018	0.9300	0.0121	0.9531
LH.st	-202	0.0688	0.0913	0.0110	0.9634	0.2089	1.5865	0.3301
logLH.st	-201	0.0690	0.0910	0.0095	0.7949	0.3693	0.8512	0.5048
FSH.st	-205	0.0679	0.0908	0.0170	0.9654	0.1105	4.5662	0.0773
logFSH.st	-204	0.0678	0.0915	0.0172	0.9077	0.1502	3.8282	0.0989
E2.st	-203	0.0675	0.0875	0.0213	-0.4623	0.2683	2.6229	0.1874
logE2.st	-205	0.0673	0.0874	0.0217	-0.5047	0.2877	4.4003	0.0734
PROG.st	-201	0.0689	0.0915	0.0083	-0.9370	0.3243	0.9713	0.4698
logPROG.st	-202	0.0687	0.0919	0.0113	-0.9643	0.2132	1.5499	0.3369
CORT.st	-200	0.0696	0.0903	0.0010	0.0208	0.9427	0.0052	0.9701
CY.BS	-201	0.0692	0.0998	0.0145	-0.9626	0.3846	0.7559	0.5349
CYL.BS.tr.st	-201	0.0694	0.0901	0.0069	-0.8444	0.4848	0.4882	0.6341
CYL.mean.tr.st	-200	0.0695	0.0901	0.0066	0.6792	0.6275	0.2355	0.7582
LPL.mean.st	-200	0.0696	0.0903	0.0002	0.0077	0.9778	0.0008	0.9887
ov.tot.PR.st	-200	0.0696	0.0903	0.0015	0.0564	0.9857	0.0003	0.9928
anov.long.PR.st	-200	0.0696	0.0903	0.0005	-0.0199	0.9765	0.0009	0.9880

Table 11.2.: Properties of a linear mixed model of CTX where a random slope of the respective covariate was included to the random intercept model (11.2). $pvalue$ and $pvalue.cor$ correspond to asymptotic likelihood ratio tests of $H_0 : d_1^2 = d_{01} = 0$ and $H_0 : d_{01} = 0$, respectively. The values less than 0.1 are marked.

Given the random intercept model, a random slope is significant for the covariates **FSH.st**, **logFSH.st** and **logE2.st**. A random slope of the non-transformed variable **FSH.st** turns out to be more significant, since the corresponding p-value is smaller. Hence we will focus on the non-transformed version. We may note that a random slope of **E2.st** is not significant and the p-value for testing the need of the respective random slope is lowest for **logE2.st**. This is why we initially presented the testing procedure for the standardized version of the logarithm of the estradiol serum concentration. Further it can be seen that

for none of these covariates a correlation between the random intercept and the random slope is needed to explain the variation in the data.

Clearly, we may continue by testing if, besides a random intercept, random effects of more than one covariate are needed. However, corresponding null hypothesis can not be rejected for any of these asymptotic likelihood ratio tests. Probably there are not enough data available so that models turn out to be too complex when including more than one random slope.

Hence, we now may concentrate on selecting significant fixed effects. Since the p-values of `FSH.st` and `logE2.st` in the last column of Table 11.2 do not highly differ and the corresponding models also have the same *AIC*, we perform the "model reduction" for each of the two random structures in Subsection 11.3.1 and Subsection 11.3.2, respectively. In Subsection 11.3.4 we then compare the reduced models.

11.3.1. Reduced model with random slope of `logE2.st`

In this Subsection we focus on selecting an appropriate structure of fixed effects for the linear mixed model with random intercept and random slope of `logE2.st`. In Chapter 6 there were presented various tests for verifying the significance of fixed effects. There is no single strategy how exactly to proceed for choosing an appropriate structure of fixed effects. Therefore we applied several approaches to this end (forward selection, backward selection and stepwise regression with respect to the different applicable tests) which basically all indicated the same set of influential covariates.

A sequential forward selection of covariates with respect to an asymptotic F-test turned out to be most appropriate. That is, we started with a fixed effects structure consisting only out of an intercept. Then we sequentially included this variable in the model with lowest p-value with respect to an asymptotic F-test (see Section 6.2). This procedure was repeated until all p-values of remaining covariates for selection were greater than 10%. We also considered logarithm-transformations of some covariates and if a dummy version of certain covariates might be more influential. Table 11.3 shows exemplarily the p-values and estimated coefficients of the covariates corresponding to the first level of selection. At first, a fixed effect of `FSH.st` is included in the model, since the respective p-value is lowest for this covariate.

This procedure is repeated until the lowest p-value is greater than 10%. The model which we obtain that way, defined as `mod.CTX.logE2`, can be briefly formulated as follows:

	$\hat{\beta}_1$	pvalue
AGE.st	-0.028	0.054
QCT.1.st	0.006	0.691
BMI.25.st	-0.015	0.307
BMI.0.st	-0.022	0.109
X.BMI.0.25.st	-0.011	0.455
LH.st	0.027	0.001
logLH.st	0.021	0.007
FSH.st	0.027	0.000
logFSH.st	0.025	0.001
E2.st	-0.005	0.485
logE2.st	-0.016	0.062
PROG.st	-0.015	0.052
logPROG.st	-0.024	0.003
CORT.st	0.007	0.338
CY.BS	-0.032	0.041
CYL.BS.tr.st	-0.019	0.013
CYL.mean.tr.st	-0.011	0.196
LPL.mean.st	0.003	0.745
ov.tot.PR.st	0.002	0.796
anov.long.PR.st	0.002	0.769
BMI0high	-0.029	0.337
BMI25high	-0.032	0.396
FSHlow	-0.059	0.001
QCT1low120	0.019	0.608
QCT1low150	-0.050	0.097
PROGhigh6	-0.044	0.005
PROGhigh12	-0.004	0.775
E2high200	0.005	0.778
E2high150	0.006	0.696

Table 11.3.: Table of $\hat{\beta}_1$ in the linear mixed model $\text{CTX}_{it} = \beta_0 + \beta_1 \cdot x_{i(t)} + \gamma_{i0} + \gamma_{i1} \cdot \log\text{E2.st}_{it} + \epsilon_{it}$ and the calculated p-value corresponding to an asymptotic F-test of $H_0 : \beta_1 = 0$. $x_{i(t)}$ represents the respective covariate given in the different rows. The p-values less than 0.1 are marked.

For participant $i = 1, \dots, 46$ and blood sample number t :

$$\begin{aligned} \text{CTX}_{it} = & \beta_0 + \beta_1 \text{FSH.st}_{it} + \beta_2 \text{AGE.st}_i + \beta_3 \text{QCT1low150}_i \\ & + \beta_4 \text{logPROG.st}_{it} + \beta_5 \text{ov.tot.PR.st}_{it} \\ & + \gamma_{i0} + \gamma_{i1} \text{logE2.st}_{it} + \epsilon_{it} \end{aligned} \quad (11.3)$$

$$\text{with } \begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N \left(\mathbf{0}, \begin{pmatrix} d_0^2 & 0 \\ 0 & d_1^2 \end{pmatrix} \right) \quad \text{and} \quad \epsilon \sim N(\mathbf{0}, \sigma^2 I) .$$

Performing the next level of the sequential forward selection shows that the lowest p-value is 0.129 corresponding to `QCT.1.st`. This is why we stop the selection procedure at model (11.3). From the corresponding summary table we may extract the estimates of the single model parameters which are given by the following table.

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	\hat{d}_0	\hat{d}_1	$\hat{\sigma}$
0.2854	0.0200	-0.0319	-0.0573	-0.0249	0.0173	0.0833	0.0176	0.0661

Furthermore, the summary table shows that the covariates are significant with respect to an asymptotic t-test even almost at level 5%:

```
> summary(mod.CTX.logE2)

Linear mixed-effects model fit by REML
Data: NULL
   AIC   BIC logLik
-294 -267    156

Random effects:
Formula: ~logE2.st | PAT.ID
Structure: Diagonal
      (Intercept) logE2.st Residual
StdDev:      0.0833    0.0176    0.0661

Fixed effects: CTX ~ FSH.st + AGE.st + QCT1low150 +
                logPROG.st + ov.tot.PR.st

              Value Std.Error   DF t-value p-value
(Intercept)   0.285    0.024 121    12.01  0.000
FSH.st         0.020    0.010 121     2.08  0.039
AGE.st        -0.032    0.014  43    -2.28  0.027
QCT1low1501   -0.057    0.029  43    -1.97  0.055
logPROG.st    -0.025    0.011 121    -2.28  0.024
ov.tot.PR.st   0.017    0.009 121     1.93  0.056

Correlation:
      (Intr) FSH.st AGE.st QCT115 1PROG.
FSH.st      0.029
AGE.st      0.012  0.000
QCT1low1501 -0.822 -0.024 -0.050
```

```
logPROG.st    -0.060  0.567  0.105  0.066
ov.tot.PR.st   0.021 -0.003  0.071 -0.006 -0.414
```

Standardized Within-Group Residuals:

```
      Min      Q1      Med      Q3      Max
-2.0688 -0.4651 -0.0811  0.5138  3.6029
```

Number of Observations: 170

Number of Groups: 46

Inclusion of an interaction

As suggested by Figure 11.8, we further investigate if an interaction of FSH with the level of estradiol might be necessary. To this end we adjust a model `mod.CTX.logE2.int`,

$$\begin{aligned} \text{CTX}_{it} = & \beta_0 + \beta_1 \text{FSH.st}_{it} + \beta_2 \text{AGE.st}_i + \beta_3 \text{QCT1low150}_i \\ & + \beta_4 \text{logPROG.st}_{it} + \beta_5 \text{ov.tot.PR.st}_{it} + \beta_6 \text{FSH.st}_{it} \cdot \text{E2high150}_{it} \\ & + \gamma_{i0} + \gamma_{i1} \text{logE2.st}_{it} + \epsilon_{it}, \end{aligned}$$

and perform an asymptotic likelihood ratio test, introduced in Subsection 6.1.1, to verify $H_0 : \beta_6 = 0$. The corresponding test statistic satisfies

$$\text{LRT} \xrightarrow{d} \chi_1^2.$$

This test can be performed by the `anova` function, but to get the LRT and not the RLRT statistic we have to adjust `"method=ML"` in the `lme` function (see Section 8.1 and Subsection 8.3.1).

```
> mod.CTX.logE2.ML = update( mod.CTX.logE2, method="ML")

> mod.CTX.logE2.int.ML = lme( CTX ~ FSH.st + AGE.st + QCT1low150 +
  logPROG.st + ov.tot.PR.st + FSH.st:E2high150,
  random = list(PAT.ID = pdDiag(~logE2.st)) , method="ML" )

> anova(mod.CTX.logE2.int.ML, mod.CTX.logE2.ML)
              Model df   AIC   BIC logLik   Test L.Ratio p-value
mod.CTX.logE2.int.ML      1 10 -337 -306    178
mod.CTX.logE2.ML          2  9 -336 -308    177 1 vs 2     2.74  0.1279
```

However, due to the p-value of almost 13%, we can not reject the null hypothesis and prefer model (11.3). Also an asymptotic F-test (p-value 0.119) and an asymptotic t-test (p-value 0.117) did not show the significance of this interaction. For the sake of completeness we extracted the respective line of the summary table of `mod.CTX.logE2.int`:

	Value	Std.Error	DF	t-value	p-value
FSH.st:E2high150high	-0.040	0.025	120	-1.63	0.117 .

Hence we suggest that the different shapes of the regression lines in Figure 11.8 can be explained by random effects, particularly by the participant-specific slope of `logE2.st`.

Model diagnostics

To decide whether model (11.3) is appropriate for explaining CTX, it is crucial to check if the model assumptions appear valid. That is, we have to verify whether

(1) $\epsilon \sim N(\mathbf{0}, \sigma^2 I)$ and

(2) $\begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} d_0^2 & 0 \\ 0 & d_1^2 \end{pmatrix}\right)$

are satisfied. Therefore we consider graphics of the fitted values, the standardized residuals and the estimated random effects as suggested in Section 7.3 and basically referring to Pinheiro and Bates (2000). The graphics shown in the following can be easily generated in R by plotting a `lme` object which is provided by the `nlme` library. As already mentioned in Section 7.3, to verify distributional assumptions made about the residual vector we consider the conditional Pearson residuals, defined as

$$\left(\text{CTX}_{it} - \hat{\beta}_0 - \hat{\beta}_1 \text{FSH.st}_{it} - \hat{\beta}_2 \text{AGE.st}_i - \hat{\beta}_3 \text{QCT1low150}_i - \hat{\beta}_4 \text{logPROG.st}_{it} - \hat{\beta}_5 \text{ov.tot.PR.st}_{it} - \hat{\gamma}_{i0} - \hat{\gamma}_{i1} \text{logE2.st}_{it} \right) / \hat{\sigma}.$$

According to Subsection 8.2.1, these residuals can be accessed in R by

```
> residuals(mod.CTX.logE2, type="p") .
```

These conditional Pearson residuals are in R also referred to as *standardized residuals*. As generally described in Section 7.3, to investigate whether the model is successful in explaining the data, the observed response values may be drawn against the fitted ones. This is shown by Figure 11.13. The points lie along the bisecting line and there are only 4 outliers, i.e. measurements with conditional Pearson residuals greater than 1.96 in absolute value. That is the $(1 - 0.05/2)$ -quantile of the standard normal distribution. The outliers are labeled by the number of the corresponding participant. Hence, this graphic does not indicate any lack of fit of model `mod.CTX.logE2`.

The following code shows exemplarily how Figure 11.13 was generated. The remaining graphics of this section have been achieved in similar manner.

```
> library(lattice)
> plot(mod.CTX.logE2, CTX ~ fitted(.), id=0.05, adj=-0.5, cex=0.8, pch=16,
+ main="observed vs. fitted values", grid=F, col="black", abline=c(0,1), lty=2 )
```

To check the distributional assumption about ϵ , Figure 11.14 shows various graphics of the conditional Pearson residuals which are seen as estimates of the error terms. First of all, a scatter plot of the fitted values and the residuals is shown. The points vary constantly around zero without forming a specific pattern. Thought away the outlying observations, variability does not seem to increase highly.

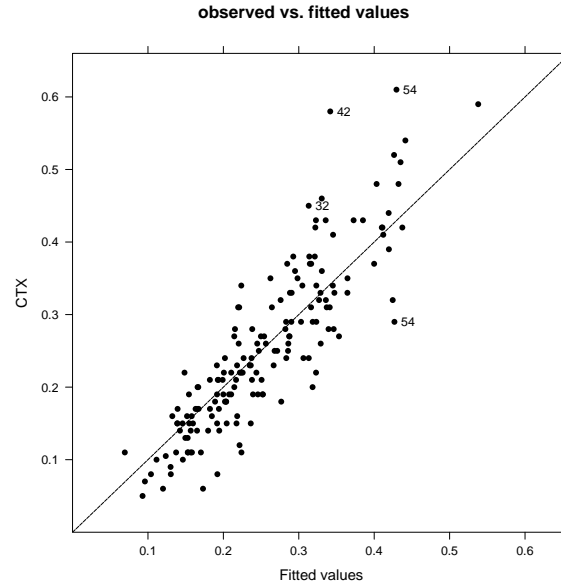


Figure 11.13.: Scatter plot of CTX and the fitted values of `mod.CTX.logE2` (11.3)

The left graphic at the top shows a boxplot of the conditional Pearson residuals for the 13 participants with 5 complete measurements. The graphic in the second line shows individual scatter plots of the fitted values and the residuals. It can be seen that depending on whether how strong CTX varies over time (see Figure 11.1), the residuals vary for each woman. For example, CTX concentrations of participant number 13 are nearly constant in all of the 5 blood samples and also the residuals lie in the same range shown by a short box in the second graphic and points close to each other in the third graphic. The outlying observations belong to participants who also had an outlying CTX observation during the course of the study. Figure 11.1 shows that the CTX concentrations of participant number 54 lie in around the same range at blood test 1,4 and 5. In blood test 2 it was extremely high and in blood test 3 it was lowest. Hence, the variability is not constant between the individuals. But as has been mentioned already in Section 11.1, we can not rely too much on individual plots in our case, since there are too little measurements for each participant available.

The last row of Figure 11.1 shows scatter plots of each of the continuous covariates in model `mod.CTX.logE2` and the residuals to verify whether the variability of the residuals is constant as a function of the covariates. The first covariate-residual plot illustrates that FSH measurements are not homogeneously distributed but concentrated to the lower part of their range (*right-skewed*). This was already noticed in the data description of Figure 9.9. However, there are also high values of residuals for high values of FSH. Thus, again thought away the outliers, the variability can be considered as constant and we do not prefer a logarithm-transformation in that case.

Altogether, the zero-mean assumption as well as the assumption about a constant variance parameter σ seems to be valid.

Finally, we have to check the assumption of normality made on ϵ and the random effects. Therefore, Figure 11.15 shows Q-Q plots of the conditional Pearson residuals and the estimated random effects $\hat{\gamma}_{i0}$ and $\hat{\gamma}_{i1}$. The distribution of the residuals seems to be slightly skewed to the right and the Q-Q plot of the estimated random slope values suggests that the corresponding distribution is somehow heavy-tailed. The normality assumption about the random intercept seems to be valid since the points lie almost exactly on a line. However, we do not place a very strong emphasis on these graphics because Q-Q plots generally need a very large number of observations to have a high explanatory power. Thus, we suspect that the normality assumptions are correct, also for the residuals and the random slope.

For the sake of completeness, Figure 11.16 shows a scatter plot of the estimated random effects. It does not seem that there is any relationship between them. Thus the assumption of uncorrelated random effects appears to be valid. to verify whether the assumption

Consequently, `mod.CTX.logE2` provides a first approach to explain the serum concentration of the bone resorption marker CTX appropriately. Conclusions and interpretations of this model we give later in this section. At first we concentrate on adjusting the structure of fixed effects for the linear mixed model with random intercept and random slope of `FSH.st`.

11.3.2. Reduced model with random slope of `FSH.st`

Table 11.2 showed that, besides a random slope of `logE2.st`, a random slope of `FSH.st` is significant in a linear mixed model with random intercept and saturated fixed effects structure. Therefore we perform the "model reduction" process also for this model.

To this end, we apply the same procedures as in the the previous subsection, basically concentrating on a sequential forward selection of covariates with respect to an asymptotic F-test. All approaches indicate exactly the same set of influential covariates as for the model with random slope of `logE2.st`. This emphasize that these covariates are the ones which influence CTX mostly.

We declare this model in R by `mod.CTX.FSH`. It can be briefly formulated as follows:

For participant $i = 1, \dots, 46$ and blood sample number t :

$$\begin{aligned} \text{CTX}_{it} = & \beta_0 + \beta_1 \text{FSH.st}_{it} + \beta_2 \text{AGE.st}_i + \beta_3 \text{QCT1low150}_i \\ & + \beta_4 \text{logPROG.st}_{it} + \beta_5 \text{ov.tot.PR.st}_{it} \\ & + \gamma_{i0} + \gamma_{i1} \text{FSH.st}_{it} + \epsilon_{it} \end{aligned} \quad (11.4)$$

$$\text{with } \begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N \left(\mathbf{0}, \begin{pmatrix} d_0^2 & 0 \\ 0 & d_1^2 \end{pmatrix} \right) \quad \text{and} \quad \epsilon \sim N(\mathbf{0}, \sigma^2 I) .$$

Sequential asymptotic F-tests about the significance of fixed effects can be performed by

anova (see 8.3.2). Hence, by

```
> anova(mod.CTX.FSH)
```

it can be asked for the corresponding p-values at which the covariates were included in model (11.4):

	numDF	denDF	F-value	p-value
(Intercept)	1	121	323	<.0001
FSH.st	1	121	12	0.0007
AGE.st	1	43	6	0.0215
QCT1low150	1	43	3	0.0765
logPROG.st	1	121	3	0.0936
ov.tot.PR.st	1	121	4	0.0487

As before, performing one step more of the sequential forward selection shows that the lowest p-value is 0.130 corresponding to QCT.1.st. To analyze also this model we consider the corresponding summary table:

```
> summary(mod.CTX.FSH)
```

Linear mixed-effects model fit by REML

Data: NULL

AIC	BIC	logLik
-295	-267	156

Random effects:

Formula: ~FSH.st | PAT.ID

Structure: Diagonal

	(Intercept)	FSH.st	Residual
StdDev:	0.0834	0.0174	0.0663

StdDev: 0.0834 0.0174 0.0663

Fixed effects: CTX ~ FSH.st + AGE.st + QCT1low150 +
logPROG.st + ov.tot.PR.st

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.285	0.024	121	11.98	0.000
FSH.st	0.021	0.010	121	2.06	0.041
AGE.st	-0.033	0.014	43	-2.34	0.024
QCT1low1501	-0.057	0.029	43	-1.95	0.058
logPROG.st	-0.026	0.011	121	-2.37	0.019
ov.tot.PR.st	0.018	0.009	121	1.99	0.049

Correlation:

	(Intr)	FSH.st	AGE.st	QCT115	1PROG.
FSH.st	0.037				
AGE.st	0.011	-0.015			
QCT1low1501	-0.822	-0.017	-0.050		
logPROG.st	-0.059	0.549	0.100	0.070	
ov.tot.PR.st	0.026	0.005	0.072	-0.009	-0.420

FSH.st 0.037

AGE.st 0.011 -0.015

QCT1low1501 -0.822 -0.017 -0.050

logPROG.st -0.059 0.549 0.100 0.070

ov.tot.PR.st 0.026 0.005 0.072 -0.009 -0.420

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.0694	-0.4802	-0.0894	0.5066	3.6600

Number of Observations: 170

Number of Groups: 46

It can be seen that all covariates are significant at level 6% with respect to an asymptotic t-test. The estimates for the fixed effects as well as for the variance components are almost equal to model (11.4). Compared with the summary table of that model, the *AIC* drops a bit but the maximal value of the conditional Pearson residuals increases to 3.66.

To decide which of the two adjusted models, `mod.CTX.logE2` or `mod.CTX.FSH`, is more appropriate for explaining `CTX`, we compare the corresponding graphics for verifying the model assumptions (see Figure 11.17).

11.3.3. Model comparisons

Finally, we aim to select this model, which seems best in fitting the data and explaining `CTX`. Therefore we compare the models (11.3) and (11.4) which do only differ in their random slope.

As a third model, we also reduced the random intercept model `mod.ri` (11.2) by selecting only significant fixed effects. Again, exactly the same set of fixed effects was suggested:

For participant $i = 1, \dots, 46$ and blood sample number t :

$$\begin{aligned} \text{CTX}_{it} = & \beta_0 + \beta_1 \text{FSH.st}_{it} + \beta_2 \text{AGE.st}_i + \beta_3 \text{QCT1low150}_i \\ & + \beta_4 \text{logPROG.st}_{it} + \beta_5 \text{ov.tot.PR.st}_{it} + \gamma_{i0} + \epsilon_{it} \end{aligned} \quad (11.5)$$

with $\gamma_{i0} \sim N(0, d_0^2)$ and $\epsilon \sim N(\mathbf{0}, \sigma^2 I)$.

We call this reduced random intercept model `mod.CTX.ri.red`.

Figure 11.17 shows graphics concerning the model analysis of each of the three models. For `mod.CTX.logE2`, these plots were already considered in Subsection 11.3.1. Altogether, it may be concluded that the models do not differ highly in their goodness of fit. In `mod.CTX.FSH` and `mod.CTX.ri.red` there is one outlier more than in `mod.CTX.logE2`. That is an observation with conditional Pearson residual greater in absolute value than the $(1 - 0.05/2)$ -quantile of the standard normal distribution. Hence, the portion of variability explained by the random slope is probably not very high. Clearly, this fact is also indicated by the very low values of \hat{d}_1 in (11.3) and 11.4.

Since these graphical model diagnostics do not clearly suggest a "best model" and basically indicate that each of these models is appropriate for explaining the data, we refer to our

initial investigations. To recapitulate, using an asymptotic likelihood ratio test we found out that random slopes of `logE2.st` and `FSH.st` are significant (see Table 11.2). The p-value corresponding to this test was even lower for `logE2.st`. Therefore, we propose `mod.CTX.logE2` as final model.

11.3.4. The final model for CTX

Using the model (11.3) we might predict CTX for the women who took part at the study as follows:

$$\begin{aligned}\widehat{\text{CTX}}_{it} = & 0.29 + 0.02 \cdot \text{FSH.st}_{it} - 0.03 \cdot \text{AGE.st}_i - 0.06 \cdot \text{QCT1low150}_i \\ & - 0.02 \cdot \text{logPROG.st}_{it} + 0.02 \cdot \text{ov.tot.PR.st}_{it} \\ & + \hat{\gamma}_{i0} + \hat{\gamma}_{i1} \cdot \text{logE2.st}_{it}\end{aligned}$$

The participant-specific effects were estimated as shown in the following table:

	2	3	5	8	9	10	11	12	13	14
$\hat{\gamma}_{i0}$	-0.0774	0.0043	-0.0715	-0.0265	0.0253	0.1580	-0.1630	0.0068	-0.0522	0.1224
$\hat{\gamma}_{i1}$	0.0082	0.0000	0.0056	-0.0018	-0.0006	0.0039	0.0064	-0.0033	-0.0014	0.0001

	15	16	17	19	20	21	24	25	26	27
$\hat{\gamma}_{i0}$	0.0057	0.0565	-0.0739	-0.0327	-0.0974	-0.0108	0.0223	0.0801	0.0963	0.0399
$\hat{\gamma}_{i1}$	-0.0200	0.0009	0.0003	-0.0002	-0.0002	-0.0008	-0.0007	0.0056	-0.0207	0.0023

	29	30	31	32	34	36	37	38	39	41
$\hat{\gamma}_{i0}$	-0.0337	-0.0139	-0.0795	0.0449	0.0539	-0.0120	-0.0994	-0.1099	-0.0459	-0.0676
$\hat{\gamma}_{i1}$	-0.0003	-0.0001	-0.0024	-0.0028	-0.0098	0.0006	0.0048	-0.0060	0.0022	0.0017

	42	43	44	45	46	48	51	52	53	54
$\hat{\gamma}_{i0}$	0.0975	0.0520	-0.0190	0.1022	0.1100	-0.1249	0.0194	-0.0110	-0.0209	0.1236
$\hat{\gamma}_{i1}$	0.0119	-0.0020	0.0052	-0.0002	-0.0025	0.0070	0.0039	-0.0000	0.0062	0.0009

	56	58	59	60	61	63
$\hat{\gamma}_{i0}$	-0.0188	-0.1045	0.0451	0.0611	0.0301	0.0090
$\hat{\gamma}_{i1}$	-0.0005	-0.0035	0.0037	0.0012	-0.0020	0.0051

To summarize the obtained result, we first note that among all continuous variables, the female sex hormone FSH has the highest influence on the bone resorption marker CTX (remember that this is true since we standardized the covariates initially). The higher the FSH serum concentration, the lower the concentration of CTX. Furthermore we get:

- If the bone density at study onset is less than 150 mgCa-Ha/ml, the mean level of CTX is 0.23 ng/ml, otherwise 0.29.
- The older the women at the begin of the study, the lower CTX.
- A high serum concentration of progesterone provoke low serum concentration of CTX.
- A high percentage of ovulatory cycles leads to higher values of CTX.

11. Linear mixed models to explain the bone markers (Longitudinal analysis)

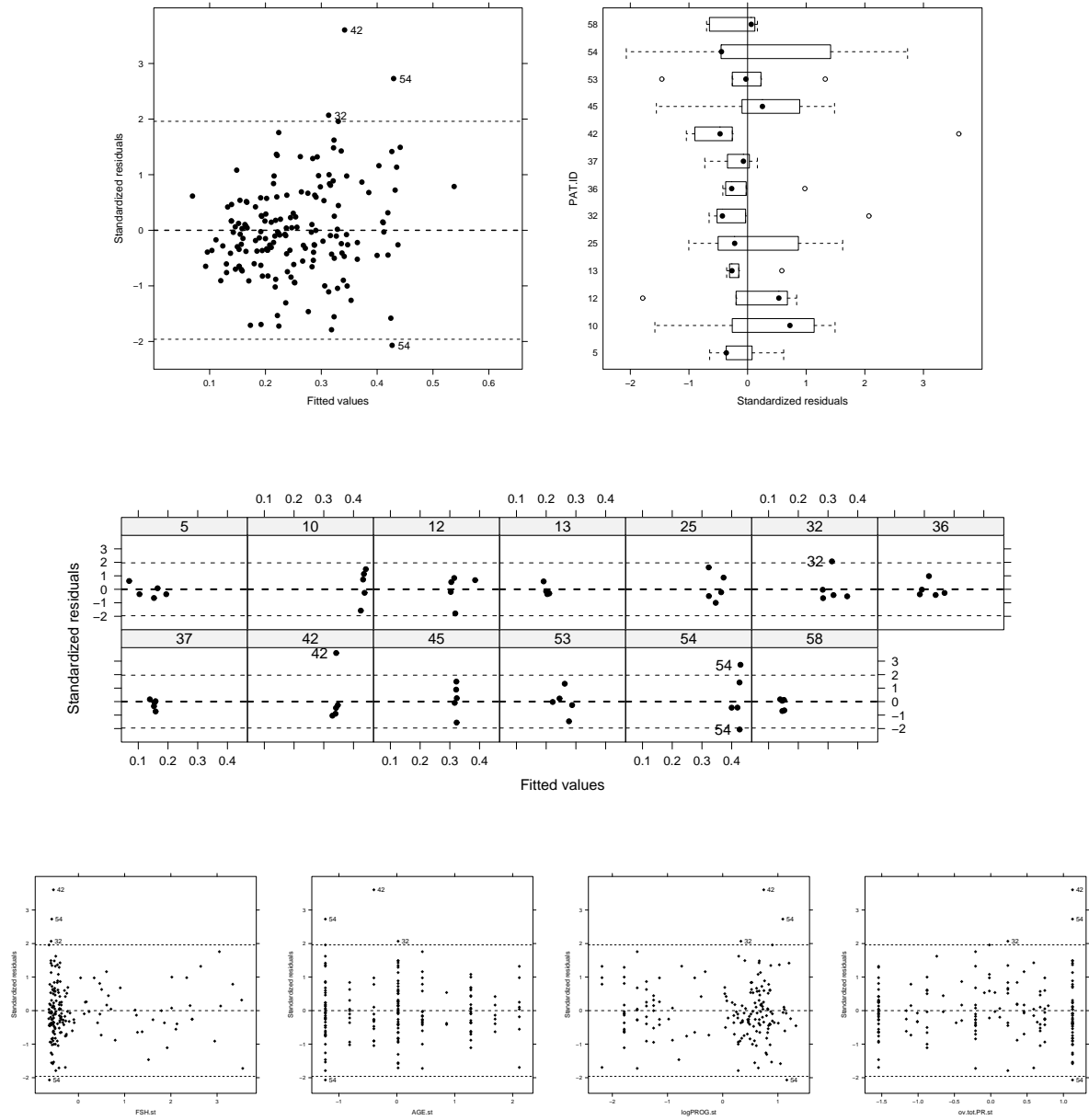


Figure 11.14.: Diagnostics of the conditional Pearson residuals of `mod.CTX.logE2` (11.3)

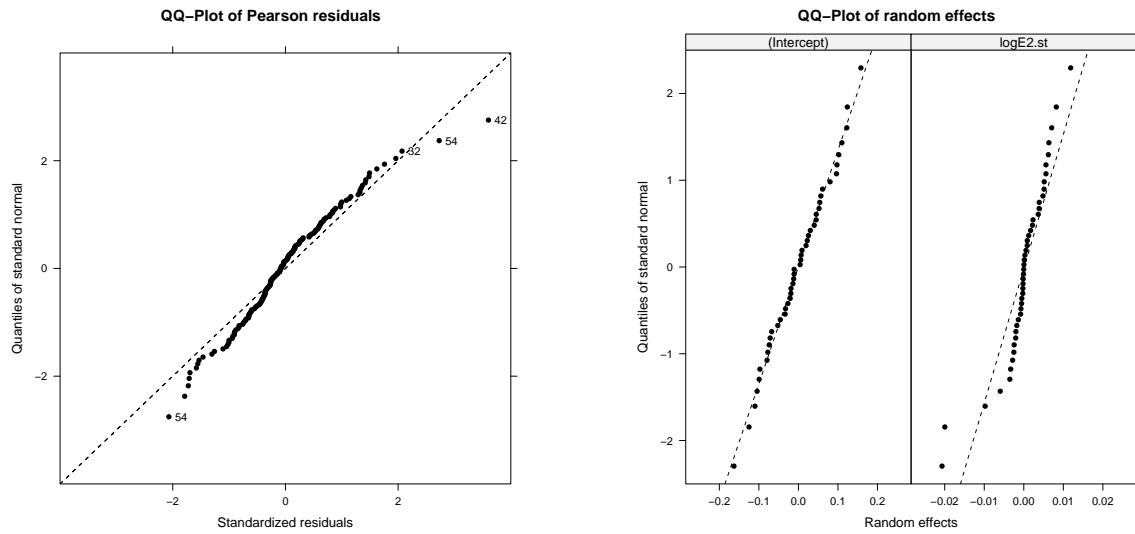


Figure 11.15.: Assessing assumptions of normality for `mod.CTX.logE2` (11.3)

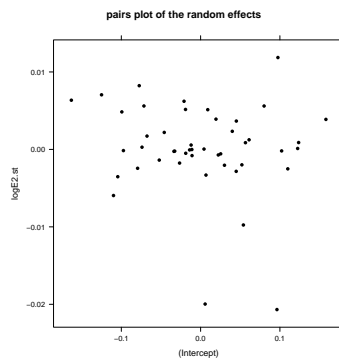


Figure 11.16.: Scatter plot of the estimated random intercept $\hat{\gamma}_{i0}$ and the estimated random slope $\hat{\gamma}_{i1}$ of `mod.CTX.logE2` (11.3)

11. Linear mixed models to explain the bone markers (Longitudinal analysis)

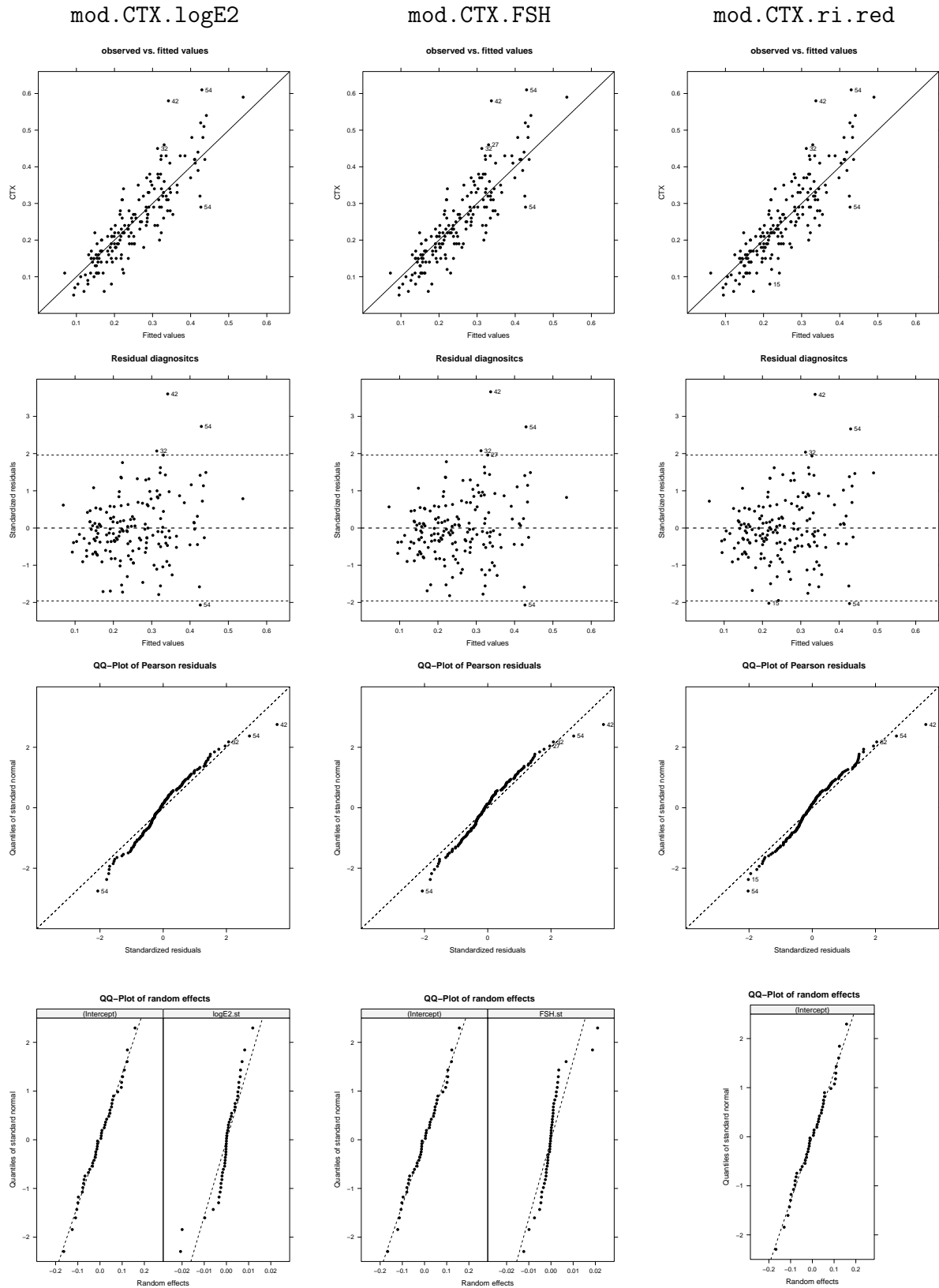


Figure 11.17.: Model diagnostics of model (11.3), (11.4) and (11.5)

11.4. Modeling the bone formation marker osteocalcin

For modeling the bone formation marker osteocalcin (OST), we proceeded exactly in the same manner as for modeling CTX. Therefore we outline only briefly the single steps in selecting the "best" model and refer to Chapter 11.3 for a detailed description of the applied approach.

We start by adjusting the saturated fixed effects model, which can be briefly formulated as

$$\text{OST}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \epsilon_{it} \quad \text{with} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I) .$$

According to the previous notations, OST_{it} indicates the serum concentration of osteocalcin measured for participant i , $i = 1, \dots, 46$ in the t^{th} blood sample. \mathbf{x}_{it}^T denotes the corresponding covariate vector consisting out of the 15 variables listed in Table 11.1 and a 1-entry corresponding to an intercept. The R^2 for this standard linear model is 0.196 and the residual standard error σ is estimated to 4.1.

Including a random intercept in this model, i.e.

$$\text{OST}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \gamma_{i0} + \epsilon_{it} \quad \text{with} \quad \gamma_{i0} \sim N(0, d_0^2) \quad \text{and} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I) , \quad (11.6)$$

the standard deviation of the residuals σ is estimated to 2.31 and the standard deviation of the random intercept d_0 is estimated to 3.65. Hence, also in this case, variability is explained by the random effect.

To verify this observation, i.e. $H_0 : d_0^2 = 0$, we perform an exact likelihood ratio test (as before). This yields

```
simulated finite sample distribution of RLRT.
(p-value based on 10000 simulated values)
```

```
data:
```

```
RLRT = 88.9, p-value < 2.2e-16
```

Hence, as expected, the random intercept is highly significant.

Further, we verify if additionally a random slope should be included in (11.6). To this end we proceed analogously to Section 11.3. Table 11.4 summarizes the results of our investigations. Only for the standardized percentage of anovulatory long cycles a random slope is significant at level 10%. It is noticeable that the null hypothesis whether the random intercept and the random slope of `anov.long.PR.st` are uncorrelated, can be rejected (`pvalue.cor=0.0420`). Hence we incorporate here also a covariance coefficient in the covariance matrix of the random effects.

Again applying several "model reduction" procedures, we got always the same set of covariates with significant fixed effects. Thus we propose the following model, declared as `mod.OST.anovlong`:

	AIC	$\hat{\sigma}$	\hat{d}_0	\hat{d}_1	cor.ri.rs	pvalue.cor	L.Ratio	pvalue
AGE.st	891	2.3117	3.6122	0.7884	-0.8728	0.1742	1.8464	0.2857
QCT.1.st	892	2.3121	3.6188	0.4613	-0.3248	0.7289	0.1201	0.8353
BMI.25.st	892	2.3115	3.3354	1.5051	-0.4172	0.3888	0.7455	0.5384
BMI.0.st	890	2.3177	1.1097	4.2697	0.1579	0.8275	1.9936	0.2635
X.BMI.0.25.st	891	2.3113	3.5797	0.6307	0.9188	0.3020	1.2688	0.3951
LH.st	892	2.3075	3.6282	0.2757	-0.6470	0.6103	0.2598	0.7442
logLH.st	892	2.2849	3.6860	0.3878	0.3077	0.7101	0.1540	0.8103
FSH.st	892	2.3001	3.6470	0.3206	-0.8963	0.3319	0.9417	0.4782
logFSH.st	892	2.3116	3.6528	0.0587	0.0270	0.9449	0.0048	0.9713
E2.st	892	2.3085	3.6544	0.1378	0.1004	0.9210	0.0098	0.9580
logE2.st	892	2.3118	3.6507	0.0663	0.0582	0.9436	0.0050	0.9706
PROG.st	892	2.2964	3.6521	0.3318	-0.5560	0.5404	0.3862	0.6793
logPROG.st	892	2.2357	3.6877	0.7457	-0.0430	0.9267	0.7866	0.5250
CORT.st	892	2.2612	3.6902	0.5913	0.4477	0.4919	0.7648	0.5320
CY.BS	892	2.2831	3.7017	0.8868	-0.1376	0.8676	0.0854	0.8641
CYL.BS.tr.st	892	2.3059	3.6388	0.2636	0.5206	0.6629	0.1901	0.7861
CYL.mean.tr.st	890	2.2693	3.5563	0.8453	0.6718	0.1202	2.6204	0.1876
LPL.mean.st	892	2.2996	3.6717	0.0700	-0.1090	0.9776	0.0426	0.9077
ov.tot.PR.st	892	2.2804	3.6530	0.5280	0.2886	0.6027	0.3929	0.6762
anov.long.PR.st	888	2.2380	3.5649	0.8965	-0.7755	0.0420	4.9098	0.0563

Table 11.4.: Properties of a linear mixed model of OST where a random slope of the respective covariate was included to the random intercept model (11.6). *pvalue* and *pvalue.cor* correspond to asymptotic likelihood ratio tests of $H_0 : d_1^2 = d_{01} = 0$ and $H_0 : d_{01} = 0$, respectively. The values less than 0.1 are marked. For details see Section 11.3.

For participant $i = 1, \dots, 46$ and blood sample number t :

$$\text{OST}_{it} = \beta_0 + \beta_1 \text{BMI.0.st}_i + \beta_2 \text{AGE.st}_i + \beta_3 \text{E2high200}_{it} + \gamma_{i0} + \gamma_{i1} \text{anov.long.PR.st}_{it} + \epsilon_{it} \quad (11.7)$$

$$\text{with } \begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N \left(\mathbf{0}, \begin{pmatrix} d_0^2 & d_{01} \\ d_{01} & d_1^2 \end{pmatrix} \right) \quad \text{and} \quad \epsilon \sim N(\mathbf{0}, \sigma^2 I) .$$

The summary table of this model shows that all covariates are significant at level 10% with respect to an asymptotic t-test. Figure 11.18 shows

Linear mixed-effects model fit by REML

Data: NULL

AIC BIC logLik

870 895 -427

Random effects:

Formula: ~anov.long.PR.st | PAT.ID

11. Linear mixed models to explain the bone markers (Longitudinal analysis)

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	3.49	(Intr)
anov.long.PR.st	1.12	-0.676
Residual	2.15	

Fixed effects: OST ~ BMI.0.st + AGE.st + E2high200

	Value	Std.Error	DF	t-value	p-value
(Intercept)	16.74	0.544	123	30.76	0.000
BMI.0.st	-1.06	0.493	43	-2.16	0.036
AGE.st	-0.95	0.522	43	-1.83	0.075
E2high2001	-0.80	0.484	123	-1.64	0.097

Correlation:

	(Intr)	BMI.0.	AGE.st
BMI.0.st	-0.023		
AGE.st	-0.211	-0.162	
E2high2001	-0.225	0.071	0.074

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.496	-0.518	-0.104	0.542	2.261

Number of Observations: 170

Number of Groups: 46

Osteocalcin concentrations might be predicted at population level, i.e. for any participant by

$$\hat{OST}_{it} = 16.74 - 1.06 \text{ BMI.0.st}_i - 0.95 \text{ AGE.st}_i - 0.8 \text{ E2high200}_{it}.$$

Thus, high weight and age at the begin of the study, as well as high estradiol concentrations lead to lower values of osteocalcin. By classifying the covariates due to the given table which shows the 25%, the 50% and the 75% quantiles,

	25%	50%	75%
BMI.0	21.48	22.83	25.77
AGE	45.25	48.00	49.00

osteocalcin concentrations can be predicted by applying the following table:

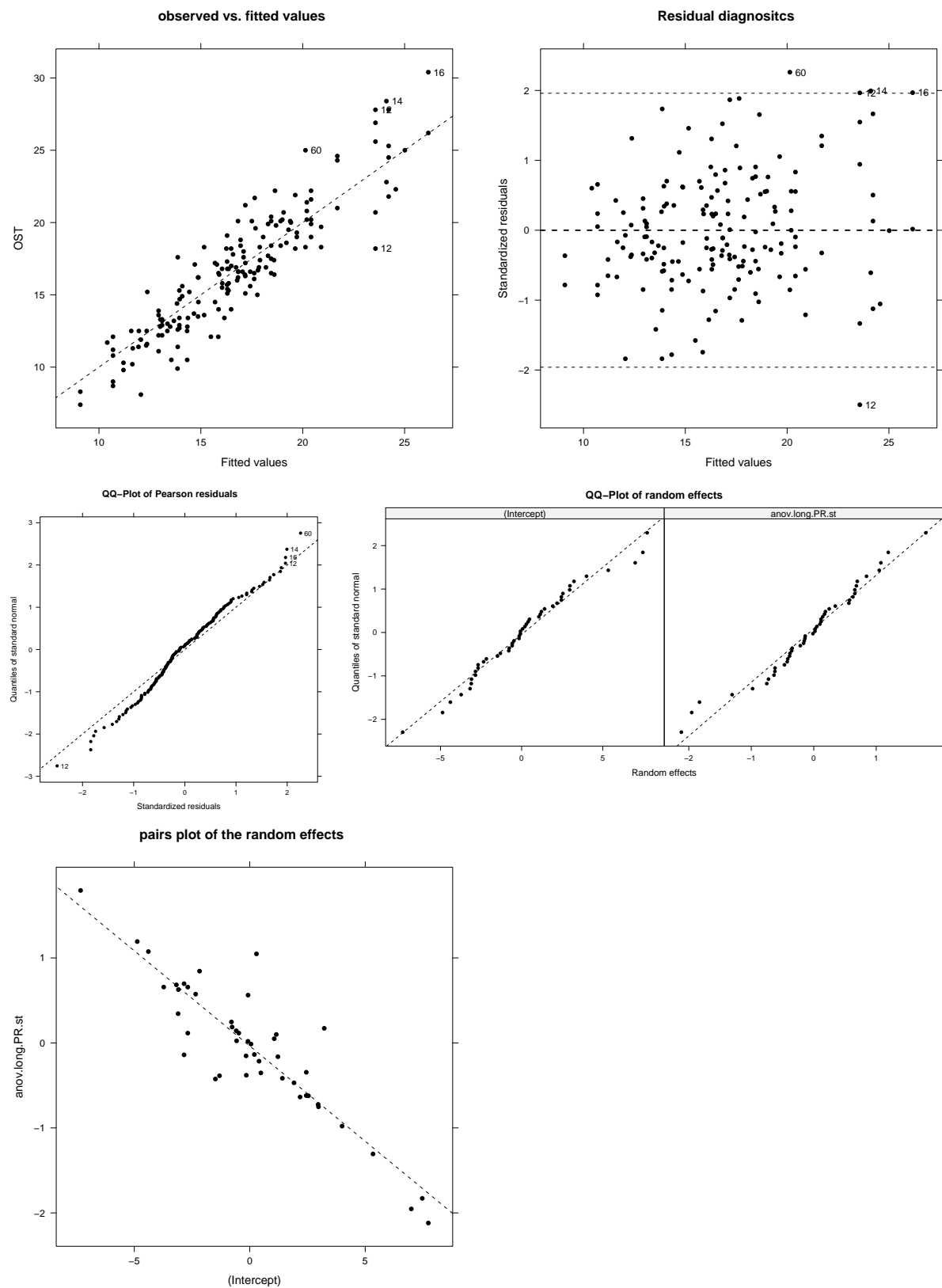


Figure 11.18.: Model diagnostics for `mod.OST.anovlong`

11. Linear mixed models to explain the bone markers (Longitudinal analysis)

BMI.0	AGE	E2high200	pred.values.OST
low	low	high	17.73
low	low	low	18.54
low	middle	high	16.66
low	middle	low	17.47
low	high	high	16.27
low	high	low	17.08
middle	low	high	17.30
middle	low	low	18.11
middle	middle	high	16.23
middle	middle	low	17.04
middle	high	high	15.84
middle	high	low	16.65
high	low	high	16.37
high	low	low	17.18
high	middle	high	15.30
high	middle	low	16.11
high	high	high	14.91
high	high	low	15.72

11.5. Modeling the bone formation marker BAP

Finally, we aim to explain the serum concentration of the bone formation marker BAP using a linear mixed model. Again, we proceed similarly as in Section 11.3 where the model building process was illustrated in detail for modeling the bone resorption marker CTX.

The saturated fixed effects model for BAP can be briefly formulated as follows: For participant $i = 1, \dots, 46$ and blood sample number t :

$$\text{BAP}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \epsilon_{it} \quad \text{with} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I) .$$

As in the previous sections, \mathbf{x}_{it}^T symbolizes the corresponding covariate vector consisting out of the 15 variables listed in Table 11.1 and a 1-entry corresponding to an intercept. The R^2 of this standard linear model is 0.21 and σ is estimated to 2.3.

In order to verify whether a random intercept model,

$$\text{OST}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \gamma_{i0} + \epsilon_{it} \quad \text{with} \quad \gamma_{i0} \sim N(0, d_0^2) \quad \text{and} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I) . \quad (11.8)$$

is preferable, we perform an exact likelihood ratio test and get

simulated finite sample distribution of RLRT.
(p-value based on 10000 simulated values)

data:

RLRT = 133, p-value < 2.2e-16 .

Hence, the null hypothesis $H_0 : d_0^2 = 0$ can be rejected, indicating that the random intercept in 11.8 is significant. It's standard deviation is estimated to 2.21 and the residual standard error is estimated to 1.05.

As for modeling CTX and osteocalcin, we investigate next if additionally a random slope should be included. Table 11.5 summarizes the results of this analysis.

The table shows that a random slope is significant for the variables QCT.1, BMI.25.st and CYL.mean. We figured out that a linear mixed model with random slope of CYL.mean fits the data best. For the sake of completeness, graphics to verify model assumptions in a reduced linear mixed model with random slope of QCT.1 are given in the Appendix (B.6).

The further selection process of fixed effects showed, that the following model `mod.BAP.CYL` is most appropriate.:

	AIC	$\hat{\sigma}$	\hat{d}_0	\hat{d}_1	cor.ri.rs	pvalue.cor	L.Ratio	pvalue
AGE.st	670	1.0503	2.1735	0.4099	0.9186	0.2761	1.1862	0.4144
QCT.1.st	663	1.0499	1.3705	1.8426	-0.4878	0.1276	7.7647	0.0130
BMI.25.st	667	1.0490	1.6397	1.4827	-0.4056	0.1769	4.0427	0.0884
BMI.0.st	667	1.0494	1.7377	1.2799	-0.2390	0.4734	3.4842	0.1186
X.BMI.0.25.st	671	1.0510	2.2089	0.0950	-0.0620	0.9736	0.0011	0.9865
LH.st	669	1.0522	2.1200	0.1792	0.7381	0.1716	1.8837	0.2799
logLH.st	667	1.0490	2.0772	0.2801	0.9318	0.0537	3.8309	0.0988
FSH.st	670	1.0522	2.1559	0.1316	0.5794	0.3792	0.7732	0.5293
logFSH.st	669	1.0504	2.1352	0.1825	0.7300	0.2185	1.5486	0.3372
E2.st	671	1.0504	2.2076	0.0500	-0.1006	0.9877	0.0002	0.9938
logE2.st	670	1.0508	2.1824	0.1051	-0.4383	0.6115	0.2579	0.7453
PROG.st	669	1.0478	2.1505	0.2023	-0.9311	0.1738	1.8499	0.2852
logPROG.st	669	1.0496	2.1384	0.2070	-0.8252	0.1798	1.7989	0.2933
CORT.st	669	1.0475	2.1679	0.1853	0.9404	0.1442	2.1330	0.2442
CY.BS	671	1.0509	2.2111	0.0294	-0.0131	0.9526	0.0035	0.9754
CYL.BS.tr.st	667	1.0346	2.1155	0.3151	-0.6730	0.1141	3.3226	0.1291
CYL.mean.tr.st	665	1.0073	2.0536	0.5392	-0.5313	0.0861	5.5141	0.0412
LPL.mean.st	670	1.0469	2.3150	0.0480	-0.5620	0.4513	0.5674	0.6021
ov.tot.PR.st	670	1.0539	2.1156	0.2102	-0.7118	0.2975	1.0854	0.4393
anov.long.PR.st	670	1.0482	2.1479	0.1894	0.6036	0.3038	1.1721	0.4177

Table 11.5.: Properties of a linear mixed model of BAP where a random slope of the respective covariate was included to the random intercept model (11.8). *pvalue* and *pvalue.cor* correspond to asymptotic likelihood ratio tests of $H_0 : d_1^2 = d_{01} = 0$ and $H_0 : d_{01} = 0$, respectively. The values less than 0.1 are marked. For details see Section 11.3.

For participant $i = 1, \dots, 46$ and blood sample number t :

$$\text{BAP}_{it} = \beta_0 + \beta_1 \text{PROGhigh6}_{it} + \beta_2 \text{BMI0high}_i + \beta_3 \text{AGE.st}_i + \gamma_{i0} + \gamma_{i1} \text{CYL.mean.tr.st}_{it} + \epsilon_{it} \quad (11.9)$$

$$\text{with } \begin{pmatrix} \gamma_{i0} \\ \gamma_{i1} \end{pmatrix} \sim N \left(\mathbf{0}, \begin{pmatrix} d_0^2 & d_{01} \\ d_{01} & d_1^2 \end{pmatrix} \right) \quad \text{and} \quad \epsilon \sim N(\mathbf{0}, \sigma^2 I) .$$

Applying the anova function, one may derive the p-values corresponding to sequential asymptotic F-test.

```
> anova(mod.BAP.CYL)
              numDF denDF F-value p-value
(Intercept)      1   123     914  <.0001
PROGhigh6         1   123      10  0.0016
BMI0high          1    43       5  0.0394
AGE.st            1    43       3  0.0697
```

The summary table shows that all covariates are significant with respect to an asymptotic t-Test. Referring to Figure 11.19, the model seems to fit the data adequately.

```
> summary(mod.BAP.CYL)
```

Linear mixed-effects model fit by REML

Data: NULL

AIC BIC logLik

628 653 -306

Random effects:

Formula: ~CYL.mean.tr.st | PAT.ID

Structure: General positive-definite, Log-Cholesky parametrization

StdDev Corr

(Intercept) 1.987 (Intr)

CYL.mean.tr.st 0.443 -0.695

Residual 1.009

Fixed effects: BAP ~ PROGhigh6 + BMI0high + AGE.st

	Value	Std.Error	DF	t-value	p-value
(Intercept)	8.94	0.406	123	22.01	0.000
PROGhigh61	-0.78	0.241	123	-3.26	0.001
BMI0high1	1.56	0.637	43	2.44	0.019
AGE.st	-0.57	0.309	43	-1.86	0.070

Correlation:

(Intr) PROG61 BMI0h1

PROGhigh61 -0.466

BMI0high1 -0.524 0.026

AGE.st 0.078 0.067 -0.193

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.2004	-0.5455	-0.0369	0.4659	2.4052

Number of Observations: 170

Number of Groups: 46

Hence we conclude, that also for the bone formation marker BAP, the age at study onset of the respective participant has a significant influence. High values of progesterone and *no* overweight at study onset provoke BAP to be low.

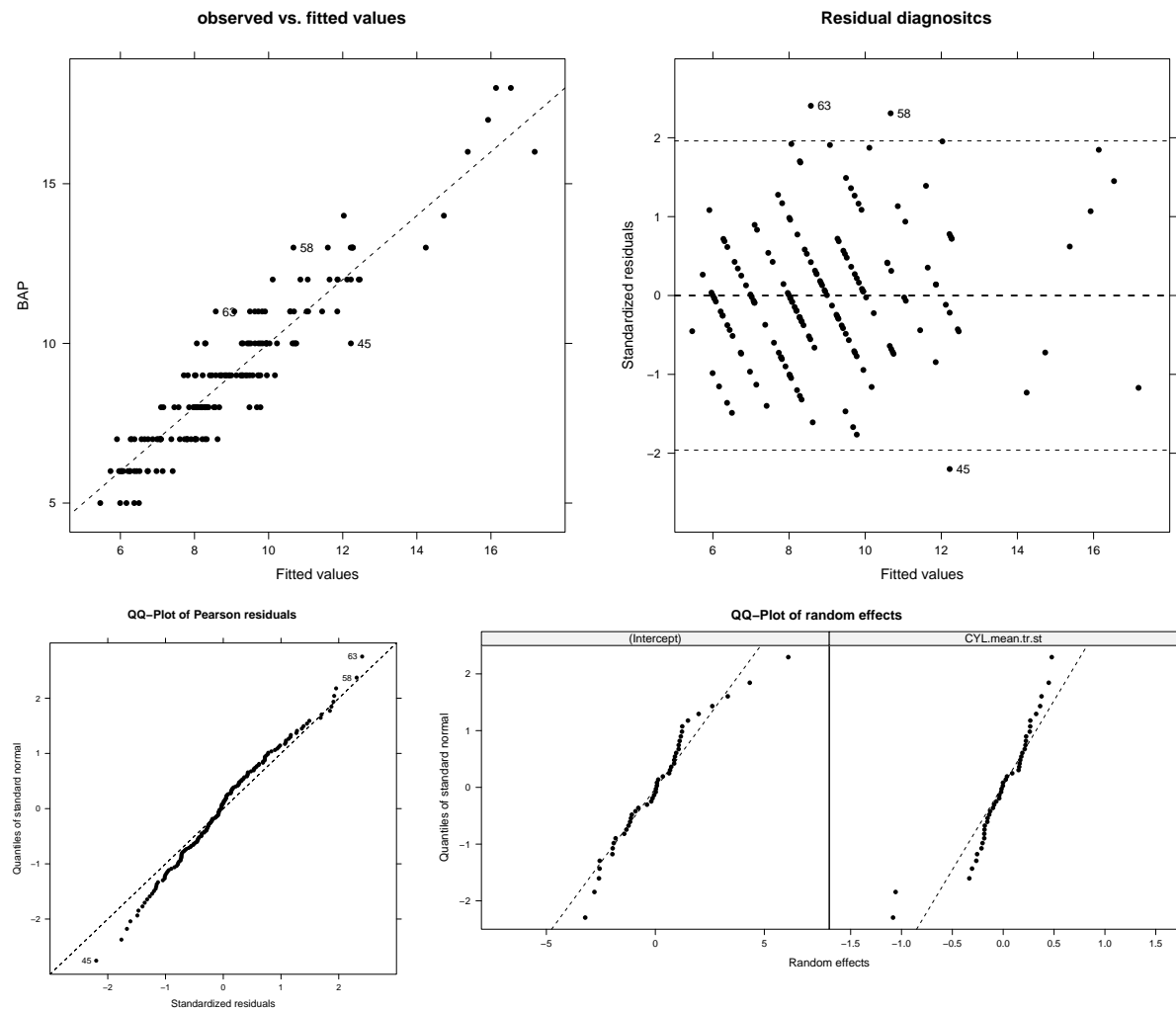


Figure 11.19.: Model diagnostics for `mod.BAP.CYL`

11.6. Correlation between the bone markers

As has been mentioned already briefly in Chapter 9, bone formation and bone resorption processes always take place at the same time (Seifert-Klauss et al. (2011)). By considering Figure 9.10(b) we noticed, that in particular the bone resorption marker CTX and the bone formation marker osteocalcin are highly correlated. Although the percental increase of CTX was illustrated already to be on average higher than the increase of the bone formation markers when BMD is lost (see Figure bpXMarkPR), we aim to investigate the occuring correlations among the bone markers further.

To this end, we consider the standardized residuals, i.e. the conditional Pearson residuals, of the linear mixed models (11.3), (11.7) and (11.9) which were adjusted in the previous sections to explain CTX, OST and BAP, respectively. These standardized residuals can be assumed to be approximately independent and identically (normally) distributed (see 7.3). We declare the respective residual vectors of the different models by the acronym "r.":

```
> r.CTX = residuals(mod.CTX.logE2, type="p")
> r.OST = residuals(mod.OST.anovlong, type="p")
> r.BAP = residuals(mod.BAP.CYL, type="p")
```

Comparing the Kendall tau rank correlation coefficients of the different bone markers,

```
> cor(cbind(CTX, OST, BAP), method = "kendall" )
      CTX  OST  BAP
CTX 1.000 0.511 0.219
OST 0.511 1.000 0.246
BAP 0.219 0.246 1.000
```

and the corresponding residual vectors,

```
> cor(cbind(r.CTX, r.OST, r.BAP), method = "kendall" )
      r.CTX r.OST r.BAP
r.CTX 1.000 0.324 0.059
r.OST 0.324 1.000 0.129
r.BAP 0.059 0.129 1.000
```

indicates that correlation is reduced, when explaining the data by the respective linear mixed model. However, especially CTX and OST are still highly correlated, indicated by a Kendall correlation coefficient of 0.324 of r.CTX and r.OST. Figure 11.20 shows the corresponding pairs plots illustrating these findings.

In addition, Figure 11.21 shows contour plots of the bivariate empirical distributions of the standardized residuals, generated in R using the function `BiCopMetaContour` of the R package `CDvine`. One may note that the (inner) contour lines of the middle graphic nearly form circles which emphasizes that CTX and BAP are not highly correlated. The basic form of the left and the right pattern is not of a circular kind of shape, indicating the correlation among the respective variables.

Furthermore, it can be concluded that the joint distributions are probably not Gaussian, since the contour lines are not elliptic.

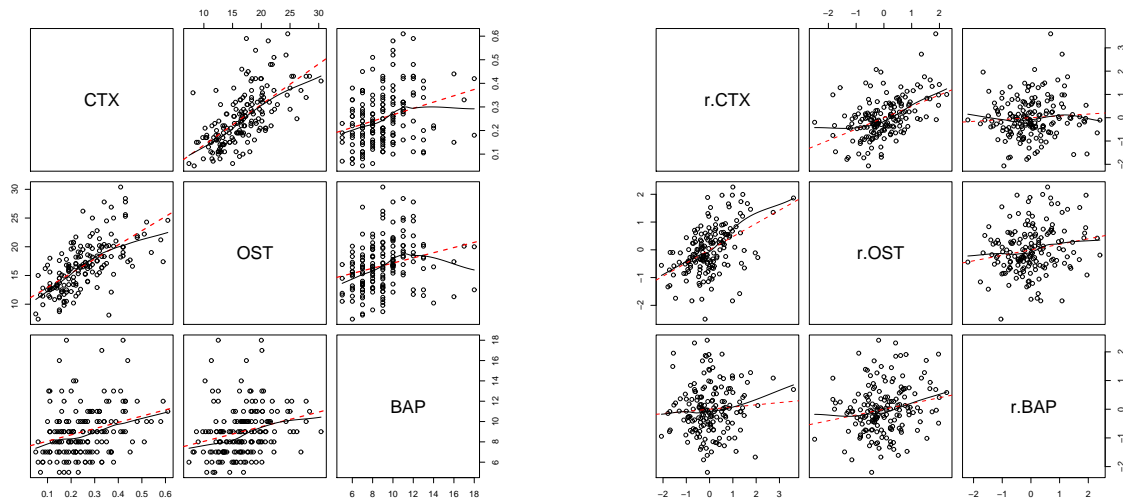


Figure 11.20.: Pairs plots of the bone markers and their conditional Pearson residuals corresponding to models (11.3), (11.7) and (11.9) with linear regression and LOESS lines

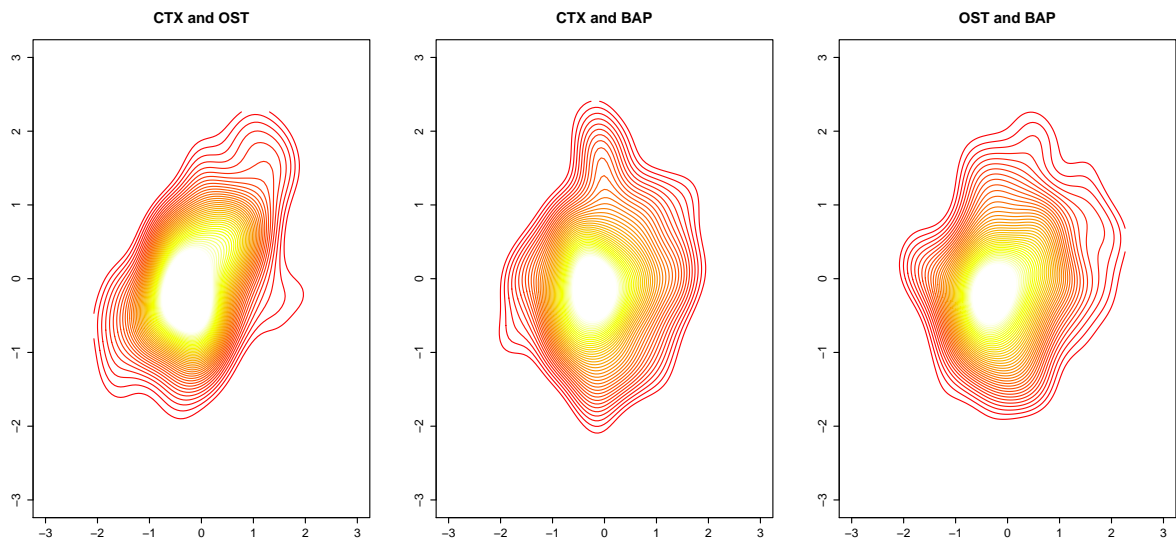


Figure 11.21.: Contour plots of the bivariate empirical distributions of the standardized residuals of the bone markers CTX, OST and BAP corresponding to models (11.3), (11.7) and (11.9)

12. Conclusions and outlook

Due to the high correlations among the covariates and the comparatively small data set, modeling the change of bone density during the course of the study has not been straightforward. However, the assumption that high mean serum concentrations of FSH provoke a higher loss of BMD, could be clearly confirmed by the *PeKnO-Study* (see Table 10.1). Furthermore, high mean serum concentrations of estradiol were shown to have a significant positive effect on the change of bone density. Unfortunately, a similar statement for the mean concentration of progesterone can not be made.

A high percentage of ovulatory cycles suggests that less BMD will be lost and vice versa for anovulatory cycles, in particular for long anovulatory cycles. Finally, high mean lengths of the cycles were detected to have a strong negative effect on the change of bone density. This was found out by transforming the corresponding variable. An appropriate model to explain and predict the change of bone density over the course of two years is given by (10.2). This model is able to explain 42% of the arising variability in the data.

As explained in Chapter 9, one unique feature of the *PeKnO-Study* is that participants applied a cycle monitor which allows to classify cycles as being ovulatory or anovulatory and to derive the (likely) lengths of corresponding luteal phases. The latter variable was shown to be significant at level 10% with respect to a t-test in a simple standard linear model. Using this model, 6.5% of the variation within the data can be explained (again referring to Table 10.1).

On the basis of a graphical analysis at the end of Chapter 9 we saw that the loss of bone density is accompanied by an increase of the bone marker concentrations. The percental increase of the bone markers, however, was seen to be highest for the bone resorption marker CTX. The bone formation markers osteocalcin and BAP showed a smaller increase on average.

In Chapter 11 we found out that the bone turnover markers can be appropriately explained using linear mixed models. Keeping in mind that the data set consists out of relatively few observations, the model diagnostics have basically shown no lack of fit. It turned out that the age of the women at study onset is one determinant factor for the bone marker concentrations. A fixed effect of this variable was shown to be significant in linear mixed model of each single bone marker in the way, that the older the women have been at the begin of the study, the lower their levels of CTX, osteocalcin and BAP.

Furthermore, when modeling CTX and BAP, progesterone was detected to have a significant negative effect on these variables. That is, high concentrations of progesterone lead to lower values of CTX and BAP.

When we considered the correlations between the bone markers in Section 11.6, we already outlined briefly that the corresponding bivariate distributions are probably not Gaussian. Therefore, and for the reason of too little data, a further investigation of the bone markers applying multivariate linear mixed models (see Boik (1988)) can hardly be justified.

However, a further investigation of the joint bone marker distributions using copulas or vines might reveal some new insights.

Appendix

A. Mathematical tools

A.1. Matrix Algebra

In this section we summarize some basic results concerning matrix calculations which frequently are used in this thesis to make clear which property we use in the single computations. Basic and commonly known propositions are listed without giving a proof. For further reading or proofs please consult some standard literature on linear algebra.

A.1.1. Properties of the determinant

- $A \in \mathbb{R}^n$, then $|A| \neq 0 \Leftrightarrow \text{rank}(A) = n$
- $|AB| = |A| |B|$
- $|A^{-1}| = |A|^{-1}$

The determinant of quadratic block matrices:

(1) Let A and D be quadratic matrices, then it holds:

$$\left| \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} \right| = \left| \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} \right| = |A| |D|$$

(2) Let M be a nonsingular, quadratic block matrix $\begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix}$, where M_{11} and M_{22} are quadratic and M_{22} is invertible. Then M can be written as

$$M = \begin{pmatrix} I & M_{12} \\ 0 & M_{22} \end{pmatrix} \begin{pmatrix} M_{11} - M_{12}M_{22}^{-1}M_{12}^T & 0 \\ M_{22}^{-1}M_{12}^T & I \end{pmatrix}$$

Hence, the determinant of M may be calculated as

$$|M| = |M_{22}| |M_{11} - M_{12}M_{22}^{-1}M_{12}^T|$$

A.1.2. Properties of the trace of a matrix

- $\text{tr}(A) = \text{tr}(A^T)$
- $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
- $\text{tr}(A \cdot B) = \text{tr}(BA)$

- $\text{tr}(kA) = k \text{tr}(A), k \in \mathbb{R}$
- $\text{tr}(\mathbf{x}\mathbf{y}^T) = \text{tr}(\mathbf{y}\mathbf{x}^T) = \text{tr}(\mathbf{x}^T\mathbf{y}) = \mathbf{x}^T\mathbf{y}$
- A positive definite and B positive semidefinite $\Rightarrow \text{tr}(AB) \geq 0$

Mean of a quadratic form

Let \mathbf{Y} be a random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix V and let A be a symmetric matrix, then

$$\mathbb{E}[\mathbf{Y}^T A \mathbf{Y}] = \text{tr}(AV) + \boldsymbol{\mu}^T A \boldsymbol{\mu} \quad (\text{A.1})$$

Proof:

$$\mathbf{Y}^T A \mathbf{Y} = \text{tr}(\mathbf{Y}^T A \mathbf{Y}) = \text{tr}(A \boldsymbol{\mu} \boldsymbol{\mu}^T)$$

$$\begin{aligned} \Rightarrow \mathbb{E}[\mathbf{Y}^T A \mathbf{Y}] &= \mathbb{E}[\text{tr}(A \mathbf{Y} \mathbf{Y}^T)] = \text{tr}(A \mathbb{E}[\mathbf{Y} \mathbf{Y}^T]) = \text{tr}(A(V + \boldsymbol{\mu} \boldsymbol{\mu}^T)) \\ &= \text{tr}(AV) + \text{tr}(A \boldsymbol{\mu} \boldsymbol{\mu}^T) = \text{tr}(AV) + \boldsymbol{\mu}^T A \boldsymbol{\mu} \end{aligned}$$

■

A.1.3. Differentiation of vector valued functions

Let $\mathbf{y} \in \mathbb{R}^n$ and $f(\mathbf{y})$ be a differentiable function of the n elements of \mathbf{y} , then the gradient of $f(\mathbf{y})$ is

$$\begin{aligned} \frac{df(\mathbf{y})}{d\mathbf{y}} &= \begin{pmatrix} \frac{\partial f(\mathbf{y})}{\partial y_1} \\ \vdots \\ \frac{\partial f(\mathbf{y})}{\partial y_n} \end{pmatrix} \text{ and} \\ \frac{d^2 f(\mathbf{y})}{d\mathbf{y} d\mathbf{y}^T} &= \left[\frac{\partial^2 f(\mathbf{y})}{\partial y_i \partial y_j} \right]_{i,j=1,\dots,n} \end{aligned}$$

denotes the corresponding Hessian matrix.

Then, for $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^n$:

- $\frac{dA\mathbf{y}}{d\mathbf{y}} = A$
- $\frac{d\mathbf{x}^T \mathbf{y}}{d\mathbf{y}} = \mathbf{x}$
- $\frac{d\mathbf{y}^T A \mathbf{y}}{d\mathbf{y}} = (A + A^T)\mathbf{y}$
- A symmetric, then $\frac{d\mathbf{y}^T A \mathbf{y}}{d\mathbf{y}} = 2A\mathbf{y}$.

A.1.4. Overview of properties of special matrices in a linear mixed model

One can easily verify that the following matrices auxiliary introduced in Chapter 5 satisfy the listed statements:

$$(1) \quad T := X(X^T V^{-1} X)^{-1} X^T V^{-1}$$

- T is idempotent
- $TVT^T = TV = VT^T$
since $TV = X(X^T V^{-1} X)^{-1} X^T = VT^T$ and $TV = TTV = TVT^T$
- $X\hat{\beta} = Ty$

$$(2) \quad Q := I - T$$

- Q is idempotent
- $QVQ^T = QV = VQ^T$
- $TVQ^T = QVT^T = 0$
since $TVQ^T = TV(I - T)^T = TV - TVT^T \stackrel{(1)}{=} TV - TV = 0 = QVT^T$

$$(3) \quad P := V^{-1}Q = V^{-1}(I - T) = V^{-1} - V^{-1}X(X^T V^{-1} X)^{-1} X^T V^{-1}$$

- P is symmetric
- $PT = T^T P = 0$
- $PVP = P$
- $P\mathbf{y} = V^{-1}(\mathbf{y} - X\hat{\beta})$
- $\mathbf{y}^T P \mathbf{y} = (\mathbf{y} - X\hat{\beta})^T V^{-1}(\mathbf{y} - X\hat{\beta})$
since $\mathbf{y}^T P \mathbf{y} = \mathbf{y}^T V^{-1} \mathbf{y} - \mathbf{y}^T V^{-1} X \hat{\beta} = \mathbf{y}^T V^{-1} \mathbf{y} - 2\mathbf{y}^T V^{-1} X \hat{\beta} + \underbrace{\mathbf{y}^T V^{-1} X}_{= \hat{\beta}^T (X^T V^{-1} X)} \hat{\beta}$

A.2. Statistical tools

A.2.1. The multivariate normal distribution

Let $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, then the density of \mathbf{X} is

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

The conditional distribution

$$\begin{aligned} \mathbf{X} &= \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \Sigma_y & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_z \end{pmatrix} \\ \Rightarrow \quad \mathbf{Z} \mid \mathbf{Y} &\sim N(\boldsymbol{\mu}_z + \Sigma_{zy} \Sigma_y^{-1} (\mathbf{y} - \boldsymbol{\mu}_y), \Sigma_z - \Sigma_{zy} \Sigma_y^{-1} \Sigma_{yz}) \end{aligned}$$

A.2.2. The mean squared error

The MSE

Let β be a fixed but unknown scalar coming from a certain parameter space and let $\hat{\beta}$ denote an estimator for β . Then the mean squared error of $\hat{\beta}$ is defined as:

$$\text{MSE}(\hat{\beta}) := \mathbb{E}[(\hat{\beta} - \beta)^2]$$

Using that β is constant, one easily may show that

$$\text{MSE}(\hat{\beta}) = \text{Var}(\hat{\beta}) + \mathbb{E}[\hat{\beta} - \beta]^2$$

where the second term is the bias of $\hat{\beta}$.

In case that $\hat{\beta}$ is unbiased, meaning that $\mathbb{E}[\hat{\beta}] = \beta$:

$$\text{MSE}(\hat{\beta}) = \text{Var}(\hat{\beta} - \beta) = \text{Var}(\hat{\beta}) .$$

The MSPE

Let γ be a random variable underlying a certain distribution. The mean squared prediction error of a predictor $\hat{\gamma}$ for γ is defined as:

$$\text{MSPE}(\hat{\gamma}) := \mathbb{E}[(\hat{\gamma} - \gamma)^2] .$$

In case that $\hat{\gamma}$ is unbiased, meaning that $\mathbb{E}[\hat{\gamma}] = \mathbb{E}[\gamma]$:

$$\text{MSPE}(\hat{\gamma}) = \text{Var}(\hat{\gamma} - \gamma) .$$

Note that in both cases unbiasedness of an estimator and a predictor includes $\mathbb{E}[\hat{\beta} - \beta] = 0$ and $\mathbb{E}[\hat{\gamma} - \gamma] = 0$, respectively. But since γ is a random variable, to ask for $\mathbb{E}[\hat{\gamma}] = \gamma$ does not make sense. Neither we are able to split the MSPE of $\hat{\gamma}$ into the sum of its variance and its bias as in the case above.

A.2.3. Asymptotic normality of the MLE

Using the notations given in A.1.3 for differentiation of a vector valued function, one may formulate the well known result:

Theorem A.2.1 Bickel and Doksum (2001, Theorem 6.2.2.)

Let Y_1, \dots, Y_n be independent and identical distributed random variables following a distribution which depends on a unknown model parameter $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_r)^T$, where

$$\boldsymbol{\vartheta} \in \Lambda \subset \mathbb{R}^r, \Lambda \text{ open.}$$

Let $f(y, \boldsymbol{\vartheta})$ denote the density function of Y_i , $i = 1, \dots, n$, and $l(y, \boldsymbol{\vartheta}) := \ln f(y, \boldsymbol{\vartheta})$.

In addition, let the following regularity conditions being satisfied:

- (1) $\frac{\partial l(y, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}$ and $\frac{\partial^2 l(y, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T}$ are well defined
- (2) $E_{\mathbb{P}_{\boldsymbol{\vartheta}}} \left[\frac{\partial l(Y_1, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right] = \mathbf{0}$
- (3) $E_{\mathbb{P}_{\boldsymbol{\vartheta}}} \left[\left\| \frac{\partial l(Y_1, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right\|^2 \right] < \infty$
- (4) $E_{\mathbb{P}_{\boldsymbol{\vartheta}}} \left[\left\| \frac{\partial^2 l(Y_1, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} \right\| \right] < \infty$ and $E_{\mathbb{P}_{\boldsymbol{\vartheta}}} \left[\frac{\partial^2 l(Y_1, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} \right]$ is nonsingular
- (5) $\sup_{\|\mathbf{t} - \boldsymbol{\vartheta}\| \leq \epsilon} \left\| \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^2 l(Y_i, \mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^T} - \frac{\partial^2 l(Y_i, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} \right) \right\| \xrightarrow{\mathbb{P}} 0$ if $\epsilon \rightarrow 0$
- (6) $-E_{\mathbb{P}_{\boldsymbol{\vartheta}}} \left[\frac{\partial^2 l(Y_1, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} \right] = E_{\mathbb{P}_{\boldsymbol{\vartheta}}} \left[\left(\frac{\partial l(Y_1, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right) \left(\frac{\partial l(Y_1, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right)^T \right],$
 where $E_{\mathbb{P}_{\boldsymbol{\vartheta}}} \left[\left(\frac{\partial l(Y_1, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right) \left(\frac{\partial l(Y_1, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right)^T \right] = I(\boldsymbol{\vartheta})$ is the Fisher information matrix.

Let $\hat{\boldsymbol{\vartheta}}$ be the MLE of $\boldsymbol{\vartheta}$.

Then,

$$\sqrt{n} (\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \xrightarrow{d} N(0, I(\boldsymbol{\vartheta})^{-1}).$$

A.3. Proofs of some used results

Lemma A.3.1

Let $X \in \mathbb{R}^{n \times p}$ with $\text{rank}(X) = p$ and $V \in \mathbb{R}^{n \times n}$ be a covariance matrix, i.e. symmetric and positive definite. Then $X^T V^{-1} X \in \mathbb{R}^{p \times p}$ is positive definite, in particular regular.

Proof: The positive definiteness of V implies that V^{-1} is positive definite, i.e. for any non-zero vector $\mathbf{a} \in \mathbb{R}^n$ it holds:

$$\mathbf{a}^T V^{-1} \mathbf{a} > 0.$$

To check this property for $X^T V^{-1} X$, we consider for any non-zero vector $\mathbf{c} \in \mathbb{R}^p$:

$$\begin{aligned} \mathbf{c}^T (X^T V^{-1} X) \mathbf{c} &= \\ (X \mathbf{c})^T V^{-1} (X \mathbf{c}) &> 0 \end{aligned}$$

since $X \mathbf{c}$ is a non-zero vector due to the regularity of X and applying that V^{-1} is positive definite. Hence we can follow that $X^T V^{-1} X$ is positive definite and therefore invertible. ■

Lemma A.3.2

Let $X \in \mathbb{R}^{n \times p}$ with $\text{rank}(X) = p$ and $V \in \mathbb{R}^{n \times n}$ be a covariance matrix, i.e. sym-

metric and positive definite. Further, let $A \in \mathbb{R}^{n \times (n-p)}$ with $\text{rank}(A) = n - p$ and $A^T X = 0$. Then $(V^{-1}X, A) \in \mathbb{R}^{n \times n}$ is regular.

Proof: Since V^{-1} is symmetric and positive definite, there is a decomposition

$$T^T V^{-1} T = D = \begin{pmatrix} \lambda_1 & \cdots & \cdots & 0 \\ \vdots & \lambda_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_n \end{pmatrix} \text{ with } \lambda_i > 0, i = 1, \dots, m \text{ and } T^T T = I .$$

Since T is regular

$$\begin{aligned} \text{rank}(V^{-1}X, A) = n &\Leftrightarrow \text{rank}(T^T (V^{-1}X, A)) = n \\ &\Leftrightarrow \text{rank}(T^T V^{-1} T T^T X, T^T A) = n \\ &\Leftrightarrow \text{rank}(D \underbrace{T^T X}_{X'}, \underbrace{T^T A}_{A'}) = n . \end{aligned}$$

X' and A' have full rank and $A'^T X' = A'^T T T^T X = A'^T X = 0$. Hence it is enough to consider the case that V^{-1} has diagonal form.

Let \mathbf{x} be a linear combination of columns of X' and \mathbf{a} be a linear combination of columns of A' , with

$$D\mathbf{x} = (\lambda_1 x_1, \dots, \lambda_n x_n)^T = \mathbf{a} .$$

The columns of A' and X' are orthogonal, so we get

$$\begin{aligned} 0 &= \langle \mathbf{x}, \mathbf{a} \rangle \\ &= \frac{1}{\lambda_1} \lambda_1 x_1 a_1 + \cdots + \frac{1}{\lambda_n} \lambda_n x_n a_n \\ &= \frac{1}{\lambda_1} a_1^2 + \cdots + \frac{1}{\lambda_n} a_n^2 \\ &= a_1^2 / \lambda_1 + \cdots + a_n^2 / \lambda_n . \end{aligned}$$

Hence $\mathbf{a} = \mathbf{0}$ and since D is regular $\mathbf{x} = \mathbf{0}$. Thus scaling X' by D does not produce linear dependence between the columns of DX' and A , i.e. $\text{rank}(DX', A) = \text{rank}(X', A) = n$ which proves the lemma. ■

B. Additional information on the PeKnO-Study

B.1. Summary of variables

The following table provides an overview of variables observed in the *PeKnO-Study*. It specifies their acronyms used in this thesis as well as a compact characterization which summarizes the elaborate descriptions of Chapter 9. The last two columns indicate the means and standard deviations of the variables corresponding to available data of all 51 participants. Only these serum concentrations of the female sex hormones, cortisol and the bone turnover markers are incorporated where blood samples are classified as taken in the *correct* time interval (luteal phase blood tests).

To indicate whether a variable depends only on the participant or also on time (i.e. on the number of the blood sample), the corresponding acronyms are not labeled or labeled by t , respectively. Thus, the subscript indicates that there were more than one measurements per participant recorded. These variables are considered in the longitudinal analysis of Chapter 11. Variables are basically categorized according to Chapter 9. At first the variables characterizing the bone metabolism which are used as response variables in the adjusted models of this thesis are presented.

ACRONYM	MEANING	DEF	MEAN	SD
	Response variables			
X.BMD	Difference of the bone density at the end and the begin of the study in mgCa-Ha/ml	QCT.2 – QCT.1	-6.1	8.8
CTX _t	Serum concentration of the bone resorption marker CTX in ng/ml	-	0.3	0.1
OST _t	Serum concentration of the bone formation marker osteocalcin in ng/ml	-	16.8	4.4
BAP _t	Serum concentration of the bone formation marker BAP in $\mu\text{g/l}$ given on a ordinal scale with usual values from 4 to 21	-	9.0	2.6

Basic characterizations of participants				
PAT.ID	Number of the participant, irregular ranged between 2 and 63	-	-	-
AGE	Age at the begin of the study in years	-	48.2	2.6
BMI.25	Body mass index in kg/m ² at 25 years of age	$\left\{ \frac{\text{weight}[\text{kg}]}{(\text{height}[\text{m}])^2} \right\}$	21.8	3.1
BMI.0	Body mass index in kg/m ² at the begin of the study		24.0	3.5
BMI.1	Body mass index in kg/m ² at the end of the study		24.3	3.7
X.BMI.0.25	Difference of the body mass index at the begin of the study and at 25 years of age in kg/m ²	BMI.0 – BMI.25	2.2	2.9
X.BMI.1.25	Difference of the body mass index at the end of the study and at 25 years of age in kg/m ²	BMI.1 – BMI.25	2.5	3.2
X.BMI.1.0	Difference of the body mass index at the end and the begin of the study in kg/m ²	BMI.1 – BMI.0	0.3	1.2
Measurements of the bone density				
QCT.1	Bone mineral density in mgCa-Ha/ml at the begin of the study	-	139.3	24.2
QCT.2	Bone mineral density in mgCa-Ha/ml at the end of the study	-	134.0	24.2
X.BMD	<i>see above</i>			
X.BMD.PR	Percental change of bone density with respect to the begin of the study	$\frac{X.BMD}{QCT.1} \cdot 100\%$	-4.2	6.3
BMD.GROUP	Category of the change of bone density: <ul style="list-style-type: none"> • = 11, if QCT.1 > 120 and QCT.2 > 120 • = 12, if QCT.1 > 120 and QCT.2 < 120 • = 22, if QCT.1 < 120 and QCT.2 < 120 	-	-	-
Blood sample associated variables				
NO.BS _t	Number of the blood sample	-	-	-
Female sex hormones				
LH _t	Serum concentration of the luteinizing hormone (LH) in IU/l	-	10.9	12.2
LH.m	Mean serum concentration of LH in IU/l	-	11.0	9.1
FSH _t	Serum concentration of the follicle-stimulating hormone (FSH) in IU/l	-	15.5	21.8
FSH.m	Mean serum concentration of FSH in IU/l	-	15.8	15.6
E2 _t	Serum concentration of (17β-) estradiol in pg/ml	-	170.9	132.9

E2.m	Mean serum concentration of (17 β -) estradiol in pg/ml	-	164.6	78.8
PROG _t	Serum concentration of progesterone in ng/ml	-	10.1	8.7
PROG.m	Mean serum concentration of progesterone in ng/ml	-	9.6	6.8
Cortisol				
CORT _t	Serum concentration of cortisol basal (stress hormone) in μ g/dl	-	16.4	5.2
CORT.m	Mean serum concentration of cortisol basal (stress hormone) in μ g/dl	-	16.1	4.1
Bone turnover markers				
CTX _t	<i>see above</i>			
CTX.m	Mean serum concentration of the bone resorption marker CTX in ng/ml	-	0.2	0.1
X.CTX	Difference between the bone resorption marker CTX at blood sample 5 (CTX.2) and blood sample 1 (CTX.1) in ng/ml	CTX.2 – CTX.1	0.04	0.12
X.CTX.PR	Percental change of CTX with respect to the first blood test	$\frac{X.CTX}{CTX.1} \cdot 100\%$	32.70	88.38
OST _t	<i>see above</i>			
OST.m	Mean serum concentration of the bone formation marker osteocalcin in ng/ml	-	16.5	4.1
X.OST	Difference between the bone formation marker osteocalcin at blood sample 5 (OST.2) and blood sample 1 (OST.1) in ng/ml	OST.2 – OST.1	1.11	4.67
X.OST.PR	Percental change of osteocalcin with respect to the first blood test	$\frac{X.OST}{OST.1} \cdot 100\%$	11.66	32.71
BAP _t	<i>see above</i>			
BAP.m	Mean serum concentration of the bone resorption marker BAP in μ g/l	-	9.0	2.4
X.BAP	Difference between the bone formation marker BAP at blood sample 5 (BAP.2) and blood sample 1 (BAP.1) in μ g/l	BAP.2 – BAP.1	1.24	2.57
X.BAP.PR	Percental change of BAP with respect to the first blood test	$\frac{X.BAP}{BAP.1} \cdot 100\%$	17.97	36.26
Cycle associated variables				
CY.BS _t	Type of the cycle where the blood test was performed <ul style="list-style-type: none"> • = 1, if cycle is ovulatory • = 0, if cycle is anovulatory 			
CYL.BS _t	Length of the cycle at which the blood test was performed	-	51.3	67.1

CYL.BS.tr _t	<i>transformation</i>	$\frac{1}{\text{CYL.BS}^2}$	0.0011	0.0006
LPL.BS _t	(Likely) length of the luteal phase of the cycle at which the blood test was performed	(9.3)	6.6	6.1
CYL.mean _t	Mean length of the cycles since the last blood test	-	44.0	56.4
CYL.mean.tr _t	<i>transformation</i>	$\frac{1}{\text{CYL.mean}^2}$	0.0011	0.0005
CYL.m	Mean length of the cycles during the whole study	-	45.0	38.0
CYL.m.tr	<i>transformation</i>	$\frac{1}{\text{CYL.m}^2}$	0.0010	0.0005
LPL.mean _t	Mean of the (likely) length of luteal phases of cycles since the last blood test	(9.3)	6.2	4.6
LPL.m	Mean of the (likely) length of luteal phases of cycles during the whole study	(9.3)	6.1	3.5
Classification of cycles				
ov.tot.PR _t	Percentage of ovulatory cycles with respect to all evaluable cycles since the last blood test	(9.1)	54.3	37.8
ov.tot.s.PR	Percentage of ovulatory cycles with respect to all evaluable cycles during the whole study	(9.2)	52.1	27.3
anov.tot.PR _t	Percentage of anovulatory cycles with respect to all evaluable cycles since the last blood test	1-ov.tot.PR	45.7	37.8
anov.tot.s.PR	Percentage of anovulatory cycles with respect to all evaluable cycles during the whole study	1 - ov.tot.s.PR	47.9	27.3
anov.short.PR _t	Percentage of short anovulatory cycles (cycle length < 20d) with respect to all evaluable cycles since the last blood test	similarly as (9.1)	4.2	10.9
anov.short.s.PR	Percentage of short anovulatory cycles with respect to all evaluable cycles during the whole study	similarly as (9.2)	5.2	6.9
anov.norm.PR _t	Percentage of anovulatory cycles with normal cycle (20d ≤ cycle length ≤ 42d) length with respect to all evaluable cycles since the last blood test	similarly as (9.1)	25.2	28.6
anov.norm.s.PR	Percentage of anovulatory cycles with normal cycle with respect to all evaluable cycles during the whole study	similarly as (9.2)	26.5	17.2
anov.long.PR _t	Percentage of long anovulatory cycles (cycle length > 42d) with respect to all evaluable cycles since the last blood test	similarly as (9.1)	16.3	27.5
anov.long.s.PR	Percentage of long anovulatory cycles with respect to all evaluable cycles during the whole study	similarly as (9.2)	16.2	18.2
Transformations of variables				
-	Basic form: <i>function (log/sqrt) - acronym</i> e.g.	-	-	-
sqrtPROG.m	<i>transformation</i>	$\sqrt{\text{PROG.m}}$	2.9	1.2

	Standardized version of variables			
<i>acronym.st</i>	Respective variable minus its mean divided by its standard deviation	e.g. (11.1)	-	-
	Dummy version of variables			
-	Basic form: <i>acronym - condition (high/low) - cutoff value for satisfied condition</i> , variable takes on the values <ul style="list-style-type: none"> • = 1, if the condition is satisfied • = 0, if the condition is not satisfied 	e.g. p. 93	-	-

Table B.1.: Overview of variables

B.2. Summary of models

ACRONYM	RESPONSE	REFERENCE	PAGE
<i>mod.fin</i>	X.BMD	(10.1)	98
<i>mod.CTX.logE2</i>	CTX	(11.3)	123
<i>mod.CTX.FSH</i>	CTX	(11.4)	127
<i>mod.CTX.ri.red</i>	CTX	(11.5)	129
<i>mod.OST.anovlong</i>	OST	(11.7)	135
<i>mod.BAP.CYL</i>	BAP	(11.9)	140

Table B.2.: Overview of models

B.3. Additional graphics

Further illustrative graphics corresponding to investigations in Chapter 9 and Chapter 10

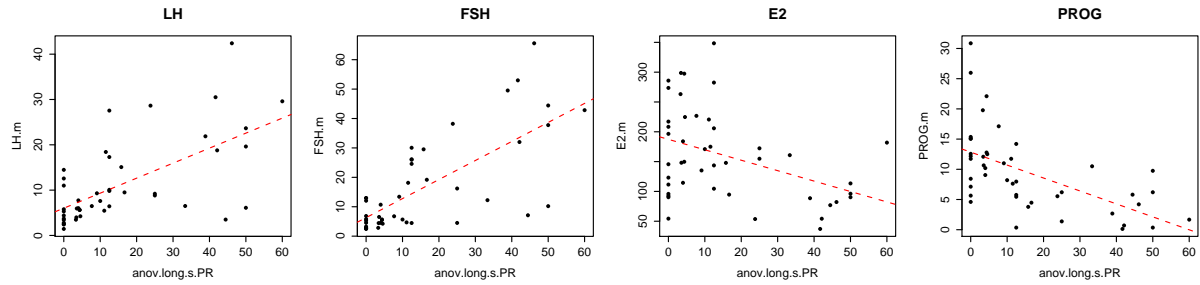


Figure B.1.: Scatter plot of the percentage of long anovulatory cycles during the whole study and the mean values of the luteal phase serum concentrations of the female sex hormones

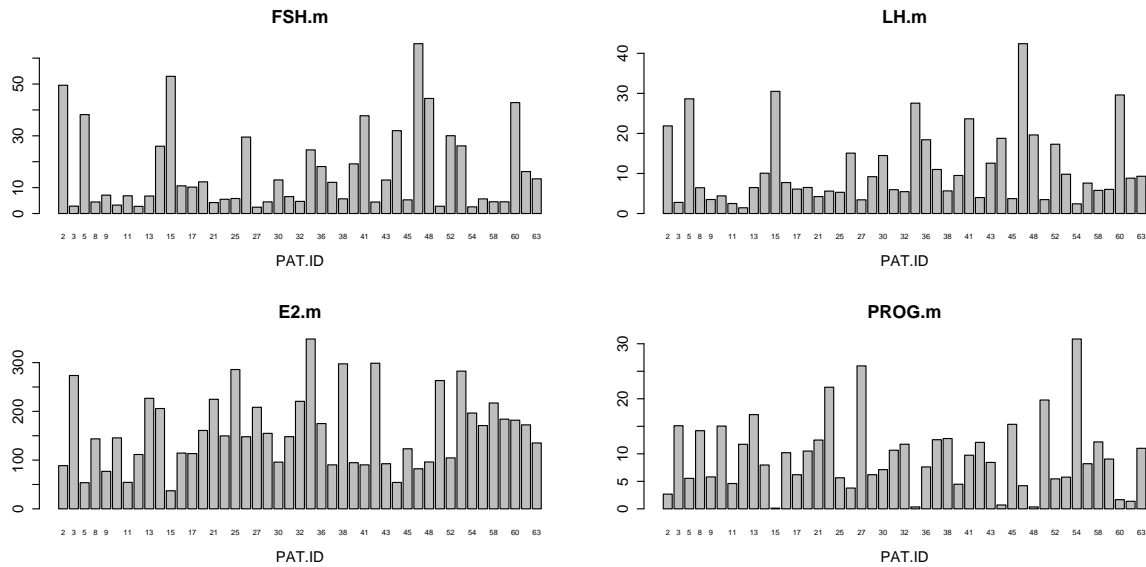


Figure B.2.: Barplots of individual mean values of the luteal phase serum concentrations of the female sex hormones

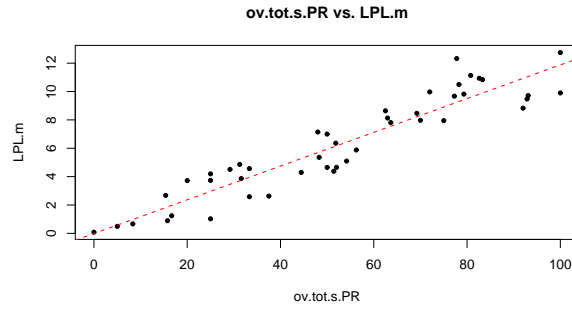


Figure B.3.: Scatter plot of the percentage of ovulatory cycles during the whole study and the mean length of luteal phases LPL.m

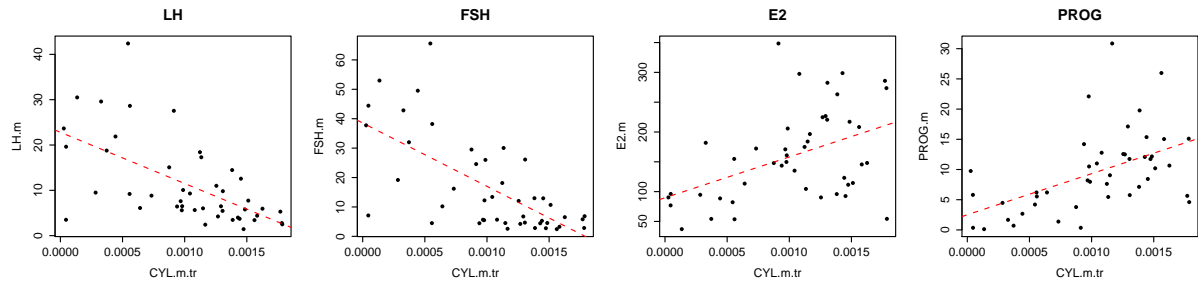


Figure B.4.: Scatter plot of the transformed mean length of cycles and the mean values of the luteal phase serum concentrations of the female sex hormones

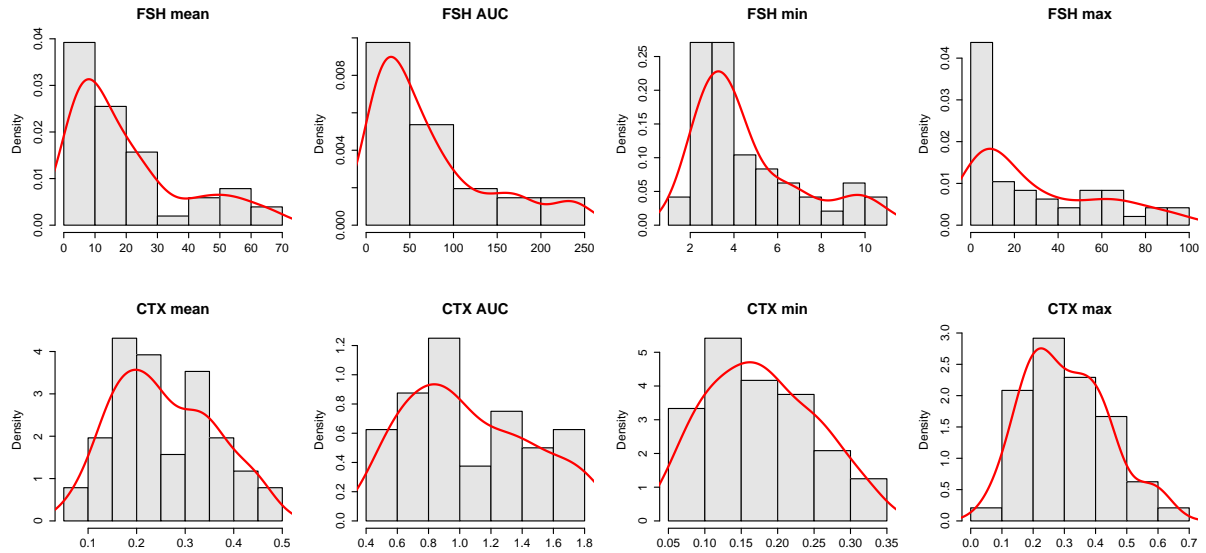


Figure B.5.: Histograms of summarized data exemplarily for FSH and CTX to compare different approaches (*mean*, *AUC*, *min*, *max*) for summarizing longitudinal data

Further illustrative graphics corresponding to investigations in Chapter 11

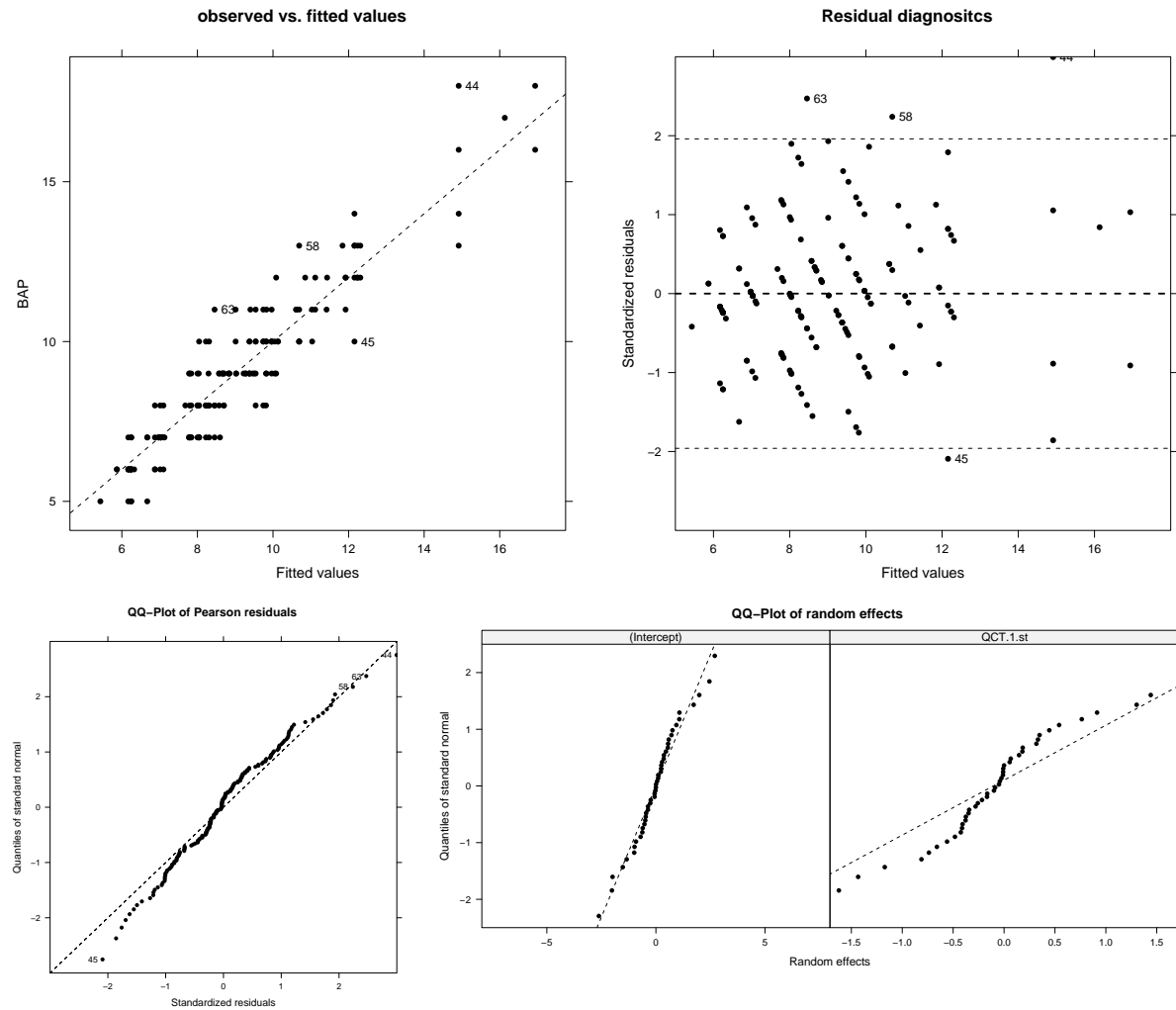


Figure B.6.: Model diagnostics for an alternative linear mixed model for modeling BAP with random slope QCT.1.st. The model can briefly formulated as: $BAP_{it} = \beta_0 + \beta_1 \text{PROGhigh6}_{it} + \beta_2 \text{BMI0high}_i + \beta_3 \text{AGE.st}_i + \gamma_{i0} + \gamma_{i1} \text{QCT.1.st}_{it} + \epsilon_{it}$

References

- Aitken, A. C. (1934). On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh, Sec A* 55, 42–47.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19, 716–723.
- Bartl, R. (2009). *Osteoporosis - diagnosis, prevention, therapy*. Berlin and Heidelberg: Springer.
- Bickel, P. J. and K. A. Doksum (2001). *Mathematical Statistics*. Upper Saddle River, New Jersey: Prentice Hall. Basic ideas and selected topics, Volume I, Second Edition.
- Boik, R. J. (1988). The mixed model for multivariate repeated measures: validity conditions and an approximate test. *Psychometrika* 53(4), 469–486.
- Clearblue (2007). Clearblue®easy fertility monitor - Instruction booklet. <http://www.clearblueeasy.com/pdfs/clearblue-easy-fertility-monitor-pdf.php>.
- Cnaan, A., N. M. Laird, and P. Slasor (1997). Tutorial in Biostatistics, Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in medicine* 16, 2349–2380.
- Crainiceanu, C. M. and D. Ruppert (2003). Proofs of theorems for the paper ”likelihood ratio tests in linear mixed models with one variance component”. Technical report, Department of Statistics, Cornell University, Ithaca.
- Crainiceanu, C. M. and D. Ruppert (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B Stat. Methodol.* 66(1), 165–185.
- Dempster, A. P., D. B. Rubin, and R. K. Tsutakawa (1981). Estimation in covariance components models. *J. Amer. Statist. Assoc.* 76(374), 341–353.
- Fahrmeir, L., T. Kneib, and S. Lang (2007). *Regression: Modelle, Methoden und Anwendungen*. Berlin: Springer.
- Fitzmaurice, G. M., N. M. Laird, and J. H. Ware (2004). *Applied longitudinal analysis*. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons].
- Fox, J. (2002). Linear Mixed Models. <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-mixed-models.pdf>. Appendix to *An R and S-PLUS Companion to Applied Regression*.
- Giampaoli, V. and J. M. Singer (2009). Likelihood ratio tests for variance components in linear mixed models. *J. Statist. Plann. Inference* 139(4), 1435–1448.

- Gruhn, J. and R. Kazer (1989). *Hormonal regulation of the menstrual cycle - The evolution of concepts*. New York and London: Plenum Medical Book Company.
- Harville, D. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* 61(2), 383–385.
- Harville, D. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Ann. Statist.* 4(2), 384–395.
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* 72(358), 320–340.
- Henderson, C. R. (1950). Estimation of genetic parameters. *Ann. Math. Statist.* 21, 309–310.
- Henderson, C. R. (1963). Selection Index and Expected Genetic Advance. In *Statistical Genetics and Plant Breeding*, pp. 141–163. National Academy of Sciences - National Research Council Publication No. 982.
- Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192–218.
- Kenward, M. and J. Roger (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53, 983–987.
- Laird, N. M. and J. H. Ware (1982). Random-Effects Models for Longitudinal Data. *Biometrics* 38(4), 963–974.
- LaMotte, L. R. (2007). A direct derivation of the REML likelihood function. *Statist. Papers* 48(2), 321–327.
- Lindstrom, M. J. and D. M. Bates (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Amer. Statist. Assoc.* 83(404), 1014–1022.
- McCulloch, C. E. and S. R. Searle (2001). *Generalized, linear, and mixed models*. New York: Wiley-Interscience [John Wiley & Sons].
- Morrell, C. H. (1998). Likelihood Ratio Testing of Variance Components in the Linear Mixed-Effects Model Using Restricted Maximum Likelihood. *Biometrics* 54(4), 1560–1568.
- Myers, R. H. (1986). *Classical and modern regression with applications (2nd Ed.)*. Boston, MA: PWS-Kent Publishing Company.
- Patterson, H. D. and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
- Patterson, H. D. and R. Thompson (1974). Maximum Likelihood Estimation of Components of Variance. In *Proceedings of the 8th International Biometric Conference*, pp. 197–207.
- Pinheiro, J. and D. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

- Satterthwaite, F. E. (1941). Synthesis of Variance. *Psychometrika* 6, 309–316.
- Schabenberger, O. (2004). Mixed Model Influence Diagnostics. In *Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference*. Paper 189-29, Cary, NC: SAS Institute.
- Scheipl, F. (2007). Testing for nonparametric terms and random effects in structured additive regression. <http://www.statistik.lmu.de/~scheipl/downloads/DIPLOM.zip>. Diploma thesis.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* 6(2), 461–464.
- Searle, S. R. (1988). Mixed models and unbalanced data: wherefrom, whereat and whereto? *Comm. Statist. Theory Methods* 17(4), 935–968.
- Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance components*. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons].
- Seifert-Klauss, V. (2007). Knochenstoffwechsel in der Perimenopause - endokrinologische und prognostische Aspekte. Professorial dissertation.
- Seifert-Klauss, V., A. Ehle, and T. Wimmer (2010,2011). Personal conversation.
- Seifert-Klauss, V., T. Link, C. Heumann, P. Lupp, M. Haseitl, J. Laakmann, J. Rattenhuber, and M. Kiechle (2006). Influence of pattern of menopausal transition on the amount of trabecular bone loss - results from a 6-year prospective longitudinal study. *Maturitas* 55, 317–324.
- Seifert-Klauss, V. and J. Prior (2010). Progesterone and bone: Actions promoting bone health in women. *Journal of Osteoporosis* 2010.
- Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* 82(398), 605–610.
- Stram, D. O. and J. W. Lee (1994). Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics* 50(4), 1171–1177.
- Verbeke, G. and G. Molenberghs (2009). *Linear mixed models for longitudinal data*. New York: Springer.
- Vu, H. T. V. and S. Zhou (1997). Generalization of likelihood ratio tests under non-standard conditions. *Ann. Statist.* 25(2), 897–916.
- Wald, A. (1947). A note on regression analysis. *Ann. Math. Statistics* 18, 586–589.
- West, B. T., K. B. Welch, and A. T. Galecki (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton: Chapman and Hall/CRC.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* 9, 60–62.