Technische Universität München

Zentrum Mathematik

The Mixed Model for Multivariate Repeated Measures

Diplomarbeit

von

Marion Sperling

Themenstellerin: Prof. Donna Ankerst, Ph.D
Betreuer: Dr. Stephan Haug
Abgabetermin: 01.Juni 2010

Technische Universität München

Zentrum Mathematik

The Mixed Model for Multivariate Repeated Measures

Diplomarbeit

von

Marion Sperling

Themenstellerin: Prof. Donna Ankerst, Ph.D
Betreuer: Dr. Stephan Haug
Abgabetermin: 01.Juni 2010

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

Garching, den 31. Mai 2010

Marion Sperling

Contents

1. Introduction **1**
1.1. Notation . . . 2

2. The Multivariate Linear Model and the Univariate Linear Mixed Model **3**
2.1. The Multivariate Linear Regression Model . . . 3
2.1.1. Estimation in Multivariate Linear Regression Models . . . 4
2.1.2. Hypothesis Tests in Multivariate Linear Regression Models . . . 7
2.2. The Univariate Linear Mixed Model . . . 12
2.2.1. Mixed ANOVA Models . . . 16

3. The Multivariate Linear Mixed Model **21**
3.1. The Multivariate Linear Mixed Model . . . 21
3.1.1. Model Specification . . . 21
3.1.2. The Multivariate Mixed Model Analysis for Testing Hypotheses . . . 24
3.2. Repeated Measurements Models . . . 26
3.2.1. The Multivariate Mixed Model in Repeated Measurements Experiments . . 26
3.2.2. Hypotheses Tests for Repeated Measures Experiments . . . 28

4. Multivariate Sphericity **34**
4.1. Multivariate Sphericity as Necessary and Sufficient Condition for the Multivariate Mixed Model Analysis . . . 35
4.2. A Likelihood Ratio Test for Multivariate Sphericity . . . 43
4.3. An Adjusted Test when Multivariate Sphericity is not Satisfied . . . 44

5. Implementation **47**
5.1. Dental Data . . . 47
5.2. Simulated Data . . . 51

6. Concluding Remarks **56**

A. Appendix **57**
A.1. Matrix Algebra . . . 57
A.2. Multivariate Distributions . . . 59

1. Introduction

Multivariate linear models are used to relate a set of variables of interest called dependent variables to another set of independent variables. Best known are models where the parameters used to describe the relation and the independent variables are fixed. It is, however, also possible to consider multivariate linear models with both, fixed and random parameters. These models form the class of multivariate linear mixed models. They are often used in social, biological or economic sciences due to the fact that they are very flexible and allow to model individual influences. Because of the random parameters, estimation and hypotheses testing do often get more complicated in mixed models. It is for example usually not possible to give exact distributions of test statistics.

A main focus of this thesis lies on the multivariate linear mixed model in the analysis of data from repeated measurements experiments, especially on hypotheses tests. In such experiments the variables of interest are observed at several points in time. It is one of the most important fields of application for linear mixed models.

There are many ways to analyze data from repeated measures experiments. It is for example possible to use a doubly multivariate model, as considered in [1]. Traditionally, one uses a univariate linear mixed model to model repeated measurement setups with only one response variable. So Thomas suggested in [27] to use a multivariate mixed model for the analysis as a natural extension to the univariate case. If the dependent variables are uncorrelated, then it is also possible to perform separate univariate analyses. This thesis concentrates on the multivariate mixed model approach.

As mentioned above, an analysis is often more complicated in mixed models. This is also the case in repeated measures experiments. The usual validity conditions for a multivariate analysis are not enough to ensure that the corresponding multivariate mixed model analysis is valid. Thomas derived a sufficient condition for validity of the multivariate mixed model analysis for repeated measurements experiments. He also gave a test to determine, whether the condition is satisfied or not. Boik showed in [1] that the condition is also necessary. This condition is called multivariate sphericity. If multivariate sphericity is satisfied then a multivariate mixed model analysis should be more powerful than other analyses, because it takes the additional information into account.

In Chapter 2 the multivariate linear regression model and the univariate linear mixed model are introduced in order to make the understanding of the multivariate mixed model easier. A likelihood ration test for hypotheses in multivariate linear models is given, because the same techniques are applied to find the corresponding likelihood ratio test statistic in the mixed model analysis of repeated measures data. Chapter 3 introduces the multivariate linear mixed model and gives model specifications and hypotheses tests for multivariate repeated measurements experiments. In Chapter 4 multivariate sphericity is shown to be a necessary and sufficient condition for a multivariate mixed model analysis if multivariate normality is satisfied. A test for multivariate sphericity is given. This enables us to test, whether a multivariate mixed model analysis can be applied or not. A correction factor for cases where multivariate sphericity is not satisfied is introduced. In Chapter 5 the multivariate mixed model and the adjusted analysis are applied to a repeated measures data set introduced by Timm [25] and on simulated data.

1.1. Notation

Preliminary, the notation used throughout the whole thesis is given.

Vector and Matrix Notation

Vectors are represented by bold lower case characters, for example

$$\boldsymbol{b} \in \mathbb{R}^n \quad , \quad \boldsymbol{\beta} \in \mathbb{R}^m.$$

Transpose vectors are denoted by a superscript T. A commonly used vector is the vector containing only ones:

$$\mathbf{1}_n := (\underbrace{1, 1, ..., 1}_{n})^T \in \mathbb{R}^n.$$

Matrices are represented by bold capital letters, for example

$$\boldsymbol{A} \in \mathbb{R}^{n \times m} \quad , \quad \boldsymbol{\Omega} \in \mathbb{R}^{n \times n}.$$

The identity matrix often occurs and is expressed by a capital $\boldsymbol{I}$ with subscript n that represents the dimension:

$$\boldsymbol{I}_n := \underbrace{\begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & & 1 & 0 \\ 0 & \cdots & & 0 & 1 \end{bmatrix}}_{n} \in \mathbb{R}^{n \times n}$$

Furthermore, following matrix spaces are defined:

Definition 1.1 (Matrix Spaces)**.** Let $n, m \in \mathbb{N}$.

- $\mathcal{M}_{n,m} := \mathbb{R}^{n \times m}$, the space containing all $n \times m$ matrices with entries in $\mathbb{R}$.
- $\mathcal{S}_n$ denotes the space of all symmetric matrices $\boldsymbol{A} \in \mathcal{M}_{n,n}$.
- $\mathcal{S}_n^+$ denotes the space of all positive semidefinite symmetric matrices $\boldsymbol{A} \in \mathcal{M}_{n,n}$.

Let $\boldsymbol{A} \in \mathcal{M}_{n,m}$. Then

- The transpose of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^T$.
- The generalized inverse or g-inverse of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^-$.
- If $n = m$ and $\boldsymbol{A}$ is non-singular, then the inverse of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^{-1}$.
- If $n = m$ the determinant of $\boldsymbol{A}$ is denoted by $|\boldsymbol{A}|$.

2. The Multivariate Linear Model and the Univariate Linear Mixed Model

The multivariate linear mixed model is of interest in many applications. It is a multivariate extension of the univariate linear mixed model, which has been known for a long time, especially in connection with repeated measurements experiments.The term linear is often omitted for the purpose of abbreviation. So whenever it says the univariate or multivariate mixed model in the following, the linear versions of these models are referred to. Formulation and interpretation as well as notation of both, univariate and multivariate mixed model are very similar. Thus, for a better understanding of the multivariate case, the univariate mixed model is discussed in this section. To illustrate the difference between the multivariate linear mixed model and the multivariate linear regression model, the latter one is also introduced.

2.1. The Multivariate Linear Regression Model

The linear regression model is a very basic concept in the analysis of both univariate and multivariate data. Here we are going to concentrate on the multivariate case. The aim of a multivariate linear regression model is to find a linear relationship of a matrix of variables one is interested in to a set of linearly independent fixed variables.

Suppose one is interested in r variables observed on n individuals or experimental units. Denote them as *dependent variables* or *response variables*. We assume them to be related to p linearly independent variables that are called *independent variables*, *predictor variables* or *covariates*.

Let $\boldsymbol{y}_i \in \mathbb{R}^n$, $i = 1, ..., r$, denote the vector of observations of the ith response variable. Furthermore, let $\boldsymbol{x}_j \in \mathbb{R}^n$, $j = 1, ..., p$ denote the p predictor variables. The relation of $\boldsymbol{y}_i$ to the independent variables can be expressed in the following way:

$$\boldsymbol{y}_i = \beta_{0i}\mathbf{1}_n + \beta_{1i}\boldsymbol{x}_1 + ... + \beta_{pi}\boldsymbol{x}_p + \boldsymbol{e}_i, \tag{2.1}$$

where $\boldsymbol{e}_i \in \mathbb{R}^n$, $i = 1, ..., r$, represents the *experimental error*. The $\beta_{0i}, ..., \beta_{pi}$ are the $p + 1$ parameters for the ith response variable. They are called *regression coefficients* or *effects*. For each response variable, the so called *intercept* term β_{0i} is included in the set of parameters.

This model can be represented in matrix form in the following way:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \tag{2.2}$$

where

$$\boldsymbol{Y} = [\boldsymbol{y}_1, ..., \boldsymbol{y}_r],$$
$$\boldsymbol{X} = [\boldsymbol{1}_n, \boldsymbol{x}_1, ..., \boldsymbol{x}_p],$$
$$\boldsymbol{B} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0r} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1r} \\ \vdots & \vdots & & \vdots \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pr} \end{bmatrix} = [\boldsymbol{\beta}_1, ...\boldsymbol{\beta}_r] \quad \text{and}$$
$$\boldsymbol{E} = [\boldsymbol{e}_1, ..., \boldsymbol{e}_r].$$

Thus, $\boldsymbol{Y} \in \mathcal{M}_{n,r}$ is the observed data matrix of the r response variables on the n individuals, $\boldsymbol{X} \in \mathcal{M}_{n,p+1}$ denotes a known nonrandom design matrix of the p independent variables observed on each of the n experimental units. The rank of $\boldsymbol{X}$ is denoted by $k \leq p+1$. $\boldsymbol{B} \in \mathcal{M}_{p+1,r}$ is a matrix containing unknown fixed regression parameters and $\boldsymbol{E} \in \mathcal{M}_{n,r}$ is a matrix of random errors. Note, that the model is linear in its parameters.

Model Assumptions. The following assumptions are usually made:

- The rows of $\boldsymbol{E}$ for given $\boldsymbol{X}$ are uncorrelated.
- Each row of $\boldsymbol{E}$ has mean $\boldsymbol{0}$ and common unknown covariance matrix $\boldsymbol{\Sigma}$.
- A further usual assumption that is made in most applications is that the rows of $\boldsymbol{E}$ follow a multivariate normal distribution.

Hence, $E[\boldsymbol{Y}] = \boldsymbol{XB}$. So the intercept terms times $\boldsymbol{1}_n$ are the expected values of the $\boldsymbol{y}_i$ given that the $\boldsymbol{x}_j$'s are all equal to zero, $i = 1, ..., r$, $j = 1, ..., p$. If the intercept terms $\beta_{0,i}$, $i = 1, ..., r$ are left out, we have a *no-intercept* regression.

Remark 2.1.1. Note that $\boldsymbol{X}$ and $\boldsymbol{B}$ are nonrandom. Thus the only random part in the model comes from the error term $\boldsymbol{E}$.

Later on we will be interested in testing hypotheses in multivariate mixed models in repeated measures experiments. The test statistics are obtained in the same way as in the multivariate linear regression model. For this reason, the derivation of maximum likelihood estimators and likelihood ratio test statistics are given here.

In the context of this thesis we are only interested in the maximum likelihood approach. It is, however, also possible to estimate the parameters using for example least square methods or to test hypotheses with union intersection tests. For a more detailed examination of the multivariate linear regression model, estimation and testing, please see Timm [26] or Mardia, Kent and Bibby [18].

2.1.1. Estimation in Multivariate Linear Regression Models

There are two parameters that have to be estimated, namely the regression parameter matrix $\boldsymbol{B}$ and the covariance matrix $\boldsymbol{\Sigma}$. Here the maximum likelihood estimates for $n \geq r + p + 1$ shall be given.

Generally, suppose that $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ are observations in a random sample from a population with probability density function $f(\boldsymbol{y}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector or matrix. Let $\boldsymbol{Y} = [\boldsymbol{y}_1, ..., \boldsymbol{y}_n]$. The likelihood function of the whole sample is:

$$L(\boldsymbol{Y};\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{y}_i;\boldsymbol{\theta}). \tag{2.3}$$

The $\hat{\boldsymbol{\theta}}$ that maximizes the likelihood function is the maximum likelihood estimator for the true parameter $\boldsymbol{\theta}$. It is the value of $\boldsymbol{\theta}$ for which the observed random sample $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ has the highest probability, or in other words the highest likelihood.

It is often convenient to take the logarithm of the likelihood function $L(\boldsymbol{Y},\boldsymbol{\theta})$. This leads to the log-likelihood function.

$$l(\boldsymbol{Y};\boldsymbol{\theta}) := \log(L(\boldsymbol{Y};\boldsymbol{\theta})) = \sum_{i=1}^{n} \log(f(\boldsymbol{Y};\boldsymbol{\theta})). \tag{2.4}$$

The logarithm function is strictly increasing, and thus the value that maximizes the log-likelihood also maximizes the likelihood function. However, calculations often simplify for the log-likelihood.

For the calculation of the maximum likelihood estimators of the parameters in the multivariate linear regression, first define

$$\boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \quad \in \mathcal{M}_{n,n}. \tag{2.5}$$

This matrix will often occur in the following. Therefore some essential properties of $\boldsymbol{P}$ are given:

Remark 2.1.2 (Mardia, Kent and Bibby, [18], Section 6.2.1)**.** $\boldsymbol{P}$ is a symmetric, idempotent matrix of rank $n-k$.

Mardia, Kent and Bibby do not prove the properties of $\boldsymbol{P}$. Therefore, we give the proof here.

Proof: Throughout the proof, the general calculation rules for the transpose are used.

Idempotence:

$$\begin{aligned}\boldsymbol{P}\boldsymbol{P} &= (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T)(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T)\\ &= \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T + \boldsymbol{X}\underbrace{(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-}_{=(\boldsymbol{X}^T\boldsymbol{X})^-}\boldsymbol{X}^T\\ &= \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\\ &= \boldsymbol{P}.\end{aligned}$$

Symmetry:

$$\begin{aligned}\boldsymbol{P}^T &= (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T)^T\\ &= \boldsymbol{I}_n^T - (\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T)^T\\ &= \boldsymbol{I}_n - (\boldsymbol{X}^T)^T((\boldsymbol{X}^T\boldsymbol{X})^-)^T\boldsymbol{X}^T\\ &= \boldsymbol{I}_n - \boldsymbol{X}((\boldsymbol{X}^T\boldsymbol{X})^T)^-\boldsymbol{X}^T\\ &= \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T(\boldsymbol{X}^T)^T)^-\boldsymbol{X}^T\\ &= \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\\ &= \boldsymbol{P}.\end{aligned}$$

Rank: Note that for an idempotent matrix $\boldsymbol{A}$, $rank(\boldsymbol{A}) = tr(\boldsymbol{A})$ (see Theorem A.11). Thus

$$
\begin{aligned}
rank(\boldsymbol{P}) &= tr(\boldsymbol{P}) \\
&= tr(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T) \\
&= tr(\boldsymbol{I}_n) - tr(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T).
\end{aligned}
$$

In the last step we used Theorem (A.1).
Moreover, $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$, thus $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$ is idempotent and therefore $tr(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T) = rank(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T) = rank(\boldsymbol{X}) = k$. This yields:

$$
\begin{aligned}
rank(\boldsymbol{P}) &= tr(\boldsymbol{I}_n) - rank(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T) \\
&= n - k.
\end{aligned}
$$

□

In order to calculate the maximum likelihood estimates for $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$, we need to know the likelihood function or rather the log-likelihood function. In the case of multivariate linear regression, we consider n independent observations $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$. They are assumed to be drawn from a multivariate normal distribution $\mathcal{N}_r(\boldsymbol{0}, \boldsymbol{\Sigma})$. Let $\boldsymbol{x}_i$, $i = 1, ..., n$, denote the rows of $\boldsymbol{X}$. Then the likelihood function defined in (2.3) is:

$$
\begin{aligned}
L(\boldsymbol{Y}; \boldsymbol{B}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} |2\pi\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{B})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{B})^T\right\} \\
&= |2\pi\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{B})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{B})^T\right\}.
\end{aligned}
$$

We are interested in the log-likelihood function, because it simplifies the calculations, especially because of the exponential function. The usual calculation rules for the logarithm are applied.

$$
\begin{aligned}
l(\boldsymbol{Y}; \boldsymbol{B}, \boldsymbol{\Sigma}) &= \log(|2\pi\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{B})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{B})^T\right\}) \\
&= -\frac{n}{2}\log(|2\pi\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{B})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{B})^T.
\end{aligned}
$$

The following theorem gives the maximum likelihood estimators for $\boldsymbol{B}$ and for $\boldsymbol{\Sigma}$.

Theorem 2.1 (Mardia,Kent and Bibby, [18], Theorem 6.2.1)**.** *For the log-likelihood function given above, the maximum likelihood estimators for $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ are*

$$
\widehat{\boldsymbol{B}} = (\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{Y}
$$

and

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Y}^T\boldsymbol{P}\boldsymbol{Y}.$$

A proof is given in the book.

Remark 2.1.3. Let $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\Sigma}}$ be defined as above. Then the *fitted* value of $\boldsymbol{Y}$ is given by

$$\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{B}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{Y}$$

and the *fitted* error matrix is

$$\begin{aligned}\widehat{\boldsymbol{E}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} &= \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}} = \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{Y} \\ &= (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T)\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{Y}.\end{aligned}$$

They are the values one obtains, when using the multivariate linear regression model (2.2), the estimator for $\boldsymbol{B}$ and the observed data $\boldsymbol{X}$ and $\boldsymbol{Y}$ for calculations.

Remark 2.1.4 (Mardia, Kent and Bibby, [18], Section 6.2.6)**.** The maximum value of the likelihood is given by

$$l(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}}) = -\frac{1}{2}\log|2\pi\widehat{\boldsymbol{\Sigma}}| - \frac{1}{2}np. \tag{2.6}$$

2.1.2. Hypothesis Tests in Multivariate Linear Regression Models

It is of interest to test hypotheses of the form:

$$H_0 : \boldsymbol{C}_1\boldsymbol{B}\boldsymbol{M}_1 = \boldsymbol{D} \tag{2.7}$$

against

$$H_1 : \boldsymbol{C}_1\boldsymbol{B}\boldsymbol{M}_1 \neq \boldsymbol{D}, \tag{2.8}$$

where $\boldsymbol{C}_1 \in \mathcal{M}_{s,p+1}$, $\boldsymbol{M}_1 \in \mathcal{M}_{r,q}$ and $\boldsymbol{D} \in \mathcal{M}_{s,q}$ are given matrices and $\boldsymbol{C}_1$ and $\boldsymbol{M}_1$ have rank s and q, respectively. The rows of $\boldsymbol{C}_1$ make assertions about the effect on the regression from linear combinations of the independent variables, whereas the columns of $\boldsymbol{M}_1$ focus on linear combinations of the dependent variables. Often, $\boldsymbol{D}$ is zero and $\boldsymbol{M}_1$ is chosen to be the identity matrix $\boldsymbol{I}_r$. Then, (2.7) reduces to

$$H_0 : \boldsymbol{C}_1\boldsymbol{B} = \boldsymbol{0} \tag{2.9}$$

against

$$H_1 : \boldsymbol{C}_1\boldsymbol{B} \neq \boldsymbol{0} \tag{2.10}$$

However, in experiments with repeated measurements it can be of interest to consider hypotheses where $\boldsymbol{M}_1$ is not equal to the identity matrix, as we will see later. Examples are given in Chapter 5. There are also many examples given to illustrate how to choose $\boldsymbol{C}$ and $\boldsymbol{M}$ for certain hypotheses in [26] and [27].

In the following we will introduce a likelihood ratio test for the hypotheses above.

The Likelihood Ration Test

The idea behind a likelihood ratio test is to maximize the likelihood under the null hypothesis H_0 and to maximize the likelihood under the alternative hypothesis H_1, too:

Definition 2.2 (The Likelihood Ration Test Statistic)**.** Let the distribution of a random sample $\boldsymbol{Y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_n)^T$ depend upon a certain parameter $\boldsymbol{\theta}$ and let $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Omega}_0$ and $H_1 : \boldsymbol{\theta} \in \boldsymbol{\Omega}_1$ be any two hypotheses. Then the *likelihood ratio statistic* for testing H_0 against H_1 is defined as

$$\lambda(\boldsymbol{Y}) = \frac{\boldsymbol{L^*}_0}{\boldsymbol{L^*}_1}, \tag{2.11}$$

where $\boldsymbol{L^*}_i$ is the largest value which the likelihood function takes in the region Ω_i, $i = 0, 1$.

It is also possible to consider an equivalent test statistic using log-likelihood functions:

$$-2\log(\lambda) = 2[\log(\boldsymbol{L^*}_1) - \log(\boldsymbol{L^*}_0)].$$

When the likelihood test statistic has a low value, one tends to favor H_1, if the likelihood ration test statistic is high, one favors H_0. One can define the following test procedure based on the likelihood ration test statistic:

Definition 2.3. The likelihood ration test of *size* α for testing H_0 against H_1 has a rejection region

$$R = \{\boldsymbol{Y}|\lambda(\boldsymbol{Y}) < c\}, \tag{2.12}$$

where c is determined such that

$$\sup_{\boldsymbol{\theta}\in\Omega_0} P_{\boldsymbol{\theta}}(\boldsymbol{Y} \in R) = \alpha.$$

In the case of multivariate linear regression, we are interested in hypotheses of the form

$$H_0 : \boldsymbol{C}_1\boldsymbol{B}\boldsymbol{M}_1 = \boldsymbol{D};$$
$$H_1 : \boldsymbol{C}_1\boldsymbol{B}\boldsymbol{M}_1 \neq \boldsymbol{D},$$

as already mentioned above. Now we want to derive a likelihood ratio test statistic for hypotheses of this form. Therefore we first concentrate on the hypothesis $H_0 : \boldsymbol{C}_1\boldsymbol{B} = \boldsymbol{D}$ and the alternative hypothesis $H_1 : \boldsymbol{C}_1\boldsymbol{B} \neq \boldsymbol{D}$ (thus $\boldsymbol{M}_1 = \boldsymbol{I}_r$) and use the results to derive the statistic for the general case. The derivation is an extension to the construction of the test statistic given by Mardia, Kent and Bibby.

First, define the following matrices that will be used in the construction of the test statistic: Let $\boldsymbol{C}_2 \in \mathcal{M}_{p+1-s,p+1}$ be a matrix such that $\boldsymbol{C}^T = [\boldsymbol{C}_1^T, \boldsymbol{C}_2^T] \in \mathcal{M}_{p+1,p+1}$ is a non-singular matrix and let $\boldsymbol{B}_0 \in \mathcal{M}_{p+1,r}$ be any matrix such that $\boldsymbol{C}_1\boldsymbol{B}_0 = \boldsymbol{D}$.

Now we want to rewrite model (2.2) such that we can find an easier formulation for H_0. Start with the linear regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}.$$

As $\boldsymbol{C}^T$ was constructed to be non-singular, $\boldsymbol{C}$ is non-singular and thus there exists a matrix $\boldsymbol{C}^{-1} \in \mathcal{M}_{p+1,p+1}$ such that $\boldsymbol{C}^{-1}\boldsymbol{C} = \boldsymbol{I}_{p+1}$. Thus we can rewrite the model above as follows:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{XB} + \boldsymbol{E} \\ &= \boldsymbol{XI}_{p+1}\boldsymbol{B} + \boldsymbol{E} \\ &= \boldsymbol{XC}^{-1}\boldsymbol{CB} + \boldsymbol{E}. \end{aligned}$$

To simplify notation, let $\boldsymbol{Z} := \boldsymbol{XC}^{-1} \in \mathcal{M}_{n,p+1}$, thus we have:

$$\boldsymbol{Y} = \boldsymbol{ZCB} + \boldsymbol{E}.$$

Furthermore, one can add $(-\boldsymbol{XB}_0 + \boldsymbol{XB}_0) = \boldsymbol{0}$:

$$\boldsymbol{Y} = \boldsymbol{ZCB} - \boldsymbol{XB}_0 + \boldsymbol{XB}_0 + \boldsymbol{E}$$

Again, use that $\boldsymbol{C}^{-1}\boldsymbol{C} = \boldsymbol{I}_{p+1}$:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{ZCB} - \boldsymbol{XI}_{p+1}\boldsymbol{B}_0 + \boldsymbol{XB}_0 + \boldsymbol{E} \\ &= \boldsymbol{ZCB} - \underbrace{\boldsymbol{XC}^{-1}}_{=\boldsymbol{Z}}\boldsymbol{CB}_0 + \boldsymbol{XB}_0 + \boldsymbol{E} \\ &= \boldsymbol{ZCB} - \boldsymbol{ZCB}_0 + \boldsymbol{XB}_0 + \boldsymbol{E}. \end{aligned}$$

The first two terms on the right-hand side can be summarized:

$$\boldsymbol{Y} = \boldsymbol{ZC}(\boldsymbol{B} - \boldsymbol{B}_0) + \boldsymbol{XB}_0 + \boldsymbol{E}.$$

Let now $\boldsymbol{\Delta} := \boldsymbol{C}(\boldsymbol{B} - \boldsymbol{B}_0) \in \mathcal{M}_{p+1,r}$. Note that $\boldsymbol{\Delta} = [\boldsymbol{\Delta_1}, \boldsymbol{\Delta_2}] = [\boldsymbol{C}_1, \boldsymbol{C}_2](\boldsymbol{B} - \boldsymbol{B}_0)$. This yields:

$$\boldsymbol{Y} = \boldsymbol{Z\Delta} + \boldsymbol{XB}_0 + \boldsymbol{E}.$$

Finally, subtract $\boldsymbol{XB}_0$ on both sides of the equation:

$$\boldsymbol{Y} - \boldsymbol{XB}_0 = \boldsymbol{Z\Delta} + \boldsymbol{E}.$$

By setting $\boldsymbol{Y}_+ = \boldsymbol{Y} - \boldsymbol{XB}_0 \in \mathcal{M}_{n,r}$, one gets the rewritten linear regression model

$$\boldsymbol{Y}_+ = \boldsymbol{Z\Delta} + \boldsymbol{E}. \tag{2.13}$$

Note that this model is just another way to write model (2.2). It is convenient to consider model (2.13), because the hypothesis $\boldsymbol{C}_1\boldsymbol{B} = \boldsymbol{D}$ becomes $\boldsymbol{\Delta}_1 = 0$, which is shown below. Recall that $\boldsymbol{\Delta}_1 = \boldsymbol{C}_1\boldsymbol{B}$.

$$\begin{aligned} & \boldsymbol{C}_1\boldsymbol{B} = \boldsymbol{D} \\ \Longleftrightarrow \quad & \boldsymbol{C_1B} - \boldsymbol{D} = \boldsymbol{0} \end{aligned}$$

Now use that $\boldsymbol{B}_0$ was chosen to satisfy $\boldsymbol{CB_0} = \boldsymbol{D}$. This yields:

$$\begin{aligned} & \boldsymbol{C}_1\boldsymbol{B} - \boldsymbol{C}_1\boldsymbol{B}_0 = \mathbf{0} \\ \iff \quad & \boldsymbol{C}_1(\boldsymbol{B} - \boldsymbol{B}_0) = \mathbf{0} \\ \iff \quad & \boldsymbol{\Delta}_1 = \mathbf{0}, \end{aligned}$$

and thus

$$\boldsymbol{C}_1\boldsymbol{B} = \boldsymbol{D} \quad \iff \quad \boldsymbol{\Delta}_1 = 0. \tag{2.14}$$

Therefore we can consider the hypotheses $H_0 : \boldsymbol{\Delta}_1 = \mathbf{0}$ and $H_1 : \boldsymbol{\Delta}_1 \neq \mathbf{0}$. Partition $\boldsymbol{C}^{-1} = [\boldsymbol{C}^{(1)}, \boldsymbol{C}^{(2)}]$, such that $\boldsymbol{C}^{(1)} \in \mathcal{M}_{p+1,s}$. Then under the null hypothesis, model (2.13) reduces to:

$$\boldsymbol{Y}_+ = \boldsymbol{X}\boldsymbol{C}^{(2)}\boldsymbol{\Delta}_2 + \boldsymbol{E}. \tag{2.15}$$

Recall that we defined the idempotent matrix $\boldsymbol{P}$ for the unrestricted multivariate regression model (2.2) as:

$$\boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T.$$

For the model under the null hypothesis the design matrix is $\boldsymbol{X}\boldsymbol{C}^{(2)}$, and thus the idempotent matrix for the restricted model, (i.e. the model under the null hypothesis), is

$$\boldsymbol{P}_0 = \boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{C}^{(2)}[(\boldsymbol{C}^{(2)})^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{C}^{(2)}]^{-1}(\boldsymbol{C}^{(2)})^T\boldsymbol{X}^T. \tag{2.16}$$

Recall, that as we have seen in (2.1.4) the maximum value of the log-likelihood is given by:

$$l(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}}) = -\frac{1}{2}\log|2\pi\widehat{\boldsymbol{\Sigma}}| - \frac{1}{2}nr.$$

and thus

$$\begin{aligned} L(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}}) &= \exp(l(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}})) \\ &= \exp(-\frac{1}{2}\log|2\pi\widehat{\boldsymbol{\Sigma}}| - \frac{1}{2}nr) \\ &= \exp(\log(|2\pi\widehat{\boldsymbol{\Sigma}}|^{-\frac{1}{2}}))\exp(-\frac{1}{2}nr) \\ &= |2\pi\widehat{\boldsymbol{\Sigma}}|^{-\frac{1}{2}}\exp(-\frac{1}{2}nr). \end{aligned}$$

It does only depend on $\widehat{\boldsymbol{\Sigma}}$, the maximum likelihood estimator for $\boldsymbol{\Sigma}$. Therefore we need the values of $\widehat{\boldsymbol{\Sigma}}$ under the null hypothesis and under the alternative hypothesis. Recall, that

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Y}_+^T\boldsymbol{P}\boldsymbol{Y}_+.$$

Under the null hypothesis, the estimator is

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Y}_+^T\boldsymbol{P}_0\boldsymbol{Y}_+.$$

Accordingly, the maximum likelihoods under the null hypothesis and under the alternative hypothesis are

$$|2\pi n^{-1}\boldsymbol{Y}_+^T\boldsymbol{P}_0\boldsymbol{Y}_+|^{-\frac{n}{2}}\exp(-\frac{1}{2}nr) \quad \text{and} \quad |2\pi n^{-1}\boldsymbol{Y}_+^T\boldsymbol{P}\boldsymbol{Y}_+|^{-\frac{n}{2}}\exp(-\frac{1}{2}nr), \tag{2.17}$$

respectively. Thus the likelihood ratio test statistic is given by:

$$\lambda^{\frac{2}{n}} = \frac{|\boldsymbol{Y}_+^T\boldsymbol{P}\boldsymbol{Y}_+|}{|\boldsymbol{Y}_+^T\boldsymbol{P}_0\boldsymbol{Y}_+|}. \tag{2.18}$$

There is another way to express this test statistic that allows to derive its distribution. It is given in the theorem below. Note that $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{0}$ and thus $\boldsymbol{P}\boldsymbol{Y}_+ = \boldsymbol{P}\boldsymbol{Y}$.

Theorem 2.4 (Mardia, Kent and Bibby, [18], Theorem 6.3.1)**.** *The likelihood ratio test of the hypothesis $\boldsymbol{C}_1\boldsymbol{B} = \boldsymbol{D}$ for the multivariate linear regression model has the statistic*

$$\lambda^{\frac{2}{n}} = \frac{|\boldsymbol{Y}^T\boldsymbol{P}\boldsymbol{Y}|}{|\boldsymbol{Y}^T\boldsymbol{P}\boldsymbol{Y} + \boldsymbol{Y}_+^T\boldsymbol{P}_1\boldsymbol{Y}_+|}, \tag{2.19}$$

where $\boldsymbol{P}_1 := \boldsymbol{P}_0 - \boldsymbol{P}$

$$\boldsymbol{P}_1 = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}_1^T[\boldsymbol{C}_1(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}_1^T]^{-1}\boldsymbol{C}_1(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$$

Under the null hypothesis it has a $\boldsymbol{\Lambda}(r, n-(p+1), s)$ distribution.

A proof is given in the book. The $\boldsymbol{\Lambda}$ distribution is defined in Definition A.22.

Remark 2.1.5. $\boldsymbol{P}_1 = \boldsymbol{P}_0 - \boldsymbol{P}$ is also idempotent and symmetric.

Now we want to consider hypotheses of the more general form $H : \boldsymbol{C}_1\boldsymbol{B}\boldsymbol{M}_1 = \boldsymbol{D}$ for the multivariate linear regression model (2.2).

It is equivalent to consider the hypothesis $H : \boldsymbol{C}_1\boldsymbol{B}\boldsymbol{M}_1 = \boldsymbol{D}$ for the model

$$\boldsymbol{Y}\boldsymbol{M}_1 = \boldsymbol{X}\boldsymbol{B}\boldsymbol{M}_1 + \boldsymbol{E}\boldsymbol{M}_1. \tag{2.20}$$

$\boldsymbol{E}\boldsymbol{M}_1 \in \mathcal{M}_{n,q}$ is a matrix whose rows follow a $\mathcal{N}_q(\boldsymbol{0}, \boldsymbol{M}_1^T\boldsymbol{\Sigma}\boldsymbol{M}_1)$ distribution.

Remark 2.1.6. Model (2.20) is a *reduced* model. Instead of r response variables, we now consider q linear functions of those response variables.

Define

$$\boldsymbol{S}_H = \boldsymbol{M}_1^T\boldsymbol{Y}_+^T\underbrace{\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}_1^T[\boldsymbol{C}_1(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}_1^T]^{-1}\boldsymbol{C}_1(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T}_{=\boldsymbol{P}_1}\boldsymbol{Y}_+\boldsymbol{M}_1 \tag{2.21}$$

and

$$\boldsymbol{S}_E = \boldsymbol{M}_1^T\boldsymbol{Y}^T\underbrace{[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T]}_{=\boldsymbol{P}}\boldsymbol{Y}\boldsymbol{M}_1 \tag{2.22}$$

to be the Sums of Squares and Cross Products (SSCP) matrix due to the regression and the residual, respectively. The likelihood ratio test statistic is given by the following theorem:

Theorem 2.5 (Mardia, Kent and Bibby, [18], Theorem 6.3.2)**.** *The likelihood ratio test of the hypothesis* $H_0 : \boldsymbol{C}_1\boldsymbol{B}\boldsymbol{M}_1 = \boldsymbol{D}$ *for the multivariate linear regression model* (2.2) *is given by the test statistic*

$$\lambda^{\frac{2}{n}} = \frac{|\boldsymbol{S}_E|}{|\boldsymbol{S}_H + \boldsymbol{S}_E|} \sim \boldsymbol{\Lambda}(q, n-(p+1), s). \tag{2.23}$$

Proof: Apply Theorem 2.4 to the reduced model (2.20):

$$\boldsymbol{M}_1^T\boldsymbol{Y}^T\boldsymbol{P}\boldsymbol{Y}\boldsymbol{M}_1 = \boldsymbol{S}_E,$$

$$\boldsymbol{M}_1^T\boldsymbol{Y}_+^T\boldsymbol{P}_1\boldsymbol{Y}_+\boldsymbol{M}_1 = \boldsymbol{S}_H.$$

Thus indeed,

$$\lambda^{\frac{2}{n}} = \frac{|\boldsymbol{S}_E|}{|\boldsymbol{S}_H + \boldsymbol{S}_E|} \tag{2.24}$$

□

Remark 2.1.7. Note that two assumptions we made in the multivariate linear regression model (2.2) were necessary to derive the test statistic:

Multivariate Normality: The rows of $\boldsymbol{MY}$ follow a multivariate normal distribution.

Homogeneity of Variance: We assume that the rows of $\boldsymbol{Y}\boldsymbol{M}_1$ do all have common covariance matrix $\boldsymbol{M}^T\boldsymbol{\Sigma}\boldsymbol{M}$. This property is also called *homoscedasticity*.

Both properties were used to obtain the likelihood function and the corresponding maximizing values.

2.2. The Univariate Linear Mixed Model

In this section a short introduction of univariate mixed models is given to make the understanding and interpretation of the multivariate case easier. It is, however, only a brief overview and does not make a claim to be complete. The model allows to regard a lot of different situations. We constrain the consideration to the introduction of several model specifications.

Verbeke and Molenberghs [28] and Fahrmeir, Kneib and Lang [6] give a lot of illustrative examples. The reader is recommended to consult Khuri, Mathew and Sinha [13] or [6] for testing, estimation and prediction in univariate mixed models.

In the multivariate linear model (2.2) in section 2.1 the only random component is the error term $\boldsymbol{E}$. The parameters are all fixed. We say that all effects in the model are *fixed* and the model is called a *fixed effects* model.

It may, however be of interest to consider models where some of the parameters are random. Consider therefore the following general univariate linear model:

$$\begin{aligned}\boldsymbol{y} &= \widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{e} \\ &= [\boldsymbol{1}_n, \boldsymbol{x}_1, ..., \boldsymbol{x}_p] \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \boldsymbol{e},\end{aligned}$$

where $\boldsymbol{y} \in \mathbb{R}^n$ is the response variable, $\boldsymbol{e} = (e_1, ..., e_n)^T \in \mathbb{R}^n$ represents the error term, $\widetilde{\boldsymbol{X}} = [\mathbf{1}_n, \boldsymbol{x}_1, ..., \boldsymbol{x}_p]$ is the design matrix and $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^T \in \mathbb{R}^{p+1}$ is the vector of regression parameters.

Now assume that some of the effects are random. Without loss of generality, assume that the first $d+1$ effects $\beta_0, ..., \beta_d$, $0 < d < p$, are fixed and rest of them is random. Define the vector

$$\boldsymbol{u} = \begin{pmatrix} \beta_{d+1} \\ \vdots \\ \beta_p \end{pmatrix} \in \mathbb{R}^{p-(d+1)}.$$

It containts the random effects, thus it is assumed to be a random vector.

By separating fixed effects and random effects one obtains:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{g} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}, \tag{2.25}$$

where

$$\boldsymbol{X} = [\mathbf{1}_n, \boldsymbol{x}_1, ..., \boldsymbol{x}_d] \in \mathcal{M}_{n,d+1},$$

$$\boldsymbol{g} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_d \end{pmatrix} \in \mathbb{R}^{d+1},$$

and

$$\boldsymbol{Z} = [\boldsymbol{x}_{d+1}, ..., \boldsymbol{x}_p] \in \mathcal{M}_{n,p-(d+1)}.$$

$\boldsymbol{X}\boldsymbol{g}$ is the fixed part of the model, $\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$ represents the random portion of the model. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are design matrices. The model contains fixed effects and random effects and is therefore called a *mixed effects model.*

If d equals zero,

$$\boldsymbol{Y} = \mathbf{1}_n\beta_0 + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e},$$

where $\boldsymbol{Z} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_p] \in \mathcal{M}_{n,p}$ and $\boldsymbol{u} = (\beta_1, ..., \beta_p)^T \in \mathbb{R}^p$. Thus, all effects except for the intercept term are random and the model is said to be a *random model.*

Accordingly, if $d = p$,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{g} + \boldsymbol{e},$$

where $\boldsymbol{X} = \widetilde{\boldsymbol{X}} \in \mathcal{M}_{n,p+1}$ and $\boldsymbol{g} = \boldsymbol{\beta} \in \mathbb{R}^{p+1}$. All effects aside from the error term are fixed and the model is the the fixed univariate linear regression model. The model does not necessarily have an intercept term.

Model Assumptions. Following distribution assumptions are usually made:

- $\boldsymbol{u}$ and $\boldsymbol{e}$ are uncorrelated.
- $\boldsymbol{u} \sim \mathcal{N}_{p-(d+1)}(\mathbf{0}, \boldsymbol{\Sigma}_1)$.
- $\boldsymbol{e} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma})$.

However, it is also possible to consider generalized univariate linear models without the normality assumptions, which is for instance done by Breslow in [2].

Remark 2.2.1. $cov(\boldsymbol{y}) = \boldsymbol{Z}\boldsymbol{\Sigma}_1\boldsymbol{Z}^T + \boldsymbol{\Sigma}$.

The model is very flexible and allows to regard a lot of different experimental setups. Random effects can be used to model individual influences or as surrogates for variables that are not observable. The model can contain both, metric and categorial variables. In the most general setup, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}$ and even $\boldsymbol{Z}$ are considered to depend on a parameter vector $\boldsymbol{\vartheta}$ that cannot be observed (see [7]).

Below are some possibilities how to specify the model for different applications. More examples are given in [26] and [7].

Models with Random Coefficients

Consider a model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e},$$

where

$$\boldsymbol{\beta} = \boldsymbol{g} + \boldsymbol{u}.$$

Assume that $\boldsymbol{u}$ is a random vector. Then

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{g} + \boldsymbol{X}\boldsymbol{u} + \boldsymbol{e}.$$

This can be interpreted as follows: the influence of the regressors is random. There is an average influence factor $\boldsymbol{g}$, but it is modified by an unknown and not measurable factor $\boldsymbol{u}$. Note that

$$\boldsymbol{Z} = \boldsymbol{X}.$$

More generally, the columns of $\boldsymbol{Z}$ can be a subset of the columns of $\boldsymbol{X}$. This can be used to model an individual specific influence and does often occur in repeated measurement experiments.

The Univariate Mixed Model for Repeated Measurements

A very important field of application for the univariate mixed model is the analysis of repeated measurements. In such setups, the dependent variables are repeatedly observed under different treatment condition. The treatment conditions can for example either be different points in time or the individuals could be chosen from different groups. The first example is a longitudinal study, the second an experiment with clustered data.

An experiment with longitudinal data is characterized by observations of individuals or experimental units at several points in time. An example is the observation of the body height of children at different ages. In studies of clustered data, subjects are chosen from different clusters at random. Then the data concerning the variables of interest are gathered on those subjects. The clusters or groups can for example be schools, companies or hospitals and the corresponding chosen subjects might be pupils, products and patients. We give the model for longitudinal data, but the same notation can be used to model clustered data.

In both cases, there are several (repeated) observations for each of the n individuals. Let t_i denote the number of observations for individual i, $i = 1, ..., n$. Then the observations of the response variable are given by

$$(y_{i1}, ..., y_{it_i}) \qquad i = 1, ..., n$$

and the corresponding observations of the covariates are

$$(\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{it_i}) \qquad i = 1, ..., n.$$

y_{ij} represents the observation on individual i at point in time t_{ij}.

If the number of observations per individual is equal, thus $t_i = t \; \forall i = 1, ..., n$, the model is said to be *balanced*. Otherwise it is *unbalanced*. Calculations often simplify in balanced models, as we will see later.

The mixed model allows to model and predict individual or cluster specific influences by representing those through one or more random effects:

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{g} + \boldsymbol{Z}_i \boldsymbol{u}_i + \boldsymbol{e}_i, \tag{2.26}$$

and thus the responses of each individual or cluster is represented by a mixed model. Here, the vector $\boldsymbol{u}$ is the vector of cluster or individual specific random effects.

Model Assumptions. For $\boldsymbol{u}_i$ and $\boldsymbol{e}_i$, $i = 1, ..., n$, assume

- $\boldsymbol{u}_1, ..., \boldsymbol{u}_n, \boldsymbol{e}_1, ..., \boldsymbol{e}_n$ are uncorrelated.
- $\boldsymbol{u}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}_i)$
- $\boldsymbol{e}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$

Note that in the setup of clustered or longitudinal data the column vectors of $\boldsymbol{Z}_i$ are often a subset of the column vectors of $\boldsymbol{X}_i$. This regards the fact that the value of the response is not merely influenced by the covariates but also by the interaction of the covariate with the individual or cluster. For example the body height of an individual might be determined by a common intercept term and by an individual specific influence:

$$\boldsymbol{y}_i = \boldsymbol{1}_{t_i} \beta_0 + \boldsymbol{1}_{t_i} u_{i0}.$$

Model (2.26) can be written in a more compact way. One stacks the column vectors $\boldsymbol{y}_i$ and thus obtains a vector $\boldsymbol{y} \in \mathbb{R}^{\sum_{i=1}^n t_i}$, and the vectors $\boldsymbol{u}$ and $\boldsymbol{b}$ are constructed in the same way:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix}, \qquad \boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_n \end{pmatrix}, \qquad \boldsymbol{e} = \begin{pmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{pmatrix}.$$

Note that within the response vector $\boldsymbol{y}$ the measurements are ordered according to individual and within this according to time.

By the same manner, we obtain a matrix $\boldsymbol{X}$ through stacking the matrices $\boldsymbol{X}_i$. The design matrix $\boldsymbol{Z}$ of the random effects is a block diagonal matrix with the $\boldsymbol{Z}_i$'s on the diagonal:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_n \end{bmatrix} \qquad \boldsymbol{Z} = diag(\boldsymbol{Z}_1, ..., \boldsymbol{Z}_n) = \begin{bmatrix} \boldsymbol{Z}_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{Z}_n \end{bmatrix}.$$

This yields

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{g} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}.$$

Note that the block diagonal structure of $\boldsymbol{Z}$ implies that the measurments of one individual do only depend on the corresponding random effects.

Here, $\boldsymbol{u} \sim \mathcal{N}_{\sum_{i=1}^n t_i}(\boldsymbol{0}, \boldsymbol{D})$ and $\boldsymbol{e} \sim \mathcal{N}_{\sum_{i=1}^n t_i}(\boldsymbol{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{D} = diag(\boldsymbol{D}_1, ..., \boldsymbol{D}_n) \quad \text{and} \quad \boldsymbol{\Sigma} = diag(\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_n),$$

respectively.

2.2.1. Mixed ANOVA Models

In most applications of mixed models one is interested in experiments with different (treatment)-groups. The corresponding models are called ANOVA-models. ANOVA stands for analysis of variance and refers to the fact that the variance is partitioned into different components. These correspond to different sources of variation. ANOVA models are well known and described. Mixed ANOVA models are for example introduced in [30].

First, consider the following One-Way ANOVA or One-Way classification model: let $i = 1, ..., p$ be the number of treatment groups and let m_i be the number of observations per group. For example, the treatment groups could be patients who take different medications.

Let y_{ij} denote the measurement of the jth experimental unit in the ith group. Then we assume the following model (compare [23]):

$$y_{ij} = \beta_0 + \beta_i + e_{ij}, \qquad i = 1, ...p, \ j = 1, ..., , m_i. \tag{2.27}$$

β_0 is called the *overall mean*, β_i is the effect of the ith group and e_{ij} is the *within-group error*.

Model Assumptions. We assume that the e_{ij} are independently distributed and follow a normal distribution with mean 0 and common variance σ^2.

We can write (2.27) in matrix-vector form. Define

$$\boldsymbol{y} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1,m_1} \\ \vdots \\ y_{p1} \\ \vdots \\ y_{pm_p} \end{pmatrix} \in \mathbb{R}^{\sum_{i=1}^p m_i}.$$

Let $N := \sum_{i=1}^p m_i$ denote the total number of observations. Consider a design matrix $\boldsymbol{X}$ whose entries are defined as:

$$x_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is from group } i \\ 0 & \text{otherwhise.} \end{cases}$$

and

$$\boldsymbol{x}_i = (x_{i1}, ..., x_{im_i})^T \qquad i = 1, ..., p.$$

Note that for each group i, $x_{ij} = 1$ for exactly one $j \in \{1, ..., m_i\}$.

Let

$$\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_p] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \in \mathcal{M}_{N,p}.$$

Furthermore, let $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T \in \mathbb{R}^p$ and let $\boldsymbol{e} = (e_{11}, ..., e_{1m_1}, ..., e_{p,m_p})^T \in \mathbb{R}^N$. Then,

$$\boldsymbol{y} = \beta_0 \mathbf{1}_N + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}. \tag{2.28}$$

This is a univariate linear model with overall mean β_0 as intercept term.

Recall that a model with p treatment groups is said to be balanced, if the number of replications of each treatment is the same. This implies that the number of ones in each column of $\boldsymbol{X}$ is equal.

Remark 2.2.2. In the balanced case $\boldsymbol{X}$ has Kronecker product structure:

$$\boldsymbol{X} = \boldsymbol{I}_p \otimes \mathbf{1}_m. \tag{2.29}$$

The Kronecker product is defined in Definition A.4.

In model (2.27) there is only one categorial covariate that provides the grouping. It is called *factor* or *grouping variable*. The levels of this factor are here denoted by $\beta_1, ..., \beta_p$. They are called *effects*. In theory as well as in applications one is interested in models with more than one grouping variable. This leads to factorial ANOVA models, as for example discussed in [26] and [13]. There are differences for balanced and unbalanced models. In order to simplify the discussion, they are treated separately.

The Balanced Mixed ANOVA Model

As mentioned above, one has an equal number of observations in each of the experimental treatments in an experiment with balanced data. In the following we will assume balanced models.

Instead of one grouping factor we are now interested in models with ν grouping factors. Consider for instance patients taking part in a study where they are assigned to different medical treatment groups. The characteristics are observed at different points in time. Then one grouping factor is time and the other is the medical treatment. Thus, in this case $\nu = 2$.

The parameters belonging to one factor are combined in one parameter vector, each. They are called the levels of the factor. This yields the following model:

$$\boldsymbol{y} = \beta_0 \mathbf{1}_N + \sum_{i=1}^{\nu} \boldsymbol{H}_i \boldsymbol{\beta}_i + \boldsymbol{e}. \tag{2.30}$$

By setting $\boldsymbol{\beta}_\nu := \boldsymbol{e}$, $\boldsymbol{H}_{\nu+1} := \boldsymbol{I}_N$ and $\boldsymbol{H}_0 = \mathbf{1}_N$, the model can be rewritten:

$$\boldsymbol{y} = \sum_{i=1}^{\nu+1} \boldsymbol{H}_i \boldsymbol{\beta}_i.$$

The model is balanced when the number of observations is equal for each cell, i.e. for each combination of grouping factors. In this case each matrix $\boldsymbol{H}_i$ has full column rank and can be written as kronecker product of ν matrices that are either the identity matrix or the vector containing ones (compare [13] or [26]).

Model Assumptions. $\boldsymbol{e} \sim \mathcal{N}_n(\mathbf{0}, \sigma \boldsymbol{I}_n)$.

Now suppose the model is mixed, i.e. there are fixed effects as well as random effects in addition to the error term. Again assume without loss of generality that the first $d+1$ summands are fixed and the rest are random, $0 < d < \nu + 1$. Fixed part and random part of the model can be separated in the following way:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{g} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}, \tag{2.31}$$

where

$$\boldsymbol{X}\boldsymbol{g} = \sum_{i=0}^{\nu-d} \boldsymbol{H}_i \boldsymbol{\beta}_i$$

represents the fixed part of the model and

$$\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e} = \sum_{i=\nu-d+1}^{\nu+1} \boldsymbol{H}_i \boldsymbol{\beta}_i$$

is the random portion of the model.

Model Assumptions. Let c_i be the number of columns of $\boldsymbol{H}_i$. Following assumptions are made in addition to the assumptions of the general balanced ANOVA models:

- $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_d$ are unknown fixed parameter vectors.
- $\boldsymbol{\beta}_{d+1}, ..., \boldsymbol{\beta}_{\nu+1}$ are independent normally distributed random vectors with means $\mathbf{0}$ and covariance matrices $Cov(\boldsymbol{\beta_i}) = \sigma_i^2 \boldsymbol{I}_{c_i}$

In applications models with two grouping factors are very common, especially in experiments with repeated measurements. They are called Two-Way ANOVA models. Therefore an example balanced Two-Way ANOVA models is given in the following.

Example 2.2.1 (Two Way Crossed Classification With Interaction)**.** This model is usually applied to experiments with two grouping factors, referred to as *treatments* and *blocks* and an interaction term which is often referred to as *block-by-treatment* interaction. $\boldsymbol{y}_{ij}$ denotes the measurements from the jth treatment in the ith block. One has the model:

$$y_{ijk} = \beta_0 + a_i + b_j + (ab)_{ij} + e_{ijk}, \qquad i = 1, ..., p,\ j = 1, ..., m,\ k = 1, ..., t. \tag{2.32}$$

Assume that β_0 and a_i are fixed and that b_j, the interactions $(ab)_{ij}$ and the error terms e_{ijk} are random variables. The model is balanced because there is the same number of observations for each ij combination.

This model can be written in matrix vector form:

$$\boldsymbol{y} = \boldsymbol{H}_0\boldsymbol{\beta}_0 + \boldsymbol{H}_1\boldsymbol{\beta}_1 + \boldsymbol{H}_2\boldsymbol{\beta}_2 + \boldsymbol{H}_3\boldsymbol{\beta}_3 + \boldsymbol{H}_4\boldsymbol{\beta}_4,$$

where

$$\begin{aligned}
\boldsymbol{H}_0 &= \mathbf{1}_p \otimes \mathbf{1}_m \otimes \mathbf{1}_t, & \boldsymbol{\beta}_0 &= \beta_0,\\
\boldsymbol{H}_1 &= \boldsymbol{I}_p \otimes \mathbf{1}_m \otimes \mathbf{1}_t, & \boldsymbol{\beta}_1 &= (a_1, ..., a_p)^T,\\
\boldsymbol{H}_2 &= \mathbf{1}_p \otimes \boldsymbol{I}_m \otimes \mathbf{1}_t, & \boldsymbol{\beta}_2 &= (b_1, ..., b_m)^T,\\
\boldsymbol{H}_3 &= \boldsymbol{I}_p \otimes \boldsymbol{I}_m \otimes \mathbf{1}_t, & \boldsymbol{\beta}_3 &= ((ab)_{11}, (ab)_{12}, ..., (ab)_{pm})^T,\\
\boldsymbol{H}_4 &= \boldsymbol{I}_p \otimes \boldsymbol{I}_m \otimes \boldsymbol{I}_t, & \boldsymbol{\beta}_4 &= (e_{111}, e_{112}, ..., e_{pmt)})^T.
\end{aligned}$$

The fixed portion of the model is given by

$$\boldsymbol{Xg} = \boldsymbol{H}_0\boldsymbol{\beta}_0 + \boldsymbol{H}_1\boldsymbol{\beta}_1$$

and the random portion of the model is given by

$$\boldsymbol{Zu} + \boldsymbol{e} = \boldsymbol{H}_2\boldsymbol{\beta}_2 + \boldsymbol{H}_3\boldsymbol{\beta}_3 + \boldsymbol{H}_4\boldsymbol{\beta}_4.$$

□

The univariate linear mixed model was first introduced by Scheffé in [22]. He considered such a two factor setup with interaction effects.

It depends on the experimental setup, which factors are said to be fixed, and which are considered to be random. Fixed factors are those, whose levels are of special interest to the experimenter. Thus, only those levels of particular concern are considered. Therefore the factors are said to have fixed effects. The corresponding levels are typically called treatments. If, however, the levels of a factor that are used in the experiment are chosen at random from a (typically large) population of potential levels, they are said to have random effects and are called random factors (see [14]).

Example 2.2.2. As an example, Scheffé discusses a model with two grouping factors. The experiment consists of p different makes of machines and m workers working on them. The workers that take actually part in the experiment are considered to be a sample from a large population of workers. The factor of interest is the first one. The experimenter wants to compare the means of the different makes of machines. The workers can be considered as random, as they were selected at random from a large population. The experimenter could want to know, whether there exists a significant variation among the workers. This possible variation should be taken into account before comparing the means of the different kinds of machines.

Thus, the machines correspond to the levels of the fixed factor and the workers depend to the levels of the random factor in Scheffés model. We consider a Two-Way crossed classification with interaction effects. Thus a_i and b_j are the corresponding main effects of the factors and $(ab)_{ij}$ is the interaction effect of worker j with machine i. The model corresponds to (2.32).

The Unbalanced Mixed ANOVA Model

In this section, the unbalanced univariate mixed model is roughly discussed. It is just a very short reference that shall help to illustrate the difference to balanced models. For a more detailed consideration, the reader is recommended to consult [7] or [26]. Tests and estimation in unbalanced models as well as examples are given in [13] and in [17].

In an experimental design with unbalanced data, the imbalance can be caused by several reasons, as described by Shaw and Mitchell-Olds in [21]. There may be unequal numbers of observations for the treatment combinations, thus the sample size is unequal. Some treatment combinations, denoted as *cells* may be missing altogether. In a multivariate setup it is also possible that some of the experimental units have only been measured on a subset of the responses. Unbalanced data do often occur in practice, especially, when longitudinal data are concerned. In longitudinal studies it is not possible to control all the circumstances of the measurements, as Laird and Ware describe in [16].

We consider the following model:

$$\boldsymbol{y} = \sum_{i=0}^{d1} \boldsymbol{X}_i \boldsymbol{g}_i + \sum_{j=1}^{d2} \boldsymbol{Z}_i \boldsymbol{u}_i + \boldsymbol{e}. \tag{2.33}$$

Let N denote the total number of observations. Again, $\boldsymbol{y} \in \mathbb{R}^N$ is the response variable, $\boldsymbol{X}_i$'s and the $\boldsymbol{Z}_i$'s are design matrices, the $\boldsymbol{g}_i$'s are the vectors of fixed effects and the $\boldsymbol{u}_i$'s are vectors of random effects. The model assumptions are the same as for the balanced model.

Recall that in the balanced case, we had an equal number of observations per cell. This does no longer hold in the unbalanced case. For example, in an experiment with patients that are observed at different points in time, some subjects might have less observations than others because they leave the study before it ends. Or, in an experiment with several medical treatments, there might be a different number of patients assigned to each group.

The entries of the design matrices are still either zero or one, depending on whether the the subject is in the level of the corresponding factor. But in unbalanced models it is no longer possible to write the design matrices as kronecker product of identity matrices or vectors of ones. This complicates calculations in some cases.

3. The Multivariate Linear Mixed Model

3.1. The Multivariate Linear Mixed Model

The multivariate linear mixed model is a very useful concept, especially in context of repeated measures. It is an extension to the univariate mixed model. In this chapter a description of the model and its specification in a repeated measures design are given. We are particularly interested in hypotheses tests for experiments with repeated measurements.

3.1.1. Model Specification

As seen above, in the multivariate linear regression model the design matrix $\boldsymbol{X}$ and the parameter matrix $\boldsymbol{B}$ are nonrandom and thus the only random component in the model is the error matrix $\boldsymbol{E}$. This is different for the multivariate mixed model. We have already seen in the univariate case that there are also models with both, fixed and random effects. This is the same for multivariate mixed models. Therefore, the model can be formulated in a similar way as in the univariate case. The main difference is that we are interested in r response variables instead of 1.

First, consider a general multivariate linear model:

$$\boldsymbol{Y} = \widetilde{\boldsymbol{X}}\boldsymbol{B} + \boldsymbol{E},$$

where

$$\begin{aligned}
\boldsymbol{Y} &= [\boldsymbol{y}_1, ..., \boldsymbol{y}_r] \in \mathcal{M}_{n,r}, \\
\boldsymbol{X} &= [\mathbf{1}_n, \boldsymbol{x}_1, ..., \boldsymbol{x}_p] \in \mathcal{M}_{n,p+1}, \\
\boldsymbol{B} &= [\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_r] \in \mathcal{M}_{p+1,r} \text{ and} \\
\boldsymbol{E} &= [\boldsymbol{e}_1, ..., \boldsymbol{e}_r] \in \mathcal{M}_{n,r}.
\end{aligned}$$

Now assume that the first $d+1$ effects are fixed and the rest are random and separate the effects:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{G} + \boldsymbol{Z}\boldsymbol{U} + \boldsymbol{E}, \tag{3.1}$$

where

$$\boldsymbol{X} = [\mathbf{1}_n, \boldsymbol{x}_1, ..., \boldsymbol{x}_d] \in \mathcal{M}_{n,d+1} \quad , \quad \boldsymbol{Z} = [\boldsymbol{x}_{d+1}, ..., \boldsymbol{x}_p] \in \mathcal{M}_{n,p-(d+1)}$$

$$\boldsymbol{G} = \begin{bmatrix} \beta_{01} & \cdots & \beta_{0r} \\ \vdots & & \\ & & \vdots \\ \beta_{d1} & \cdots & \beta_{dr} \end{bmatrix} \in \mathcal{M}_{d+1,r} \quad \text{and} \quad \boldsymbol{u} = \begin{bmatrix} \beta_{(d+1)1} & \cdots & \beta_{(d+1)r} \\ \vdots & & \\ & & \vdots \\ \beta_{p1} & \cdots & \beta_{pr} \end{bmatrix} \in \mathcal{M}_{p-(d+1),r}.$$

Again, we are especially interested in balanced and unbalanced MANOVA models (as extension to the corresponding ANOVA models).

Balanced Data MANOVA Models

We want to consider the multivariate analogon to the univariate balanced mixed ANOVA model discussed in 2.2.1. Again, we consider ν grouping factors and an equal number of observations per cell. Instead of one response variable we are now interested in r responses that are supposed to be somehow correlated. For instance, we might not only be interested in the body height of children but also in their weight and in their size of shoe. We assume that these responses are influenced by one or more factors.

So in the case of multivariate responses, observations of the experimental units, error terms and the various effects of the model are expressed in terms of matrices.

The following multivariate model is considered:

$$\boldsymbol{Y} = \sum_{i=0}^{\nu} \boldsymbol{H}_i \boldsymbol{B}_i + \boldsymbol{E}. \tag{3.2}$$

$\boldsymbol{Y}$ is an $n \times r$ data matrix. The matrices $\boldsymbol{H}_i \in \mathcal{M}_{n,c_i}$ are known matrices of rank c_i. $\boldsymbol{B}_i \in \mathcal{M}_{c_i,r}$ is the ith effect matrix and $\boldsymbol{E} \in \mathcal{M}_{n,r}$ is the matrix of random errors.

The observations of the jth response are given by the jth column of $\boldsymbol{Y}$. This column is denoted by $\boldsymbol{y_j}$ and has a representation by a univariate model:

$$\boldsymbol{y}_j = \sum_{i=0}^{\nu} \boldsymbol{H}_i \boldsymbol{\beta}_{ij} + \boldsymbol{e}_j \qquad , \qquad j = 1, ..., r. \tag{3.3}$$

$\boldsymbol{\beta}_{ij}$ is the jth column of $\boldsymbol{B}_i$ and $\boldsymbol{e}_j$ is the jth column of $\boldsymbol{E}$. Model (3.3) is a univariate mixed ANOVA model. Model (3.2) is said to be *balanced*, if model (3.3) is balanced for $j = 1, ..., r$. In this case, the matrices $\boldsymbol{H}_i$ are a Kronecker product of identity matrices and vectors of ones (see [13]).

Again we want to consider mixed models, so we suppose the first $d+1$ effects to be fixed and the rest to be random, where $0 \le d \le \nu$. Then we obtain the balanced multivariate mixed model by separating fixed and random effects:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{G} + \boldsymbol{Z}\boldsymbol{U} + \boldsymbol{E}, \tag{3.4}$$

where the fixed portion of the model is

$$\boldsymbol{X}\boldsymbol{G} = \sum_{i=0}^{d} \boldsymbol{H}_i \boldsymbol{B}_i$$

and the random portion besides the experimental error is

$$\boldsymbol{Z}\boldsymbol{U} = \sum_{i=d+1}^{\nu} \boldsymbol{H}_i \boldsymbol{B}_i.$$

For $i = 0, ..., d$ the $\boldsymbol{B}_i$'s are fixed, but unknown parameters. For $i = d+1, ..., \nu$ the $\boldsymbol{B}_i$'s are random variables.

Model Assumptions.

- The rows of $\boldsymbol{E}$ are uncorrelated r-variate normal vectors with mean $\boldsymbol{0}$ and common covariance matrix $\boldsymbol{\Sigma}_{\nu+1}$

- The rows of $\boldsymbol{B}_i$ are uncorrelated and also independent of $\boldsymbol{E}$.
- The rows od $\boldsymbol{B}_i$ are normally distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_i$, $i = d+1, ..., \nu$

Sometimes the normality assumptions are omitted.

Example 3.1.1 (Two Way Crossed Classification With Interaction)**.** Again we consider a model with blocks and treatments and their interaction. Let $\widetilde{\boldsymbol{y}}_{ijk} = (y_{1ijk}, ..., y_{rijk})$ denote the observations of the r dependent variables under the ith level of the first factor combined with the jth level of the second one. This yields the following model:

$$\widetilde{\boldsymbol{y}}_{ijk} = \boldsymbol{\beta}_0 + \boldsymbol{a}_i + \boldsymbol{b}_j + \boldsymbol{ab}_{ij} + \boldsymbol{e}_{ijk}, \quad i = 1, ..., p;\ j = 1, ..., m, k = 1, ..., t. \tag{3.5}$$

$\boldsymbol{\beta}_0$ is a vector of unknown constants. $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$ represent the main levels of the two factors, $\boldsymbol{ab}_{ij}$ is the vector of interaction effects and $\boldsymbol{e}_{ijk}$ represents the random error.

The univariate two way model for the lth element of $\widetilde{\boldsymbol{y}}_{ijk}$ is given by

$$y_{lijk} = \beta_{0l} + a_{il} + b_{jl} + c_{ijl} + e_{ijkl}, \quad l = 1, ..., r. \tag{3.6}$$

Now define a new vector $\boldsymbol{y}_{lij}$ as $(y_{lij1}, ..., y_{lijt})'$. Let $\boldsymbol{y}_l$ be the vector one obtains by placing the $\boldsymbol{y}_{lij}$'s after one another for $i = 1, ..., p$, $j = 1, ..., m$. $\boldsymbol{e}_j$ is defined in the same way. Then the model for $\boldsymbol{y}_l$ is

$$\boldsymbol{y}_l = \boldsymbol{H}_0\boldsymbol{\beta}_{0l} + \boldsymbol{H}_1\boldsymbol{\beta}_{2l} + \boldsymbol{H}_3\boldsymbol{\beta}_{3l} + \boldsymbol{e}_l, \quad l = 1, ..., r, \tag{3.7}$$

where

$$\begin{aligned}
\boldsymbol{H}_0 &= \mathbf{1}_p \otimes \mathbf{1}_m \otimes \mathbf{1}_t \quad , \quad & \boldsymbol{\beta}_{0l} &= \beta_{0l}, \\
\boldsymbol{H}_1 &= \boldsymbol{I}_p \otimes \mathbf{1}_m \otimes \mathbf{1}_t \quad , \quad & \boldsymbol{\beta}_{1l} &= (\alpha_{1l}, ..., \alpha_{pl})^T, \\
\boldsymbol{H}_2 &= \mathbf{1}_p \otimes \boldsymbol{I}_m \otimes \mathbf{1}_t \quad , \quad & \boldsymbol{\beta}_{2l} &= (b_{1l}, ..., b_{ml})^T, \\
\boldsymbol{H}_3 &= \boldsymbol{I}_p \otimes \boldsymbol{I}_m \otimes \mathbf{1}_t \quad , \quad & \boldsymbol{\beta}_{3l} &= (c_{11l}, c_{12l}, ..., c_{pml})^T.
\end{aligned}$$

To obtain a single multivariate model, the $\boldsymbol{y}_l$'s are combined into one matrix $\boldsymbol{Y}$. This results in

$$\boldsymbol{Y} = \boldsymbol{H}_0\boldsymbol{B}_0 + \boldsymbol{H}_1\boldsymbol{B}_1 + \boldsymbol{H}_2\boldsymbol{B}_2 + \boldsymbol{H}_3\boldsymbol{B}_3 + \boldsymbol{E}. \tag{3.8}$$

The matrix $\boldsymbol{B}_i$ is the one that has $\boldsymbol{\beta}_{ij}$ as its jth column, $i = 1, 2, 3$, $j = 1, ..., r$. Model (3.8) has the same form as model (3.2). The second factor and the interaction of both factors are assumed to be random. Model (3.8) has the same form as model (3.2). The fixed portion is represented by

$$\boldsymbol{XG} := \boldsymbol{H}_0\boldsymbol{B}_0 + \boldsymbol{H}_1\boldsymbol{B}_1,$$

whereas the random effects are represented by

$$\boldsymbol{ZU} := \boldsymbol{H}_2\boldsymbol{B}_2 + \boldsymbol{H}_3\boldsymbol{B}_3.$$

□

Unbalanced Data

Now, the unbalanced case is considered. Thus, the number of observations per cell can be unequal. The model is the multivariate extension to model (2.33):

$$\boldsymbol{Y} = \sum_{i=0}^{d1} \boldsymbol{X_i B_i} + \sum_{j=1}^{d2} \boldsymbol{Z}_j \boldsymbol{U}_j + \boldsymbol{E}, \tag{3.9}$$

where $\boldsymbol{Y} \in \mathcal{M}_{n,r}$ is a data matrix, the $\boldsymbol{X_i}$ are known design matrices, the $\boldsymbol{B}_i$ are matrices of fixed but unknown effects, $\boldsymbol{Z}_j$ are design matrices of full column rank, $\boldsymbol{U}_j$ are random effects matrices and $\boldsymbol{E}$ is again an $n \times r$ matrix of unknown random errors.

As in the univariate unbalanced model, the entries of the design matrices are either equal to one or to zero, but the matrices are not a kronecker product of identity matrices or vectors of ones.

The model assumptions are the same as for the balanced model.

Example 3.1.2. Thomas suggested the following setup in [27]: the subjects are each randomly assigned to one of the m treatment groups. It is possible that the groups have a different number of subjects. r different characteristics are measured on each subject on each of m groups under given treatment conditions (time) k, $k = 1, ..., t$. Now $\boldsymbol{y}_{ij}$ denotes the vector that represents the r observations of the ith subject in group j under treatment condition k, where $i = 1, ..., n_j$, $j = 1, ..., g$ and $k = 1, ..., t$. Thomas assumed the following model for y_{ijk}:

$$\boldsymbol{y}_{ijk} = \boldsymbol{\beta}_0 + \boldsymbol{a}_i + \boldsymbol{b}_j + (\boldsymbol{ab})_{ij} + \boldsymbol{e}_{ijk}$$

This model looks very similar to model (3.5). There are treatment effect, group effects and a groups by treatment interaction effect. However, we have different number of subjects in each group here, which is caused by the fact that the subjects are assigned to the groups randomly. Thus the data are unbalanced. The design matrices cannot be given in kronecker product form.

□

3.1.2. The Multivariate Mixed Model Analysis for Testing Hypotheses

We are interested in parameter estimation and hypotheses tests in multivariate mixed models and want to concentrate on the maximum likelihood approach. This section shall only give an impression of the peculiarities of estimation and testing in multivariate mixed models. The results mentioned here require a lot of theory. We only give a rough summary and recommend [13] and [19] for further reading. Later we will concentrate on the special case of hypotheses tests in experiments with repeated measurements.

We have already seen maximum likelihood estimates and likelihood ratio test statistics for multivariate linear regression models in Section 2.1. Note that the multivariate linear mixed model can be written as

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{\zeta},$$

where

$$\boldsymbol{\zeta} = \boldsymbol{ZU} + \boldsymbol{E} \quad \in \mathcal{M}_{n,r}$$

is the random part of the model. This looks similar to the multivariate linear regression model. Estimators and test statistics can in principle be derived by the same techniques. However, the distributions for these are different, because the error matrix is a sum of random matrices. This complicates the derivation of distributions.

Recall that in Section 2.1 we calculated the SSCP matrices due to regression and residual in order to find the likelihood ratio test statistic. In general we want to partition the variance into components corresponding to different sources of variation. In the multivariate setup, it is a multivariate analysis of variance (MANOVA). We have already considered MANOVA models that have their name from this technique. It is possible to use a MANOVA to gain information about distributions of estimators and test statistics in a multivariate mixed model.

Mathew suggested the following definition for a MANOVA in [19]:

Definition 3.1. A MANOVA in a multivariate mixed model is a partition $\boldsymbol{Y}^T\boldsymbol{Y} = \sum_{i=1}^{\kappa} \boldsymbol{Y}^T\boldsymbol{R}_i\boldsymbol{Y}$, where the $\boldsymbol{R}_i$ are positive semidefinite matrices for $i = 1, ..., \kappa$, satisfying

- $\boldsymbol{Y}^T\boldsymbol{R}_i\boldsymbol{Y}$ are independent,
- $\boldsymbol{Y}^T\boldsymbol{R}_i\boldsymbol{Y}$ have (possibly non central) Wishart distribution,
- $\boldsymbol{Y}^T\boldsymbol{R}_i\boldsymbol{Y}$ have expected values that are functions of the parameters.

The last property can be used to find estimates based on the moments.

Now let $\boldsymbol{Q} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$ and $\boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{Q}$. Furthermore, suppose that $\boldsymbol{P}_1, ..., \boldsymbol{P}_t, \boldsymbol{Q}_1, ..., \boldsymbol{Q}_s$ are orthogonal, idempotent matrices such that $\boldsymbol{P} = \sum_{i=1}^{t} \boldsymbol{P}_i$ and $\boldsymbol{Q} = \sum_{j=1}^{s} \boldsymbol{Q}_j$. Then clearly,

$$\boldsymbol{Y}^T\boldsymbol{Y} = \boldsymbol{Y}^T(\boldsymbol{Q}+\boldsymbol{P})\boldsymbol{Y} = \sum_{j=1}^{s} \boldsymbol{Y}^T\boldsymbol{Q}_i\boldsymbol{Y} + \sum_{i=1}^{t} \boldsymbol{Y}^T\boldsymbol{P}_i\boldsymbol{Y}.$$

Mathew gives a conditions for the existence of such a MANOVA and the parameters for the Wishart distribution and Khuri, Mathew and Sinha show in [13], that such a partition can always be found for balanced MANOVA models. They show that in this case the matrices $\boldsymbol{P}_i$ and $\boldsymbol{Q}_j$ can be constructed and that $\boldsymbol{Y}^T\boldsymbol{Q}_i\boldsymbol{Y}$ and $\boldsymbol{Y}^T\boldsymbol{P}_i\boldsymbol{Y}$ are the SSCP matrices due to the fixed and random effects, respectively.

There are several ways to use a MANOVA for estimation as well as for testing. As mentioned above, moments can be used for estimation or the projection matrices can be used to find least squares estimates. If a MANOVA is given, then the components of variance due to the different effects can be estimated by using $\boldsymbol{Y}^T\boldsymbol{P}_i\boldsymbol{Y}$ and $\boldsymbol{Y}^T\boldsymbol{Q}_i\boldsymbol{Y}$.

For hypotheses tests in multivariate mixed models one can calculate the same likelihood ratio test statistic as in 2.5. Recall that $\boldsymbol{S}_E$ was defined as

$$\boldsymbol{S}_E = \boldsymbol{Y}^T\boldsymbol{P}\boldsymbol{Y} = \boldsymbol{Y}^T(\sum_{i=1}^{t} \boldsymbol{P}_i)\boldsymbol{Y}.$$

The $\boldsymbol{P}_i$'s follow Wishart distributions, but do not necessarily have the same covariance matrix. The MANOVA can be used to determine the (approximate) distributions of $\boldsymbol{S}_E$ and $\boldsymbol{S}_H$. Khuri, Mathew and Sinha use the MANOVA to show in [13], that $\boldsymbol{S}_E$ and $\boldsymbol{S}_H$ are both asymptotically distributed as Wishart matrices in the case of a balanced MANOVA model. They also given another way to use a MANOVA for hypotheses tests, but we are only interested in the maximum likelihood approach here.

3.2. Repeated Measurements Models

Univariate repeated measurements designs are often analyzed by the univariate mixed model, described in 2.2. It is also possible to analyze such a design by a multivariate analysis, if the number of observations is sufficient. Now it is of interest to look at data, where the responses of the repeated measures experiments are multivariate. It is possible to analyze such a setup by a doubly multivariate model, as for instance suggested in [1]. However, in this thesis the multivariate mixed model analysis is considered. It was suggested by Thomas in [27] and by Boik in [1]. A similar approach was also suggested by Reinsel in [20]. Later we shall see that a special covariance structure called multivariate sphericity is needed for the validity of the multivariate mixed model in repeated measurement designs.

The application of mixed models to repeated measurements setups was already discussed in 2.2 in the context of longitudinal studies. Recall that the univariate mixed model particularly came in useful in experiments with one dependent variable that was observed at several points in time, as it allowed to model individual specific influences on the dependent variable. Those were represented by random factors.

The following chapters concentrate on the analysis of repeated measurements experiments. It is therefore important to understand, how the multivariate mixed model can be used in context of such experimental setups which is derived in the next section. The derivation starts with a linear regression model. This more general approach allows a more detailed theoretical consideration of the model.

3.2.1. The Multivariate Mixed Model in Repeated Measurements Experiments

We consider a multivariate repeated measures experiment, where r distinct characteristics are examined, each of the n observed subjects gives a r-dimensional response on each of t occasions. Note that the number of observations is equal for each subject. Let y_{ijk} denote the jth observation on the ith subject at the kth point in time, $i = 1, ..., n$, $j = 1, ..., r$, $k = 1, ..., t$. The response variable is supposed to be linearly dependent on p covariates, thus consider the following linear model:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{\zeta}. \tag{3.10}$$

$$\boldsymbol{Y} = \begin{bmatrix} y_{111} & \cdots & y_{1r1} & \cdots & y_{1rt} \\ \vdots & & & & \vdots \\ y_{n11} & \cdots & y_{nr1} & \cdots & y_{nrt} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_1^T \\ \vdots \\ \boldsymbol{y}_n^T \end{bmatrix} \in \mathcal{M}_{n,rt}$$

is the response matrix. The ith row of $\boldsymbol{Y}$ represents the rt measurements of the response variables of the jth subject. As indicated above, the responses are ordered within each row according to time and within time according to dependent variable.

$$\boldsymbol{\zeta} = \begin{bmatrix} \boldsymbol{\zeta}_1^T \\ \vdots \\ \boldsymbol{\zeta}_n^T \end{bmatrix} \in \mathcal{M}_{n,rt}$$

is the matrix of random errors. Its ith row $\boldsymbol{\zeta}_i^T$ is the vector of random errors corresponding to the ith subject. The $\boldsymbol{\zeta}$ is used here instead of the common $\boldsymbol{E}$ in order to distinguish the multivariate linear model with repeated measurements from other linear models. In the following this symbol will always be used in the context of repeated measurements.

Furthermore, $\boldsymbol{X} \in \mathcal{M}_{n,p}$ is a known design matrix with rank $k \leq p$ and $\boldsymbol{B} \in \mathcal{M}_{p,rt}$ is the matrix of unknown (fixed) parameters. The model is also known as the doubly multivariate model and the corresponding analysis as doubly multivariate analysis.

Model Assumptions. Again it is assumed that $\boldsymbol{\zeta}_i \sim$ iid $\mathcal{N}_{rt}(\boldsymbol{0}, \boldsymbol{\Sigma})$, $i = 1, ..., n$, where $\boldsymbol{\Sigma} \in \mathcal{S}_{rt}^{+}$. Note that

$$vec(\boldsymbol{\zeta}^T) \sim \mathcal{N}_{nrt}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}).$$

The experimental conditions on each of the t occasions are given, but may be different. Multivariate models with repeated measurements occur in social sciences as well as in medicine or biology. As mentioned above, there are several possibilities to analyze such a dataset. One has to choose between a multivariate or a univariate approach and between a doubly multivariate and a mixed model. If the r dependent variables are uncorrelated, it may be sufficient to perform r separate analyses, for instance by using the univariate mixed model. However, if there is a correlation, a multivariate analysis is necessary.

Model (3.10) can also be analyzed by a multivariate approach, considering a multivariate linear regression model. No prior assumptions about the correlation structure of the repeated measurements are made. Thus, this information is not taken into account.

The multivariate mixed model analysis does take the special structure into account. To perform it, the data have to be rearranged. Boik [1] shows how to specify the model in the repeated measurements setup accordingly. First, the response matrix $\boldsymbol{Y}$ has to be rearranged.

Multivariate Mixed Model Analysis

As mentioned above, each of the i subjects has a rt-dimensional vector of observations. This vector is the ith row vector of $\boldsymbol{Y}$ in model (3.10). It was denoted by $\boldsymbol{y}_i \in \mathbb{R}^{rt}$. Now we rearrange $\boldsymbol{y}_i$ in such a way that we obtain a $t \times r$ matrix $\boldsymbol{Y}_i^*$ such that $vec((\boldsymbol{Y}_i^*)^T) = \boldsymbol{y}_i^T$, $i = 1, ..., n$, where the $vec()$-operator stacks the columns of $(\boldsymbol{Y}_i^*)^T$. So

$$\boldsymbol{Y}_i^* = \begin{bmatrix} y_{i11} & \cdots & y_{ir1} \\ y_{i12} & \cdots & y_{ir2} \\ \vdots & & \vdots \\ y_{i1t}, & \cdots & y_{irt} \end{bmatrix}$$

The columns of $\boldsymbol{Y}_j^*$ correspond to dependent variables, whereas the rows represent the occasions.

Now define the rearranged response matrix $\boldsymbol{Y}^*$ as follows:

$$\boldsymbol{Y}^* = \begin{bmatrix} \boldsymbol{Y}_1^* \\ \vdots \\ \boldsymbol{Y}_n^* \end{bmatrix} \in \mathcal{M}_{tn,r}.$$

We want to find a linear model for $\boldsymbol{Y}^*$. As one already has a linear model for $\boldsymbol{Y}$ and $\boldsymbol{Y}^*$ was constructed from the first, it is obvious to use the already known matrices $\boldsymbol{X}$, $\boldsymbol{B}$ and $\boldsymbol{\zeta}$ and rearrange them in a similar manner as above which induces new design -, parameter - and error matrices. The procedure of the construction follows now.

Let $\boldsymbol{b}_i \in \mathbb{R}^{rt}$ such that $\boldsymbol{b}_i^T$ is the ith row of $\boldsymbol{B}$, $i = 1, ..., p$ and let $\boldsymbol{\zeta}_j^T$, $j = 1, ..., n$ denote the jth row of $\boldsymbol{\zeta}$, as above . Then, $\boldsymbol{B}_i^*$ and $\boldsymbol{\zeta}_j^*$ are constructed in the same way as $\boldsymbol{Y}_i^*$:

$$\boldsymbol{B}_i^* = \begin{bmatrix} b_{i11} & \cdots & b_{ir1} \\ b_{i12} & \cdots & y_{ir2} \\ \vdots & & \vdots \\ b_{i1t}, & \cdots & b_{irt} \end{bmatrix} \in \mathcal{M}_{t,r} \qquad \boldsymbol{\zeta}_i^* = \begin{bmatrix} \zeta_{i11} & \cdots & \zeta_{ir1} \\ \zeta_{i12} & \cdots & \zeta_{ir2} \\ \vdots & & \vdots \\ \zeta_{i1t}, & \cdots & \zeta_{irt} \end{bmatrix} \in \mathcal{M}_{t,r}$$

and, accordingly

$$\boldsymbol{B}^* = \begin{bmatrix} \boldsymbol{B}_1^* \\ \vdots \\ \boldsymbol{B}_p^* \end{bmatrix} \in \mathcal{M}_{tp,r}, \qquad \boldsymbol{\zeta}^* = \begin{bmatrix} \boldsymbol{\zeta}_1^* \\ \vdots \\ \boldsymbol{\zeta}_p^* \end{bmatrix} \in \mathcal{M}_{tn,r}.$$

In this manner we obtain a $tp \times r$ matrix of unknown parameters and a $tn \times r$ matrix of random errors. Note that $vec((\boldsymbol{Y}^*)^T) = vec((\boldsymbol{Y})^T)$, $vec((\boldsymbol{B}^*)^T) = vec((\boldsymbol{B})^T)$ and $vec((\boldsymbol{\zeta}^*)^T) = vec((\boldsymbol{\zeta})^T)$.

This yields the following modified model:

$$\boldsymbol{Y}^* = (\boldsymbol{X} \otimes \boldsymbol{I}_t)\boldsymbol{B}^* + \boldsymbol{\zeta}^*. \tag{3.11}$$

Remark 3.2.1. $(\boldsymbol{X} \otimes \boldsymbol{I}_t) \in \mathcal{M}_{nt,pt}$ is a matrix whose $t \times t$ blocks are diagonal matrices by definition of the Kronecker product (see Definition A.4):

$$(\boldsymbol{X} \otimes \boldsymbol{I}_t) = \begin{pmatrix} diag(x_{11}) & diag(x_{12}) & \cdots & diag(x_{1p}) \\ \vdots & & & \vdots \\ diag(x_{n1}) & diag(x_{n2}) & \cdots & diag(x_{np}) \end{pmatrix}$$

The structure of the new design matrix preserves the dependency. This can easily be seen by multiplying out. For example, for the entry y_{ijk} of $\boldsymbol{Y}^*$ we get:

$$y_{ijk} = x_{i1}b_{1jk} + x_{i2}b_{2jk} + \cdots + x_{ip}b_{pjk}.$$

But this is exactly the same as in (3.10).

Remark 3.2.2. (3.11) is a mixed model because the special structure of $\boldsymbol{\zeta}^*$. Note that the rows of $\boldsymbol{\zeta}^*$ are still normally distributed, but need no longer to be independent. In fact, the rows of $\boldsymbol{\zeta}^*$ corresponding to one individual can be correlated. This implies individual influences on the outcome of the response variable.

3.2.2. Hypotheses Tests for Repeated Measures Experiments

Again, consider a multivariate repeated measurements experiment. First, the original form considered in (3.10), thus

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{\zeta}.$$

Most hypotheses of interest can be put in the following form:

$$H_0 : \boldsymbol{C}\boldsymbol{B}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) = \boldsymbol{0}, \tag{3.12}$$

Recall that we have already seen hypotheses of this form in section 2.1.

$\boldsymbol{M} \in \mathcal{M}_{q,t}$ has $rank(\boldsymbol{M}) = q$ and $\boldsymbol{C} \in \mathcal{M}_{s,p}$ has $rank(\boldsymbol{C}) = s$. The rows of $\boldsymbol{M}$ consist of q linear functions of the t time periods. Without loss of generality we assume that $\boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{I}_q$. The rows of $\boldsymbol{C}$ consist of coefficients of s estimable between group functions. For $\boldsymbol{CB}$ to be estimable, $\boldsymbol{C}$ must be contained in the column range of $\boldsymbol{X}$. Examples for $\boldsymbol{M}$ and $\boldsymbol{C}$ for different hypotheses are given in Chapter 5.

As we have seen in Section 2.1.2 in Chapter 2 and discussed in Section 3.1.2, the common likelihood ratio test statistic for a hypothesis of the form (3.12) is a function of

$$\frac{|\boldsymbol{S}_E|}{|\boldsymbol{S}_E + \boldsymbol{S}_H|},$$

where

$$\boldsymbol{S}_E = (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{Y}^T(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T)\boldsymbol{Y}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) \tag{3.13}$$

and

$$\boldsymbol{S}_H = (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{Y}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{Y}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r). \tag{3.14}$$

Note that $\boldsymbol{Y}_+ = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}_0 = \boldsymbol{Y}$, because $\boldsymbol{B}_0$ can be chosen to be $\boldsymbol{0}$ and that by Theorem A.5 $(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)^T = (\boldsymbol{M} \otimes \boldsymbol{I}_r)$.

To be able to perform the test we need to know the distributions of $\boldsymbol{S}_E$ and $\boldsymbol{S}_H$.

Theorem 3.2. *$\boldsymbol{S}_E$ and $\boldsymbol{S}_H$ are independently distributed as Wishart matrices:*

$$\boldsymbol{S}_E \sim W_{rq}((n-k), \boldsymbol{\Omega}, \boldsymbol{0})$$

and

$$\boldsymbol{S}_H \sim W_{rq}(s, \boldsymbol{\Omega}, \boldsymbol{\Xi}),$$

where

$$\boldsymbol{\Omega} = (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) \tag{3.15}$$

and

$$\boldsymbol{\Xi} = \boldsymbol{\Psi}\boldsymbol{\Omega}^{-1} \tag{3.16}$$

for

$$\boldsymbol{\Psi} = (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{B}^T\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}]^{-1}\boldsymbol{C}\boldsymbol{B}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r). \tag{3.17}$$

The proof of Theorem 3.2 is not given by Boik, hence we prove it.

Proof. **Independence:** The rows of $\boldsymbol{Y}$ are distributed as multivariate normal vectors and thus the rows of $\boldsymbol{Y}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)$ follow a multivariate normal distribution, as well: let $\boldsymbol{y}_i \in \mathbb{R}^{rt}$, $i = 1, ..., n$ such that $\boldsymbol{y}_i^T$ is the ith row of $\boldsymbol{Y}$. $\boldsymbol{y}_i$ follows a multivariate normal distribution and according to Theorem (A.14) $(\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{y}_i$ follows a multivariate normal distribution. But the transpose of this is the ith row vector of $\boldsymbol{Y}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)$.

Let $\boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$ and $\boldsymbol{P}_1 = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$.

According to Theorem (A.16),

$$\boldsymbol{S}_E = (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{Y}^T\boldsymbol{P}\boldsymbol{Y}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)$$

and

$$\boldsymbol{S}_H = (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{Y}^T\boldsymbol{P}_1\boldsymbol{Y}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)$$

are distributed independently, if and only if

$$\boldsymbol{P}\boldsymbol{P}_1 = \boldsymbol{0}.$$

But this is the case as:

$$\begin{aligned}
\boldsymbol{P}\boldsymbol{P}_1 = &\ (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T)(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T) \\
= &\ \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \\
&- \boldsymbol{X}\underbrace{(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-}_{=(\boldsymbol{X}^T\boldsymbol{X})^-}\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \\
= &\ \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \\
&- \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \\
= &\ \boldsymbol{P}_1 - \boldsymbol{P}_1 \\
= &\ \boldsymbol{0}.
\end{aligned}$$

Thus, $\boldsymbol{S}_E$ and $\boldsymbol{S}_H$ are independently distributed.

Wishartness: We want to apply Theorem (A.20). First, we have to show that all the required conditions are fulfilled. Let $\boldsymbol{A} := \boldsymbol{Y}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)$, hence

$$\boldsymbol{S}_E = \boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A}.$$

and

$$\boldsymbol{S}_H = \boldsymbol{A}^T\boldsymbol{P}_1\boldsymbol{A}.$$

As above, the rows of $\boldsymbol{A}$ follow a multivariate normal distribution. The means are the rows of the matrix $\boldsymbol{X}\boldsymbol{B}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)$ and the covariance matrix is $\underbrace{(\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)}_{=\boldsymbol{\Omega}}$.

$\boldsymbol{P}$ and $\boldsymbol{P}_1$ are idempotent. In Remark (2.1.2) we have seen, that $rank(\boldsymbol{P}) = n - k$ and $rank(\boldsymbol{P}_1) = s$, respectively.

From Theorem (A.20) it follows that

$$\boldsymbol{S}_E \sim W_{rq}((n-k), \boldsymbol{\Omega}, \boldsymbol{0}).$$

$$\boldsymbol{S}_H \sim W_{rq}(s, \boldsymbol{\Omega}, \boldsymbol{\Xi}),$$

where the parameters are given as in (3.2).

□

As already mentioned above, a multivariate mixed model analysis is only valid under special circumstances. We will see later that a special structure of $\boldsymbol{\Omega}$ as defined in (3.15) called multivariate sphericity is needed. Since we want to consider tests in mixed models, the SSCP matrices for regression and residual, respectively, shall now be given for the rearranged model (3.11). Thus

$$\boldsymbol{Y}^* = (\boldsymbol{X} \otimes \boldsymbol{I}_t)\boldsymbol{B}^* + \boldsymbol{\zeta}^*.$$

The null hypothesis in this case is given by

$$H_0 : (\boldsymbol{C} \otimes \boldsymbol{M})\boldsymbol{B}^* = \boldsymbol{0}, \tag{3.18}$$

Hypothesis (3.18) and (3.12) are identical, which can be verified by multiplying out and comparing the entries of the matrices.

Again, the likelihood ratio test statistic is a function of $\frac{|\boldsymbol{S}_E^*|}{|\boldsymbol{S}_E^*+\boldsymbol{S}_H^*|}$, where

$$\boldsymbol{S}_E^* = (\boldsymbol{Y^*})^T \left\{[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^- \boldsymbol{X}^T] \otimes \boldsymbol{M}^T\boldsymbol{M}\right\} \boldsymbol{Y^*} \tag{3.19}$$

and

$$\boldsymbol{S}_H^* = (\boldsymbol{Y^*})^T \left\{\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^- \boldsymbol{C}[\boldsymbol{C}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \otimes \boldsymbol{M}^T\boldsymbol{M}\right\} \boldsymbol{Y^*}. \tag{3.20}$$

The derivation of these is not given by Boik and shall therefore be given here. To get the matrices, we consider the reduced linear model

$$(\boldsymbol{I}_n \otimes \boldsymbol{M})\boldsymbol{Y}^* = \underbrace{(\boldsymbol{I}_n \otimes \boldsymbol{M})(\boldsymbol{X} \otimes \boldsymbol{I}_t)}_{=(\boldsymbol{X}\otimes\boldsymbol{M})}\boldsymbol{B}^* + (\boldsymbol{I}_n \otimes \boldsymbol{M})\boldsymbol{\zeta}^*. \tag{3.21}$$

Note: $(\boldsymbol{X} \otimes \boldsymbol{M}) = (\boldsymbol{X} \otimes \boldsymbol{I}_q)(\boldsymbol{I}_p \otimes \boldsymbol{M})$. Thus

$$(\boldsymbol{I}_n \otimes \boldsymbol{M})\boldsymbol{Y}^* = (\boldsymbol{X} \otimes \boldsymbol{I}_q)(\boldsymbol{I}_p \otimes \boldsymbol{M})\boldsymbol{B}^* + (\boldsymbol{I}_n \otimes \boldsymbol{M})\boldsymbol{\zeta}^*.$$

and

$$H_0 : (\boldsymbol{C} \otimes \boldsymbol{I}_q)(\boldsymbol{I}_p \otimes \boldsymbol{M})\boldsymbol{B}^* = \boldsymbol{0}.$$

Recall that we have already considered a reduced model in (2.1.6). Here, $(\boldsymbol{I}_n \otimes \boldsymbol{M})$ is multiplied to the model.

Now we derive $\boldsymbol{S}_E^*$ and $\boldsymbol{S}_H^*$ from this model in the same way as in (2.21) and (2.22). The calculations are only given explicitly for $\boldsymbol{S}_E^*$.

$$\boldsymbol{S}_E^* = (\boldsymbol{Y}^*)^T(\boldsymbol{I}_n \otimes \boldsymbol{M}^T)(\boldsymbol{I}_n - (\boldsymbol{X} \otimes \boldsymbol{I}_q)[(\boldsymbol{X}^T \otimes \boldsymbol{I}_q^T)(\boldsymbol{X} \otimes \boldsymbol{I}_q)]^-(\boldsymbol{X}^T \otimes \boldsymbol{I}_q^T))(\boldsymbol{I}_n \otimes \boldsymbol{M})\boldsymbol{Y}^*$$

We have used that for $\boldsymbol{A} \in \mathcal{M}_{n1,m1}$ and $\boldsymbol{B} \in \mathcal{M}_{n2,m2}$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^T = \boldsymbol{A}^T \otimes \boldsymbol{B}^T.$$

Furthermore, we will use that

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}$$

for $\boldsymbol{C} \in \mathcal{M}_{m1,d1}$ and $\boldsymbol{D} \in \mathcal{M}_{m2,d2}$. Both properties of the Kronecker product are given in Theorem (A.5). This yields:

$$\begin{aligned}
\boldsymbol{S}_E^* &= (\boldsymbol{Y}^*)^T(\boldsymbol{I}_n \otimes \boldsymbol{M}^T)\left\{\boldsymbol{I}_n - (\boldsymbol{X} \otimes \boldsymbol{I}_q)[\boldsymbol{X}^T\boldsymbol{X} \otimes \boldsymbol{I}_q]^-(\boldsymbol{X}^T \otimes \boldsymbol{I}_q)\right\}(\boldsymbol{I}_n \otimes \boldsymbol{M})\boldsymbol{Y}^* \\
&= (\boldsymbol{Y}^*)^T(\boldsymbol{I}_n \otimes \boldsymbol{M}^T)\left\{\boldsymbol{I}_n - (\boldsymbol{X} \otimes \boldsymbol{I}_q)[(\boldsymbol{X}^T\boldsymbol{X})^- \otimes \boldsymbol{I}_q](\boldsymbol{X}^T \otimes \boldsymbol{I}_q)\right\}(\boldsymbol{I}_n \otimes \boldsymbol{M})\boldsymbol{Y}^* \\
&= (\boldsymbol{Y}^*)^T(\boldsymbol{I}_n \otimes \boldsymbol{M}^T)\left\{\boldsymbol{I}_n - [\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^- \otimes \boldsymbol{I}_q](\boldsymbol{X}^T \otimes \boldsymbol{I}_q)\right\}(\boldsymbol{I}_n \otimes \boldsymbol{M})\boldsymbol{Y}^* \\
&= (\boldsymbol{Y}^*)^T(\boldsymbol{I}_n \otimes \boldsymbol{M}^T)\left\{\boldsymbol{I}_n - [\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \otimes \boldsymbol{I}_q]\right\}(\boldsymbol{I}_n \otimes \boldsymbol{M})\boldsymbol{Y}^* \\
&= (\boldsymbol{Y}^*)^T(\boldsymbol{I}_n \otimes \boldsymbol{M}^T)\left\{(\boldsymbol{I}_n \otimes \boldsymbol{M}) - [\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \otimes \boldsymbol{I}_q\boldsymbol{M}]\right\}\boldsymbol{Y}^* \\
&= (\boldsymbol{Y}^*)^T\{(\boldsymbol{I}_n \otimes \boldsymbol{M}^T\boldsymbol{M}) - [\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \otimes \boldsymbol{M}^T\boldsymbol{M}\}\boldsymbol{Y}^* \\
&= (\boldsymbol{Y}^*)^T([\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T] \otimes \boldsymbol{M}^T\boldsymbol{M})\boldsymbol{Y}^*
\end{aligned}$$

The calculations of $\boldsymbol{S}_H^*$ are based on the same principle, but using $\boldsymbol{P}_1$ instead of $\boldsymbol{P}$.

Remark 3.2.3. Let $\boldsymbol{\zeta}_i^* \in \mathcal{M}_{t,r}$, $i = 1, ..., p$ be defined as above, and let $\boldsymbol{\Omega}_i := cov(vec(\boldsymbol{\zeta}_i^*))$.

In a multivariate linear regression model the required conditions to test hypotheses of the form

$$H_0 : \boldsymbol{CB}(\boldsymbol{M}^T \otimes \boldsymbol{I}_n) = \boldsymbol{D}$$

against

$$H_1 : \boldsymbol{CB}(\boldsymbol{M}^T \otimes \boldsymbol{I}_n) \neq \boldsymbol{D}$$

are

Multivariate Normality

$$vec((\boldsymbol{\zeta}_i^*)^T) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$$

Homogeneity

$$(\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}_i(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) = (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}_j(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) \qquad \forall i, j,$$

as we have seen in Remark 2.1.7.

Lemma 3.3. *If*

$$vec(\boldsymbol{\zeta}^T) \sim \mathcal{N}_{npt}(\boldsymbol{0}, (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})) \tag{3.22}$$

is satisfied, then multivariate normality and homogeneity are also satisfied.

Proof. Assume, $vec(\boldsymbol{\zeta}^T) \sim \mathcal{N}_{npt}(\boldsymbol{0}, (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}))$ does holds.

Multivariate Normality $vec(\boldsymbol{\zeta}_i^*)$ is a subvector of $vec(\boldsymbol{\zeta}^T)$. (Remember that $vec(\boldsymbol{\zeta}^T) = vec((\boldsymbol{\zeta}^*)^T)$.)
According to Theorem A.14

$$\boldsymbol{\zeta}_i^* \sim \mathcal{N}_{tn}(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

hence the first condition follows.

Homogeneity From the multivariate normality condition it follows, that in case of repeated measurements $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$ $\forall i, j = 1, ..., r$.

Thus, homogeneity is fulfilled as well.

□

In the following we will assume that (3.22) holds.

However, if a multivariate mixed model analysis shall be performed, $\boldsymbol{S}_H^*$ and $\boldsymbol{S}_E^*$ have to be distributed independently as Wishart matrices. Only then it is possible to determine the distribution of the likelihood ratio test statistic. In the following chapter it will be shown that the independence of $\boldsymbol{S}_E^*$ and $\boldsymbol{S}_H^*$ follows from (3.22). The condition for the Wishartness of the matrices is derived.

4. Multivariate Sphericity

As stated above, four conditions have to be satisfied for a multivariate mixed model analysis to be valid:

- Multivariate normality,
- Homogeneity,
- Independence of $\boldsymbol{S}_H^*$ and $\boldsymbol{S}_E^*$,
- Wishartness of $\boldsymbol{S}_H^*$ and $\boldsymbol{S}_E^*$.

If (3.22) holds, hence

$$vec(\boldsymbol{\zeta}^T) \sim \mathcal{N}_{npt}(\mathbf{0}, (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})),$$

then multivariate normality and homogeneity hold, as shown in the previous chapter. Furthermore:

Lemma 4.1 (Boik, [1], Section 3). *If* (3.22) *holds* $\boldsymbol{S}_H^*$ *and* $\boldsymbol{S}_E^*$ *are independent.*

Proof. Let $\boldsymbol{S}_E^*$ and $\boldsymbol{S}_H^*$ denote the SSCP matrices of the mixed model analysis and $\boldsymbol{S}_E$ and $\boldsymbol{S}_H$ those of the multivariate linear model analysis, respectively. Partition $\boldsymbol{S}_E$ and $\boldsymbol{S}_H$ into $r \times r$ submatrices such that

$$\boldsymbol{S}_E = \{(\boldsymbol{S}_E)_{ij}\} \qquad \text{and} \qquad \boldsymbol{S}_H = \{(\boldsymbol{S}_H)_{ij}\} \qquad \forall i, j = 1, ..., q.$$

Then,

$$\boldsymbol{S}_E^* = \sum_{i=1}^{q} (\boldsymbol{S}_E)_{ii}, \tag{4.1}$$

and

$$\boldsymbol{S}_H^* = \sum_{i=1}^{q} (\boldsymbol{S}_H)_{ii}. \tag{4.2}$$

Hence, $\boldsymbol{S}_H^*$ is a matrix function of $\boldsymbol{S}_H$ and $\boldsymbol{S}_E^*$ is a matrix function of $\boldsymbol{S}_E$. If (3.22) is satisfied, then $\boldsymbol{S}_E$ and $\boldsymbol{S}_H$ are independent. It follows from (4.1) and (4.2) that $\boldsymbol{S}_E^*$ and $\boldsymbol{S}_H^*$ are also independent. □

[29] and [25] give examples of the SSCP matrices and there relations.

Thus, only Wishartness of $\boldsymbol{S}_E$ and $\boldsymbol{S}_H$ is required additionally. In the next section it will be shown that multivariate sphericity is both a necessary and sufficient condition for the SSCP matrices to be Wishart. It is worthwhile to show this because there is a test to check if multivariate sphericity is satisfied. So this test provides us with a possibility to find out whether multivariate sphericity is given and thus, whether a multivariate mixed model analysis is valid.

4.1. Multivariate Sphericity as Necessary and Sufficient Condition for the Multivariate Mixed Model Analysis

In this section we will again assume that

$$cov(vec(\boldsymbol{\zeta}^T)) \sim \mathcal{N}_{npt}(\mathbf{0}, (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})).$$

The following theorem gives a necessary and sufficient condition for a matrix quadratic form to be distributed according to a Wishart distribution. It is later on applied to $\boldsymbol{S}_E^*$ and $\boldsymbol{S}_H^*$.

Theorem 4.2 (Boik, [1], Theorem 1). *Let $\boldsymbol{U} \in \mathcal{M}_{n,m}$ have distribution $vec(\boldsymbol{U}^T) \sim \mathcal{N}_{nm}(\boldsymbol{\mu}, \boldsymbol{\Delta})$, where $\boldsymbol{\Delta} \in \mathcal{S}_{nm}^+$ and $\boldsymbol{\mu} = vec(\hat{\boldsymbol{M}}^T)$, $\hat{\boldsymbol{M}} \in \mathcal{M}_{n,m}$. Let $\boldsymbol{A} \in \mathcal{S}_n$ be a matrix of known constants. Then $\boldsymbol{Q} = \boldsymbol{U}^T\boldsymbol{A}\boldsymbol{U}$ has distribution $\boldsymbol{Q} \sim W_m(f, \boldsymbol{\Upsilon}, \boldsymbol{\Xi})$, if and only if $\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$ is idempotent.*
The parameters $f, \boldsymbol{\Upsilon}$ and $\boldsymbol{\Xi}$ are:

$$\begin{aligned} f &= rank(\boldsymbol{A}) \\ \boldsymbol{\Upsilon} &= [tr(\boldsymbol{A})^{-1}] \sum_{i=1}^{n} [(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m)]_{ii} \in \mathcal{S}_m^+ \\ \boldsymbol{\Xi} &= \hat{\boldsymbol{M}}^T \boldsymbol{A} \hat{\boldsymbol{M}} \boldsymbol{\Upsilon}^{-1} \end{aligned}$$

and $[(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m)]_{ii} \in \mathcal{M}_{m,m}$ is the ith diagonal block of $[(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m)] \in \mathcal{M}_{nm,nm}$, $i = 1, ..., n$.

Proof. **Sufficiency:** Idempotence of $\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$ implies that

$$(\boldsymbol{A} \otimes \boldsymbol{I}_m)^T \boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m) = (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}), \tag{4.3}$$

because:

$$\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1}) = \boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1}) \qquad \text{(Def. Idempotence, see A.9)}$$

Hence,

$$(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1}) = (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1}).$$

Now multiply the last equation with $(\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon})$ from the right-hand side and use Theorem A.3. This yields:

$$\begin{aligned} & (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})(\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}) = (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})(\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}) \\ \Leftrightarrow \quad & (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})\boldsymbol{\Delta}(\boldsymbol{A}\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}\boldsymbol{\Upsilon}) = (\boldsymbol{A}\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}\boldsymbol{\Upsilon}) \\ \Leftrightarrow \quad & (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m) = (\boldsymbol{A} \otimes \boldsymbol{I}_m). \end{aligned}$$

By multiplying $(\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon})$ from the left-hand side and using Theorem A.3 again, we get:

$$\begin{aligned} & (\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon})(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m) = (\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon})(\boldsymbol{A} \otimes \boldsymbol{I}_m) \\ \Leftrightarrow \quad & (\boldsymbol{I}_n\boldsymbol{A} \otimes \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{-1})\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m) = (\boldsymbol{I}_n\boldsymbol{A} \otimes \boldsymbol{\Upsilon}\boldsymbol{I}_m) \\ \Leftrightarrow \quad & (\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m) = (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}). \end{aligned}$$

Next we want to show that $(\boldsymbol{A} \otimes \boldsymbol{\Upsilon})$ is positive semidefinite. We use (4.3) and show that $(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m)$ is positive semidefinite. As $\boldsymbol{A}$ and $\boldsymbol{I}_m$ are symmetric,

$$\begin{aligned} & (\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m) = (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}) \\ \Leftrightarrow \quad & (\boldsymbol{A}^T \otimes \boldsymbol{I}_m^T)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m) = (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}) \end{aligned}$$

Finally, apply Theorem A.5:

$$(\boldsymbol{A}^T \otimes \boldsymbol{I}_m^T) = (\boldsymbol{A} \otimes \boldsymbol{I}_m)^T$$

and hence

$$(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}) = (\boldsymbol{A} \otimes \boldsymbol{I}_m)^T\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m).$$

This implies that for any vector $\boldsymbol{x} \in \mathbb{R}^{nm}$

$$\begin{aligned} & \boldsymbol{x}^T(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})\boldsymbol{x} \\ = \quad & \boldsymbol{x}^T(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{x} \\ = \quad & \boldsymbol{x}^T(\boldsymbol{A} \otimes \boldsymbol{I}_m)^T\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{x} \\ = \quad & ((\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{x})^T\boldsymbol{\Delta}\underbrace{(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{x}}_{=:\boldsymbol{y}} \\ = \quad & \boldsymbol{y}^T\boldsymbol{\Delta}\boldsymbol{y} \end{aligned}$$

Note that $\boldsymbol{y} \in \mathbb{R}^{nm}$. $\boldsymbol{\Delta}$ is positive definite. Thus

$$\boldsymbol{y}^T\boldsymbol{\Delta}\boldsymbol{y} > 0 \qquad \forall \boldsymbol{y} \in \mathbb{R}^{nm} \setminus \{\boldsymbol{0}\}$$

and therefore

$$\boldsymbol{x}^T(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})\boldsymbol{x} \geq 0 \qquad \forall \boldsymbol{x} \in \mathbb{R}^{nm}.$$

It follows that $(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$ is positive semidefinite.

(Note that $\boldsymbol{x}^T(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{x} = 0$ for $x \neq 0$, if $\boldsymbol{x}^T(\boldsymbol{A} \otimes \boldsymbol{I}_m) = 0$. Therefore $(\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m)$ is not necessarily positive definite.)

$\boldsymbol{\Upsilon}$ is positive definite as a sum of positive definite matrices. Thus, $\boldsymbol{A}$ is positive semidefinite.

Let f denote the rank of $\boldsymbol{A}$. Since $\boldsymbol{A}$ is positive semidefinite and symmetric, it can be factorized:

$$\boldsymbol{A} = \boldsymbol{G}^T\boldsymbol{G},$$

where $\boldsymbol{G} \in \mathcal{M}_{f,n}$, $rank(\boldsymbol{G}) = f$. $(\boldsymbol{G}\boldsymbol{G}^T) \in \mathcal{M}_{f,f}$ has also rank f. Thus, $(\boldsymbol{G}\boldsymbol{G}^T)$ has full rank and does have an inverse. The existence of $(\boldsymbol{G}\boldsymbol{G}^T)^{-1}$ and the idempotence of $\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$ is used to show that

$$(\boldsymbol{I}_f \otimes \boldsymbol{\Upsilon}) = (\boldsymbol{G} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{G}^T \otimes \boldsymbol{I}_m). \tag{4.4}$$

We start with (4.3) and perform calculations similar to those in the derivation of (4.3) by multiplying with kronecker products of suitable matrices from the right-hand and left-hand side.

$$\begin{aligned}
& (\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{I}_m) = (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}) \\
\Leftrightarrow \quad & (\boldsymbol{G}^T\boldsymbol{G} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{G}^T\boldsymbol{G} \otimes \boldsymbol{I}_m) = (\boldsymbol{G}^T\boldsymbol{G} \otimes \boldsymbol{\Upsilon}) \\
\Leftrightarrow \quad & (\boldsymbol{G}^T\boldsymbol{G} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{G}^T\boldsymbol{G}\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G}^T)^{-1} \otimes \boldsymbol{I}_m) = (\boldsymbol{G}^T\boldsymbol{G}\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G}^T)^{-1} \otimes \boldsymbol{\Upsilon}) \\
\Leftrightarrow \quad & (\boldsymbol{G}^T\boldsymbol{G} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{G}^T \otimes \boldsymbol{I}_m) = (\boldsymbol{G}^T \otimes \boldsymbol{\Upsilon}) \\
\Leftrightarrow \quad & ((\boldsymbol{G}\boldsymbol{G}^T)^{-1}\boldsymbol{G}\boldsymbol{G}^T\boldsymbol{G} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{G}^T \otimes \boldsymbol{I}_m) = ((\boldsymbol{G}\boldsymbol{G}^T)^{-1}\boldsymbol{G}\boldsymbol{G}^T \otimes \boldsymbol{\Upsilon}) \\
\Leftrightarrow \quad & (\boldsymbol{G} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{G}^T \otimes \boldsymbol{I}_m) = (\boldsymbol{I}_f \otimes \boldsymbol{\Upsilon}).
\end{aligned}$$

Recall that $\boldsymbol{Q}$ was defined as

$$\boldsymbol{Q} = \boldsymbol{U}^T\boldsymbol{A}\boldsymbol{U}.$$

Inserting $\boldsymbol{A} = \boldsymbol{G}^T\boldsymbol{G}$ yields

$$\boldsymbol{Q} = \boldsymbol{U}^T(\boldsymbol{G}^T\boldsymbol{G})\boldsymbol{U}.$$

$vec(\boldsymbol{U}^T) \sim \mathcal{N}_{nm}(\boldsymbol{\mu}, \boldsymbol{\Delta})$ and thus

$$vec((\boldsymbol{G}\boldsymbol{U})^T) \sim \mathcal{N}_{fm}((\boldsymbol{G} \otimes \boldsymbol{I}_m)\boldsymbol{\mu}, (\boldsymbol{G} \otimes \boldsymbol{I}_m)\boldsymbol{\Delta}(\boldsymbol{G} \otimes \boldsymbol{I}_m)^T).$$

One can simplify this by using (4.4):

$$vec((\boldsymbol{G}\boldsymbol{U})^T) \sim \mathcal{N}_{fm}((\boldsymbol{G} \otimes \boldsymbol{I}_m)\boldsymbol{\mu}, (\boldsymbol{I}_f \otimes \boldsymbol{\Upsilon})).$$

This implies that the rows of $\boldsymbol{G}\boldsymbol{U}$ follow a normal distribution with mean $\boldsymbol{G}\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Upsilon}$, according to Theorem A.14. Applying A.17 and Definition A.18, it follows that

$$\boldsymbol{Q} \sim W_m(f, \boldsymbol{\Upsilon}, \boldsymbol{\Xi}), \tag{4.5}$$

where

$$\begin{aligned}\boldsymbol{\Xi} &= \hat{\boldsymbol{M}}^T \boldsymbol{G}^T \boldsymbol{G} \hat{\boldsymbol{M}} \boldsymbol{\Upsilon}^{-1} \\ &= \hat{\boldsymbol{M}}^T \boldsymbol{A} \hat{\boldsymbol{M}} \boldsymbol{\Upsilon}^{-1}.\end{aligned}$$

The equation for $\boldsymbol{\Upsilon}$ follows from (4.3):

$$(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}) = (\boldsymbol{A} \otimes \boldsymbol{I}_m)^T \boldsymbol{\Delta} (\boldsymbol{A} \otimes \boldsymbol{I}_m)$$

Necessity: The characteristic function of $\boldsymbol{Q}$ is given by

$$\varphi(\boldsymbol{T}) = E[\exp(i \cdot tr(\boldsymbol{T}\boldsymbol{U}^T \boldsymbol{A}\boldsymbol{U}))]$$

for a real symmetric matrix $\boldsymbol{T}$ by definition of the characteristic function for random matrices. i denotes the imaginary unit, thus $\boldsymbol{i}^2 = -1$. The trace of the matrix is the same as the trace of its transpose and $\boldsymbol{A}$ and $\boldsymbol{T}$ are symmetric. Thus we can rewrite the characteristic function.

$$\varphi(\boldsymbol{T}) = E[\exp(i \cdot tr(\boldsymbol{U}^T \boldsymbol{A}\boldsymbol{U}\boldsymbol{T}))]$$

Using Theorem A.8, the exponent in the equation has another representation:

$$\begin{aligned} i \cdot tr(\boldsymbol{U}^T \boldsymbol{A}\boldsymbol{U}\boldsymbol{T}) &= i \cdot [vec((\boldsymbol{U}^T)^T)^T (\boldsymbol{T}^T \otimes \boldsymbol{A}) vec(\boldsymbol{U})] \\ &= i \cdot [vec(\boldsymbol{U})^T (\boldsymbol{T} \otimes \boldsymbol{A}) vec(\boldsymbol{U})] \end{aligned}$$

In the second step the symmetry of $\boldsymbol{T}$ was used. This is a scalar quadratic form in $vec(\boldsymbol{U})$. Boik gives another form of the characteristic function of $\boldsymbol{Q}$ in [1]:

$$\varphi(\boldsymbol{T}) = |\boldsymbol{I}_{nm} - 2i\Delta(\boldsymbol{A} \otimes \boldsymbol{T})|^{-\frac{1}{2}} \cdot \exp(i\boldsymbol{\mu}^T (\boldsymbol{A} \otimes \boldsymbol{T})[\boldsymbol{I}_{nm} - 2i\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{T})]^{-1} \boldsymbol{\mu}) \tag{4.6}$$

Furthermore it can also be obtained by using Theorem A.21.

$$\varphi(\boldsymbol{T}) = |\boldsymbol{I}_m - 2i\boldsymbol{\Upsilon}\boldsymbol{T}|^{-\frac{f}{2}} \cdot \exp(i\boldsymbol{\mu}^T [\boldsymbol{I}_n \otimes \boldsymbol{T}(\boldsymbol{I}_m - 2i\boldsymbol{\Upsilon}\boldsymbol{T})^{-1}] \boldsymbol{\mu}). \tag{4.7}$$

Equating (4.6) and (4.7) yields

$$\begin{aligned} & |\boldsymbol{I}_{nm} - 2i\Delta(\boldsymbol{A} \otimes \boldsymbol{T})|^{-\frac{1}{2}} \cdot \exp(i\boldsymbol{\mu}^T (\boldsymbol{A} \otimes \boldsymbol{T})[\boldsymbol{I}_{nm} - 2i\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{T})]^{-1} \boldsymbol{\mu}) \\ = \quad & |\boldsymbol{I}_m - 2i\boldsymbol{\Upsilon}\boldsymbol{T}|^{-\frac{f}{2}} \cdot \exp(i\boldsymbol{\mu}^T [\boldsymbol{I}_n \otimes \boldsymbol{T}(\boldsymbol{I}_m - 2i\boldsymbol{\Upsilon}\boldsymbol{T})^{-1}] \boldsymbol{\mu}) \\ \Leftrightarrow \quad & \frac{|\boldsymbol{I}_{nm} - 2i\Delta(\boldsymbol{A} \otimes \boldsymbol{T})|^{-\frac{1}{2}}}{|\boldsymbol{I}_m - 2i\boldsymbol{\Upsilon}\boldsymbol{T}|^{-\frac{f}{2}}} \\ = \quad & \frac{\exp(i\boldsymbol{\mu}^T [\boldsymbol{I}_n \otimes \boldsymbol{T}(\boldsymbol{I}_m - 2i\boldsymbol{\Upsilon}\boldsymbol{T})^{-1}] \boldsymbol{\mu})}{\exp(i\boldsymbol{\mu}^T (\boldsymbol{A} \otimes \boldsymbol{T})[\boldsymbol{I}_{nm} - 2i\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{T})]^{-1} \boldsymbol{\mu})} \end{aligned}$$

Using the common calculation rules for the exponential function we obtain

$$\begin{aligned}
&\frac{|\boldsymbol{I}_{nm} - 2i\Delta(\boldsymbol{A}\otimes\boldsymbol{T})|^{-\frac{1}{2}}}{|\boldsymbol{I}_m - 2i\boldsymbol{\Upsilon}\boldsymbol{T}|^{-\frac{f}{2}}} \\
= \;& \exp(i\boldsymbol{\mu}^T \left\{[\boldsymbol{I}_n \otimes \boldsymbol{T}(\boldsymbol{I}_m - 2i\boldsymbol{\Upsilon}\boldsymbol{T})^{-1}] - (\boldsymbol{A}\otimes\boldsymbol{T})[\boldsymbol{I}_{nm} - 2i\boldsymbol{\Delta}(\boldsymbol{A}\otimes\boldsymbol{T})]^{-1}\right\}\boldsymbol{\mu}). \qquad (4.8)
\end{aligned}$$

(4.8) has to be satisfied for all $\boldsymbol{T}$, in particular for $\boldsymbol{T} = t\boldsymbol{\Upsilon}^{-1}$, where $t \in \mathbb{R}$ is any real scalar. Inserting this into (4.8) yields:

$$\begin{aligned}
&\frac{|\boldsymbol{I}_{nm} - 2it\Delta(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})|^{-\frac{1}{2}}}{|\boldsymbol{I}_m - 2it\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{-1}|^{-\frac{f}{2}}} \\
= \;& \exp(it\boldsymbol{\mu}^T \left\{[\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}(\boldsymbol{I}_m - 2it\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{-1})^{-1}] - (\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})[\boldsymbol{I}_{nm} - 2it\boldsymbol{\Delta}(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})]^{-1}\right\}\boldsymbol{\mu}) \\
\Leftrightarrow \;& \\
&\frac{|\boldsymbol{I}_{nm} - 2it\Delta(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})|^{-\frac{1}{2}}}{|\boldsymbol{I}_m - 2it\boldsymbol{I}_m|^{-\frac{f}{2}}} \\
= \;& \exp(it\boldsymbol{\mu}^T \left\{[\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}(\boldsymbol{I}_m - 2it\boldsymbol{I}_m)^{-1}] - (\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})[\boldsymbol{I}_{nm} - 2it\boldsymbol{\Delta}(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})]^{-1}\right\}\boldsymbol{\mu}) \\
\Leftrightarrow \;& \\
&\frac{|\boldsymbol{I}_{nm} - 2it\Delta(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})|^{-\frac{1}{2}}}{|(1-2it)\boldsymbol{I}_n|^{-\frac{f}{2}}} \\
= \;& \exp(it\boldsymbol{\mu}^T \left\{[\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}((1-2it)\boldsymbol{I}_m)^{-1}] - (\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})[\boldsymbol{I}_{nm} - 2it\boldsymbol{\Delta}(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})]^{-1}\right\}\boldsymbol{\mu})
\end{aligned}$$

Now use that $|(1-2it)\boldsymbol{I}_m|^{-\frac{f}{2}} = ((1-2it)^m|\boldsymbol{I}_m|)^{-\frac{f}{2}}$. Furthermore, the determinant of $\boldsymbol{I}_m$ is equal to one. Thus

$$|(1-2it)\boldsymbol{I}_m|^{-\frac{f}{2}} = (1-2it)^{-\frac{fm}{2}}.$$

Hence,

$$\begin{aligned}
&\frac{|\boldsymbol{I}_{nm} - 2it\Delta(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})|^{-\frac{1}{2}}}{(1-2it)^{-\frac{fm}{2}}} \\
= \;& \exp(it\boldsymbol{\mu}^T \left\{[(1-2it)^{-1}\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}] - (\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})[\boldsymbol{I}_{nm} - 2it\boldsymbol{\Delta}(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})]^{-1}\right\}\boldsymbol{\mu}) \qquad (4.9)
\end{aligned}$$

Theorem A.12 shall be applied to (4.9). Therefore first note that the left-hand side of (4.9),

$$\frac{|\boldsymbol{I}_{dm} - 2it\Delta(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})|^{-\frac{1}{2}}}{(1-2it)^{-\frac{fd}{2}}}$$

is a ratio of polynomials of finite degree in t. Using the notation of Theorem A.12, $R(it) = |\boldsymbol{I}_{dm} - 2it\Delta(\boldsymbol{A}\otimes\boldsymbol{\Upsilon}^{-1})|^{-\frac{1}{2}}$ (the determinant is a polynomial in t) and $S(it) = (1-2it)^{-\frac{fd}{2}}$. Here, $t_1 = t_2 = t$. Furthermore,

$$\begin{aligned}
& it\boldsymbol{\mu}^T \left\{[(1-2it)^{-1}\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}] - (\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})[\boldsymbol{I}_{nm} - 2it\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})]^{-1}\right\} \boldsymbol{\mu} \\
= \quad & it(1-2it)^{-1}\boldsymbol{\mu}^T[\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}]\boldsymbol{\mu} - it\boldsymbol{\mu}^T(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})[\boldsymbol{I}_{nm} - 2it\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})]^{-1}\boldsymbol{\mu} \\
= \quad & (1-2it)^{-1}\left\{it(\boldsymbol{\mu}^T[\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}]\boldsymbol{\mu} - (1-2it)\boldsymbol{\mu}^T(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})[\boldsymbol{I}_{nm} - 2it\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})]^{-1}\boldsymbol{\mu})\right\}.
\end{aligned}$$

Set

$$P(it) = it(\boldsymbol{\mu}^T[\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}]\boldsymbol{\mu} - (1-2it)\boldsymbol{\mu}^T(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})[\boldsymbol{I}_{nm} - 2it\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})]^{-1}\boldsymbol{\mu})$$

and

$$Q(it) = 1 - 2it.$$

P and Q are both finite polynomials in t. (Note that again $t_1 = t_2 = t$.) Then,

$$it(1-2it)^{-1}\boldsymbol{\mu}^T[\boldsymbol{I}_n \otimes \boldsymbol{\Upsilon}^{-1}]\boldsymbol{\mu} - it\boldsymbol{\mu}^T(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})[\boldsymbol{I}_{nm} - 2it\boldsymbol{\Delta}(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})]^{-1}\boldsymbol{\mu} = \frac{P(it)}{Q(it)}.$$

Thus one can rewrite (4.9) in terms of P, Q, R and S:

$$\frac{R(it)}{S(it)} = \exp(\frac{P(it)}{Q(it)}). \tag{4.10}$$

Theorem A.12 can be applied and thus we obtain that equality (4.10) holds if and only if $\frac{R(it)}{S(it)} = 1$ and $\frac{P(it)}{Q(it)} = 0$. Thus the left-hand side of (4.9) is equated to 1:

$$\begin{aligned}
1 &= \frac{|\boldsymbol{I}_{nm} - 2it\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})|^{-\frac{1}{2}}}{(1-2it)^{-\frac{fm}{2}}} \\
\Leftrightarrow \quad (1-2it)^{-\frac{fm}{2}} &= |\boldsymbol{I}_{nm} - 2it\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})|^{-\frac{1}{2}} \\
\Rightarrow \quad (1-2it)^{fm} &= |\boldsymbol{I}_{nm} - 2it\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})| \\
\Rightarrow \quad (1-2it)^{fm}t^{-mn} &= |\frac{1}{t}\boldsymbol{I}_{nm} - 2i\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})| \\
\Rightarrow \quad (\frac{1}{t} - 2i)^{fm}(\frac{1}{t})^{m(n-f)} &= |\frac{1}{t}\boldsymbol{I}_{nm} - 2i\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})|
\end{aligned}$$

$|\frac{1}{t}\boldsymbol{I}_{nm} - 2i\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})|$ is the characteristic polynomial of $2i\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$. Its roots are the eigenvalues of $2i\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$. There are exactly fm eigenvalues equal to $2i$ and $m(n-f)$ eigenvalues equal to 0. But the eigenvalues of $\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$ are $(2i)^{-1}$ times the eigenvalues of $2i\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$. Thus $\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$ has exactly fm eigenvalues equal to one and $m(n-f)$ eigenvalues equal to zero. We know from Theorem A.10 that $\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})$ is idempotent. Moreover it follows from Theorem A.2 that $rank(\Delta(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})) = fm$, because fm eigenvalues are not equal to zero.

Concluding, the *rank* of $\boldsymbol{A}$ shall be derived by applying Theorem A.3. Note that $\boldsymbol{\Delta}$ is positive definite and its eigenvalues are all positive. Therefore $rank(\boldsymbol{\Delta}) = nm$ according to Theorem A.3. All in all we know that

$$\begin{aligned} & nm + rank(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1}) - nm \leq fm \leq \min(nm, rank(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})) \\ \Rightarrow \quad & rank((\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})) \leq fm \leq rank(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1}). \end{aligned}$$

Note that it was used that if $a \leq \min(b, c)$, then $a \leq b$ and $a \leq c$. It follows that $rank((\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1})) = fm$. Furthermore we know from (A.6) that

$$rank(\boldsymbol{A} \otimes \boldsymbol{\Upsilon}^{-1}) = rank(\boldsymbol{A}) rank(\boldsymbol{\Upsilon}^{-1}).$$

$\boldsymbol{\Upsilon}^{-1}$ is nonsingular, thus it has full rank m. It follows that $rank(\boldsymbol{A}) = f$.

□

Theorem 4.2 is now applied to find necessary and sufficient conditions for $\boldsymbol{S}_E^*$ and $\boldsymbol{S}_H^*$ to be Wishart:

Corollary 4.3. *The error matrix* $\boldsymbol{S}_E^*$ *defined in* (3.19)

$$\boldsymbol{S}_E^* = \boldsymbol{Y}^{*T} \left\{ [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T] \otimes \boldsymbol{M}^T\boldsymbol{M} \right\} \boldsymbol{Y}^*$$

is distributed as

$$\boldsymbol{S}_e^* \sim W_r((n-k)q, \boldsymbol{\Upsilon}, \boldsymbol{0}), \tag{4.11}$$

if and only if

$$\boldsymbol{\Omega} = (\boldsymbol{I}_q \otimes \boldsymbol{\Upsilon}). \tag{4.12}$$

Condition (4.12) is called *multivariate sphericity.*

Proof. We want to apply Theorem 4.2. Therefore, set $\boldsymbol{U} = \boldsymbol{Y}^*$,
$\boldsymbol{A} = \left\{ [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T] \otimes \boldsymbol{M}^T\boldsymbol{M} \right\}$, $n = nt$, $m = r$, $\boldsymbol{\Delta} = (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ and $\boldsymbol{M} = (\boldsymbol{X} \otimes \boldsymbol{I}_t)\boldsymbol{B}^*$. Then $\boldsymbol{S}_E^*$ has a Wishart distribution if and only if

$$(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) \left\{ [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T] \otimes \boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1} \right\}$$

is idempotent.

Apply Theorem A.5

$$(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) \left\{ [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T] \otimes \boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1} \right\} = \left\{ [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T] \otimes \boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] \right\}$$

We want to show that

$$\left\{ [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T] \otimes \boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] \right\}^2 = \left\{ [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T] \otimes \boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] \right\}$$

Note that by (A.3)

$$\left\{ [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T] \otimes \boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] \right\}^2 = \left\{ [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T]^2 \otimes (\boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}])^2 \right\}$$

$[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T]$ is idempotent. Thus, $\left\{[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T] \otimes \boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M}^T \otimes \boldsymbol{\Upsilon}^{-1}]\right\}$ is idempotent if and only if $\boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}]$ is idempotent. We show that

$$\boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] \cdot \boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] = \boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}]$$

is equivalent to multivariate sphericity.

$\boldsymbol{\Sigma} \in \mathcal{S}_n^+$ in nonsingular, hence the inverse $\boldsymbol{\Sigma}^{-1}$ does exist. It is multiplied from the left to both sides of the equations. This yields:

$$\begin{aligned} & \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] \cdot \boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] \\ \Leftrightarrow \quad & [\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] \cdot \boldsymbol{\Sigma}[\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] = [\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}] \end{aligned}$$

Now multiply with $(\boldsymbol{M} \otimes \boldsymbol{\Upsilon})$ and apply Theorem (A.3):

$$\begin{aligned} & (\boldsymbol{M} \otimes \boldsymbol{\Upsilon})(\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}) \cdot \ \boldsymbol{\Sigma}(\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}) = (\boldsymbol{M} \otimes \boldsymbol{\Upsilon})(\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}) \\ \Leftrightarrow \quad & (\boldsymbol{M}\boldsymbol{M}^T\boldsymbol{M}) \otimes (\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{-1}) \cdot \ \boldsymbol{\Sigma}(\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}) = (\boldsymbol{M}\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{-1}). \end{aligned}$$

Recall that $\boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{I}_q$. This yields:

$$(\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1}) = (\boldsymbol{M} \otimes \boldsymbol{I}_r)$$

$(\boldsymbol{M}^T \otimes \boldsymbol{\Upsilon})$ is multiplied from the right-hand side. The calculations are the same:

$$\begin{aligned} & (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T\boldsymbol{M} \otimes \boldsymbol{\Upsilon}^{-1})(\boldsymbol{M}^T \otimes \boldsymbol{\Upsilon}) = (\boldsymbol{M}^T \otimes \boldsymbol{I}_r)(\boldsymbol{M}^T \otimes \boldsymbol{\Upsilon}) \\ \Leftrightarrow \quad & (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{M}^T \otimes \boldsymbol{\Upsilon}^{-1}\boldsymbol{\Upsilon}) = (\boldsymbol{M}\boldsymbol{M}^T \otimes \boldsymbol{I}_r\boldsymbol{\Upsilon}) \\ \Leftrightarrow \quad & (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) = (\boldsymbol{I}_q \otimes \boldsymbol{\Upsilon}). \end{aligned}$$

But

$$(\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) = \boldsymbol{\Omega}$$

by definition.

Thus $\boldsymbol{S}_E^*$ is distributed as in the Corollary, if and only if

$$\boldsymbol{\Omega} = \boldsymbol{I}_q \otimes \boldsymbol{\Upsilon},$$

thus if and only if multivariate sphericity is fulfilled.

Finally, note that $\hat{\boldsymbol{M}}\boldsymbol{A} = \boldsymbol{0}$ and that $rank(\boldsymbol{A}) = rank(\boldsymbol{P} \otimes \boldsymbol{M}^T\boldsymbol{M}) = rank(\boldsymbol{P})rank(\boldsymbol{M}^T\boldsymbol{M}) = (n-k)q$. □

Corollary 4.4. *The SSCP matrix due to the regression $\boldsymbol{S}_H^*$ defined in* (3.20)*,*

$$\boldsymbol{S}_H^* = (\boldsymbol{Y}^*)^T \left\{\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T \otimes \boldsymbol{M}^T\boldsymbol{M}\right\} \boldsymbol{Y}^*.$$

is distributed as

$$\boldsymbol{S}_H^* \sim W_r(sq, \boldsymbol{\Upsilon}, \boldsymbol{\Xi}^*), \tag{4.13}$$

where

$$\boldsymbol{\Xi}^* = (\boldsymbol{B}^*)^T \left\{\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C} \otimes \boldsymbol{M}^T\boldsymbol{M}\right\} \boldsymbol{B}^*\boldsymbol{\Upsilon}^{-1}. \tag{4.14}$$

if and only if multivariate sphericity is satisfied.

Proof. Let $\boldsymbol{A} = \left\{\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T]^{-1}\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X} \otimes \boldsymbol{M}^T\boldsymbol{M}\right\}$. Define all other terms as in the proof of Corollary 4.3 and follow the proof of Corollary 4.3. □

Remark 4.1.1. Multivariate sphericity is a condition for the structure of $\boldsymbol{\Omega}$. Recall that ***Omega*** is the covariance matrix of $\boldsymbol{Y}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)$, and thus of the restricted data matrix, where we consider q functions of the points in time instead of t points in time. The q $r \times r$ diagonal blocks of $\boldsymbol{\Omega}$ correspond to the variation of the combinations of time levels. Multivariate sphericity implies that these blocks are all equal and that the q subvectors of each row of $\boldsymbol{Y}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)$ are independent. Thus multivariate sphericity is a condition concerning the variation of the differences between time levels.

The Corollaries are used to derive the distribution of the likelihood ratio test statistic. As we have seen in Theorem 2.5, the test statistic $(\lambda^*)^{\frac{2}{nt}} = (\frac{|\boldsymbol{S}_E^*|}{|\boldsymbol{S}_E^*+\boldsymbol{S}_H^*|})^{\frac{2}{nt}}$ is distributed as Wilks' $\boldsymbol{\lambda}$. However, it is common use to the following F-approximation for the distribution (see Timm [26]):

$$F = ((\lambda^*)^{-a} - 1)(\frac{\nu 1}{\nu 2}) \;\dot{\sim} \boldsymbol{F}(\nu_1, \nu_2),$$

where

$$\begin{aligned}
d &= r, \\
m_1 &= sq, \\
m_2 &= (n-k)q, \\
a &= \left(\frac{d^2 + m_1^2 - 5}{(m_1 d)^2 - 4}\right)^{\frac{1}{2}}, \\
\nu_1 &= m_1 d, \\
\nu_2 &= a^{-1}[m_2 - \frac{1}{2}(d - m_1 + 1)] - \frac{1}{2}(m_1 d - 2).
\end{aligned}$$

Multivariate sphericity is a very strict condition which is often not fulfilled in repeated measurements experiments. Thus it is important to have a test to see whether the condition is satisfied or not.

4.2. A Likelihood Ratio Test for Multivariate Sphericity

To check, whether multivariate sphericity is satisfied, Boik suggests a test that was first introduced by Thomas in [27]. Details of the derivation can be found in [1]. We want to test the null hypothesis

$$H_0 : (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) = (\boldsymbol{I}_q \otimes \boldsymbol{\Upsilon}) \tag{4.15}$$

against the alternative

$$H_1 : (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) \text{ is positive definite.} \tag{4.16}$$

The likelihood ratio test statistic for testing H_0 was given by Thomas as

$$\lambda = \frac{|\boldsymbol{S}_E|^{\frac{n}{2}}}{|\frac{1}{q}\sum_{i=1}^{q}(\boldsymbol{S}_E)_{ii}|^{\frac{nq}{2}}}. \tag{4.17}$$

It can be shown that

$$-2\ln(\lambda) \dot{\sim} \chi^2_f \tag{4.18}$$

under the null hypothesis, i.e. that under the null its asymptotic distribution is that of a chi-squared random variable. The degrees of freedom f are given by

$$f = \frac{r(q-1)(rq+r+1)}{2}. \tag{4.19}$$

However, Boik notes that the asymptotic distribution may be a poor approximation, in particular if r or q are large in comparison to n. On this account he gives a more accurate approximation.

Theorem 4.5 (Boik, [1], Theorem 3)**.** *Let* $W := -(\frac{2\rho\nu}{n})\ln(\lambda)$*, where* λ *is defined in* (4.17) *and* $\nu = (n-k)$*. Under the null hypothesis* H_0 *defined in* (4.15)*, the following holds:*

- *If* $q = 1$ *and* $\rho \neq 0$*, then* $P(W = 0) = 1$.
- *If* $q \geq 2$ *and*

$$\rho = 1 - (\frac{r}{12qf\nu})[2r^2(q^4-1) + 3r(q^3-1) - (q^2-1)],$$

then

$$P(W \leq w) = P(G_f \leq w) + \omega[P(G_{f+4} \leq w) - P(G_f \leq w)] + \mathcal{O}(\nu^{-3}),$$

where $G_f \sim \chi^2_f$*,* f *as defined in* (4.19) *and*

$$\omega = (2\rho^2)^{-1}\left\{\left[\frac{(rq-1)rq(rq+1)(rq+2)}{24\nu}\right] - \left[\frac{(r-1)r(r+1)(r+2)}{24q^2\nu}\right] - \left[\frac{f(r-1)}{2}\right]\right\}.$$

Remark 4.2.1 (Boik, [1])**.** If the term that involves ω is omitted, then the error in the approximation is $\mathcal{O}(\nu^{-2})$ instead of $\mathcal{O}(\nu^{-3})$.

4.3. An Adjusted Test when Multivariate Sphericity is not Satisfied

When Multivariate Sphericity is not satisfied, a multivariate mixed model analysis can still be performed. We do no longer know the exact distribution of $\boldsymbol{S}^*_E$ and $\boldsymbol{S}^*_H$, therefore we need to find an approximation. As mentioned in Section 3.1.2, their asymptotic distribution should be Wishart. Boik introduces a correction factor in order to find degrees of freedom f_1 and f_2, covariance matrix $\boldsymbol{\Upsilon}$ and non-centrality parameter $\boldsymbol{\Xi}^*$, such that

$$\boldsymbol{S}_H^* \dot{\sim} W_r(f_1, \boldsymbol{\Upsilon}, \boldsymbol{\Xi}^*)$$

and

$$\boldsymbol{S}_E^* \dot{\sim} W_r(f_2, \boldsymbol{\Upsilon}).$$

Boik uses the moments of Wishart distributed random variables to proof the following theorem:

Theorem 4.6 (Boik, [1], Theorem 5)**.** *Let* $vec((\zeta)^T) \sim \mathcal{N}_{nrt}(\boldsymbol{0}, (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}))$ *and recall that* $\boldsymbol{\Omega} = (\boldsymbol{M} \otimes \boldsymbol{I}_r)\boldsymbol{\Sigma}(\boldsymbol{M}^T \otimes \boldsymbol{I}_r)$*.* $\boldsymbol{S}_E^*$ *and* $\boldsymbol{S}_H^*$ *are asymptoticly distributed as*

$$\boldsymbol{S}_E^* \dot{\sim} W_r(\epsilon q(n-k), \boldsymbol{\Upsilon})$$

and

$$\boldsymbol{S}_H^* \dot{\sim} W_r(\epsilon q(s + 2q\epsilon\delta), \boldsymbol{\Upsilon}, \boldsymbol{\Xi}^*),$$

where

$$\epsilon = \frac{tr\left[(\sum_{i=1}^q \boldsymbol{\Omega}_{ii})^2\right] + \left[tr\left(\sum_{i=1}^q \boldsymbol{\Omega}_{ii}\right)\right]^2}{q\left\{\sum_{i=1}^q \sum_{j=1}[(tr(\boldsymbol{\Omega}_{ij})^2 + tr(\boldsymbol{\Omega}_{ij}^2)]\right\}},$$

$$\delta = \frac{\delta_1}{\delta_2},$$

$$\delta_1 = \frac{tr(\sum_{i=1}^q \boldsymbol{\Omega}_{ii})tr(\sum_{i=1}^q \boldsymbol{\Psi}_{ii}) + tr\left[(\sum_{i=1}^q \boldsymbol{\Omega}_{ii})(\sum_{i=1}^q \boldsymbol{\Psi}_{ii})\right]}{q\epsilon} - \sum_{i=1}^q \sum_{j=1}^q \left[tr(\boldsymbol{\Omega}_{ij})tr(\boldsymbol{\Psi}_{ij}) + tr(\boldsymbol{\Omega}_{ij}\boldsymbol{\Psi}_{ij})\right],$$

$$\delta_2 = tr\left((\sum_{i=1}^q \boldsymbol{\Omega}_{ii})^2\right) + \left[tr\left(\sum_{i=1}^q \boldsymbol{\Omega}_{ii}\right)\right]^2,$$

$$\boldsymbol{\Upsilon} = \frac{1}{q\epsilon}\sum_{i=1}^q \boldsymbol{\Omega}_{ii}$$

and

$$\boldsymbol{\Xi}^* = q\epsilon\left(\sum_{i=1}^q \boldsymbol{\Omega}_{ii}\right)^{-1}\sum_{i=1}^q \boldsymbol{\Psi}_{ii} - 2(q\epsilon)^2\delta\boldsymbol{I}_r.$$

A proof of the theorem is given in the Appendix of [1].

Note that under the null hypothesis $\boldsymbol{\Psi} = \boldsymbol{0}$ and thus $\delta = 0$. This yields the following corollary:

Corollary 4.7 (Boik, [1], Corollary 5.1)**.** *If the null hypothesis is true, then*

$$\boldsymbol{S}_E^* \dot{\sim} W_r(\epsilon q(n-k), \boldsymbol{\Upsilon})$$

and

$$\boldsymbol{S}_H^* \dot{\sim} W_r(\epsilon qs, \boldsymbol{\Upsilon}, \boldsymbol{0}).$$

If multivariate sphericity is satisfied, $\epsilon = 1$ and $\delta = 0$, as $\mathbf{\Omega}_{ii} = \mathbf{\Upsilon}$ and $\mathbf{\Omega}_{ij} = \mathbf{0}$ for $i \neq j$. Then the distributions are exact and as defined Corollary (4.3) and Corollary (4.4). Again, we can use the F-approximation for the distribution of $\frac{|\boldsymbol{S}_E|}{|\boldsymbol{S}_E + \boldsymbol{S}_H|}$.

ϵ is unknown and has to be estimated from the data. The unknown quantity in the definition of ϵ is $\mathbf{\Omega}$. For our purpose it suffices to replace $\mathbf{\Omega}$ by its unbiased estimate $\hat{\mathbf{\Omega}} = \frac{\boldsymbol{S}_E}{n-k}$.

5. Implementation

To illustrate multivariate sphericity, the multivariate mixed model analysis and the $\hat{\epsilon}$-adjusted multivariate mixed model analysis, the tests introduced in the previous two chapters shall now be applied to data sets. First, we are going to analyze a data set that Timm introduced in [25]. In the second section of this chapter we shall consider simulated data and contemplate the differences of a multivariate mixed model analysis and an $\hat{\epsilon}$-adjusted multivariate mixed model analysis, if multivariate sphericity is not satisfied.

5.1. Dental Data

Timm used the following data set containing vertical positions of the mandible in order to illustrate a multivariate analysis in experiments with repeated measurements (see [25]). The data were also analyzed by Boik in [1] and by Chaganty and Naik in [4] and were provided by Dr. Tom Zullo in the School of Dental Medicine at the University at Pittsburgh.

In the experiment there were two orthopedic adjustments of the mandible called activator treatments. The study was concerned in the effectiveness of the treatments. Nine subjects were assigned to each of the two treatment groups. Thus the total number of subjects $n = 18$.

Three response variables were analyzed, namely SOr-Me (in mm), ANS-Me (in mm) and Pal-MP angle (in degrees). These characteristics were measured in order to appraise the change of the vertical position of the mandible over time. Therefore the measurements were taken at three points in time.

Summarized we get:

Design matrix X	$\boldsymbol{I}_2 \otimes \boldsymbol{1}_9$
Total number of subjects n	18
Number of response variables r	3
Number of treatments k ($= rank(\boldsymbol{X})$)	2
Number of points in time t	3

The data set and the means of the measurements of the response variables at each of the three points in time are given in the table 5.1. The observations are arranged as described in Section 3.2.1.

We are mainly interested in three hypotheses. We want to test if there is a difference between the treatment groups, if there is a difference among the conditions (time) and if there is a interaction between time and treatment. All three hypotheses can be put in the form

$$H_0 : \boldsymbol{CBM} = \boldsymbol{0}.$$

As mentioned above, $\boldsymbol{C} \in \mathcal{M}_{s,r}$ induces s linear functions of the treatments and $(\boldsymbol{M}^T \otimes \boldsymbol{I}_r) \in \mathcal{M}_{tr,qr}$ generates qr linear combinations of the t occasions. Recall that we chose $\boldsymbol{M}$ to satisfy $\boldsymbol{MM}^T = \boldsymbol{I}_q$.

Group	t1			t2			t3		
	SOr-Me (mm)	ANS-Me (mm)	Pal-MP (dgs)	SOr-Me (mm)	ANS-Me (mm)	Pal-MP (dgs)	SOr-Me (mm)	ANS-Me (mm)	Pal-MP (dgs)
T1	117.0	59.0	10.5	117.5	59.0	16.5	118.5	60.0	16.5
T1	109.0	60.0	30.5	110.5	61.5	30.5	111.0	61.5	30.5
T1	117.0	60.0	23.5	120.0	61.5	23.5	120.5	62.0	23.5
T1	122.0	67.5	33.0	126.0	70.5	32.0	127.0	71.5	32.5
T1	116.0	61.5	24.5	118.5	62.5	24.5	119.5	63.5	24.5
T1	123.0	65.6	22.0	126.0	61.5	22.0	127.0	67.5	22.0
T1	130.5	68.5	33.0	132.5	69.5	32.5	134.5	71.0	32.0
T1	126.5	69.0	20.0	128.5	71.0	20.0	130.5	73.0	20.0
T1	113.0	58.0	25.0	116.5	59.0	25.0	118.0	60.5	24.5
Means T1	119.3	63.22	24.76	121.72	64.0	25.17	122.94	65.61	25.11
T2	128.0	67.0	24.0	129.0	67.5	24.0	131.5	69.0	24.0
T2	116.5	63.5	28.5	120.0	65.0	29.5	121.5	66.0	29.5
T2	121.5	64.5	26.5	125.5	67.5	27.0	127.0	69.0	27.0
T2	109.5	54.0	18.0	112.0	55.5	18.5	114.0	57.0	19.0
T2	133.0	72.0	34.5	136.0	73.5	34.5	137.5	75.5	34.5
T2	120.0	62.5	26.0	124.5	65.0	26.0	126.0	66.0	26.0
T2	129.5	65.0	18.5	133.5	68.0	18.5	134.5	69.0	18.5
T2	122.0	64.5	18.5	124.0	65.5	18.5	125.5	66.0	18.5
T2	125.0	65.5	21.5	127.0	66.5	21.5	128.0	67.0	21.6
Means T2	122.78	64.28	24.00	125.72	66.0	24.22	127.28	67.17	24.29

Table 5.1.: Individual Measurements on the Vertical Position of the Mandible at three Points in Time

The matrices for the three different tests are for instance given by Timm in [25]. They are

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{M} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \qquad \text{for the treatment test (note: } q = 1, s = 1\text{);}$$

$$\boldsymbol{C} = \begin{pmatrix} 1 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{M} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{pmatrix} \qquad \text{for the time test } (q = 2, s = 1);$$

and

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{M} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{pmatrix} \qquad \text{for the time} \times \text{treatment test } (q = 2, s = 1).$$

First, we want to make sure that it is reasonable to perform a multivariate analysis instead of three univariate analyses. Therefore we want to know if the three response variables are correlated. This assumption seems reasonable because the three characteristics are all measurements to determine the position of the mandible. Therefore we look at the scatter plot.

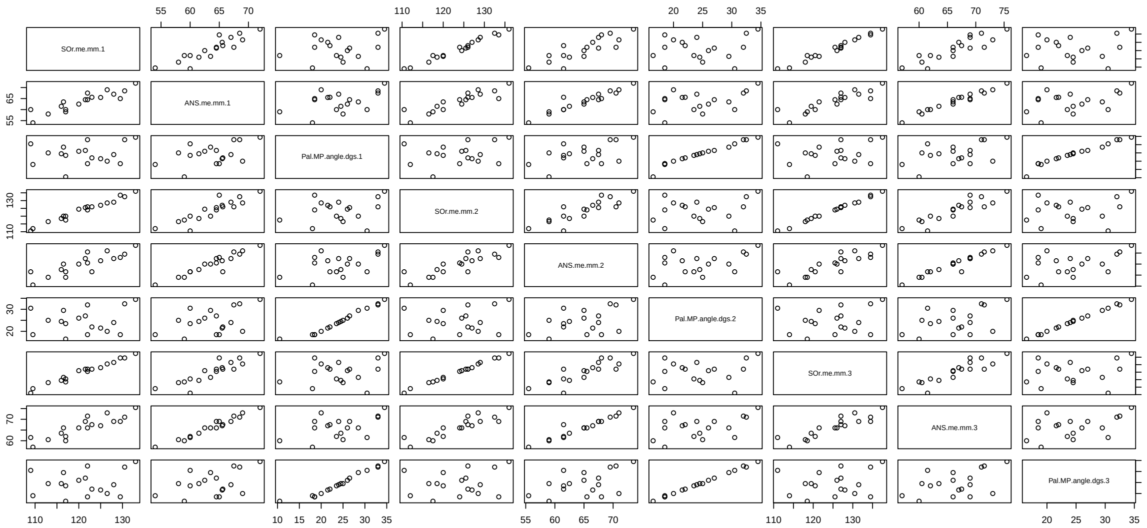

Figure 5.1.: Scatterplot

One can see that there seems to be a linear dependency of the measurements of ANS-Me at the three points in time. The same seems to hold for the measurements of SOr-Me. This is reasonable because in both cases it is the same response variable measured at different points in time, so the values should be linearly dependent.

There also seems to be a linear relation between the values of ANS-Me and SOr-Me. The scatter plot for the 18×9 data matrix is very impractical because it is hard to see anything. We are mainly interested in a correlation of the three orthopedic characteristics. Therefore we now have a look at the scatter plot for the rearranged data matrix $\boldsymbol{Y}^*$, as defined in 3.1.2.

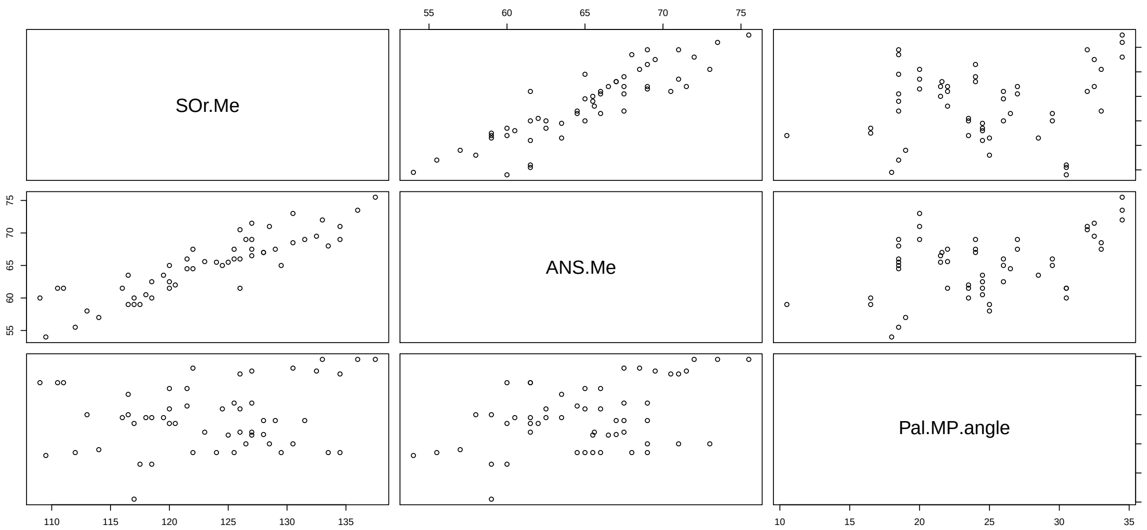

Figure 5.2.: Scatter plot for the Rearranged Data

There seems to be a linear relation between ANS-Me and SOr-Me. However, Pal-MP angle seems to be only little correlated to the other two response variables. Maybe it should be analyzed separately. When we have a look at the values in Table 5.1 it seems doubtful that there is much of a difference between times or treatments at all. Here we shall include Pal-MP angle in the analysis, as Boik and Timm have.

To get a first impression, we consider Figures 5.3 - 5.5, showing the means of the treatments at each point in time for the three responses:

There seems to be no time×treatment interaction effect for any response. However, there could be differences in treatments and time. As already suggested before, there seems to be hardly any effect on the response PAL-MP angle.

To see whether the rows of the data matrix follow a multivariate normal distribution, we create a qq plot of the squared Mahalanobis distance between the rows and their mean. It is displayed in Figure 5.6. If the points on the plot form a straight line, then multivariate normality can be assumed. For a detailed consideration of the detection of normality using qq plots of the squared Mahalanobis distance, see [3]. In our case, the points do not form a straight line. However, the two points at the edge of the plot could be outliers. Still, there are departures from the line, so multivariate normality might not be satisfied. Nevertheless we are going to analyze the data, in order to compare the results with Boiks analysis.

Next, we are interested, in whether multivariate sphericity is satisfied or not. Thus we have to test the hypothesis:

$$H_0 : (\boldsymbol{M} \otimes \boldsymbol{I}_3)\Sigma(\boldsymbol{M}^T \otimes \boldsymbol{I}_3) = (\boldsymbol{I}_q \otimes \boldsymbol{\Upsilon}).$$

First, we calculate $\boldsymbol{S}_E$ as defined in (3.13) in order to estimate $\boldsymbol{\Omega} = (\boldsymbol{M} \otimes \boldsymbol{I}_3)\Sigma(\boldsymbol{M}^T \otimes \boldsymbol{I}_3)$. Recall that $\hat{\boldsymbol{\Omega}} = \frac{\boldsymbol{S}_E}{n-k}$. For time and time×treatment interaction tests, $\boldsymbol{M}$ is the same. We get:

```
 Se
          [,1]      [,2]        [,3]      [,4]       [,5]        [,6]
[1,]  9.694444  7.286111  -6.7972222 -4.538615 -0.6623491   3.7255130
[2,]  7.286111  8.932222  -4.4361111 -3.202690 -3.5179876   2.9108076
[3,] -6.797222 -4.436111  18.6155556  2.994203  0.8708367 -10.1363462
[4,] -4.538615 -3.202690   2.9942026  4.824074  1.4694444  -1.9472222
[5,] -0.662349 -3.517988   0.8708367  1.469444 18.6774074  -0.7675926
[6,]  3.725513  2.910808 -10.1363462 -1.947222 -0.7675926   6.1274074
```

and

```
Omega.hat
            [,1]       [,2]        [,3]        [,4]        [,5]        [,6]
[1,]  0.60590278  0.4553819 -0.42482639 -0.28366341 -0.04139682  0.23284456
[2,]  0.45538194  0.5582639 -0.27725694 -0.20016814 -0.21987423  0.18192548
[3,] -0.42482639 -0.2772569  1.16347222  0.18713767  0.05442729 -0.63352164
[4,] -0.28366341 -0.2001681  0.18713767  0.30150463  0.09184028 -0.12170139
[5,] -0.04139682 -0.2198742  0.05442729  0.09184028  1.16733796 -0.04797454
[6,]  0.23284456  0.1819255 -0.63352164 -0.12170139 -0.04797454  0.38296296
```

The values of the constants given in Theorem (4.5) are calculated. We get $f = 15$, $\omega = 0.03426486$ and $\rho = 0.828125$. The corrected test statistic $-2\frac{\rho\nu}{n}\ln(\lambda)$ is 55.73919. We use Theorem 4.5 to calculate the p-value. It is approximately 1.9×10^{-6}. Hence the null hypothesis is rejected at any reasonable significance level α. We reject the hypothesis that multivariate sphericity is satisfied.

Nevertheless we want to use both, multivariate mixed model analysis and $\hat{\epsilon}$-adjusted multivariate mixed model analysis to test for differences in time and for time×treatment interaction and compare the results.

In the tables 5.2 and 5.3 below the values for the likelihood ratio test statistic $\lambda^* = \frac{|\boldsymbol{S}_E^*|}{|\boldsymbol{S}_E^* + \boldsymbol{S}_H^*|}$, corrected test statistic, degrees of freedom and p-values are given for each of the three hypotheses.

Source	λ^*	F	ν_1	ν_2	p-value
Treatments	0.885793	0.60168	3	14	0.62
Time	0.056966	31.89778	6	60	1.1×10^{-16}
Time×Treatment	0.82448	1.013109	6	60	0.43

Table 5.2.: Multivariate Mixed Model Analysis of the Dental Data

So only the null hypothesis for the difference in time is rejected at a significance level $\alpha = 0.05$ or $\alpha = 0.1$. The other two null hypotheses cannot be rejected. Thus, the two activator treatments did, on average, not differ. There was also no time×treatment interaction, thus the two treatments did not behave differently in time. But we can conclude that both treatments did have an effect over time, although there was no distinguishable difference between the effects.

Source	λ^*	F	ν_1	ν_2	p-value
Treatments	0.885793	0.60168	3	14	0.62
Time	0.056966	39.48886	4.4	33.9	1.1×10^{-12}
Time×Treatment	0.82448	1.000334	4.4	33.9	0.426

Table 5.3.: $\hat{\epsilon}$-Adjusted Multivariate Mixed Model Analysis of the Dental Data

For the treatment test, $q = 1$, and thus multivariate sphericity is evidently satisfied. Then $\hat{\epsilon} = 1$ and the analysis is the same as the multivariate mixed model analysis. For the other two tests, $q = 2$ and $\hat{\epsilon} = 0.733612$. The results are qualitatively the same as in the multivariate mixed model analysis.

So in this case both analyses yielded qualitatively the same results, even though the hypothesis that multivariate sphericity is satisfied was clearly rejected for time and time×treatment interaction tests. The result of our analysis are also similar to the ones obtained by Boik. We want to get a better idea about the differences between the analyses and are therefore analyzing simulated data next.

5.2. Simulated Data

For the simulated data, we assume the same design as in the dental data example. Thus, we have two treatments, three response variables and 3 points in time. We also generate a balanced data set, as in the example, thus we have the same number of observations for each of the two groups. We generate 10000 simulation models of multivariate normally distributed data for 18 and for 100 subjects and perform both, a multivariate mixed model analysis and an $\hat{\epsilon}$-adjusted multivariate mixed model analysis. Then we compare the results.

We concentrate on the test for difference in time. Thus,

$$H_0 : (\boldsymbol{C} \otimes \boldsymbol{M})\boldsymbol{B}^* = \boldsymbol{0},$$

where

$$\boldsymbol{C} = \begin{pmatrix} 1 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{M} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{pmatrix}.$$

The matrix $\boldsymbol{\Omega}$ is chosen in a way that it does not satisfy multivariate sphericity. However we chose it such that the test for multivariate sphericity does not reject the null hypothesis at a significance level $\alpha = 0.05$ in most of the cases. The matrix was constructed from the estimated matrix $\hat{\boldsymbol{\Omega}}$ from the dental data set. We averaged the two 3×3 blocks and changes two entries in the first block. Furthermore the elements that are not in the diagonal blocks are devided by two.

```
Omega = Se/16

Omega = Omega/2

Upsilon = Omega[1:3,1:3] + Omega[4:6,4:6]

Omega[1:3,1:3] = Upsilon
Omega[4:6,4:6] = Upsilon

Omega[2,3] = 0.8*Omega[2,3]
Omega[3,2] = Omega[2,3]

Omega
            [,1]        [,2]        [,3]        [,4]        [,5]        [,6]
[1,]  0.45370370  0.27361111 -0.27326389 -0.14183171 -0.02069841  0.11642228
[2,]  0.27361111  0.86280093 -0.13009259 -0.10008407 -0.10993711  0.09096274
[3,] -0.27326389 -0.13009259  0.77321759  0.09356883  0.02721365 -0.31676082
[4,] -0.14183171 -0.10008407  0.09356883  0.45370370  0.27361111 -0.27326389
[5,] -0.02069841 -0.10993711  0.02721365  0.27361111  0.86280093 -0.16261574
[6,]  0.11642228  0.09096274 -0.31676082 -0.27326389 -0.16261574  0.77321759
```

Furthermore we chose $\boldsymbol{B}^* = \boldsymbol{0}$ and thus there are no time effects. Hence the null hypothesis

$$(\boldsymbol{C} \otimes \boldsymbol{M})\boldsymbol{B}^* = \boldsymbol{0}$$

should not be rejected.

10000 simulations were run and both analyses were done on each of the simulated observation matrices. Thus the results are comparable. The numbers of cases were the null hypothesis was rejected are given in the following table.

number of observations	Multivariate Mixed Model	$\hat{\epsilon}$-Adjusted Analysis
18	582 (= 5.82%)	491 (= 4.91%)
100	601 (= 6.01%)	540 (= 5.4%)

Table 5.4.: Results of the Multivariate Mixed Model Analysis and $\hat{\epsilon}$-Adjusted Analysis

For both, $n = 18$ and $n = 100$, the null hypothesis was rejected by the multivariate mixed model analysis in each of the cases where it was also rejected by the $\hat{\epsilon}$-adjusted analysis. Thus the multivariate mixed model analysis made the same errors as the $\hat{\epsilon}$-adjusted analysis and some more errors in addition. Note that while the $\hat{\epsilon}$-adjusted analysis performs slightly better, the error rates of both analyses are close to the significance level $\alpha = 0.05$.

We are also interested in the differences between the p-values. We want to know if the results are qualitatively the same for both analyses. To gain an impression, we compare the resulsts of the first 10 simulations.

```
n = 18:

pvaluesMMM
0.975 0.826 0.766 0.149 0.880 0.793 0.517 0.098 0.397 0.703
pvaluesad
0.958 0.813 0.751 0.151 0.866 0.744 0.508 0.107 0.399 0.685


n = 100:

pvaluesMMM
0.545 0.832 0.244 0.936 0.966 0.151 0.157 0.724 0.674 0.002

pvaluesad
0.535 0.829 0.247 0.926 0.957 0.157 0.162 0.702 0.658 0.002
```

Indeed it looks like the results are qualitatively the same. We take the averages of the differences of all 10000 p-values to get a better picture. For $n = 18$ it is 0.0035 and for $n = 100$ it is 0.0044. This difference is very small.

Finally note that multivariate sphericity was rejected in only 14.92% of the cases for $n = 18$, but in 96.7% of the cases for $n = 100$. So while the results are qualitatively the same for the time difference test for $n = 18$ and $n = 100$, there is a huge difference in the results of the tests for sphericity.

The simulation was also run for $n = 18$ and $n = 100$ with the $\mathbf{\Omega}$ we estimated for the dental data. Recall that multivariate sphericity was clearly rejected.

number of observations	Multivariate Mixed Model	$\hat{\epsilon}$-Adjusted Analysis
18	924 (= 9.24%)	609 (= 6.09%)
100	829 (= 8.29%)	604 (= 6.04%)

Table 5.5.: Results of the Multivariate Mixed Model Analysis and $\hat{\epsilon}$-Adjusted Analysis 2

For both, $n = 18$ and $n = 100$, multivariate sphericity was rejected in 100% of the cases. It seems that the size of the adjusted test is better than the one of the multivariate mixed model analysis if multivariate sphericity is clearly not satisfied. The differences between the p-values of the two analyses do also slightly increase. For $n = 18$, the mean of the differences is 0.011, for $n = 100$ it is 0.019. The differences are still pretty small, but may be crucial in critical cases. Note that for both $\mathbf{\Omega}$, the difference was larger for $n = 100$. This could be due to the fact, that the lack of multivariate sphericity is more apparent for a larger number of observations.

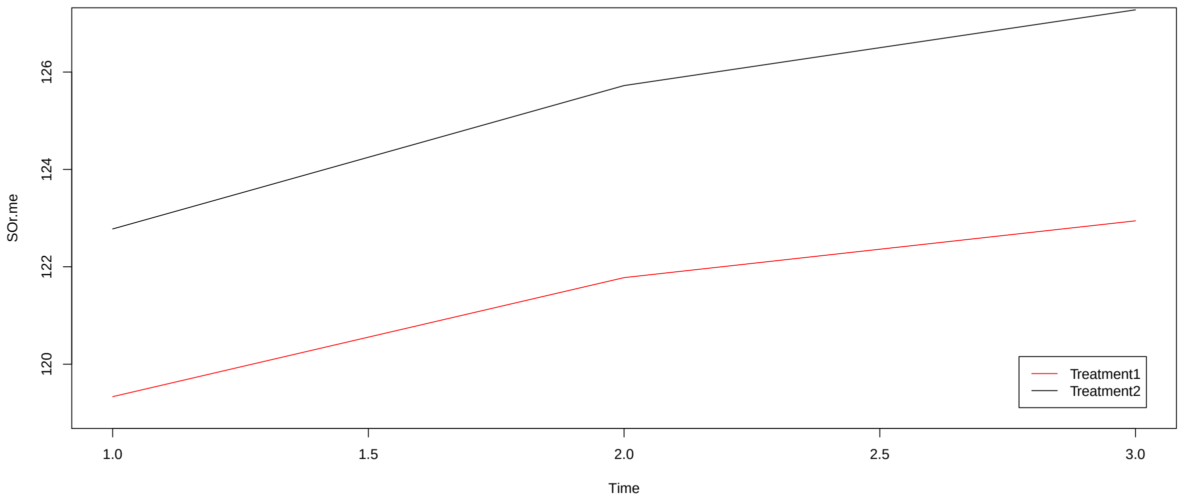

Figure 5.3.: Plot for SOr-me

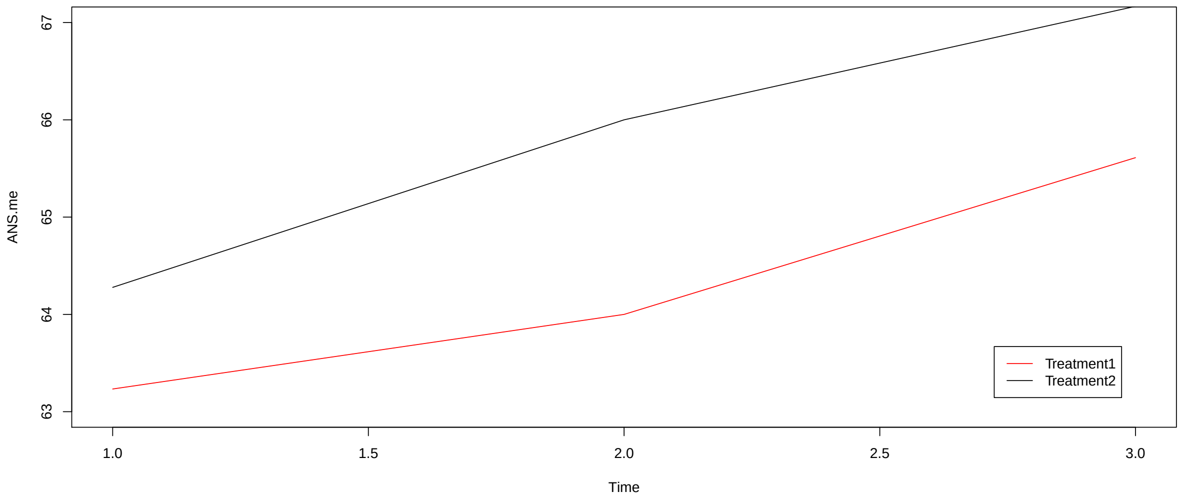

Figure 5.4.: Plot for ANS-me

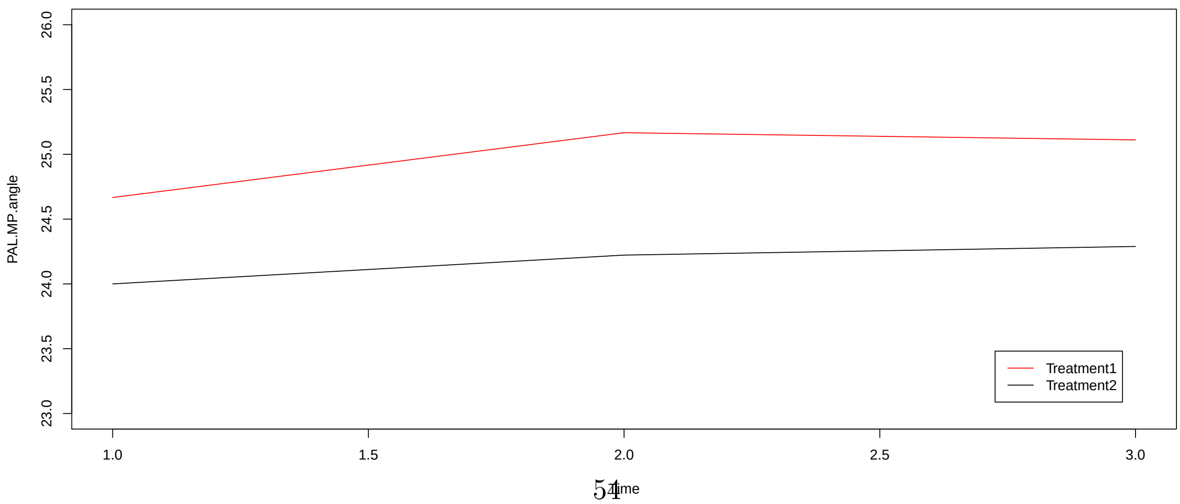

Figure 5.5.: Plot for PAL-MP angle

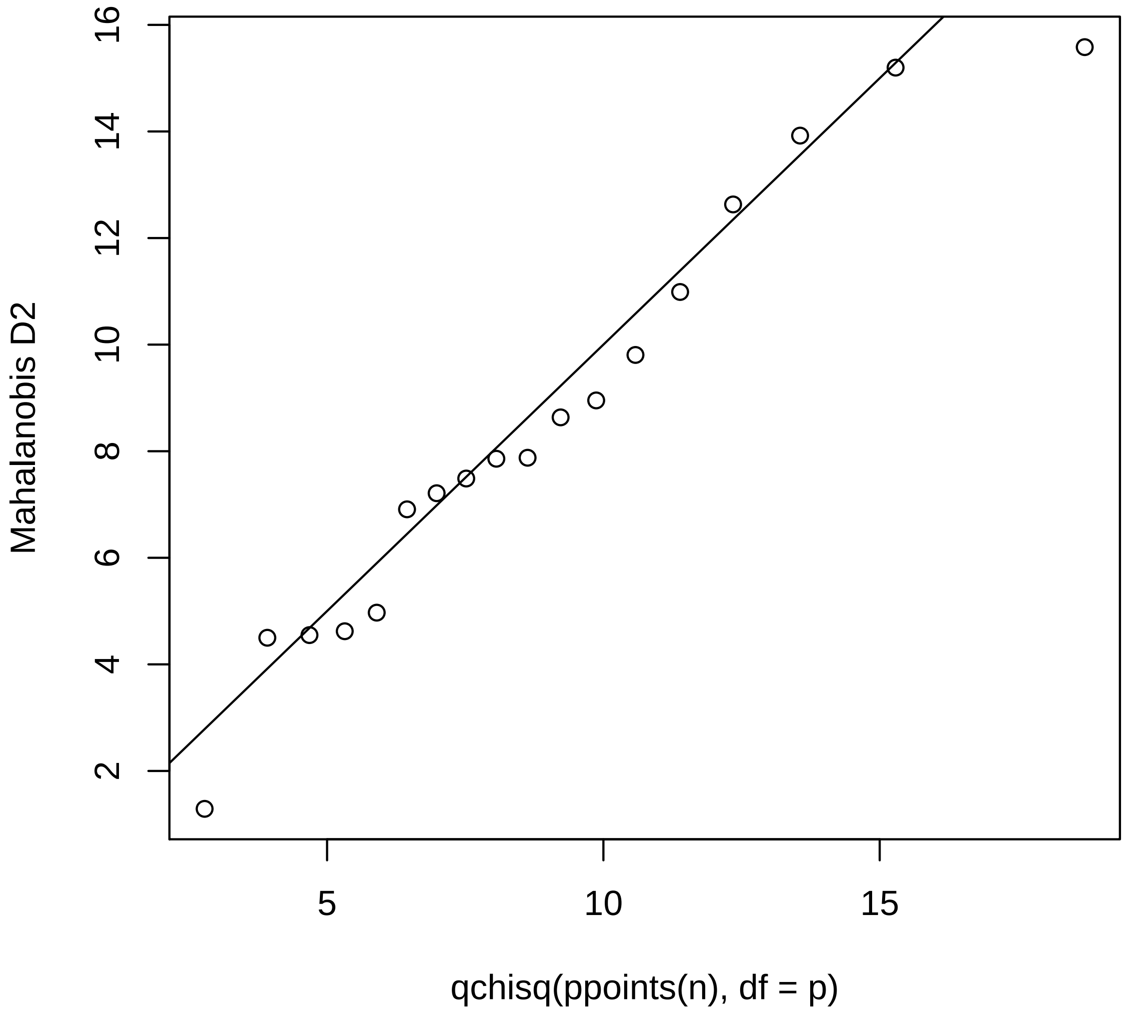

Figure 5.6.: QQ plot assesing multivariate normality

6. Concluding Remarks

Data of repeated measurements experiments are often analyzed with multivariate mixed models, even though a rather strict condition, namely multivariate sphericity, is required for its validity. We have seen an $\hat{\epsilon}$-adjusted analysis that can be performed if multivariate sphericity is not satisfied. Both analyses seem to yield qualitatively similar results, even if multivariate sphericity is not satisfied. However, the multivariate mixed model analysis test rejected a true null hypothesis more often than the adjusted test. The difference seems to increase with an increasing departure from multivariate sphericity. In both of the cases we considered, the size of the ϵ-adjusted test was close to the chosen significance level, but this was not true for the multivariate mixed model analysis. When we used an $\boldsymbol{\Omega}$ that was clearly not satisfying multivariate sphericity, the size was way above the chosen significance level. If it is not clear, whether multivariate sphericity is satisfied or not, the adjusted test should be chosen, even if the calculations are more complicated. Note that in our simulation it was harder to detect that multivariate sphericity is not satisfied for a small number of subjects. But in repeated measurements experiments, the number of experimental units is usually small. One should be especially careful in such cases.

It would be interesting to perform power studies and compare the power of the multivariate mixed model analysis to the power of other multivariate analyses if multivariate sphericity is satisfied. Boik analyzes the Timm data by a doubly multivariate model and compares the results to the ones of the mixed model analysis and the ϵ-adjusted analysis. He also gives a rough index of the relative power of doubly multivariate mixed model analysis and the ϵ-adjusted mixed model analysis. It can be taken into account when one wants to choose between a the analyses (see [1]). According to Boik, neither test is uniformly most powerfull. As the power index suggested by Boik is also often unknown, it is hard to chose the right model for a testing problem. If there are any doubts and if it is possible, then it would be advisable to performe an (ϵ-adjusted) mixed model analysis and a doubly multivariate analysis and compare the results. So the analysis of multivariate repeated measures is rather complicated.

Note that we have only considered balanced models. But unbalanced models do very often occur in experiments with repeated measures. Fieuws and Verbeke consider unbalanced bivariate models for longitudinal studies in [8]. It would interesting to know whether or not there is a validity condition for a multivariate mixed model analysis in the unbalanced case.

A. Appendix

A.1. Matrix Algebra

The following is a collection of results of matrix algebra that are used in several proofs and remarks throughout the thesis and are therefore given here. For a detailed consideration of the concepts as well as for all proofs of the given theorems the reader is recommended to result for example [18] or [26].

Trace and Rank

Trace

Lemma A.1. *The trace function* $tr(\boldsymbol{A}) = \sum a_{ii}$ *satisfies the following properties for the matrices* $\boldsymbol{A} \in \mathcal{M}_{n,n}$*,* $\boldsymbol{B} \in \mathcal{M}_{n,n}$*,* $\boldsymbol{C} \in \mathcal{M}_{p,n}$*,* $\boldsymbol{D} \in \mathcal{M}_{n,p}$ *and a scalar* $\alpha \in \mathbb{R}$*:*

- $tr(\alpha) = \alpha$,
- $tr(\boldsymbol{A} \pm \boldsymbol{B}) = tr(\boldsymbol{A}) \pm tr(\boldsymbol{B})$,
- $tr(\boldsymbol{CD}) = tr(\boldsymbol{DC}) = \sum_{ij} c_{ij} d_{ji}$,
- *In particular,* $tr(\boldsymbol{C}^T\boldsymbol{C}) = tr(\boldsymbol{C}\boldsymbol{C}^T) = \sum_{ij} c_{ij}^2$.

Rank

The following result is given by Timm in [26] in Theorem 2.6.6:

Theorem A.2. *For a symmetric matrix* $\boldsymbol{A} \in \mathcal{S}_n$*,* $rank(A)$ *equals the number of nonzero eigenvalues* λ_i*.*

Fischer proves the following assessment of the rank of a product matrix in Section 2.5.5 in [9]:

Theorem A.3. *Let* $\boldsymbol{A} \in \mathcal{M}_{m,n}$ *and* $\boldsymbol{B} \in \mathcal{M}_{n,r}$*. Then*

$$rank(\boldsymbol{A}) + rank(\boldsymbol{B}) - n \leq rank(\boldsymbol{AB}) \leq \min(rank(\boldsymbol{A}, rank(\boldsymbol{B})))$$

Kronecker product and vec-operator

The following definitions and theorems can be found in [18], [26] and [10].

Definition A.4 (The Kronecker product)**.** Let $\boldsymbol{A} \in \mathcal{M}_{n,m}$, $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} \in \mathcal{M}_{p,q}$, $\boldsymbol{B} = (b_{kl})$ be two matrices. Then the *Kronecker product* of $\boldsymbol{A}$ and $\boldsymbol{B}$ is denoted by $\boldsymbol{A} \otimes \boldsymbol{B}$ and is defined as follows:

$$\boldsymbol{A} \otimes \boldsymbol{B} := \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}, \tag{A.1}$$

$\boldsymbol{A} \otimes \boldsymbol{B} \in \mathcal{M}_{mp,nq}$.

From these definition the results below follow. They are given without proof, but can be proved easily.

Theorem A.5. *Let $\boldsymbol{A} \in \mathcal{M}_{m_1,n_1}$, $\boldsymbol{B} \in \mathcal{M}_{m_2,n_2}$, $\boldsymbol{C} \in \mathcal{M}_{m_3,n_3}$ and $\boldsymbol{D} \in \mathcal{M}_{m_4,n_4}$ be matrices and let $\boldsymbol{x} \in \mathbb{R}_1^d$ and $\boldsymbol{y} \in \mathbb{R}_2^d$ be two vectors.*

(i)
$$\boldsymbol{x}^T \otimes \boldsymbol{y} = \boldsymbol{y}\boldsymbol{x}^T = \boldsymbol{y} \otimes \boldsymbol{x}^T \tag{A.2}$$

(ii)
$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC} \otimes \boldsymbol{BD}) \tag{A.3}$$

(iii)
$$\boldsymbol{I}_d \otimes \boldsymbol{A} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A} \end{bmatrix} \in \mathcal{M}_{dm,dn}. \tag{A.4}$$

(iv)
$$(\boldsymbol{A} \otimes \boldsymbol{B})^T = (\boldsymbol{A}^T \otimes \boldsymbol{B}^T) \tag{A.5}$$

(v)
$$rank(\boldsymbol{A} \otimes \boldsymbol{B}) = rank(\boldsymbol{A})rank(\boldsymbol{B}) \tag{A.6}$$

(vi) In general:
$$\boldsymbol{A} \otimes \boldsymbol{B} \neq \boldsymbol{B} \otimes \boldsymbol{A} \tag{A.7}$$

Definition A.6 (The *vec*-operator)**.** Let $\boldsymbol{A} \in \mathcal{M}_{n,m}$, $\boldsymbol{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_m]$, where the $\boldsymbol{a}_i$, $i = 1, ..., m$ are the columns of $\boldsymbol{A}$. The *vec-operator* stacks the columns of $\boldsymbol{A}$ sequentially to form an $(nm \times 1)$ -vector:

$$vec(\boldsymbol{A}) := \begin{bmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix} \tag{A.8}$$

Theorem A.7. *Let $\boldsymbol{X} \in \mathcal{M}_{n,r}$ be a random variable whose rows have common covariance matrix $\boldsymbol{\Sigma}$. Then*
$$Cov[vec(\boldsymbol{X}^T)] = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}. \tag{A.9}$$

The following theorem gives a connection of vec operator and Kronecker product. It is part of Theorem 9.4.7 in Timm ([26])

Theorem A.8. *Let $\boldsymbol{A} \in \mathcal{M}_{m,n}$, $\boldsymbol{B} \in \mathcal{M}_{n,d}$ and $\boldsymbol{C} \in \mathcal{M}_{d,c}$ and $\boldsymbol{D} \in \mathcal{M}_{c,g}$. Then*
$$tr((\boldsymbol{ABCD})) = (vec(\boldsymbol{A}^T))^T(\boldsymbol{D}^T \otimes \boldsymbol{B})vec(\boldsymbol{C}) \tag{A.10}$$

Idempotent Matrices

Definition A.9 (Idempotent Matrix). A matrix $\boldsymbol{A}$ is said to be *idempotent* if $\boldsymbol{A}^2 = \boldsymbol{A}$.

Timm gives the following characterization of idempotent matrices in Theorem 2.6.6 in [26]:

Theorem A.10. *$\boldsymbol{A} \in \mathcal{M}_{n,n}$ is an idempotent matrix if and only if the eigenvalues of $\boldsymbol{A}$ are either equal to one or equal to zero.*

Theorem A.11. *: Let $\boldsymbol{A} \in \mathcal{M}_{n,n}$ be an idempotent matrix. Then*

$$rank(\boldsymbol{A}) = tr(\boldsymbol{A}).$$

Ratios of Polynomial of finite degree

The following result can be found in [15] and [5] and is used to prove Theorem 4.2.

Theorem A.12. *Let $P(it_1, it_2)$, $Q(it_1, it_2)$, $R(it_1, it_2)$ and $S(it_1, it_2)$ be polynomials in t_1 and t_2. Then the relatio*

$$\frac{R(it_1, it_2)}{S(it_1, it_2)} = \exp(\frac{P(it_1, it_2)}{Q(it_1, it_2)})$$

holds for all $t_1, t_2 \in \mathbb{R}$ if and only if $\frac{R(it_1,it_2)}{S(it_1,it_2)} = 1$ and $\exp(\frac{P(it_1,it_2)}{Q(it_1,it_2)}) = 0$.

A.2. Multivariate Distributions

The Multivariate Normal Distribution

Definition A.13 (p-Variate Normal Distribution). A vector $\boldsymbol{y} \in \mathbb{R}^p$ is said to have a p-variate normal distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and covariance matrix $\boldsymbol{\Sigma} \in \mathcal{S}_p^+$, written $\boldsymbol{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if it has the joint density function given by

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}} \exp\left\{-\frac{(\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{y}-\boldsymbol{\mu})}{2}\right\}. \tag{A.11}$$

Theorem A.14. *Let $\boldsymbol{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.*

(i) If $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$, where $\boldsymbol{A} \in \mathcal{M}_{q,p}$ and $\boldsymbol{b} \in \mathbb{R}^q$, then

$$\boldsymbol{y} \sim \mathcal{N}_q(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T).$$

(ii) Let $\boldsymbol{x} = \begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$ be a partitioning, where $\boldsymbol{x}_i$, $\boldsymbol{\mu}_i \in \mathbb{R}^{p_i}$, $i = 1, 2$, $\boldsymbol{\Sigma}_{ii} \in \mathcal{S}_{p_i}^+$, $i = 1, 2$ and $\boldsymbol{\Sigma}_{i,j} \in \mathcal{M}_{p_i,p_j}$, $i, j = 1, 2, i \neq j$, $p_1 + p_2 = p$. Then the resulting subvectors of $\boldsymbol{x}$ are multivariate normally distributed in the following way:

$$\boldsymbol{x}_i \sim \mathcal{N}_{p_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii}), \qquad i = 1, 2.$$

More generally, all marginal distributions of any subset of random variables are normal. However, the converse is not generally true.

Theorem A.15. *(i) Two jointly multivariate normal vectors are independent if and only if they are uncorrelated.*

(ii) For two jointly multivariate vectors, pairwise independence of their components implies complete independence.

Theorem A.16 (Khatri 1962, [11], Theorem II.). *Let $\boldsymbol{Y} \in \mathcal{M}_{n,m}$, $\boldsymbol{A} \in \mathcal{S}_n$ and $\boldsymbol{B} \in \mathcal{S}_n$. Let the rows of $\boldsymbol{Y}$ be independent multivariate normals with common covariance matrix $\boldsymbol{\Sigma}$ and mean $\boldsymbol{\mu}_i$, $i = 1, ..., n$. Then $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$ and $\boldsymbol{X}^T\boldsymbol{B}\boldsymbol{X}$ are distributed independently if and only if*

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{0}.$$

The Wishart Distribution

The Wishart distributions is a matrix generalization of the univariate chi-squared distribution. Further details may be found in [18] and [26].

Definition A.17 (Central Wishart Distribution)**.** Let $\boldsymbol{A} \in \mathcal{M}_{p,p}$. If $\boldsymbol{A}$ can be written $\boldsymbol{A} = \boldsymbol{X}^T\boldsymbol{X}$, where $\boldsymbol{X} \in \mathcal{M}_{m,p}$ is a matrix whose rows are distributed $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, then $\boldsymbol{A}$ is said to have a *central Wishart distribution* with covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom parameter m. We write $\boldsymbol{A} \sim W_p(m, \boldsymbol{\Sigma})$. When $\boldsymbol{\Sigma} = \boldsymbol{I}_p$, the distribution is said to be in standard form.

Definition A.18 (Non-central Wishart Distribution)**.** For $E[\boldsymbol{X}] = \boldsymbol{M}$, $\boldsymbol{M} \neq \boldsymbol{0}$, $\boldsymbol{A}$ has a *non central Wishart distribution* with non centrality parameter $\boldsymbol{\Gamma} = \boldsymbol{M}^T\boldsymbol{M}\boldsymbol{\Sigma}^{-1}$.
We write $\boldsymbol{A} \sim W_n(m, \boldsymbol{\Sigma}, \boldsymbol{\Gamma})$.

Theorem A.19. *Let $\boldsymbol{A} \in \mathcal{M}_{p,p}$, $\boldsymbol{A} \sim W_p(m1, \boldsymbol{\Sigma})$, $\boldsymbol{B} \in \mathcal{M}_{p,p}$, $\boldsymbol{B} \sim W_p(m_2, \boldsymbol{\Sigma})$ and let $\boldsymbol{G} \in \mathcal{M}_{p,q}$ denote a matrix.*

(i) $\boldsymbol{G}^T\boldsymbol{A}\boldsymbol{G} \sim W_q(m_1, \boldsymbol{G}^T\boldsymbol{\Sigma}\boldsymbol{G})$.

(ii) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are independent, then $\boldsymbol{A} + \boldsymbol{B} \sim W_p(m_1 + m_2, \Sigma)$.

Theorem A.20 (Khatri, [11],Theorem I.)**.** *Let $\boldsymbol{Y} \in \mathcal{M}_{n,m}$ be a matrix whose rows are distributed independently as multivariate normals with means $\boldsymbol{\mu}_i \in \mathbb{R}^m$, $i = 1, ..., n$ and common covariance matrix $\boldsymbol{\Upsilon} \in \mathcal{S}_m^+$. Let $\boldsymbol{\mu} \in \mathcal{M}_{n,m}$ be the matrix whose rows are $\boldsymbol{\mu}_i$, $i = 1, ..., n$. Moreover, let $\boldsymbol{C} \in \mathcal{S}_n^+$. Then*

$$\boldsymbol{Y}^T\boldsymbol{C}\boldsymbol{Y} \sim W_m(r, \boldsymbol{\Upsilon}, \boldsymbol{\Xi}),$$

if and only if $\boldsymbol{C}$ is idempotent.

The following theorem is an adaption of Theorem 3.3.10 in [24] to the notation used here.

Theorem A.21. *Let $\boldsymbol{Q} = \boldsymbol{U}^T\boldsymbol{A}\boldsymbol{U} \sim W_n(m, \boldsymbol{\Upsilon}, \boldsymbol{\Xi})$. Let $\boldsymbol{\mu}$ be the mean of vec$(\boldsymbol{Q}^T)$. Then the characteristic function of $\boldsymbol{Q}$ is given by*

$$\varphi(\boldsymbol{T}) = |\boldsymbol{I}_n - 2i\boldsymbol{\Upsilon}\boldsymbol{T}|^{-\frac{n}{2}} \exp(i\boldsymbol{\mu}^T[\boldsymbol{I}_n \otimes \boldsymbol{T}(\boldsymbol{I}_m - 2i\boldsymbol{\Upsilon}\boldsymbol{T})^{-1}]\boldsymbol{\mu}),$$

where $\boldsymbol{T} \in \mathcal{S}_n$.

Wilks' Lambda Distribution

Definition A.22 (Wilks' Lambda Distribution)**.** When $\boldsymbol{A} \sim W_p(m, \boldsymbol{\Sigma})$ and $\boldsymbol{B} \sim W_p(n, \boldsymbol{\Sigma})$ are independent, $m \geq p$, we say that

$$\Lambda = \frac{|\boldsymbol{A}|}{|\boldsymbol{A}+\boldsymbol{B}|} = |\boldsymbol{I}_p + \boldsymbol{A}^{-1}\boldsymbol{B}| \sim \Lambda(p, m, n)$$

has a *Wilks' lambda distribution* with parameters p, m and n.

The Wilks' lambda distribution is a multivariate extension to the beta distribution and occurs frequently in the context of likelihood ratio tests. The parameter m usually represents the "error" degrees of freedom and n the "hypothesis" degrees of freedom. Hence, $m+n$ represents the total degrees of freedom. Wilks' lambda distribution is invariant under changes of scale parameters of $\boldsymbol{A}$ and $\boldsymbol{B}$.

List of Figures

5.1. Scatterplot 49
5.2. Scatter plot for the Rearranged Data 49
5.3. Plot for SOr-me 54
5.4. Plot for ANS-me 54
5.5. Plot for PAL-MP angle 54
5.6. QQ plot assesing multivariate normality 55

List of Tables

5.1. Individual Measurements on the Vertical Position of the Mandible at three Points in Time 48
5.2. Multivariate Mixed Model Analysis of the Dental Data 51
5.3. $\hat{\epsilon}$-Adjusted Multivariate Mixed Model Analysis of the Dental Data 51
5.4. Results of the Multivariate Mixed Model Analysis and $\hat{\epsilon}$-Adjusted Analysis 52
5.5. Results of the Multivariate Mixed Model Analysis and $\hat{\epsilon}$-Adjusted Analysis 2 . . 53

Bibliography

[1] R. Boik (1988). The mixed model for multivariate repeated measures: validity conditions and and approximate tests. *Psychometrika* 53, 469 - 482.

[2] N.E. Breslow and D.G. Clayton (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88 (441), 9 - 25.

[3] T. Burdenski (1998). Evaluating Univariate, Bivariate and Multivariate Normality Using Graphical and Statistical Procedures. *Multiple Linear Regression Viewpoints* 25, 15 - 28.

[4] N. Rao Chaganty and Dayanand N. Naik (2002). Analysis of Multivariate Longitudinal Data Using Quasi Least Squares. *Journal of Statistical Planning and Inference* 103 (1-2), 421 - 436.

[5] M.F. Driscoll and W.R. Grundberg (1986). A history of the development of Craig's theorem. *The American Statistician* 40 (1), 65 - 70.

[6] L. Fahrmeir, T. Kneib and S. Lang (2007). *Regression.* Springer.

[7] L. Fahrmeir, A. Hammerle and G. Tutz (Ed.) (1996). *Multivariate Statistische Verfahren.* de Gruyter.

[8] S. Fieuws and G. Verbeke (2006). Pairwise Fitting of Mixed Models for Joint Modelling of Multivariate Longitudinal Profiles. *Biometrics* 63, 424 - 431.

[9] G.Fischer (2002). *Lineare Algebra.* vieweg.

[10] H.V. Henderson and S.R: Saerle (1981). The vec-permutation matrix, the vec operator and Kronecker products: A review. *Linear and Multilinear Algebra*, 9(4), 271 - 288.

[11] C.G. Khatri (1962). Conditions for Wishartness and Independence of Second Degree Polynomials in a Normal Vector. *The Annals of Mathematical Statistics*, 33(3), 1002 - 1007.

[12] A.I. Khuri (1982). Direct Products: a Powerful Tool for the Analysis of Balanced Data. *Communications in Statistics - Theory and Methods* 11(25), 2903 - 2920.

[13] A.I. Khuri, T, Mathew and B.K. Sinha (1998). *Statistical Tests for Mixed Linear Models.* John Wiley & Sons.

[14] A.I. Khuri (2009). *Linear Model Methodology.* Chapman and Hall CRC.

[15] R.G. Laha (1956). On the Stochastic Independence of Two Second-Degree Polynomial Statistics in Normally Distributed Variates. *The Annals of Mathematical Statistics* 27(3), 790 - 796.

[16] M.N. Laird, and J.H. Ware (1982). Random-Effects for Longitudinal Data; *Biometrics* 38(4), 962 - 974.

[17] R.C. Littell (2002). Analysis of Unbalanced Mixed Model Data: A Case Study Comparision of ANOVA versus REML/GLS. *Journal of Agricultural, Biological and Environmental Statistics* 7(4), 472 - 490.

[18] K.V. Mardia, J.T. Kent and J.M. Bibby (2003). *Multivariate Analysis.* American Press.

[19] T. Mathew (1989). MANOVA in the Multivariate Components of Variance Model. *Journal of Multivariate Analysis* 29, 30 - 38.

[20] G. Reinsel (1982). Multivariate Repeated-Measurement or Growth Curve Models with Multivariate Random Effects Covariance Structure. *Journal of the American Statistical Association* 77(1), 190 - 195.

[21] R.G. Shaw and T. Mitchell-Olds (1993). ANOVA for Unbalanced Data: An Overview. *Ecology* 74(6), 1638 - 1645.

[22] H. Scheffé (1956). A Mixed Model for the Analysis of Variance. *Annals of Mathematical Statistics* 27(1), 23 - 36.

[23] J.R. Schott and J.G. Shaw (1984). A Multivariate One-Way Classification Model with random effects. *Journal of Multivariate Analysis* 15 (1), 1 - 12.

[24] M.S. Sivastrava and C.G. Khatri (1970). *An Introduction to Multivariate Statistics.* North Holland

[25] N.H. Timm (1980). Multivariate Analysis of repeated measurements. In P.R Keishnaiah (Ed.), *handbook of statistics*, vol.1 (Analysis of Variance), 41 - 87. North-Holland.

[26] N.H. Timm (2002). *Applied Multivariate Analysis.* Springer.

[27] D.R. Thomas (1983). Univariate Repeated Measured Techniques Applied to Multivariate Data. *Psychometrika* 48(3), 451 - 464.

[28] G. Verbeke, G. Molenberghs (2009). *Linear Mixed Models for Longitudinal Data.* Springer.

[29] C.M. Wang (1983). On the Analysis of Multivariate Repeated Measures Designs. *Communications in Statistics - Theory and Methods* 12(14), 1647 - 1659

[30] M.B. Wilk and O. Kempthorne (1955). Fixed, Linear and Mixed Models. *Journal of the American Statistical Association* 50(4), 1144 - 1167.