Linköping Studies in Science and Technology
Dissertations No. 2221

Residual Analysis in the GMANOVA-MANOVA Model

Béatrice Byukusenge

LINKÖPING
UNIVERSITY

Linköping Studies in Science and Technology. Dissertations.
No. 2221

Residual Analysis in the GMANOVA-MANOVA Model

Béatrice Byukusenge

Department of Mathematics
Linköping University, SE–581 83 Linköping, Sweden
Linköping 2022

Linköping Studies in Science and Technology. Dissertations.
No. 2221

Residual Analysis in the GMANOVA-MANOVA Model

Béatrice Byukusenge

`beatrice.byukusenge@liu.se`
`www.mai.liu.se`
Department of Mathematics
Linköping University
SE–581 83 Linköping
Sweden

ISBN 978-91-7929-284-3 (print)
978-91-7929-285-0 (PDF) ISSN 0345-7524

Printed by LiU-Tryck, Linköping, Sweden 2022

To my Family

Abstract

Béatrice Byukusenge (2022). Residual Analysis in the GMANOVA-MANOVA model. Doctoral dissertation. ISBN 978-91-7929-284-3 (print) 978-91-7929-285-0 (PDF) . ISSN 0345-7524.

This thesis focuses on the establishment and analysis of residuals in the so called GMANOVA-MANOVA model. The model is a special case of the Extended Growth Curve Model. It has two terms where one term models the profiles (growth curves) and the other the covariables of interest. This model is useful in studying growth curves in short time series in fields such as economics, biology, medicine, and epidemiology. Furthermore, in the literature, residuals have been extensively studied and used to check model adequacy in univariate linear models. This thesis contributes to the extension of the study of residuals in the GMANOVA-MANOVA model.

In this thesis, a new pair of residuals is established via the maximum likelihood estimators of the parameters in the model. One residual indicates whether an individual is far away from the group means and a second residual is used to check assumptions about the mean structure. Different properties of these residuals are verified and their interpretation is discussed. Moreover, using parametric bootstrap, the empirical distributions of the extreme elements in the residuals are derived.

Finally, testing bilinear restriction in the MANOVA model is considered. One can show that the MANOVA model with bilinear restrictions is nothing more than a GMANOVA-MANOVA model. Furthermore, the likelihood ratio test can be shown to be given as a function of the residuals to the GMANOVA-MANOVA model, which can be used to understand the appropriateness of the model and test the bilinear hypothesis.

Populärvetenskaplig sammanfattning

I den här avhandling härleds och analyseras residualer för den så kallade GMANOVA-MANOVA modellen. Modellen, som är ett specialfall av den utökade tillväxtkurvemodellen, har två delar där den ena beskriver profilerna (väntevärdesmodellen som tillväxtkurvor) för olika grupper och den andra delen i modellen de kovariater som kan vara av vikt (variabler som påverkar analysen men som inte direkt är av intresse). GMANOVA-MANOVA modellen är användbar för att studera upprepade mätningar, som korta tidsserier, inom områden som teknik, ekonomi, biologi, medicin och epidemiologi.

Residualer för en statistisk modell är ett viktigt verktyg för att studera om modellantaganden är uppfyllda. Den här avhandlingen bidrar till resiudalanalys i GMANOVA-MANOVA-modellen, genom att härleda två residualer för modellen. Den första residualen beskriver om en observation är långt borta från gruppmedelvärdet medan den andra residualen används för att kontrollera antaganden om väntevärdesmodellen (profilen) är uppfyllda. Olika egenskaper hos dessa residualer härleds och deras tolkning diskuteras. Dessutom, beskrivs med hjälp av parametrisk bootstrap, hur de empiriska fördelningarna av de extrema elementen i residualerna kan beräknas och utnyttjas i analysen.

.

Acknowledgments

I take this opportunity to express my gratitude to my supervisors Professor Dietrich von Rosen and Martin Singull for their guidance, patience, and good collaboration during my studies at Linköping University. Thank you Dietrich for teaching me a lot about writing. In the same way, I wish to thank the entire community at the Department of Mathematics, Linköping University for a good working environment and support.

Especially, I would like to thank Bengt-Ove Turesson, Björn Textorius, Theresa Lagali-Hensen, and Meaza Abebe for their invaluable assistance and support. I thank also Tomas Sjödin and Theresia Carlsson Roth. They directed me through various situations allowing me to reach this accomplishment.

I am very much grateful to Denis Ndanguza, Dean of the school of science at the University of Rwanda; Célestin Kurijyubwamwi, Head of the Department of Mathematics and Froduald Minani.

I would also like to thank fellow Ph.D. students and other staff of the Department of Mathematics for a friendly working environment. To my family, there do not exist words that can express my gratefulness for your support. A special thank you to my parents for all the sacrifices you have taken to support me.

Finally, I take this opportunity to acknowledge the sub-program of Applied Mathematics and Statistics under the Swedish International Development Cooperation Agency - funded bilateral program, The University of Rwanda - Sweden Programme for Research, Higher Education and Institutional Advancement for financial support of my Ph.D. training.

Linköping, November 3, 2022
"Béatrice Byukusenge"

Contents

1

Introduction

For a long time, residuals have played an important role in statistical modeling. Residuals which often are defined as the difference between observed and estimated mean values can be used for model diagnostics. In particular, residuals can be used to find outliers and influential observations, determine the level of multicollinearity among independent variables, and test for violation of model assumptions. The study of residuals in the multivariate analysis of variance (MANOVA) model is complicated because they are multi-dimensional. Different techniques have developed for example von Rosen (1995) studied residuals in the Growth Curve model (GCM) which is also known as Generalized Multivariate Analysis of Variance (GMANOVA) model. In this thesis, the main focus is on the establishment and analysis of residuals in the so-called GMANOVA-MANOVA model. The model is a special case of the Extended Growth Curve model, and its mean consists of two terms where one term models the profile (growth curve), and the second term models the covariables. The GCM as well as the GMANOVA-MANOVA model are useful in studying short time series in fields such as economics, biology, medicine, and epidemiology, for example.

1.1 Background

In the early nineteenth-century Legendre and Gauss pioneered linear models and least squares when applying the method to astronomical data (see Stigler, 1986). Since then, the development of linear models has occupied an important place in statistics. It all began with Galton's work, where multivariate statistical analyses were introduced (Galton (1886, 1889a)). When samples from bivariate normal distribution were taken into consideration Galton (1889b) developed concepts about regression and using plots, he gave an idea on how to determine the slope of regression lines. Pearson (1896) established the correlation coefficient to measure the relationship between subjects. Later, Fisher (1915) derived the simultaneous distribution of the sample variances and covariance when the underlying distribution is bivariate normal. Wishart (1928) extended Fisher's work by finding the simultaneous sampling distribution of the variances and covariances in samples from a $p-$variate multivariate normal population. Later, Wilks (1932) used the likelihood ratio approach (Neyman and Pearson, 1928, 1931) to perform a suitable generalization in the analysis of variance for many variables. This was further explored by Rao (1948) by whom the name Wilks' criterion was given. Different books on multivariate analysis appeared for the first time in the 1950s, for example, Roy (1957), Kendall (1957) and Anderson (2003) with its first addition Anderson, 1958. Over the years, several books on classical multivariate analysis have been published Dempster (1969), Srivastava and Khatri (1979), Mardia et al. (1979) and Muirhead (1982), among others. This thesis considers a special case of the GCM model introduced by Potthoff and Roy (1964). The GCM model is an extension of the multivariate analysis of variance model (MANOVA) and it is suitable for modelling data with repeated measurements on individuals. The GCM has different names such as the Potthoff and Roy model, the bilinear regression model, and the Generalized MANOVA (GMANOVA). Application of the GCM model lies for example in treating trends or growth curves found in epidemiology, medical research, and biostatistics (see Pan et al., 2002 and von Rosen, 2018). The GCM model is a natural generalization of the MANOVA model and different test statistics on the hypotheses of the mean structure have been developed, see for instance the book of Roy (1957) and the paper of Khatri (1966) where the bilinear hypothesis testing was formulated. Unlike the MANOVA model, the maximum likelihood estimator (MLE) of the mean parameter in the GCM model is a non-linear random expression, which poses a lot of problems when it comes to inference, e.g., see (Srivastava and Carter, 1983). The GMANOVA model has been extensively studied by many authors see for example Rao (1965, 1966, 1987). The GCM model has been reviewed by (Woolson and Leeper, 1980; Seber, 1984; von Rosen, 1991; Srivastava and von Rosen, 1999; von Rosen, 2018).

A generalization of this model, called the Extended Growth Curve model (EGCM), was studied by von Rosen (1989). Its canonical form was already considered in Gleser and Olkin (1970) and later extended by Srivastava and Khatri (1979). The EGCM model known as the sum of the profiles model was introduced by Verbyla and Venables (1988) who derived the explicit form of the MLEs. Details about the MLEs of the EGCM under the nested subspace conditions have been presented in Kollo and von Rosen (2005). Filipiak and von Rosen (2012) provided the MLEs in EGCM based on the range condition in the design matrices. The details about the EGCM model can be found in von Rosen (2018).

The thesis is mainly focused on the study of residuals in the special case of an EGCM model called the GMANOVA-MANOVA model. The model is composed of two terms where one term models the profile (growth curve) and the second term is the same as the mean in a MANOVA model which takes care of the covariables, see (Chinchilli and Elswick, 1985, von Rosen, 1989). The first published work on the GMANOVA-MANOVA model is the paper by Chinchilli and Elswick (1985). They derived MLEs and provided some significance tests.

Residuals that are obtained after subtracting the estimated mean from the observations were introduced in the GCM model by von Rosen (1995). In that paper, von Rosen studied residuals in the GCM model through the decomposition of linear spaces. He showed that residuals in the GCM are symmetrically distributed around zero and established a couple of moment relations for three different types of residuals. Hamid and von Rosen (2006) did the same study on the EGCM model. For more details about the analysis of residuals in growth models, we refer the reader to von Rosen (2018).

1.2 Aims

The main topic under investigation in this thesis is the analysis of residuals from a GMANOVA-MANOVA model, with the following specific objectives:

(i) to establish matrix residuals in the GMANOVA-MANOVA model, discuss their interpretations and verify their basic properties;

(ii) to study the distribution of the extreme elements in the residuals by using a parametric bootstrap approach;

(iii) to test and interpret a hypothesis related to bilinear restrictions on the parameter space in the MANOVA model.

1.3 Thesis outline

This thesis comprises two parts. The first part provides the background and summary of the thesis, while the second part is a compilation of four papers. Below we provide an outline of both parts.

1.3.1 Outline of Part I

The first chapter is the introduction which consists of the background, aims of the thesis, and summary of the papers. Chapter 2 presents some preliminaries including a general introduction to multivariate linear models with a focus on models used in this thesis together with MLEs for the GMANOVA-MANOVA model. Chapter 3, presents the definitions of residuals for the GMANOVA-MANOVA model, interpretation, and properties of the residuals. A parametric bootstrap algorithm is used to identify the extreme residual elements. A test statistic for bilinear restrictions in the MANOVA model is presented in this chapter. Part I ends with Chapter 4 which gives the conclusion of the thesis and suggestions for further research.

1.3.2 Outline of Part II

Part II consists of four papers. Below is a summary of each of the papers.

Paper A: On Residual Analysis in the GMANOVA-MANOVA Model

Byukusenge, B., von Rosen, D. and Singull, M. (2022). On Residual Analysis in the GMANOVA-MANOVA Model. In: Balakrishnan, N. Gil, María Á., Martín, N., Morales, D., Pardo, M.d.C. (eds) Trends in Mathematical, Information and Data Sciences. Studies in Systems, Decision, and Control, 445, 287–305, Springer, Cham.

In Paper A, the GMANOVA-MANOVA model is considered. The main achievement in this paper is the establishment of a new pair of residuals via MLEs. One residual indicates whether an individual is far away from the group means and a second residual is used to check assumptions about the mean structure. Moreover, different properties of these residuals are verified and their interpretations are discussed.

Paper B: On an Important Residual in the GMANOVA-MANOVA Model

Byukusenge, B., von Rosen, D. and Singull, M. (2022). On an Important Residual in the GMANOVA-MANOVA Model. Journal of Statistical Theory and Practice, 16:1-20.

In Paper B, by emphasizing the model formulation and validation, we focus on one of the two GMANOVA-MANOVA residuals in a special situation where some of its elements vanish. Through an example, the interpretation of the model is presented and a discussion about the vanishing residuals is given. We have found that if some of the residuals are vanishing, the model assumptions have to be investigated because some of the model parameters can not be estimated.

Paper C: On the Identification of Extreme Elements in a Residual for the GMANOVA-MANOVA Model

Byukusenge, B., von Rosen, D. and Singull, M. (2022). On the Identification of Extreme Elements in a Residual for the GMANOVA-MANOVA Model. Accepted for publication in Innovations in multivariate statistical modeling: navigating theoretical and multidisciplinary domains, Springer Emerging Topics in Statistics and Biostatistics.

Paper C is a continuation of Paper A. By using parametric bootstrap, we derived an approximative distribution of the extreme elements in the residuals. Two data sets are used and through the bootstrap techniques, we give cut-off points to identify extreme residual values. Specifically, from the residual elements, we select by absolute value the three largest elements and study the simulated distribution of these residuals using different data sets. Hence, parametric bootstrap samples are used to obtain a simulated distribution of the "largest " residuals to see how far the residuals are from the centre of the distribution.

Paper D: Test of bilinear restrictions in the MANOVA model

Byukusenge, B., von Rosen, D. and Singull, M. (2022). Test of bilinear restrictions in the MANOVA model. Report.

In Paper D, statistics for testing hypothesis in the MANOVA model with bilinear restrictions on the parameter space are formulated. The test statistic is constructed using the ratio of determinants of the estimated dispersion matrices for both MANOVA models with and without restrictions, i.e., the likelihood ratio test. Since the dispersion matrices are functions of residuals the likelihood ratio test obtained is expressed in terms of residuals. It is possible to calculate the moments of the residuals. These moments are important in interpreting the obtained test statistic because they give information on the model and can be used for checking the appropriateness of the model.

1.4 Contributions

The main contributions of the thesis are as follows.

> In Paper A two residuals for the GMANOVA-MANOVA models are derived and interpreted. Properties such that expectation, dispersion, and relevant covariances are provided.
>
> To perform data analysis, the matrix expression in one of the residuals in the GMANOVA-MANOVA model is discussed in Paper B. The suggestion here is to understand the choices and expressions of the matrices that give information on the deviating observations for a given model.
>
> In Paper C, the parametric bootstrap procedure is applied and the simulated distribution of the three largest elements of the residuals in the GMANOVA-MANOVA model is obtained. Using the simulated distributions, a cut off point for identifying extreme residuals in the statistics is obtained through the suggested quantiles of the estimated distribution to conclude if any violation of the null hypothesis exists.
>
> In Paper D, hypothesis about the bilinear restriction on the parameter spaces in a MANOVA model is formulated and a LR test is constructed. The test statistic here is the ratio between the determinants of the estimated covariance matrices from the MANOVA model with and without restrictions. We have shown that the test statistic is a function of the residuals from the MANOVA model with restrictions (which is equivalent to the GMANOVA-MANOVA model) and without restrictions (the usual MANOVA model). This makes the test statistic to be interpretable. Based on the calculated moments of those residuals, we showed that under the null hypothesis the calculated moments are independent of the unknown parameters and under the alternative hypothesis, the calculated moments are function of the unknown parameters.

2

Multivariate Models

With several response variables in a dataset, multivariate models can be useful to apply. This chapter aims to give an overview of multivariate models and their generalizations. We begin with the MANOVA model. Then, two bilinear models, namely the GCM model, also known as the GMANOVA model, and the GMANOVA-MANOVA model are presented.

2.1 Multivariate analysis of variance model

The MANOVA model is an extension of the univariate analysis of variance model. It can be used when more than one variable on each experimental unit is measured. References to the MANOVA model can be found in many textbooks on multivariate analysis, see for example (Roy, 1957; Dempster, 1969; Srivastava and Khatri, 1979; Mardia et al., 1979; Muirhead, 1982; Srivastava and Carter, 1983; Seber, 1984; Anderson, 2003; Rencher and Christensen, 2012).

Definition 2.1 (MANOVA). Let $\boldsymbol{X} : p \times n$, be a random observation matrix, $\boldsymbol{B} : p \times k$, be an unknown parameter matrix and $\boldsymbol{C} : k \times n$ be a design matrix, and $r(\boldsymbol{C}) + p \leq n$, where $r(\cdot)$ represents the rank of a matrix. Then the MANOVA model equals

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{C} + \boldsymbol{E}, \tag{2.1}$$

where $\boldsymbol{E}$ is the $p \times n$ matrix of unobservable random errors which follows a matrix normal distribution, $\boldsymbol{E} \sim N_{p,n}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_n)$, i.e., the columns are independently $p-$dimensional normally distributed with expectation $\boldsymbol{0}$ and a positive definite dispersion matrix $\boldsymbol{\Sigma}$.

The MLEs of the parameter matrices $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ are respectively given by:

$$\begin{aligned}
\widehat{\boldsymbol{B}} &= \boldsymbol{X}\boldsymbol{C}' \left(\boldsymbol{C}\boldsymbol{C}'\right)^{-} + \boldsymbol{C}^{\circ\prime}\boldsymbol{Z}, \\
n\widehat{\boldsymbol{\Sigma}} &= \left(\boldsymbol{X} - \widehat{\boldsymbol{B}}\boldsymbol{C}\right)\left(\right)' = \boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-}\boldsymbol{C}\right)\boldsymbol{X}',
\end{aligned}$$

where $\boldsymbol{Z}$ is an arbitrary matrix of proper size, $\boldsymbol{G}^-$ denotes an arbitrary generalized inverse in the sense that $\boldsymbol{G}\boldsymbol{G}^-\boldsymbol{G} = \boldsymbol{G}$ and $\boldsymbol{A}'$ denotes the transpose of a matrix $\boldsymbol{A}$. For any $\boldsymbol{H}$ the matrix $\boldsymbol{H}^\circ$ is a matrix spanning the orthogonal complement to the column space generated by $\boldsymbol{H}$. Hence $\widehat{\boldsymbol{B}}$ is not unique whereas $\widehat{\boldsymbol{\Sigma}}$ satisfies the uniqueness property. Moreover, $(\boldsymbol{Q})()'$ means $(\boldsymbol{Q})(\boldsymbol{Q})'$ for any matrix function $\boldsymbol{Q}$. When a short time series for a given study has to be considered, the MANOVA model is no longer applied. So, another model which is time-dependent has to be considered. The GCM model which is defined in the next section can be used.

2.2 Growth Curve model

The GCM model which is known as a GMANOVA model (see Potthoff and Roy, 1964), is an important model when repeated measures data and longitudinal data are considered, particularly for the analysis of balanced repeated measures data.

Definition 2.2 (Growth Curve model (GCM)). Let $\boldsymbol{X} : p \times n$, $\boldsymbol{A} : p \times m$, $m < p$, $\boldsymbol{B} : m \times r$, $\boldsymbol{C} : r \times n$, and $r(\boldsymbol{C}) + p \leq n$. The GCM model is given by:

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

where $\boldsymbol{A}$ and $\boldsymbol{C}$ are the within and between individuals design matrices respectively, $\boldsymbol{B}$ is a matrix of unknown mean parameters. $\boldsymbol{E} : p \times n$ is the error matrix whose columns are assumed to be independently distributed as a $p-$variate normal distribution with mean $\boldsymbol{0}$ and an unknown positive definite dispersion matrix $\boldsymbol{\Sigma}$, that is $\boldsymbol{E} \sim N_{p,n}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_n)$.

The between individuals design matrix $\boldsymbol{C}$ is the same design matrix as used in the model presented in Definition 2.1. The model is usually applied when repeated measurements follow a within-individual linear regression model for each independent observation vector. Assume that the design matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ are of full rank, then the MLEs of the parameter matrices $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ are respectively given by:

$$\widehat{\boldsymbol{B}} = \left(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}\right)^{-1} \boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{C}' \left(\boldsymbol{C}\boldsymbol{C}'\right)^{-1}, \quad n\widehat{\boldsymbol{\Sigma}} = \left(\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C}\right)\left(\right)', \tag{2.3}$$

where $\boldsymbol{S} = \boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-1}\boldsymbol{C}\right)\boldsymbol{X}'$, see (Khatri, 1966; Kariya, 1985; Kollo and von Rosen, 2005; von Rosen, 2018) for more details. It might happen that different individuals in the GCM model do not follow the same growth profile. In this case, the model is no longer adequate to use. For that reason, von Rosen (1989) (see also Srivastava and Khatri, 1979) extended the GCM model. The EGCM model has several mean profiles and it is given in the following definition.

Definition 2.3 (Extended Growth Curve model (EGCM)). Let $\boldsymbol{X} : p \times n$, $\boldsymbol{A}_i : p \times m_i$, $\boldsymbol{B}_i : m_i \times r_i$, $\boldsymbol{C} : r_i \times n$. Suppose that $m_i \leq p$, $r(\boldsymbol{C}_1) + p \leq n$, $\mathcal{C}\left(\boldsymbol{C}_i'\right) \subseteq \mathcal{C}\left(\boldsymbol{C}_{i-1}'\right)$, $i = 1, 2, \ldots, k$. Then the EGCM model is given by,

$$\boldsymbol{X} = \sum_{i=1}^{k} \boldsymbol{A}_i\boldsymbol{B}_i\boldsymbol{C}_i + \boldsymbol{E},$$

where the columns of $\boldsymbol{E}$ are assumed to be independently distributed as a multivariate normal distribution with mean $\mathbf{0}$ and a positive definite covariance matrix $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{E} \sim N_{p,n}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_n)$. The matrices $\boldsymbol{A}_i$ and $\boldsymbol{C}_i$ are known design matrices whereas $\boldsymbol{B}_i$ and $\boldsymbol{\Sigma}$ are unknown parameter matrices.

The model without restrictions on the subspaces was studied by Verbyla and Venables (1988) under the name of sum of profiles model. The EGCM for two different growth profiles ($k = 2$) is given by

$$\boldsymbol{X} = \boldsymbol{A}_1\boldsymbol{B}_1\boldsymbol{C}_1 + \boldsymbol{A}_2\boldsymbol{B}_2\boldsymbol{C}_2 + \boldsymbol{E}, \quad \mathcal{C}\left(\boldsymbol{C}_2'\right) \subseteq \mathcal{C}\left(\boldsymbol{C}_1'\right), \tag{2.4}$$

the subspace condition may be replaced by $\mathcal{C}\left(\boldsymbol{A}_2\right) \subseteq \mathcal{C}\left(\boldsymbol{A}_1\right)$, i.e.,

$$\boldsymbol{X} = \boldsymbol{A}_1\boldsymbol{B}_1\boldsymbol{C}_1 + \boldsymbol{A}_2\boldsymbol{B}_2\boldsymbol{C}_2 + \boldsymbol{E}, \quad \mathcal{C}\left(\boldsymbol{A}_2\right) \subseteq \mathcal{C}\left(\boldsymbol{A}_1\right). \tag{2.5}$$

This model was considered for example by Filipiak and von Rosen (2012) for $m = 3$. For more details about the model and its estimators see von Rosen (2018).

In Example 2.1, we show how the GCM model may arise in practice.

Example 2.1: Growth of Sitka Spruce data from Diggle et al. 2002)

The study objective was to assess the effect of ozone pollution on trees. As ozone pollution is common in urban areas, the impact of increased ozone concentrations on tree growth is of considerable interest. The response variable is log-tree size, where size is conventionally measured by the product of tree height and squared diameter. The data for 78 trees over two growing seasons are listed in Table 1 and 2 in Appendix A. A total of 53 trees were grown with ozone exposure at 70 ppb and 25 trees were grown in a normal environment. For each tree in the study, a measure of its height was recorded at time, 5, 8.5, 15.5, and 22.5 months. Suppose linear growth curves describe the mean growth for trees in the ozone and normal environments. Then we may use model (2.2) for the analysis of the data set. In model (2.2), the observation matrix is $\boldsymbol{X} : 4 \times 78$ in which 53 columns correspond to measurements on trees in ozone exposure which constitutes Group I and the 25 last columns correspond to measurements on trees in normal environments, say Group II. In order to handle this example the design matrices of the GCM model are given by

$$\boldsymbol{A} = \begin{pmatrix} 1 & 5 \\ 1 & 8.5 \\ 1 & 15.5 \\ 1 & 22.5 \end{pmatrix}, \quad \boldsymbol{C} = \left(\mathbf{1}_{53}' \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} : \mathbf{1}_{25}' \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right), \tag{2.6}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{1}_a$ in (2.6) stands for the vector of "a" ones. From (2.3), parameter estimates of the model equal

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} 4.55 & 4.77 \\ 0.07 & 0.07 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 1.23 & 0.18 & 0.46 & 0.29 \\ 0.18 & 0.44 & 0.38 & 0.37 \\ 0.46 & 0.38 & 0.45 & 0.40 \\ 0.29 & 0.37 & 0.40 & 0.40 \end{pmatrix}.$$

In this case, a linear growth curve is assumed. In Figure 2.1 the repeated measurements for the height of the trees have been plotted group-wise together with the estimated mean

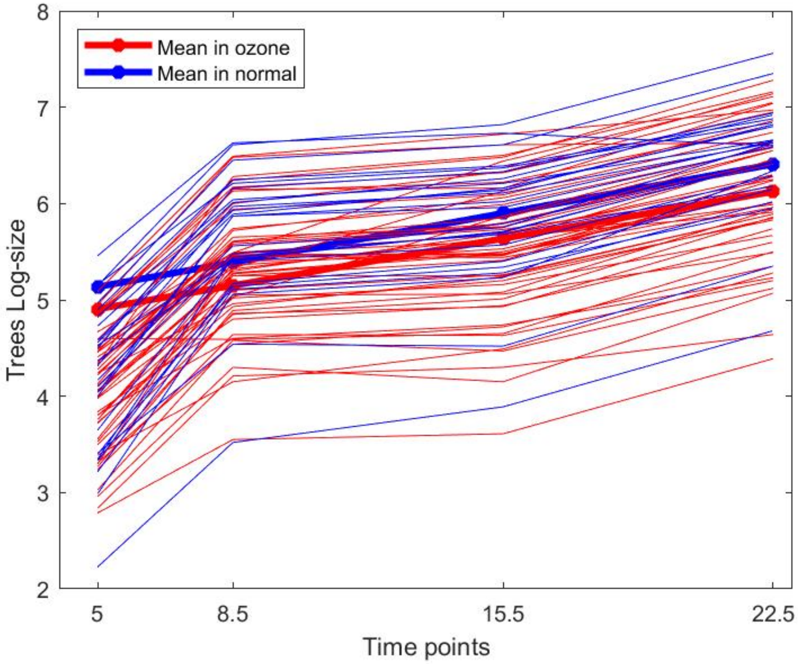

Figure 2.1: *All observations in data set of Diggle et al.* (2002) *are presented and for each group, the estimated mean profiles for the GCM model are shown.*

profiles. It seems like the trees in Group I and Group II are growing at the same rate (see also $\widehat{\boldsymbol{B}}$). This study has been extended to the case where the trees in the two groups were again placed into four chambers. i.e., two chambers for Group I and two chambers for Group II. Those chambers are considered to generate covariable information. To study the effect of the covariate in the model, a special case of the EGCM model is considered. Let for example the matrix $\boldsymbol{A}_2 = \boldsymbol{I}$ then the model (2.4) is reduced to the GMANOVA-MANOVA model defined in the next section.

2.3 GMANOVA-MANOVA model

The GMANOVA-MANOVA model is a special case of the EGCM model. It was introduced in statistics by Chinchilli and Elswick (1985) and it can be used for the analysis of balanced repeated measurements with covariates.

Definition 2.4. Let $\boldsymbol{X}$ be an $p \times n$ matrix of observations, where n represents the number of subjects, each measured at p occasions. The GMANOVA-MANOVA model is defined by

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}_1\boldsymbol{C}_1 + \boldsymbol{B}_2\boldsymbol{C}_2 + \boldsymbol{E}, \tag{2.7}$$

where $\boldsymbol{A} : p \times m$, $\boldsymbol{C}_1 : r_1 \times n$ and $\boldsymbol{C}_2 : r_2 \times n$ are known design matrices, $\boldsymbol{B}_1 : m \times r_1$, $\boldsymbol{B}_2 : p \times r_2$, are unknown parameter matrices, and the random error matrix $\boldsymbol{E} : p \times n$ is such that its columns are assumed to be independently distributed following a $p-$variate normal distribution with mean zero and an unknown positive definite dispersion matrix $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{E} \sim \mathcal{N}_{p,n}\left(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_n\right)$.

Let us consider the parameter matrices $\boldsymbol{B}_1, \boldsymbol{B}_2$ and $\boldsymbol{\Sigma}$ of the GMANOVA-MANOVA model given in Definition 2.4 and let

$$\boldsymbol{Q}_{\boldsymbol{C}_2'} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_2'}, \quad \boldsymbol{S} = \boldsymbol{X}\boldsymbol{Q}_{\boldsymbol{C}_2'}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{X}', \tag{2.8}$$

where for any matrix $\boldsymbol{A}$ of proper size, $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}'$ denotes the projector on the space $\mathcal{C}(\boldsymbol{A})$. Depending on the space generated by the design matrices, the MLEs of the mean parameters are presented in the theorems given below.

Theorem 2.1 (see Byukusenge et al. (2022a,b) for more details)
The MLEs of the parameter matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ in the GMANOVA-MANOVA model in Definition 2.4 can be represented by

$$\widehat{\boldsymbol{B}}_1 = \left(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}\right)^{-}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'\left(\boldsymbol{C}_1\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'\right)^{-} + (\boldsymbol{A}^{\circ})'\boldsymbol{Z}_1 + \boldsymbol{A}'\boldsymbol{Z}_2\left(\boldsymbol{C}_1'\boldsymbol{Q}_{\boldsymbol{C}_2'}\right)^{\circ'}, \tag{2.9}$$

$$\widehat{\boldsymbol{B}}_2 = \left(\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1\right)\boldsymbol{C}_2'\left(\boldsymbol{C}_2\boldsymbol{C}_2'\right)^{-} + \boldsymbol{Z}_3\boldsymbol{C}_2^{\circ'}, \tag{2.10}$$

where $\boldsymbol{Z}_1, \boldsymbol{Z}_2$ and $\boldsymbol{Z}_3$ are arbitrary matrices of proper size. Moreover, the dispersion matrix is uniquely estimated as

$$n\widehat{\boldsymbol{\Sigma}} = \boldsymbol{S} + \left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\right)\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\boldsymbol{X}'\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}'\right),$$

where for any pair of matrices $\boldsymbol{A}$ and $\boldsymbol{S}$, which is supposed to be positive definite, we define

$$\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}} = \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}\right)^{-}\boldsymbol{A}'\boldsymbol{S}^{-1}.$$

The matrices $\widehat{\boldsymbol{B}}_1$ and $\widehat{\boldsymbol{B}}_2$ are not unique. Furthermore, using Theorem 2.1 the predicted values are given in the next theorem.

Theorem 2.2 (see Byukusenge et al., 2022a)
Let $\widehat{\boldsymbol{B}}_1$ and $\widehat{\boldsymbol{B}}_2$ be the MLEs of $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ given in (2.9) and (2.10), respectively. Then, the estimated mean structure is given by

$$\boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2 = \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2'} + \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}}\boldsymbol{C}_1' + \left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\right)\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2'}. \tag{2.11}$$

Using Lemma 2.1 the estimated mean structure (2.11) can be rewritten as

$$\boldsymbol{A}\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2 = \left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\right)\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2'} + \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}, \tag{2.12}$$

where $\boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'} = \boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} + \boldsymbol{P}_{\boldsymbol{C}_2'}$. Hence, we see that the estimators of the mean structure are based on projections of the observations on spaces generated by the design matrices. The model (2.7) is illustrated in Figure 2.2, where a tensor space decomposition of the mean space for the GMANOVA-MANOVA model is shown. Lemma 2.1 is used in the subsequent

Lemma 2.1 (Blockwise formula)
Suppose that a matrix $\boldsymbol{C}$ can be decomposed by columns as $\boldsymbol{C} = [\boldsymbol{C}_1 : \boldsymbol{C}_2]$. Define the projection operator as $\boldsymbol{P}_{\boldsymbol{C}} = \boldsymbol{C}\,(\boldsymbol{C}'\boldsymbol{C})^{-}\,\boldsymbol{C}'$. Similarly, define the residual operator as $\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}}$. Then the projection matrix can be decomposed as (see pp. 323 or Theorem A.45 of Rao et al., 2008)*:*

$$\boldsymbol{P}_{\boldsymbol{C}} = \boldsymbol{P}_{\boldsymbol{C}_1:\boldsymbol{C}_2} = \boldsymbol{P}_{\boldsymbol{C}_1} + \boldsymbol{P}_{(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{C}_1})\boldsymbol{C}_2} \tag{2.13}$$

$$= \boldsymbol{P}_{\boldsymbol{C}_1} + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1})\,\boldsymbol{C}_2\left[\boldsymbol{C}_2'\,(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1})\,\boldsymbol{C}_2\right]^{-}\boldsymbol{C}_2'\,(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1})\,. \tag{2.14}$$

Examination of the between individuals design matrices shows that the space which generates those estimators has been decomposed in the following manner:

$$\mathcal{C}\,(\boldsymbol{C}_2' : \boldsymbol{C}_1') = \mathcal{C}\,(\boldsymbol{C}_1') \boxplus \mathcal{C}\,(\boldsymbol{C}_2' : \boldsymbol{C}_1') \cap \mathcal{C}\,(\boldsymbol{C}_1')^{\perp}, \tag{2.15}$$

where $\boxplus$ denotes the orthogonal sum of subspaces.

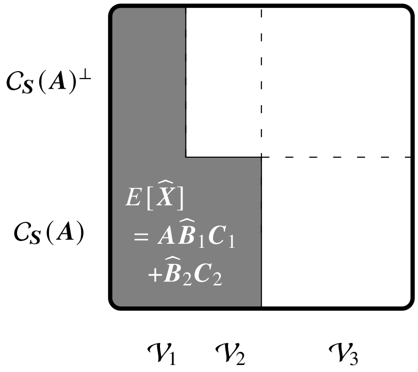

Figure 2.2: *The model in Definition* 2.4 *is illustrated. Spaces connected to the between-individuals decomposition are given by,* $\mathcal{V}_1 = \mathcal{C}(\boldsymbol{C}_2')$, $\mathcal{V}_2 = \mathcal{C}(\boldsymbol{C}_2')^{\perp} \cap \mathcal{C}(\boldsymbol{C}_1' : \boldsymbol{C}_2')$ *and* $\mathcal{V}_3 = \mathcal{C}(\boldsymbol{C}_1' : \boldsymbol{C}_2')^{\perp}$.

Theorem 2.3
Let the matrices $\boldsymbol{A}$, $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ be full rank, i.e., rank $(\boldsymbol{A}) = m_1$, $\text{rank}(\boldsymbol{C}_1) = r_1$ and $\text{rank}(\boldsymbol{C}_2) = r_2$. Then the MLEs given in Theorem 2.1 are uniquely estimated as

$$\widehat{\boldsymbol{B}}_1 = \left(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'\left(\boldsymbol{C}_1\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'\right)^{-1}, \tag{2.16}$$

$$\widehat{\boldsymbol{B}}_2 = \left(\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1\right)\boldsymbol{C}_2'\left(\boldsymbol{C}_2\boldsymbol{C}_2'\right)^{-1}. \tag{2.17}$$

The MLE of the covariance matrix equals

$$n\widehat{\boldsymbol{\Sigma}} = \left(\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 - \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2\right)\left(\right)'. \tag{2.18}$$

The estimated parameters in (2.16) and (2.17) can also be found in Byukusenge et al. (2022b). We again consider Example 2.1, to show how the model in Definition 2.4 may arise.

Example 2.2: (Example 2.1 continued)

Consider again the trees in the ozone and normal environment data set given in Table 1 and 2 in Appendix A. In addition to the controlled conditions, data are organized in four blocks, corresponding to four controlled environment chambers. The first two chambers, each containing 26 and 27 trees, have an ozone-enriched atmosphere, the remaining two, containing 12 and 13 trees respectively, were controlled. For each tree in the study, height was recorded at time 5, 8.5, 15.5, and 22.5 months. Then we may use the model in Definition 2.4 to see the effect of the four blocks. The purpose is to see how the ozone effect the trees growth in each chamber. The design matrices are

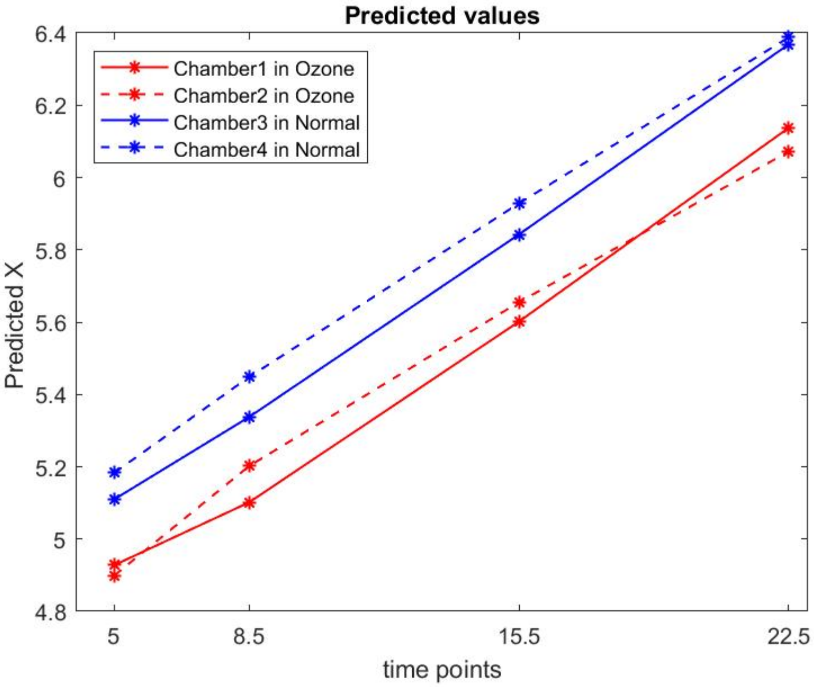

Figure 2.3: Estimated growth curves for the GMANOVA-MANOVA model where the estimated mean profiles for the four chambers are presented.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 5 \\ 1 & 8.5 \\ 1 & 15.5 \\ 1 & 22.5 \end{pmatrix}, \quad \boldsymbol{C}_1 = \left(\mathbf{1}'_{53} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} : \mathbf{1}'_{25} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right), \tag{2.19}$$

$$\boldsymbol{C}_2 = \left(\mathbf{1}'_{26} \otimes \begin{pmatrix} -1 \\ 0 \end{pmatrix} : \quad \mathbf{1}'_{27} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} : \quad \mathbf{1}'_{13} \otimes \begin{pmatrix} 0 \\ -1 \end{pmatrix} : \quad \mathbf{1}'_{12} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right),$$

and the parameter estimates (2.16), (2.17) and (2.18) of the model are given by

$$\widehat{\boldsymbol{B}}_1 = \begin{pmatrix} 4.57 & 4.79 \\ 0.06 & 0.07 \end{pmatrix}, \quad \widehat{\boldsymbol{B}}_2 = \begin{pmatrix} -0.01 & 0.04 \\ 0.05 & 0.05 \\ 0.03 & 0.04 \\ -0.03 & 0.01 \end{pmatrix}, \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 1.24 & 0.18 & 0.45 & 0.27 \\ 0.18 & 0.43 & 0.38 & 0.38 \\ 0.45 & 0.38 & 0.45 & 0.39 \\ 0.27 & 0.38 & 0.39 & 0.40 \end{pmatrix}.$$

According to $\widehat{\boldsymbol{B}}_1$, there can be a difference between Group I and Group II in terms of ozone enrichment. For the covariate estimator, $\widehat{\boldsymbol{B}}_2$ the effect of the chamber is different in Group I and Group II. In Figure 2.3 the estimated mean profiles for the four chambers are presented. It can be seen that, over the entire experiment, the trees in each of the four chambers grow. But trees in chamber one ozone environment grow faster at the end of the experiment compared to the trees in chamber 2. It seems like both environments (ozone and normal) increase the height of the trees over the 22.5 months, i.e., there is a positive treatment effect for both groups. However, according to the statistical paradigm, assumptions and results should be validated, in particular, the model seems to fit the data well. This is often carried out by studying residuals. If one looks closer at the data one can find some observations (for example trees number 23 and 60) that maybe do not follow the model and thereby can have an impact on the results and conclusions.

3

Residual analysis

3.1 Introduction

The analysis of residuals is one of the most crucial parts of any model fitting technique. Residuals indicate what is left unexplained after the estimated model has been subtracted from the observations. They are used to find outliers, verify for normality of data, and assess model appropriateness, among other things. The work about residuals started with Fisher (1915) when he was looking for the exact distribution of the product-moment correlation coefficient in a sample of observations. The work became interesting to different authors. For example, Bartlett (1934) considered the term residual in the analysis of least squares. Later, Dwyer (1941) used residuals to explore the behaviour of normality in data. Thereafter, several discussions about outliers and different types of residuals were provided and they can be found in the papers of Anscombe and Guttman (1960) and Anscombe and John (1963), for example. Moreover, residuals can be used in variable selection, Zyskind (1963).

Various approaches for the analysis of residuals have been produced. For example, through observations, different types of graphs were presented and explanations about the model behaviour were provided. Residuals were utilized to show subjects or individuals who do not follow model assumptions. For the univariate model, it is easy to use the ordinary residuals for the analysis of the model by drawing graphs. However, Behnken and Draper (1972) showed that the standardized residuals are more indicative of deviant observations than ordinary residuals.

Some other techniques without a graphical representation in the analysis of residuals have been provided, see for example Stefansky (1972), Tietjen et al. (1973), Lund (1975) and Prescott (1975). Different forms of residuals for the linear model can be found in the book by Cook and Weisberg (1982). von Rosen (1995), and Hamid and von Rosen (2006) presented residuals for the bilinear model which has later been detailed and discussed in von Rosen (2018).

The main focus of this section is on the residuals for the GMANOVA-MANOVA model. In this thesis, a new pair of residuals for the GMANOVA-MANOVA model is established. It is shown that one residual follows a matrix normal distribution whereas the distribution of the second residual can be obtained by approximation. Moreover, in this thesis, studies of the empirical distribution of the largest elements (by absolute value) of the residual, via a data set, are performed. Parametric bootstrap is used to identify thresholds so that extreme elements of the residuals can be identified.

3.2 Residuals

Using results from Byukusenge et al. (2022a) the predicted values in the GMANOVA-MANOVA model given in Theorem 2.2 can be written (see (2.12))

$$\boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2 = \left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\right)\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2'} + \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}. \tag{3.1}$$

The next lemma is of use for the forthcoming results.

Lemma 3.1 (Transpose and trace including the Kronecker product ⊗ and relationships to the vec-operator)
Let $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ be the arbitrary matrices of proper size so that all given operations are well defined. Then,

$$vec\,\boldsymbol{ABC} = (\boldsymbol{C}' \otimes \boldsymbol{A})\,vec\,\boldsymbol{B}.$$

Applying Lemma 3.1 to (3.1) we obtain the following expression,

$$\begin{aligned}
&\left(\boldsymbol{C}_1' \otimes \boldsymbol{A}\right)\mathrm{vec}(\widehat{\boldsymbol{B}}_1) + \left(\boldsymbol{C}_2' \otimes \boldsymbol{I}\right)\mathrm{vec}(\widehat{\boldsymbol{B}}_2)\\
&= \left(\boldsymbol{P}_{\boldsymbol{C}_2'} \otimes \boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\right)\mathrm{vec}\,(\boldsymbol{X}) + \left(\boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'} \otimes \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\right)\mathrm{vec}\,(\boldsymbol{X})\\
&= \left(\left(\boldsymbol{P}_{\boldsymbol{C}_2'} \otimes \boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\right) + \left(\boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'} \otimes \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\right)\right)\mathrm{vec}\,(\boldsymbol{X}),
\end{aligned}$$

where $\boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}$. Let $\boldsymbol{P} = \boldsymbol{P}_{\boldsymbol{C}_2'} \otimes \boldsymbol{P}_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}} + \boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'} \otimes \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}$, with its column vector space,

$$\mathcal{C}(\boldsymbol{P}) = \mathcal{C}(\boldsymbol{C}_2') \otimes \mathcal{C}_{\boldsymbol{S}}(\boldsymbol{A})^\perp + \mathcal{C}(\boldsymbol{C}_2' : \boldsymbol{C}_1') \otimes \mathcal{C}_{\boldsymbol{S}}(\boldsymbol{A}). \tag{3.2}$$

The space $\mathcal{C}(\boldsymbol{P})$ is known as the linear column space for the mean. Therefore the fact that the estimated mean is obtained by the projection of $\boldsymbol{X}$ on the space generated by the design matrices is in the sense of the expression (3.2). Some authors refer to the matrix $\boldsymbol{P}$ as the prediction matrix because of its role in statistical analysis (see for example Cook and Weisberg, 1982; Chatterjee and Hadi, 1988). Now, our main focus is on the space $\left(\mathcal{C}(\boldsymbol{C}_2' : \boldsymbol{C}_1') \otimes \mathcal{C}_{\boldsymbol{S}}(\boldsymbol{A}) + \mathcal{C}(\boldsymbol{C}_2') \otimes \mathcal{C}_{\boldsymbol{S}}(\boldsymbol{A})^\perp\right)^\perp$, i.e., the orthogonal complement to $\mathcal{C}(\boldsymbol{P})$. This space appears when applying the vec operation to the equation of the ordinary residuals given by

$$\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 - \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2. \tag{3.3}$$

The result is stated in the next theorem which was presented in Byukusenge et al. (2022d).

Theorem 3.1 (Residual space decomposition)
Let $\boldsymbol{P}$ and $\mathcal{C}(\boldsymbol{P})$ be given in (3.2). The space orthogonal to $\mathcal{C}(\boldsymbol{P})$ can be decomposed into three orthogonal spaces as follows:

$$\begin{aligned}&\left(\mathcal{C}(\boldsymbol{C}_2' : \boldsymbol{C}_1') \otimes \mathcal{C}_{\boldsymbol{S}}(\boldsymbol{A}) + \mathcal{C}(\boldsymbol{C}_2') \otimes \mathcal{C}_{\boldsymbol{S}}(\boldsymbol{A})^{\perp}\right)^{\perp}\\ &= \mathcal{C}(\boldsymbol{C}_2' : \boldsymbol{C}_1')^{\perp} \otimes \mathcal{C}_{\boldsymbol{S}}(\boldsymbol{A}) \boxplus \mathcal{C}(\boldsymbol{C}_2' : \boldsymbol{C}_1')^{\perp} \otimes \mathcal{C}_{\boldsymbol{S}}(\boldsymbol{A})^{\perp} \boxplus \left(\mathcal{C}(\boldsymbol{C}_2')^{\perp} \cap \mathcal{C}(\boldsymbol{C}_1')\right) \otimes \mathcal{C}_{\boldsymbol{S}}(\boldsymbol{A})^{\perp}.\end{aligned}$$

By projecting the observations matrix $\boldsymbol{X}$ onto the space presented in Theorem 3.1, the residuals defined by Byukusenge et al. (2022b) and also studied by Byukusenge et al. (2022b), are obtained.

Definition 3.1. For the GMANOVA-MANOVA model presented in Definition 2.4, the following residuals can be utilized:

$$\boldsymbol{R}_1 = \boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}\right) = \boldsymbol{R}_{11} + \boldsymbol{R}_{12}, \tag{3.4}$$

$$\boldsymbol{R}_2 = \left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\right)\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}, \tag{3.5}$$

where

$$\boldsymbol{R}_{11} = \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}\right), \tag{3.6}$$

$$\boldsymbol{R}_{12} = \left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\right)\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}\right). \tag{3.7}$$

The residual $\boldsymbol{R}_1$ can be used to determine if one or more observations differ from the rest whereas the residual $\boldsymbol{R}_2$ can then be used to evaluate assumptions about the mean structure. These ideas were applied in Byukusenge et al. (2022b,c).

3.3 Interpretation

The residuals in Definition 3.1 have a clear meaning which is going to be discussed. By considering the expression of residual $\boldsymbol{R}_1$ given in (3.4),

$$\boldsymbol{R}_1 = \boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}\right), \tag{3.8}$$

it is clear that $\boldsymbol{R}_1$ is the difference between the observations $\boldsymbol{X}$ and the "group mean" $\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}$. Moreover, $\boldsymbol{X} - \boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2'}$ means that $\boldsymbol{X}$ has been adjusted with the effect from the covariate and $\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}$ is an adjusted "mean" effect. Specifically, $\boldsymbol{R}_1$ gives information about the between individual assumptions in a given group. Therefore, it can be used to detect observations that deviate from the others without taking into account any model assumption (see Hamid and von Rosen, 2006).

In this regard, $\boldsymbol{R}_{11}$ given in (3.6), is the difference between the observations $\boldsymbol{X}$ and the mean $\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}$ relative to the within-individuals model. It can therefore be used for detecting if observations do not follow the "within-individuals" model assumptions. Similarly, $\boldsymbol{R}_{12}$ given in (3.7), is the difference between the observations $\boldsymbol{X}$ and the mean

$\boldsymbol{XP}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}$ relative to the case where the within-individuals model assumptions do not hold. For $\boldsymbol{R}_2$ in (3.5), the residual can be written as

$$\begin{aligned}\boldsymbol{R}_2 &= (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}})\, \boldsymbol{XP}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\\ &= (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}})\, \boldsymbol{XP}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} + \boldsymbol{XP}_{\boldsymbol{C}_2'} + \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{XP}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} - \left(\boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2\right)\\ &= \boldsymbol{X}\left(\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} + \boldsymbol{P}_{\boldsymbol{C}_2'}\right) - \left(\boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2\right).\end{aligned}$$

Since $\boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'} = \boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} + \boldsymbol{P}_{\boldsymbol{C}_2'}$, we have

$$\boldsymbol{R}_2 = \boldsymbol{XP}_{\boldsymbol{C}_2':\boldsymbol{C}_1'} - \left(\boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2\right), \tag{3.9}$$

which is the observed "mean" $\boldsymbol{XP}_{\boldsymbol{C}_2':\boldsymbol{C}_1'}$ minus the estimated mean structure (the model), i.e., $\boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2 = \boldsymbol{XP}_{\boldsymbol{C}_2'} + \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{XP}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}$, and therefore $\boldsymbol{R}_2$ tells us how well the estimated mean structure fits the observed mean. More specifically, it gives us information about the within individual assumptions, i.e., the model. Therefore, $\boldsymbol{R}_2$ provides information about the appropriateness of the model assumptions about the mean structure (the profile).

3.4 Properties

It is known that linear models are symmetrically distributed around zero and are uncorrelated with the estimated mean structure. Similar results are derived for the GMANOVA-MANOVA model. Byukusenge et al. (2022a) have shown that residuals given in Definition 3.1 are symmetrically distributed around zero. To establish the theorem, the next lemma presents some technical results which were used.

Lemma 3.2 (see Byukusenge et al., 2022a)
Let $\boldsymbol{C}_1$, $\boldsymbol{C}_2$ be as in Definition 2.4, let $\boldsymbol{Q}_{\boldsymbol{C}_2'}$ be defined in (2.8) and let $\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'$ be the projection of the columns of $\boldsymbol{C}_1'$ onto the orthogonal complement of $\boldsymbol{C}_2'$. Then,

$$\begin{gathered}\boldsymbol{C}_2\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} = \boldsymbol{0}, \quad \boldsymbol{Q}_{\boldsymbol{C}_2'} - \boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_2':\boldsymbol{C}_1'},\\ \mathcal{C}(\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1') = \mathcal{C}(\boldsymbol{C}_2')^{\perp} \cap \{\mathcal{C}(\boldsymbol{C}_2') + \mathcal{C}(\boldsymbol{C}_1')\}.\end{gathered}$$

The expectations of the residuals $\boldsymbol{R}_1, \boldsymbol{R}_2$ in Definition 3.1 are given in the next theorem.

Theorem 3.2 (see Byukusenge et al., 2022a)
Let $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ be the residuals defined in (3.4) and (3.5). Then

$$E\left(\boldsymbol{R}_i\right) = \boldsymbol{0}, \qquad i \in \{1, 2\}.$$

Proof: Since $\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}$ is the projection matrix on $\mathcal{C}\left(\boldsymbol{C}_1' : \boldsymbol{C}_2'\right)^{\perp}$,

$$E\left(\boldsymbol{R}_1\right) = E\left(\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right)\right) = (\boldsymbol{AB}_1\boldsymbol{C}_1 + \boldsymbol{B}_2\boldsymbol{C}_2)\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right) = \boldsymbol{0}.$$

For the residual $\boldsymbol{R}_2$, due to independence between $\boldsymbol{S}$ and $\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}$,

$$E\left(\boldsymbol{R}_2\right) = E\left(\boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right) = E\left(\boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}E\left(\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)\right).$$

Hence, using Lemma 3.2 and that $\boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\boldsymbol{A} = \boldsymbol{0}$,

$$E\left(\boldsymbol{R}_2\right) = E\left(\boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\left(\boldsymbol{A}\boldsymbol{B}_1\boldsymbol{C}_1 + \boldsymbol{B}_2\boldsymbol{C}_2\right)\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right) = \boldsymbol{0}.$$

■

In the next theorem, the dispersion matrices $D\left(\cdot\right)$, for the residuals $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are presented.

Theorem 3.3 (Byukusenge et al., 2022a)
Let $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ be the residuals respectively defined in (3.4) and (3.5). Then

$$\begin{aligned}
D\left(\boldsymbol{R}_1\right) &= \left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right) \otimes \boldsymbol{\Sigma},\\
D\left(\boldsymbol{R}_2\right) &= \boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} \otimes \left(\boldsymbol{\Sigma} - \frac{n - r\left(\boldsymbol{C}_1' : \boldsymbol{C}_2'\right) - 2\left(p - r\left(\boldsymbol{A}\right)\right) - 1}{n - r\left(\boldsymbol{C}_1' : \boldsymbol{C}_2'\right) - \left(p - r\left(\boldsymbol{A}\right)\right) - 1}\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-}\boldsymbol{A}'\right).
\end{aligned}$$

Proof: Consider $D\left(\boldsymbol{R}_1\right)$, and because $\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}$ is idempotent

$$D\left(\boldsymbol{R}_1\right) = D\left(\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right)\right) = \left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right) \otimes \boldsymbol{\Sigma}.$$

For $D\left(\boldsymbol{R}_2\right)$ it follows since $E\left(\boldsymbol{R}_2\right) = \boldsymbol{0}$

$$\begin{aligned}
D\left(\boldsymbol{R}_2\right) &= E\left(\operatorname{vec}\left(\boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)\operatorname{vec}'\left(\boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)\right)\\
&= \boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} \otimes E\left(\boldsymbol{P}'_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\boldsymbol{\Sigma}\boldsymbol{P}_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\right)\\
&= \boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'} \otimes \left(\boldsymbol{\Sigma} - \frac{n - r\left(\boldsymbol{C}_1' : \boldsymbol{C}_2'\right) - 2\left(p - r\left(\boldsymbol{A}\right)\right) - 1}{n - r\left(\boldsymbol{C}_1' : \boldsymbol{C}_2'\right) - \left(p - r\left(\boldsymbol{A}\right)\right) - 1}\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-}\boldsymbol{A}'\right),
\end{aligned}$$

where the last equality follows from the same calculations as when deriving the expectation of the MLE of the dispersion in the GCM model (see von Rosen 2018, p. 113). ■

Finally the pairs $\boldsymbol{R}_1$, and $\boldsymbol{R}_2$, $\widehat{\boldsymbol{X}}$ and $\widehat{\boldsymbol{B}}_j$, $j \in \{1, 2\}$ which are uncorrelated are presented. However, if the covariance $\operatorname{cov}(\cdot, \cdot)$ equals $\boldsymbol{0}$ this does not imply independence.

Theorem 3.4 (Byukusenge et al., 2022a)
Let $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ be the residuals defined in the (3.4) and (3.5), respectively. Then

$$Cov\left(\boldsymbol{R}_1, \boldsymbol{R}_2\right) = \boldsymbol{0}, \tag{3.10}$$

$$Cov\left(\boldsymbol{R}_1, \widehat{\boldsymbol{B}}_j\right) = \boldsymbol{0}, \qquad j \in \{1, 2\}, \tag{3.11}$$

$$Cov\left(\boldsymbol{R}_1, \widehat{\boldsymbol{X}}\right) = \boldsymbol{0}, \tag{3.12}$$

where in (3.11) it is assumed that $\widehat{\boldsymbol{B}}_j$, $j \in \{1, 2\}$, is uniquely estimated.

Proof: First, (3.10) is proven. Since $E(\boldsymbol{R}_1) = \boldsymbol{0}$

$$Cov\,(\boldsymbol{R}_1, \boldsymbol{R}_2) = E\left(\mathrm{vec}\left(\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right)\right)\mathrm{vec}'\left(\boldsymbol{P}_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)\right),$$

and uncorrelatedness follows because $\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right)$ as well as $\boldsymbol{S}$ are independently distributed of $\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}$, i.e.,

$$\begin{aligned}Cov\,(\boldsymbol{R}_1, \boldsymbol{R}_2) &= E\left(\mathrm{vec}\left(\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right)\right)\mathrm{vec}'\left(\boldsymbol{P}_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)\right)\\ &= E\left(\mathrm{vec}\left(\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right)\right)E\left(\mathrm{vec}'\left(\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)\right)\left\{\boldsymbol{I}\otimes\boldsymbol{P}_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\right\}\right),\end{aligned}$$

and $E\left(\mathrm{vec}'\left(\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)\right) = \mathrm{vec}'\left(\boldsymbol{A}\boldsymbol{B}_1\boldsymbol{C}_1\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)$, implies

$$E\left(\mathrm{vec}'\left(\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'}\right)\left\{\boldsymbol{I}\otimes\boldsymbol{P}_{\boldsymbol{A}^\circ,\boldsymbol{S}^{-1}}\right\}\right) = \boldsymbol{0},$$

which proves (3.10).

For (3.11) where it is assumed that $\boldsymbol{B}_i$, $i \in \{1, 2\}$, are uniquely estimated,

$$\begin{aligned}Cov\left(\boldsymbol{R}_1, \widehat{\boldsymbol{B}}_1\right) &= Cov\left(\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'}\right), \left(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'\left(\boldsymbol{C}_1\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1'\right)^{-1}\right)\\ &= \boldsymbol{0},\end{aligned}$$

because $E(\boldsymbol{R}_1) = \boldsymbol{0}$ and $\boldsymbol{X}\boldsymbol{Q}_{\boldsymbol{C}_2'}$ is independent of $\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}_1':\boldsymbol{C}_2'})$ and $\boldsymbol{S}$. Moreover,

$$Cov\left(\boldsymbol{R}_1, \widehat{\boldsymbol{B}}_2\right) = Cov\left(\boldsymbol{R}_1, \boldsymbol{X}\boldsymbol{C}_2'\left(\boldsymbol{C}_2\boldsymbol{C}_2'\right)^{-1}\right) - Cov\left(\boldsymbol{R}_1, \boldsymbol{A}\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1\boldsymbol{C}_2'\left(\boldsymbol{C}_2\boldsymbol{C}_2'\right)^{-1}\right) = \boldsymbol{0},$$

since $Cov\left(\boldsymbol{R}_1, \widehat{\boldsymbol{B}}_1\right) = \boldsymbol{0}$ and $\boldsymbol{R}_1$ is independently distributed of $\boldsymbol{X}\boldsymbol{C}_2'$.

Finally it is noted that $Cov\left(\boldsymbol{R}_1, \widehat{\boldsymbol{X}}\right) = \boldsymbol{0}$ because $Cov\left(\boldsymbol{R}_1, \widehat{\boldsymbol{B}}_1\right) = \boldsymbol{0}$ and $Cov\left(\boldsymbol{R}_1, \widehat{\boldsymbol{B}}_2\right) = \boldsymbol{0}$. ■

In the next example, we present numerical values of the residuals $\boldsymbol{R}_1$ $\boldsymbol{R}_2$ and we illustrate them in Figure 3.1 and 3.2.

Example 3.1: (Example 2.2 continued)

Consider again the trees in ozone and normal environment data set given in Example 2.2, using the GMANOVA-MANOVA model (2.19), the residuals $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ established in Definition 3.1 are then represented in Figure 3.1 and 3.2 and $\boldsymbol{R}_2$ equals,

$$\boldsymbol{R}_2 = \left(\boldsymbol{1}_{26}'\otimes\begin{pmatrix}-0.87\\0.16\\-0.12\\0.039\end{pmatrix} : \boldsymbol{1}_{27}'\otimes\begin{pmatrix}-0.84\\0.15\\-0.11\\0.038\end{pmatrix} : \boldsymbol{1}_{13}'\otimes\begin{pmatrix}-0.94\\0.24\\-0.04\\0.1\end{pmatrix} : \boldsymbol{1}_{12}'\otimes\begin{pmatrix}-1.028\\0.26\\-0.046\\0.011\end{pmatrix}\right). \tag{3.13}$$

We know that residual $\boldsymbol{R}_1$ indicates the observations which deviate from the rest of the observations. Looking at Figure 3.1 we can see that two observations are far away from

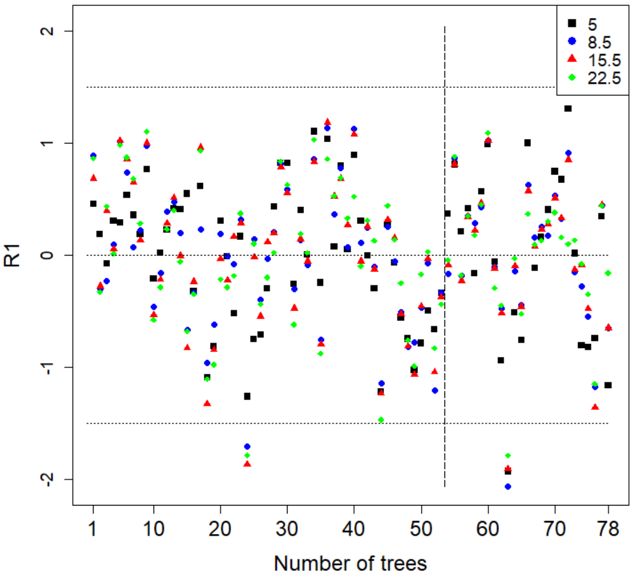

Figure 3.1: *The residuals* $\boldsymbol{R}_1$*, introduced in Definition* 3.1*, using the data in Example* 2.2*.*

the mean group at different time points. For the residual $\boldsymbol{R}_2$ which is suitable for checking assumptions about the mean structure, we need to evaluate the distribution for "large" elements of $\boldsymbol{R}_2$. This has been done by generating parametric bootstrap samples to study the distribution of the "largest" elements of the residual, which is detailed in the next section.

3.5 Residual analysis via parametric bootstrap

The bootstrap theory was introduced by Efron (1979). As time passes many authors have continued to develop the approach. For excellent introductions and summaries see for example Efron and Tibshirani (1994) or Davison and Hinkley (1997). Bootstrapping theory has been developed for performing residual analysis within the text of regression analysis, for example, see Wakefied (2013) and Weisberg (2014).

In this thesis, we use the parametric bootstrap approach to approximate the distributions for the "largest" elements of the residual $\boldsymbol{R}_2$ in (3.5) in the GMANOVA-MANOVA model. In many non-standard scenarios, the parametric bootstrap has been used (see, e.g., Cheng 2017). When studying the "largest" elements of $\boldsymbol{R}_2$ the problem is that we study extreme

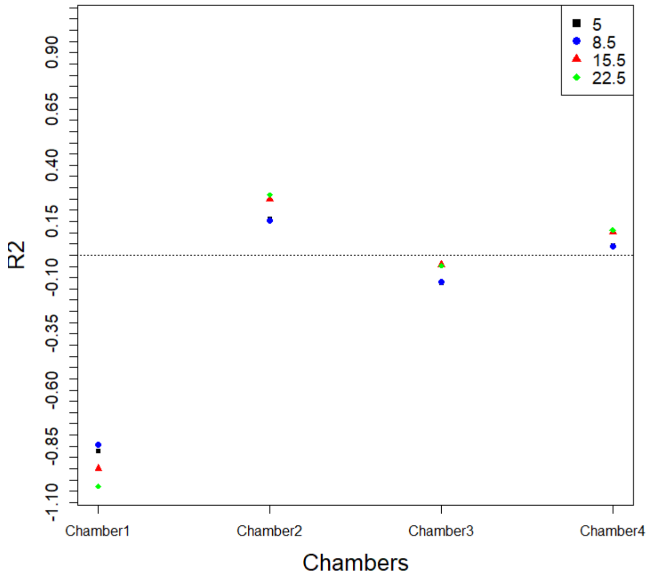

Figure 3.2: *The residuals* $\boldsymbol{R}_2$ *when using the data in Example* 2.2. *The figure shows the residuals per chamber at time points* 5, 8.5, 15.5, *and* 22.5.

values of dependent and non-identical observations which is a complicated situation. In Byukusenge et al. (2022c) article, the parametric bootstrap procedure is used to obtain an approximation of the distribution for the three "largest" values in $\boldsymbol{R}_2$ and define cut off points so that with a high probability a certain value is extreme. The MLEs $\widehat{\boldsymbol{B}}_1$, $\widehat{\boldsymbol{B}}_2$ and $\widehat{\boldsymbol{\Sigma}}$ were presented in (2.16)–(2.18). Then random numbers which are elements of $\boldsymbol{E}^* \sim N_{p,n}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}}, \boldsymbol{I}_n)$ are generated m times (m is large) and for each m,

$$\boldsymbol{X}^* = \boldsymbol{A}\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \widehat{\boldsymbol{B}}_2\boldsymbol{C}_2 + \boldsymbol{E}^*, \tag{3.14}$$

is computed. Thereafter, using Definition 3.1, $\boldsymbol{R}_2^* = (\boldsymbol{I} - \boldsymbol{P}_{A,S})\boldsymbol{X}^*\boldsymbol{P}_{\boldsymbol{Q}_{C_2'}C_1'}$ is derived. To indicate that the process is repeated m times it is written

$$\boldsymbol{R}_{2i}^* = (\boldsymbol{I} - \boldsymbol{P}_{A,S})\boldsymbol{X}^*\boldsymbol{P}_{\boldsymbol{Q}_{C_2'}C_1'}, \quad i \in \{1, \dots, m\}. \tag{3.15}$$

We are interested in "large" values of the residual and the distributions of the three "largest" residuals are of interest.

Definition 3.2. Define the largest by absolute value of $\boldsymbol{R}_2$ by ${}_1R_2$, the second "largest" by ${}_2R_2$ and the third "largest" by ${}_3R_2$, and the largest by absolute value of $\boldsymbol{R}_{2i}^*$ by ${}_1R_{2i}^*$, the second "largest" by ${}_2R_{2i}^*$ and the third "largest" by ${}_3R_{2i}^*$.

The proposed parametric bootstrap approach is described in Algorithm 1. Moreover,

Algorithm 1 Parametric bootstrap procedure

```
1: Set $m$ to be the number of required bootstrap samples.
2: For a given data set compute the parameter estimates $\widehat{\boldsymbol{B}}_1$, $\widehat{\boldsymbol{B}}_2$ and $\widehat{\boldsymbol{\Sigma}}$, presented in (2.16)-(2.18).
3: For $i \in \{1, \dots, m\}$ sample $\boldsymbol{E}^*$ from $N_{p,n}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}}, \boldsymbol{I}_n)$ and compute $\boldsymbol{X}^*$ according to (3.14).
4: Compute the residual components ${}_jR^*_{2i}$, $j \in \{1, 2, 3\}$, $i \in \{1, \dots, m\}$, via (3.15).
```

cut off points for identifying extreme residuals in these statistics are obtained through the quantiles (95%, 99%) of the estimated distribution.

Example 3.2

If we consider again our Example 3.1 we note that all residuals $\boldsymbol{R}_2$ are between -1.028 and 0.26 and we do not know at what level those residuals can deviate from the model fit to be considered as an appropriate model. To get an insight into this feature, we explored the distribution of the three largest elements of $\boldsymbol{R}_2$. Our approach is to investigate the distribution of the three largest elements of the matrix $\boldsymbol{R}_2$ through 10,000 parametric bootstrap samples obtained using Algorithm 1. The complete marginal bootstrap distributions of the "largest" elements in $\boldsymbol{R}_2$ are presented in Figure 3.3. Moreover, in Table 3.1 the 95% and 99% quantiles for the marginal parametric bootstrap estimated distribution of the "largest" elements in $\boldsymbol{R}_2$ are presented. It follows that ${}_1R_2 = 1.028$, ${}_2R_2 = 0.94$ and ${}_3R_2 = 0.87$ are above the thresholds defined via the 95% and 99% percentiles of the bootstrap distributions. This indicates that the linear model which is used in the analysis is not appropriate. One can also observe that all residuals in Chamber 1 are above the thresholds defined via the 95% and 99% percentiles of the bootstrap distributions, this indicates that it should be possible to analyse those specific trees when for example a quadratic growth is assumed to hold.

Table 3.1: *The estimated percentiles* (95% *and* 99%) *for the three largest residual elements* (*in absolute values*) *in the residual* $\boldsymbol{R}_2$ *defined in* (3.15), *are presented. These residuals are based on* 10,000 *parametric bootstrap samples.*

Percentile	${}_1R^*_{2i}$	${}_2R^*_{2i}$	${}_3R^*_{2i}$
99%	0.473	0.470	0.25
95%	0.37	0.36	0.191

3.6 Testing bilinear restrictions in the MANOVA model through residuals

In this section, bilinear restrictions in the MANOVA model are tested. The likelihood ratio (LR) test is constructed which consists of a ratio of determinants of the estimated

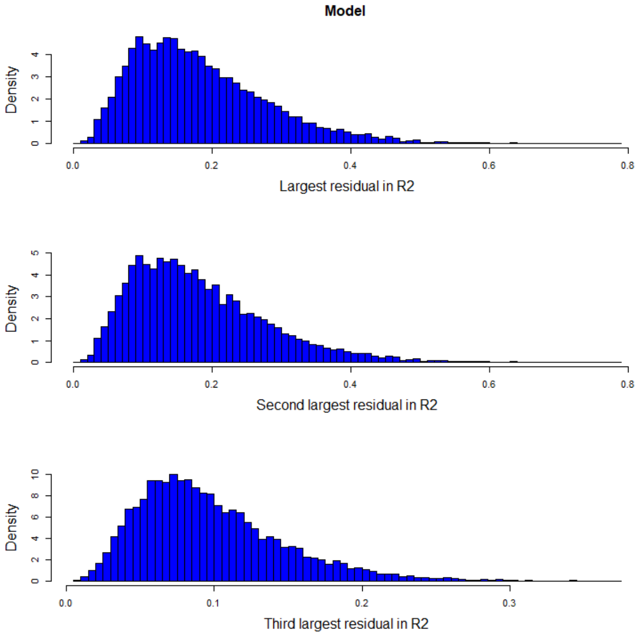

Figure 3.3: *The approximative distribution for the three "largest" elements for* $\boldsymbol{R}_2$ *in* (3.15)*, obtained when* 10,000 *bootstrap samples were generated.*

dispersions matrices. It is shown that the LR can be written in terms of the residuals. The material is based on the results given by Byukusenge et al. (2022a,d). Consider the MANOVA model in Definition 2.1 given by

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{C} + \boldsymbol{E}, \tag{3.16}$$

where $\boldsymbol{X} : p \times n$ is the observation matrix, $\boldsymbol{B} : p \times k$ the unknown parameter matrix, $\boldsymbol{C} : k \times n$ the between design matrix and $\boldsymbol{E} : p \times n$ the random error matrix. Statisticians have considered hypothesis tests in the MANOVA model (3.16), and several test statistics have been proposed. For example, Wilks' lambda test (also known as the likelihood ratio test), Roy's maximum root test, and Lawley-Hotelling's trace test (see Anderson, 2003). In this thesis, a test statistic for testing a bilinear mean structure in (3.16) has been established. A likelihood ratio test is constructed using residuals, i.e., the ratio between the determinants of residuals that come from the MANOVA model with and without restriction. The hypotheses are formulated as:

$$H_0 : \boldsymbol{M}\boldsymbol{B}\boldsymbol{G} = \boldsymbol{0} \quad \text{against} \quad H_A : \boldsymbol{M}\boldsymbol{B}\boldsymbol{G} \neq \boldsymbol{0}, \tag{3.17}$$

where $\boldsymbol{M} : j \times p$ and $\boldsymbol{G} : k \times l$ are known matrices. In this thesis, the testing hypothesis (3.17) under model (3.16) is considered. Byukusenge et al. (2022d) showed that under the null hypothesis, i.e., for $\boldsymbol{M}\boldsymbol{B}\boldsymbol{G} = \boldsymbol{0}$, the model (3.16) is equivalent to a GMANOVA-

MANOVA model of the following form

$$\boldsymbol{X} = (\boldsymbol{M}')^{\circ}\,\boldsymbol{\Theta}_1\boldsymbol{G}'\boldsymbol{C} + \boldsymbol{\Theta}_2\boldsymbol{G}^{\circ\prime}\boldsymbol{C} + \boldsymbol{E}, \tag{3.18}$$

where $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ are unknown new parameter matrices. Therefore, using results from Byukusenge et al. (2022d), the test statistic $\Lambda^{2/n}$ can be written as

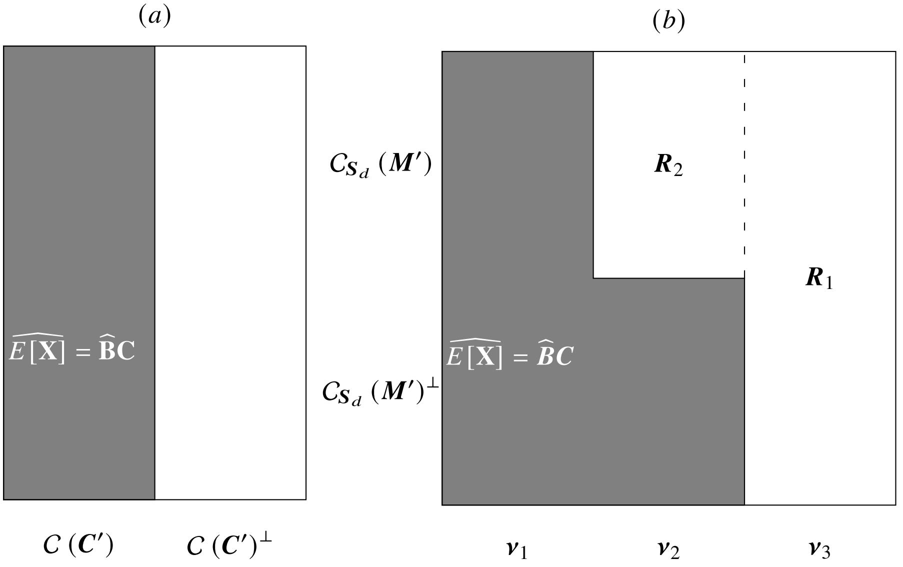

Figure 3.4: *Consider the model in Definition* 2.1 *with and without restrictions. A decomposition is presented of the whole space according to the design and restrictions. In (a), there are no restrictions and the decomposition consists of the subspaces* $\mathcal{C}(\boldsymbol{C}')$ *and* $\mathcal{C}(\boldsymbol{C}')^{\perp}$. *In (b), with the restrictions* $\boldsymbol{MBG} = \mathbf{0}$, *the spaces connected to the between-individuals decomposition are given by* $\boldsymbol{v}_1 = \mathcal{C}(\boldsymbol{C}'\boldsymbol{G}^{\circ})$, $\boldsymbol{v}_2 = \mathcal{C}(\boldsymbol{C}'\boldsymbol{G}^{\circ})^{\perp} \cap \mathcal{C}(\boldsymbol{C}')$ *and* $\boldsymbol{v}_3 = \mathcal{C}(\boldsymbol{C}')^{\perp}$. *The predicted values,* $\widehat{E[\mathbf{X}]}$, *and the residuals* $\boldsymbol{R}_1$ *and* $\boldsymbol{R}_2$ *are shown.*

$$\Lambda^{2/n} = \frac{\left|\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}'})\boldsymbol{X}' + \boldsymbol{P}'_{\boldsymbol{M}',\boldsymbol{S}_d^{-1}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{N}}\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{M}',\boldsymbol{S}_d^{-1}}\right|}{|\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}'})\boldsymbol{X}'|}, \tag{3.19}$$

where

$$\begin{aligned}
\boldsymbol{P}_{\boldsymbol{M}',\boldsymbol{S}_d^{-1}} &= \boldsymbol{M}'(\boldsymbol{M}\boldsymbol{S}_d\boldsymbol{M}')^{-1}\boldsymbol{M}\boldsymbol{S}_d,\\
\boldsymbol{P}_{\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{N}} &= \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{N}(\boldsymbol{N}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{N})^{-}\boldsymbol{N}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C},
\end{aligned}$$

are projection matrices,

$$\boldsymbol{S}_d = \boldsymbol{X}\boldsymbol{Q}_d\left(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{Q}_d\boldsymbol{C}'\boldsymbol{G}}\right)\boldsymbol{Q}_d\boldsymbol{X}', \quad \boldsymbol{Q}_d = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}'\boldsymbol{G}^{\circ}},$$

and $\boldsymbol{N}$ is any matrix such that $C(\boldsymbol{N}) = C(\boldsymbol{C}) \cap C(\boldsymbol{G})$. The projection matrix $\boldsymbol{P}_{\boldsymbol{C}'\boldsymbol{G}^{\circ}}$ is obtained when the model (3.16) with restriction is considered. Using results provided by Byukusenge et al. (2022a,c), the predicted values $\widehat{\boldsymbol{X}}$ obtained for the MANOVA model with restriction, equals

$$\widehat{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'\boldsymbol{G}^{\circ\prime}} + \boldsymbol{P}_{(\boldsymbol{M}')^{\circ},\boldsymbol{S}_d}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}_d\boldsymbol{C}'\boldsymbol{G}}.$$

The MANOVA model with and without restrictions is represented in Figure 3.4. The illustration in Figure 3.4 (a) the decomposition consists of the subspaces $C(\boldsymbol{C}')$ and $C(\boldsymbol{C}')^{\perp}$ while Figure 3.4 (b) is based on the following decompositions

$$C(\boldsymbol{C}') = C(\boldsymbol{C}'\boldsymbol{G}^{\circ}) \boxplus \left\{C(\boldsymbol{C}'\boldsymbol{G}^{\circ})^{\perp} \cap C(\boldsymbol{C}')\right\}.$$

Moreover, using the idea of Byukusenge et al. (2022a) and from Figure 3.4, the total variation under H_0 is given by the square of the residuals, i.e., $n\widehat{\boldsymbol{\Sigma}}_{H_0} = \boldsymbol{R}_1\boldsymbol{R}_1' + \boldsymbol{R}_2\boldsymbol{R}_2'$ and under H_A the total variation equals $n\widehat{\boldsymbol{\Sigma}}_{H_A} = \boldsymbol{R}_1\boldsymbol{R}_1'$. Therefore, the following proposition was obtained.

Proposition 3.1 (Byukusenge et al., 2022d)
Suppose that the MANOVA model (3.16) has been fitted to data and suppose that the hypothesis given in (3.17) is to be tested. The test statistic is given by

$$\Lambda^{2/n} = \left|\boldsymbol{I} + \boldsymbol{R}_2'\left(\boldsymbol{R}_1\boldsymbol{R}_1'\right)^{-1}\boldsymbol{R}_2\right|, \tag{3.20}$$

where $\boldsymbol{R}_1 = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{C}'})$, and $\boldsymbol{R}_2 = \boldsymbol{P}'_{\boldsymbol{M}',\boldsymbol{S}_d^{-1}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{N}}$ are the residuals from both MANOVA and GMANOVA-MANOVA models and $\boldsymbol{N}$ is any matrix such that $C(\boldsymbol{N}) = C(\boldsymbol{C}) \cap C(\boldsymbol{G})$. The hypothesis is rejected when $\Lambda^{2/n}$ is large.

The test is given in Proposition 3.1 which is a function of residuals $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ is more interpretable compared to what is obtained in Roy (1957). Since (3.20) is a function of residuals, one can calculate moments of the residuals, and based on them it is possible to interpret the result of the test. To determine how large the value of the test statistic must be for the hypothesis to be rejected, the critical point has to be calculated. This requires the knowledge of the distribution of $\Lambda^{2/n}$ which is difficult to obtain. An approximation of the distribution of the test statistic obtained is given in the next theorem. Since the test statistic in Proposition 3.1 depends on the two residuals $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, Byukusenge et al. (2022d) proposed to investigate the expression of the test statistic by calculating different moments of the residuals. It was shown that under the alternative hypothesis H_A, the expectations $E\left(\boldsymbol{R}_2\right)$, $E\left(\boldsymbol{R}_2\boldsymbol{R}_2'\right)$ and $E\left(\boldsymbol{R}_2'\left(\boldsymbol{R}_1\boldsymbol{R}_1'\right)^{-1}\boldsymbol{R}_2\right)$ are respectively given by

$$E\left(\boldsymbol{R}_2\right) = \boldsymbol{\Sigma}\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'\right)^{-}\boldsymbol{M}\boldsymbol{B}\boldsymbol{N}\Big(\boldsymbol{N}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-}\boldsymbol{N}\Big)^{-}\boldsymbol{N}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-}\boldsymbol{C}, \tag{3.21}$$

$$\begin{aligned}E\left(\boldsymbol{R}_2\boldsymbol{R}_2'\right) &= r(\boldsymbol{N})\Big(c_1\boldsymbol{\Sigma} + (1-c_1)\,\boldsymbol{\Sigma}\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'\right)^{-}\boldsymbol{M}\boldsymbol{\Sigma}\Big)\\ &+ \boldsymbol{\Sigma}\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'\right)^{-}\boldsymbol{M}\boldsymbol{B}\boldsymbol{N}\Big(\boldsymbol{N}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-}\boldsymbol{N}\Big)^{-}\boldsymbol{N}'\boldsymbol{B}'\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'\right)^{-}\boldsymbol{M}\boldsymbol{\Sigma}\\ &+ c_1 tr\Big(\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'\right)^{-1}\boldsymbol{M}\boldsymbol{B}\boldsymbol{N}\Big(\boldsymbol{N}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-}\boldsymbol{N}\Big)^{-}\boldsymbol{N}'\boldsymbol{B}'\Big)\left(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'\right)^{-}\boldsymbol{M}\boldsymbol{\Sigma}\right),\end{aligned} \tag{3.22}$$

$$\begin{aligned}E\left(\boldsymbol{R}_2'\left(\boldsymbol{R}_1\boldsymbol{R}_1'\right)^{-}\boldsymbol{R}_2\right) &= c_1 r(\boldsymbol{M})\boldsymbol{P}_{\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{N}} + \boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-}\boldsymbol{N}\\ &\times \Big(\boldsymbol{N}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-}\boldsymbol{N}\Big)^{-}\boldsymbol{N}'\boldsymbol{B}'\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'\right)^{-}\boldsymbol{M}\boldsymbol{B}\boldsymbol{N}\Big(\boldsymbol{N}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-}\boldsymbol{N}\Big)^{-}\boldsymbol{N}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-}\boldsymbol{C},\end{aligned} \tag{3.23}$$

where $c_1 = \frac{p-r(\boldsymbol{M})}{n-r(\boldsymbol{N})-p+r(\boldsymbol{M})-1}$. Under the null hypothesis H_0, the expression $\boldsymbol{M}\boldsymbol{B}\boldsymbol{N} = \boldsymbol{0}$ reduces the above expectations into the following

$$E\left(\boldsymbol{R}_2\right) = \boldsymbol{0}, \tag{3.24}$$

$$E\left(\boldsymbol{R}_2\boldsymbol{R}_2'\right) = r(\boldsymbol{N})\Big(c_1\boldsymbol{\Sigma} + (1-c_1)\,\boldsymbol{\Sigma}\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'\right)^{-}\boldsymbol{M}\boldsymbol{\Sigma}\Big), \tag{3.25}$$

$$E\left(\boldsymbol{R}_2'\left(\boldsymbol{R}_1\boldsymbol{R}_1'\right)^{-}\boldsymbol{R}_2\right) = c_1 r(\boldsymbol{M})\boldsymbol{P}_{\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{N}}. \tag{3.26}$$

Example 3.3: (Growth of Sitka Spruce data from Diggle et al. 2002)

Consider the data presented in Table 1 and 2 in the Appendix A. Let $\boldsymbol{X} = \boldsymbol{B}\boldsymbol{C} + \boldsymbol{E}$, where

$$\boldsymbol{C} = \left(\boldsymbol{1}_{53}' \otimes \begin{pmatrix}1\\0\end{pmatrix} : \;\; \boldsymbol{1}_{25}' \otimes \begin{pmatrix}0\\1\end{pmatrix}\right), \quad \boldsymbol{B} = \begin{pmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\\ \beta_{31} & \beta_{32}\\ \beta_{41} & \beta_{42}\end{pmatrix},$$

$\boldsymbol{C}$ indicates that the data consists of two different groups of independent observations. The estimated values of $\boldsymbol{B}$ equals

$$\widehat{\boldsymbol{B}} = \begin{pmatrix}4.05 & 4.15\\ 5.30 & 5.64\\ 5.50 & 5.83\\ 6.14 & 6.48\end{pmatrix}.$$

First

$$H_0 : \boldsymbol{M}\boldsymbol{B}\boldsymbol{G} = \boldsymbol{0} \quad \text{versus} \quad H_A : \boldsymbol{M}\boldsymbol{B}\boldsymbol{G} \neq \boldsymbol{0},$$

is tested where

$$\boldsymbol{M} = \begin{pmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}, \quad \boldsymbol{G} = \begin{pmatrix}-1 & 1\end{pmatrix}'.$$

If H_0 is true the test is stated as $\boldsymbol{MBG} = \mathbf{0}$, i.e., $\beta_{21} = \beta_{22}$, $\beta_{31} = \beta_{32}$ and $\beta_{41} = \beta_{42}$. According to Proposition 3.1, the LR test statistic equals $\Lambda^{2/n} = \left|\boldsymbol{I} + \boldsymbol{R}_2' \left(\boldsymbol{R}_1 \boldsymbol{R}_1'\right)^{-1} \boldsymbol{R}_2\right|$. Let Λ_0 be the observed value of Λ, and let

$$x = \frac{2}{n}\left(f - \frac{1}{2}\left(j - m + 1\right)\right) \ln \Lambda_0,$$

where $f = n - r\left(\boldsymbol{C}\right)$, $m = dim\left\{C\left(\boldsymbol{N}\right)\right\}$. From von Rosen (2018), the LR test approximately, at the level of significance α, rejects the hypothesis when

$$\begin{aligned} &P\left\{\chi^2_{jm} \geq x\right\} + a\left(1 - a\right)\left(P\left\{\chi^2_{jm+4} \geq x\right\} - P\left\{\chi^2_{jm} \geq x\right\}\right) \\ &+ b\Big(P\left\{\chi^2_{jm+8} \geq x\right\} - P\left\{\chi^2_{jm} \geq x\right\}\Big) \leq \alpha, \end{aligned}$$

where

$$\begin{aligned} a &= \frac{jm(j^2 + m^2 - 5)}{48\left(f - \frac{1}{2}\left(j - m + 1\right)\right)^2}, \\ b &= \frac{1}{2}a^2 + \frac{jm\left(3j^4 + 3m^4 + 10j^2m^2 - 50\left(j^2 + m^2\right) + 159\right)}{1920\left(f - \frac{1}{2}\left(j - m + 1\right)\right)^4}. \end{aligned}$$

In this example, $\Lambda^{2/78} = 1.116$, $j = 3$, $m = dim(C(\boldsymbol{N})) = 1$, and $f = n - r(\boldsymbol{C}) = 76$. The LR test does not reject H_0 at significance level 5% since

$$\begin{aligned} &P\left\{\chi^2_3 \geq x\right\} + a\left(1 - a\right)\left(P\left\{\chi^2_7 \geq x\right\} - P\left\{\chi^2_3 \geq x\right\}\right) \\ &+ b\Big(P\left\{\chi^2_{11} \geq x\right\} - P\left\{\chi^2_3 \geq x\right\}\Big) \approxeq P\left\{\chi^2_3 \geq x\right\} = 0.13 > \alpha = 0.05, \end{aligned}$$

where $x = 5.50$, $a \approxeq 0$ and $b \approxeq 0$ for details see Byukusenge et al. (2022d).

i) Under the null hypothesis H_0, the moments of the residuals equal

$$\begin{aligned} E\left(\boldsymbol{R}_2\right) &= \mathbf{0}, \\ \widehat{E\left(\boldsymbol{R}_2 \boldsymbol{R}_2'\right)} &= \begin{pmatrix} 0.32 & 0.35 & 0.36 & 0.32 \\ 0.35 & 0.40 & 0.40 & 0.366 \\ 0.36 & 0.40 & 0.44 & 0.40 \\ 0.32 & 0.36 & 0.40 & 0.39 \end{pmatrix}, \\ E\left(\boldsymbol{R}_2' \left(\boldsymbol{R}_1 \boldsymbol{R}_1'\right)^{-} \boldsymbol{R}_2\right) &= \begin{pmatrix} 2.41 \times 10^{-4} \otimes \boldsymbol{J}_{53\times53} & -5.12 \times 10^{-4} \otimes \boldsymbol{J}_{53\times25} \\ -5.12 \times 10^{-4} \otimes \boldsymbol{J}_{25\times53} & 10.87 \times 10^{-4} \otimes \boldsymbol{J}_{25\times25} \end{pmatrix}. \end{aligned}$$

ii) Under H_A the estimated moments of the residuals equal

$$E\widehat{(\boldsymbol{R}_2)} = \left(\mathbf{1}'_{53} \otimes \begin{pmatrix} -0.09 \\ -0.10 \\ -0.10 \\ -0.10 \end{pmatrix} : \quad \mathbf{1}'_{25} \otimes \begin{pmatrix} 0.19 \\ 0.22 \\ 0.22 \\ 0.23 \end{pmatrix}\right),$$

$$E\widehat{\left(\boldsymbol{R}_2\boldsymbol{R}_2'\right)} = \begin{pmatrix} 1.76 & 2.02 & 2.00 & 2.01 \\ 2.02 & 2.33 & 2.30 & 2.32 \\ 2.00 & 2.30 & 2.30 & 2.32 \\ 2.01 & 2.32 & 2.32 & 2.38 \end{pmatrix},$$

$$E\left(\boldsymbol{R}_2' \widehat{\left(\boldsymbol{R}_1\boldsymbol{R}_1'\right)}^{-} \boldsymbol{R}_2\right) = \begin{pmatrix} 0.03 \otimes \boldsymbol{J}_{53\times53} & -0.07 \otimes \boldsymbol{J}_{53\times25} \\ -0.07 \otimes \boldsymbol{J}_{25\times53} & 0.16 \otimes \boldsymbol{J}_{25\times25} \end{pmatrix}.$$

Looking at the results obtained in the above moments, one can see that there is not a significant difference between the estimated moments of the residuals obtained under H_A and H_0 when H_0 is not rejected. However, when H_0 is rejected there is a significant difference among estimated moments obtained in both cases. This can be seen in the next case. Let consider some elements of $\boldsymbol{X}$ to be $x_{2,65} = x_{2,70} = 12$ and $x_{3,65} = x_{3,70} = 12$. Let

$$\boldsymbol{M} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{G} = \begin{pmatrix} -1 & 1 \end{pmatrix}'.$$

The estimated values of $\boldsymbol{B}$ under H_A equals

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} 4.05 & 4.15 \\ 5.30 & 6.15 \\ 5.50 & 6.32 \\ 6.14 & 6.48 \end{pmatrix},$$

the test statistic $\Lambda^{2/78} = 1.13$, $j = 3$, $m = dim(C(\boldsymbol{N})) = 1$, and $f = n - r(\boldsymbol{C}) = 76$. The LR test rejects H_0 at significance level 5% since

$$P\left\{\chi_3^2 \geq x\right\} + a\,(1-a)\left(P\left\{\chi_7^2 \geq x\right\} - P\left\{\chi_3^2 \geq x\right\}\right)$$
$$+\, b\left(P\left\{\chi_{11}^2 \geq x\right\} - P\left\{\chi_3^2 \geq x\right\}\right) \cong P\left\{\chi_3^2 \geq x\right\} = 0.02 < \alpha = 0.05,$$

where $x = 9.10$. The expected moments of the residuals are respectively given by

a) under H_0, the moments of the residuals equal

$$E\left(\boldsymbol{R}_2\right) = \mathbf{0}$$

$$E\widehat{\left(\boldsymbol{R}_2\boldsymbol{R}_2'\right)} = \begin{pmatrix} 0.28 & 0.34 & 0.35 & 0.32 \\ 0.34 & 1.34 & 1.32 & 0.34 \\ 0.35 & 1.32 & 1.33 & 0.38 \\ 0.32 & 0.34 & 0.38 & 0.39 \end{pmatrix},$$

$$E\left(\boldsymbol{R}_2' \left(\boldsymbol{R}_1\boldsymbol{R}_1'\right)^{-1} \boldsymbol{R}_2\right) = \begin{pmatrix} 2.41 \times 10^{-3} \otimes \boldsymbol{J}_{53\times53} & -5.12 \times 10^{-3} \otimes \boldsymbol{J}_{53\times25} \\ -5.12 \times 10^{-3} \otimes \boldsymbol{J}_{25\times53} & 1.08 \times 10^{-3} \otimes \boldsymbol{J}_{25\times25} \end{pmatrix}.$$

b) Under H_A the estimated moments of the residuals equal

$$\widehat{E(\boldsymbol{R}_2)} = \left(\mathbf{1}'_{53} \otimes \begin{pmatrix} -0.10 \\ -0.26 \\ -0.26 \\ -0.10 \end{pmatrix} : \quad \mathbf{1}'_{25} \otimes \begin{pmatrix} 0.22 \\ 0.57 \\ 0.55 \\ 0.23 \end{pmatrix} \right),$$

$$\widehat{E(\boldsymbol{R}_2\boldsymbol{R}'_2)} = \begin{pmatrix} 2.11 & 5.01 & 4.91 & 2.22 \\ 5.01 & 13.37 & 13.07 & 5.23 \\ 4.91 & 13.07 & 12.80 & 5.15 \\ 2.22 & 5.23 & 5.15 & 2.38 \end{pmatrix},$$

$$E\left(\boldsymbol{R}'_2 \left(\widehat{\boldsymbol{R}_1\boldsymbol{R}'_1}\right)^{-1} \boldsymbol{R}_2\right) = \begin{pmatrix} 0.06 \otimes \boldsymbol{J}_{53\times 53} & -0.13 \otimes \boldsymbol{J}_{53\times 25} \\ -0.13 \otimes \boldsymbol{J}_{25\times 53} & 0.27 \otimes \boldsymbol{J}_{25\times 25} \end{pmatrix}.$$

Using this example, one can see that when the null hypothesis is not rejected, there is not a big difference among the estimated moments of the residuals under H_0 and H_A. However, when the null hypothesis is rejected there is a significant difference between the estimated moments of the residuals, and under H_A the moments of the residuals are larger. Therefore, it is useful to consider those moments in any statistical study, because they provide a good interpretation of how the data behaves in front of the given models.

4

Conclusions

4.1 Conclusion

In statistical theory, residual is an important quantity for studying model assumptions. Residuals which are the difference between the observed values and the predicted values can show how far an object deviates from the estimated mean. In this thesis residuals in the GMANOVA-MANOVA model are considered and based on space decomposition two residuals are established. One residual is the difference between the observations and the group mean corrected for the covariate. Hence, it gives information about the between individual assumptions in a given group. It can be used to detect observations that deviate from the others without taking into account any model assumption. This residual can be divided into two parts which are related to the within-individual model assumption. The second residual is the observed group mean, corrected for the covariate minus the fitted mean model. Thus, it describes how well the estimated mean structure fits the observed group mean, i.e., it relates to the within-individual structure and gives information about the mean model assumptions.

Properties such as expectation, dispersion, and relevant covariances are derived in Byukusenge et al. (2022a). Those properties are can be used to detect if some observations deviate from the model assumptions. Byukusenge et al. (2022a) provides an example of how residuals are affected when observations fail to follow model assumptions. It was demonstrated that residual $\boldsymbol{R}_2$ plays an interesting role in model validation. A special but interesting example can be found in the paper of Byukusenge et al. (2022b) where when some elements of $\boldsymbol{R}_2$ became exactly $\boldsymbol{0}$ it was not possible to estimate a linear trend over time. In this case, the model assumptions have to be examined. Thus, a matrix $\boldsymbol{Q}_{\boldsymbol{C}_2'}\boldsymbol{C}_1$ with none zero elements were proposed for having residuals $\boldsymbol{R}_2$ with none zero elements.

Byukusenge et al. (2022c) provided a parametric bootstrap algorithm that gives simulated distributions of the extreme elements of the residuals. Through simulated distribution, cut off points are used for identifying outliers in the given data set.

The LR test to test bilinear restrictions on the parameter space for the MANOVA model is formulated see (Byukusenge et al. (2022d)). Since the test statistic is function of residuals, it is interpretable compared to the existing LR test and it is possible to calculate the moments of residuals. Looking at the estimated moments of the residuals, they are close to each others when the hypothesis is not rejected and with a significant difference in otherwise. The estimated moments can be used to explain how the model behaves when a data set is presented.

4.2 Future research

In this thesis, different aspects of the residuals in the GMANOVA-MANOVA model are explored and a list of potential future research areas in this field is provided below.

- In this thesis, residuals for the GMANOVA-MANOVA model were studied and some properties of the residuals are given and explained. For extension of this work, we would like to see how the residuals should behave when random effects are considered in this model.

- For identifying outliers in the model we proposed parametric bootstrapping in Paper C where we set a cut off point as a critical condition. This algorithm can be developed to see if it can help to identify deviating observations in a model.

- The likelihood ratio test can be constructed when a bilinear restriction on the parameter space for the MANOVA model is considered. From the LR test one can work more on the estimated moments of the residual to explain how the model is behaving for a given data set.

Bibliography

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, Hoboken, third edition.

Anscombe, F. and Guttman, I. (1960). Rejection of outliers. *Technometrics*, 2:123–147.

Anscombe, F. J. and John, W. T. (1963). The examination and analysis of residuals. *Technometrics*, 5:141–160.

Bartlett, M. (1934). On the theory of statistical regression. *Proceedings of the Royal Society of Edinburgh*, 53:260–283.

Behnken, D. W. and Draper, N. R. (1972). Residuals and their variance patterns. *Technometrics*, 14:102–111.

Byukusenge, B., von Rosen, D., and Singull, M. (2022a). On an important residual in the GMANOVA-MANOVA model. *Journal of Statistical Theory and Practice*, 16:1–20.

Byukusenge, B., von Rosen, D., and Singull, M. (2022b). On residual analysis in the GMANOVA-MANOVA model. In Balakrishnan, N., Gil, M. Á., Martín, N., Morales, D., and Pardo, M. d. C., editors, *Trends in Mathematical, Information and Data Sciences*: *A Tribute to Leandro Pardo*, pages 287–305. Springer International Publishing, Cham.

Byukusenge, B., von Rosen, D., and Singull, M. (2022c). On the identification of extreme elements in a residual for the GMANOVA-MANOVA model. *Accepted for publication in Innovations in multivariate statistical modelling*: *navigating theoretical and multidisciplinary domains, Springer Emerging Topics in Statistics and Biostatistics*.

Byukusenge, B., von Rosen, D., and Singull, M. (2022d). Testing bilinear restriction in the GMANOVA-MAOVA model. *Report*.

Chatterjee, S. and Hadi, A. S. (1988). Impact of simultaneous omission of a variable and an observation on a linear regression equation. *Computational Statistics and Data Analysis*, 6:129 – 144.

Cheng, R. (2017). *Non-Standard Parametric Statistical Inference*. Oxford University Press, Oxford.

Chinchilli, V. M. and Elswick, R. K. (1985). A mixture of the MANOVA and GMANOVA models. *Communications in Statistics-Theory and Methods*, 14:3075–3089.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Monographs on statistics and applied probability. Chapman and Hall, London.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Dempster, A. (1969). *Elements of Continuous Multivariate Analysis*. Addison-Wesley, New York.

Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford university press, second edition.

Dwyer, P. (1941). The skewness of the residuals in linear regression theory. *Annals of Mathematical Statistics*, 12:104–110.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7:1–26.

Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Filipiak, K. and von Rosen, D. (2012). On MLEs in an extended multivariate linear growth curve model. *Metrika*, 5:1069–1092.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10:507–521.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263.

Galton, F. (1889a). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, 45:135–145.

Galton, F. (1889b). *Natural Inheritance*. Macmillan, London.

Gleser, L. and Olkin, I. (1970). Linear models in multivariate analysis. *Essays in probability and statistics*, pages 267–292.

Hamid, J. S. and von Rosen, D. (2006). Residuals in the extended growth curve model. *Scandinavian Journal of Statistics*, 33:121–138.

Kariya, T. (1985). *Testing in the Multivariate General Linear Model*. Kinokuniya, New York.

Kendall, M. G. (1957). *A Course in Multivariate Analysis. Griffin's Statistical Monographs and Courses*. Hafner Publishing Company, New York, second edition.

Khatri, C. G. (1966). A note on a manova model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics*, 86:75–86.

Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*, volume 579. Springer Press, Holboken.

Lund, R. (1975). Tables for an approximate test for outliers in linear models. *Technometrics*, 17:473–476.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis. Probability and Mathematical Statistics*: *A Series of Monographs and Textbooks*. Academic Press, New York.

Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 20A:175–240.

Neyman, J. and Pearson, E. S. (1931). On the problem of k samples. *Bulletin de Lacademie Polonaise des Sciences-Serie des Sciences de la Terre*, A:460–481.

Pan, J., Fang, K., and Fang, K. T. (2002). *Growth Curve Models and Statistical Diagnostics*. Springer Series in Statistics.

Pearson, K. (1896). Mathematical contributions to the theory of evolution.III. regression, heredity and panmixia. *Philosophical Transactions of Royal Society A*: *Mathematical, Physical and Engineering Sciences*, 187:253–318.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51:313–326.

Prescott, P. (1975). An approximate test for outliers in linear models. *Technometrics*, 17:129–132.

Rao, C. (1948). Tests of significance in multivariate analysis. *Biometrika*, 35:58–79.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the growth curve problems. *Biometrika*, 52:447–458.

Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis. *Multivariate analysis*, pages 87–103.

Rao, C. R. (1987). Prediction of future observations in polynomial growth curve models. *Statistical Science*, 2:434–447.

Rao, C. R., Helge, T., Shalabh, and Christian, H. (2008). *Linear Models and Generalizations*. Springer, Berlin, third edition.

Rencher, A. and Christensen, W. (2012). *Methods of Multivariate Analysis. Third edition*. Wiley Series in Probability and Statistics. Wiley, Hoboken.

Roy, S. (1957). *Some Aspects of Multivariate Analysis*. Indian Statistical Series. Wiley, New York.

Seber, G. A. F. (1984). *Multivariate Observations*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Srivastava, M. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*. North-Holland, New York.

Srivastava, M. S. and Carter, E. M. (1983). *Applied Multivariate Statistics*. North-Holland, New York.

Srivastava, M. S. and von Rosen, D. (1999). Growth curve models. In *Multivariate analysis, design of experiments, and survey sampling*, A series of textbooks and monographs, pages 571–602. New York:Dekker.

Stefansky, W. (1972). Rejecting outliers in factorial designs. *Technometrics*, 14:105–111.

Stigler, M. S. (1986). *The History of Statistics*: *the Measurement of Uncertainty before* 1900. Belknap Press of Havard University Press.

Tietjen, G., Moore, R. H., and Beckman, R. (1973). Testing for a single outlier in simple linear regression. *Technometrics*, 15:717–721.

Verbyla, A. P. and Venables, W. N. (1988). An extension of the growth curve model. *Biometrika*, 75:129–138.

von Rosen, D. (1989). Maximum likelihood estimators in multivariate linear normal models. *Journal of Multivariate Analysis*, 31:187–200.

von Rosen, D. (1991). The growth curve model: A review. *Communication in Statistics. Theory-Methods*, 20:2791–2822.

von Rosen, D. (1995). Residuals in the growth curve model. *Annals of the Institute of Statistical Mathematics*, 47:129–136.

von Rosen, D. (2018). *Bilinear Regression Analysis*: *An Introduction*. Lecture Notes Statistics 220. Springer-Verlag, New York.

Wakefied, J. (2013). *Bayesian and Frequentist Regression Methods*. Springer Series in Statistics. Springer, New York.

Weisberg, S. (2014). *Applied Linear Regression*. Wiley series in Probability and Statistics. Wiley, Hoboken, fourth edition.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, 31:218.

Wishart, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A:55–57.

Woolson, R. and Leeper, J. D. (1980). Growth curve analysis of complete and incomplete longitudinal data. *Communications in Statistics - Theory and Methods*, 9:1491–1513.

Zyskind, G. (1963). A note on residual analysis. *Journal of the American Statistical Association*, 58:1125–1132.

Appendix A

Appendix

Table 1: *Repeated measurements for trees in Group* 1, *i.e., the log size for sitka spruce trees grown in ozone enriched environments, together with the covariate ’chambers’* (*see Diggle et al.,* 2002).

	Time points					
Id	5	8.5	15.5	22.5	Group	Chambers
1	4.51	6.15	6.16	7.04	1	1
2	4.24	4.96	5.20	5.85	1	1
3	3.98	5.03	5.87	6.61	1	1
4	4.36	5.36	5.53	6.19	1	1
5	4.34	6.28	6.50	7.16	1	1
6	4.59	6.00	6.33	7.05	1	1
7	4.41	5.33	6.13	6.86	1	1
8	4.24	5.48	5.61	6.46	1	1
9	4.82	6.24	6.48	7.28	1	1
10	3.84	4.80	4.94	5.60	1	1
11	4.07	5.10	5.26	5.89	1	1
12	4.28	5.65	5.76	6.41	1	1
13	4.47	5.74	5.99	6.58	1	1
14	4.46	5.46	5.47	6.12	1	1
15	4.6	4.59	4.65	5.50	1	1
16	3.73	4.93	5.24	5.83	1	1
17	4.67	5.49	6.44	7.11	1	1
18	2.96	4.30	4.15	5.07	1	1
19	3.24	4.64	4.63	5.20	1	1
20	4.36	5.45	5.44	5.96	1	1
21	4.04	5.25	5.25	5.89	1	1
22	3.53	5.18	5.64	5.99	1	1
23	4.22	5.58	5.76	6.55	1	1
24	2.79	3.55	3.61	4.39	1	1
25	3.30	5.40	5.46	6.28	1	1
26	3.34	4.86	4.93	5.74	1	1
27	3.76	5.32	5.65	5.91	1	2
28	4.49	5.56	5.73	6.13	1	2
29	4.88	6.17	6.32	6.95	1	2
30	4.88	5.94	6.09	6.74	1	2
31	3.80	5.05	5.06	5.49	1	2
32	4.46	5.49	5.68	6.30	1	2
33	4.06	5.27	5.48	6.13	1	2
34	5.16	6.21	6.37	7.14	1	2
35	3.81	4.60	4.74	5.23	1	2
36	5.09	6.49	6.72	6.97	1	2
37	4.13	5.72	6.06	6.64	1	2
38	4.85	6.13	6.22	6.80	1	2
39	4.11	5.43	5.80	6.44	1	2
40	4.95	6.48	6.61	6.63	1	2
41	4.36	5.47	5.48	6.01	1	2
42	4.05	5.60	5.79	6.42	1	2
43	3.76	5.25	5.41	6.24	1	2
44	2.84	4.21	4.30	4.64	1	2
45	4.33	5.61	5.85	6.55	1	2
46	3.99	5.30	5.69	6.25	1	2
47	3.50	4.85	5.01	5.86	1	2
48	3.31	4.54	4.72	5.35	1	2
49	3.03	4.58	4.47	5.12	1	2
50	3.27	4.89	5.08	5.94	1	2
51	3.56	5.28	5.50	6.14	1	2
52	3.39	4.15	4.49	5.28	1	2
53	3.72	5.02	5.16	5.67	1	3

Table 2: *Repeated measurements for trees in Group* 2*, i.e., the log size for sitka spruce trees grown in normal environments, together with the covariate ’chambers’* (*see Diggle et al.,* 2002).

	Time points					
Id	5	8.5	15.5	22.5	Group	Chambers
54	4.53	5.42	5.71	6.43	2	3
55	4.97	6.45	6.61	7.35	2	3
56	4.37	5.40	5.57	6.29	2	3
57	4.58	5.93	6.14	6.82	2	3
58	4.00	5.87	6.02	6.65	2	3
59	4.73	6.01	6.26	6.92	2	3
60	5.15	6.61	6.82	7.56	2	3
61	4.10	5.48	5.68	6.18	2	3
62	3.22	5.11	5.28	6.02	2	3
63	2.23	3.52	3.89	4.68	2	3
64	3.65	5.44	5.70	6.44	2	3
65	3.40	5.14	5.34	5.95	2	3
66	5.16	6.21	6.37	6.84	2	3
67	4.04	5.87	5.96	6.59	2	4
68	4.32	5.97	6.11	6.63	2	4
69	4.56	5.89	6.16	6.8	2	4
70	4.90	6.25	6.39	6.88	2	4
71	4.83	6.04	6.21	6.66	2	4
72	5.46	6.63	6.73	6.60	2	4
73	4.17	5.56	5.75	6.63	2	4
74	3.35	5.44	5.79	6.42	2	4
75	3.33	5.17	5.4	6.15	2	4
76	3.41	4.54	4.52	5.35	2	4
77	4.50	6.16	6.33	6.94	2	4
78	2.99	5.06	5.23	6.34	2	4

Part I

PAPERS

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

https://doi.org/10.3384/9789179292850

FACULTY OF SCIENCE AND ENGINEERING

Linköping Studies in Science and Technology, Dissertations No. 2221, 2022
Department of Mathematics

Linköping University
SE-581 83 Linköping, Sweden

www.liu.se

LINKÖPING
UNIVERSITY