The GMANOVA Model
Modelling Pigmentation

Master Thesis

Lasse Lykke Nielsen

Aalborg University
Department of Mathematics

Department of Mathematics
Aalborg University
http://www.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
The GMANOVA Model
Modelling Pigmentation

Theme:
Multivariate Linear Models

Period:
Spring 2019

Participants:
Lasse Lykke Nielsen

Supervisor:
Poul Svante Eriksen

Copies: 2

Page Numbers: 54

Date of Completion:
May 27, 2019

Abstract:

The purpose of this project is to investigate the pigmentation of skin by the use of genetic markers. In order to do this a multivariate linear model is used, since there are repeated observations for each person on 3 different positions of the body. Of these multivariate linear models, the GMANOVA is chosen. This project will only focus on predicting pigmentation.
First the theory regarding GMANOVA is introduced and then the GMANOVA is implemented into *R*, and hereafter it is used to analyse the pigmentation using only the genetic markers. Then the results are discussed in the last chapter.
This project found four genetic markers as the most important for predicting pigmentation; *rs12913832*, *rs1408799*, *rs1800407*, *rs2470102* for a p-value of 0.05.

Dansk Referat

Dette speciale omhandler pigmentering hos 458 forskellige brasilianere. Hos disse brasilianere var målt deres pigmentering samt hvilke mutationer de har fået på forskellige pladser i deres DNA. Denne mutation kaldes for en Single Nucleotide Polymorphism(SNP), hvis mutationen ses i mere end 1% af en befolkning. I DNA findes base par, som er hydrogen bundne nucleobaser, som har koderne A,C,G,T, som beskriver de forskellige stoffer. De findes som kombinationer af 2 i dette speciale, hvor hver SNP i dette projekt kun anvender to af de fire nucleobaser. Der findes i dette speciale 23 forskellige pladser, hvor denne mutation er tænkt at være med til at bestemme pigmenteringen hos folk.

Hver person har fået målt deres pigmentering 3 gange på 3 forskellige steder. Hvordan, disse målinger fordeler sig, kan ses i Kapitel 1.

For at kunne undersøge denne sammenhæng mellem SNP'er og pigmentering er der blevet brugt Generalized Multivariate ANalysis Of VAriance(GMANOVA), som håndterer sammenhænge mellem personer, samt sammenhængene mellem de forskellige steder. Denne model er valgt på baggrund af, at der flere målinger for hver person, som skal tages højde for, samtidig med at der også er flere målinger for hvert sted. Dette kan GMANOVA modellen håndtere.

Der er lagt stor vægt på teorien bag GMANOVA modellen, da denne er teoretisk tung og mange resultater ligger bag estimatoren af parameter matricen. Først beskrives den almindelige GMANOVA, hvorefter den udvidede GMANOVA introduceres. Denne gør det muligt at skulle have mere end blot gruppering efter individ og placering af målingen. Begge disse metoder bliver beskrevet i detaljer og flere fordelingsresultater bliver udledt.

I analysen bliver der lagt vægt på, hvad resultaterne betyder for det enkelte SNP samt hvordan man ser de forskellige base par i forhold til hinanden i forhold til pigmenteringen. Det ender med at være modellen, hvor man ikke har lavet nogle forudsætninger omkring data, som er den bedste, idet den har den mindste fejl i modelleringen af data. Der bliver også lavet en udskilning af forskellige SNP'er, hvor de SNP'er, som er signifikante i forhold til at skulle modellere pigmenteringen, kan ses i tabel 4.2 for to forskellige p-værdi niveauer.

Contents

Preface

The thesis is addressed to other math students or students who already have a basic knowledge of statistics, model construction and probability theory. The project will reference to sources by [*number*] where the number corresponds to the number in bibliography. Equations are referenced by (*chapter.number*). Chapters, sections, definitions etc. are referred by the chapter and the number, e.g. Definition 1.1.

This thesis is written by one student of the mathematics education at Aalborg University. The theme is multivariate linear models. The data is regarding pigmentation of 458 Brazilians measured at three different spots. A big thanks should be given to supervisor Poul Svante Eriksen for his great contribution in the completion of this thesis.

Aalborg University, May 27, 2019

Lasse Lykke Nielsen
<lasnie13@student.aau.dk>

Introduction

Introduction

Genetic variation is the basis of human diversity and plays an important role in human diseases. Methods to screen and map genetic variability have, for more than two decades, been based on restriction fragment length polymorphism and microsatellite markers. More recent efforts have focused on the most common type of human genetic variation, single-nucleotide polymorphisms (SNPs). A position is referred to as a SNP when it exists in at least two variants with a frequency of more than 1% for the least common alternative. SNPs are distributed across the human genome by an approximate average of 1% SNP per 1000 base pairs. Base pairs are two nucleotides[1] which are bound to each other by hydrogen bonds. These base pairs are the basis of the DNA base helix. As for microsatellite markers, SNPs can be used in linkage studies for identifying disease genes, in clinical genetic testing and in forensics. The properties that make SNP analysis preferable compared to microsatellites are that SNPs are more prevalent than microsatellites and that many SNPs are located within the genes, directly affecting the gene product (protein). As the number of identified SNPs increases, there will be an increasing demand for efficient methods to type and assess the biological impact of this kind of genetic variation. This thesis will focus on the biological impact of these genetic variations.[1]

In this thesis the Generalised Multivariate ANalysis Of VAriance(GMANOVA) will be introduced and used to analyse 23 different SNPs of their ability to predict pigmentation alone. These 23 SNPs are thought to have an influence on the pigmentation. The subjects are from Brazil, because Brazilians are thought to be migrated from different parts of the world hence the genetic variation is great. This makes it more viable to analyse as there is some variations in the skin tones, which makes the model more robust.

This thesis builds upon projects made at the 8th and 9th semester of the same author at Aalborg University. In these the analysis was made with machine learning algorithms and a mixed model approach to the data. Some conclusions were drawn in these projects, which will be used to compare the conclusions of this analysis with.

This thesis introduces the data in Chapter 1 where it has been described in full detail what the data contains. Then the theory is introduced in Chapter 2 where the standard and the extended GMANOVA is introduced together with conditions of when they are unique. Furthermore their distributions are also introduced in order to obtain a Wald test for the standard GMANOVA. This has been implemented into R in Chapter 3 together

[1] Adenine(A), Cytosine(C), Guanine(G), Thymine(T)

with a likelihood ratio test. Then an analysis is made with the GMANOVA in Chapter 4 where different approaches has been taken, and then a discussion and a conclusion is made in Chapter 5.

Notation List

$\mathbf{A}$	Matrix
$\mathbf{a}$	Vector
A	Random variable
$r(\mathbf{A})$	Rank of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of the matrix $\mathbf{A}$
$\mathbf{A}^{-}$	Generalized inverse of the matrix $\mathbf{A}$
$\mathbf{A}^{o}$	Matrix that satisfies $\mathbf{A}^{oT}\mathbf{A} = \mathbf{0}$, $r(\mathbf{A}^{o}) = m - r(\mathbf{A})$ if $\mathbf{A} : m \times n$
$C(\mathbf{A})$	Column space of $\mathbf{A}$
$\mathcal{R}(A)$	Range space $\mathbf{R}(A) = \{\mathbf{x} : \mathbf{x} = A\mathbf{y}, \mathbf{y} \in \mathbb{V}\}$ where $\mathbb{V}$ is a vector space
$vec\mathbf{A}$	$\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_q) \in \mathbb{R}^{p\times q}$ then $vec\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_q \end{pmatrix} \in \mathbb{R}^{qp}$.
$(\mathbf{A}_1 : \ldots : \mathbf{A}_n)$	A matrix $\mathbf{A}$ can be partitioned into submatrices and written as $\mathbf{A} = (\mathbf{A}_1 : \ldots : \mathbf{A}_n)$
$E(\mathbf{A})$	Expected value of the matrix $\mathbf{A}$
$D(\mathbf{A})$	Dispersion of the matrix $\mathbf{A}$
$N_p(\mu, \mathbf{\Sigma})$	Multivariate normal distribution
$N_{p,n}(\mu, \mathbf{\Sigma}, \mathbf{\Psi})$	Matrix normal distribution which is equivalent to $N_{p,n}(\mu, \mathbf{\Sigma} \otimes \mathbf{\Psi})$
$W_p(\mathbf{\Sigma}, n)$	The central Wishart distribution
$W_p(\mathbf{\Sigma}, n, \mathbf{\Delta})$	The non-central Wishart distribution

Chapter 1

Data

In this Chapter the data will be introduced and described. In the data there are 458 Brazilians where their pigmentation has been measured 3 times at three different locations; buttock, arm and forehead. This results in 9 measurements for each person.

1.1 Data Description

There are 458 different subjects of which there are some subjects that contain missing values. There are 30 measurements where missing values are present. These are taken out, which leaves 428 measurements that are ready for analysis. But we also need to check that all measurements have 3 measurements at all places, which not all have. Therefore only 376 subjects are to be used. We could have imputed this, but since there are only 82 subjects that are taken out, which is 18% of the total amount of observations, we do not impute. Hence we will only be describing those measurements where all measurements are complete.

The first thing to talk about is the pigmentation. This is measured at three different sites at three different times. There is some variation where the values can be seen in Table 1.1.

	Median	Mean	Max	Min
Arm	8.95	9.38	20.3	3.7
Buttock	8	8.63	21.8	2.5
Forehead	9.4	9.84	21.2	4.1

Table 1.1: Pigmentation information

As it can be seen forehead has the highest amount of mean pigmentation, then arm and then buttock. Everything looks as to be expected, except the maximum value of buttock, which is the highest. This is surprising, but as there are no evidence that it is a faulty observation, it can not be ruled as such.

The data also contains 23 SNPs which are thought to be influential on the pigmentation of a person. These SNPs have, in this data set, 3 combinations each. It should be noted that C refers to the combination CC of the nucleotide bases and the same goes for A,G and T.

	rs1015362	rs10777129	rs10831496	rs11238349	rs12203592	rs12350739	rs12668421
1	C :141	C :236	C : 68	A : 18	C :318	A : 80	A :217
2	CT:176	CT:122	CT:178	G :196	CT: 57	G :129	AT:145
3	T : 59	T : 18	T :130	GA:162	T : 1	GA:167	T : 14

	rs12896399	rs12913832	rs13289	rs13933350	rs1408799	rs1426654	rs16891982
1	A : 18	A :204	C : 96	C :249	A : 76	A :250	C : 10
2	C :215	G : 39	G : 96	CT:119	G :115	G : 29	G :295
3	CA:143	GA:133	GC:184	T : 8	GA:185	GA: 97	GC: 71

	rs1800407	rs2031526	rs2424984	rs2470102	rs26722	rs4424881	rs4911414
1	C :326	C :241	C : 18	A :238	A : 2	C :197	G :202
2	CT: 48	CT:117	CT:115	G : 30	G :299	CT:148	GT:153
3	T : 2	T : 18	T :243	GA:108	GA: 75	T : 31	T : 21

	rs6119471	rs6742078
1	C :103	G :158
2	G : 90	GT:181
3	GC:183	T : 37

Table 1.2: All combinations of the 23 SNP

As can be seen some combinations are more represented than others for different SNPs, which is due to it being collected from the same country. Brazil is thought of as a country where many people migrated to, which explains the diversity in the SNPs.

Now that the data has been described, the theory will be introduced in Chapter 2 before it is being analysed in Chapter 4.

Chapter 2

Growth Curve Model

This chapter is based on [2].

The growth curve model proposed by Potthoff & Roy is the starting point for the theory. The growth curve model is a linear model, but the maximum likelihood estimator of its mean parameters is a non-linear expression which causes great difficulties when doing inference. Since the estimators are non-linear stochastic expressions, their distributions have to be approximated. This demands knowledge about moments which often can be given as an exact expression or at least approximated very accurately. This is based on moments of the matrix normal distribution and the Wishart distribution.

2.1 Introduction

Suppose we have an observation vector $\mathbf{y}_i \in \mathbb{R}^p$ which has a linear model

$$\mathbf{y}_i = \boldsymbol{\mu}_i + \boldsymbol{\Sigma}^{1/2}\mathbf{e}_i \tag{2.1}$$

where $\boldsymbol{\mu}_i \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ are unknown parameters, $\boldsymbol{\Sigma}^{1/2}$ is a symmetric square root of the positive definite matrix $\boldsymbol{\Sigma}$ and $\mathbf{e}_i \sim N_p(\mathbf{0}, \mathbf{I})$. It is assumed that $\boldsymbol{\mu}_i = \mathbf{X}\boldsymbol{\beta}_i$ where $\mathbf{X}$ is a $p \times q$ known matrix and $\boldsymbol{\beta}_i$ is an unknown $q \times 1$ vector. Note that the $\mathbf{X}$ matrix is the within-individuals and corresponds to the same structure for all $\mathbf{y}_i$, which will be extended in Section 2.3 to use different structure for different individuals. The observations $\mathbf{y}_i$ may contain repeated measurements on some individuals. In this project data contains repeated measurements on all individuals. This raises the situation where a between-individuals interaction is natural. Let $\mathbf{Y} \in \mathbb{R}^{p\times n}$ be a matrix where each column corresponds to one individual. Then instead of writing $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ we can write this as two matrices i.e.

$$\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n\} = \mathbf{BC} \tag{2.2}$$

where $\mathbf{C} \in \mathbb{R}^{k\times n}$ is a known between-individuals design matrix and $\mathbf{B} \in \mathbb{R}^{q\times k}$ is an unknown parameter matrix. This leads to the growth curve model definition.

Definition 2.1. Let $\mathbf{Y} : p \times n, \mathbf{X} : p \times q, q \leq p, \mathbf{B} : q \times k, \mathbf{C} : k \times n, r(\mathbf{C}) + p \leq n$ and let $\mathbf{\Sigma} : p \times p$ be a positive definite matrix. Then

$$\mathbf{Y} = \mathbf{XBC} + \mathbf{\Sigma}^{1/2}\mathbf{E} \tag{2.3}$$

defines the growth curve model, where $\mathbf{E} \sim N_{p,n}(\mathbf{0}, \mathbf{I}_p, \mathbf{I}_n)$, $\mathbf{X}, \mathbf{C}$ are known matrices and $\mathbf{B}, \mathbf{\Sigma}$ are unknown parameter matrices.

Note that the columns of $\mathbf{Y}$ are independent normally distributed p-vectors with unknown dispersion matrix $\mathbf{\Sigma}$ and expectation given by $E[\mathbf{Y}] = \mathbf{XBC}$. Note that if $\mathbf{X} = \mathbf{I}$ it reduces to an ordinary Multivariate ANalysis Of VAriance(MANOVA) model. Hence the model in (2.3) is a Generalised MANOVA or a GMANOVA. Much literature refers to the GMANOVA as growth curve statistics model, and these two terms will be used interchangeably. We can estimate the unknown parameter estimates by maximum likelihood which is given in the next section.

2.2 Maximum likelihood Estimators

We are only interested in the mean structure and hence we suppose that we have no information regarding the structure of $\mathbf{\Sigma}$. Therefore an arbitrary $\mathbf{\Sigma}$ is considered but it is assumed that $\mathbf{\Sigma}$ is positive definite. The following lemma presents a useful inequality that will be used later.

Lemma 2.2. Let $\mathbf{\Sigma}$ and $\mathbf{S}$ be positive definite matrices of size $p \times p$. Then

$$|\mathbf{\Sigma}|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}tr(\mathbf{\Sigma}^{-1}\mathbf{S})\right) \leq |\frac{1}{n}\mathbf{S}|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}np\right) \tag{2.4}$$

where equality holds if and only if $\mathbf{\Sigma} = \frac{1}{n}\mathbf{S}$.

Proof. We have from Theorem A.1 that there exists matrices $\mathbf{H}$ and $\mathbf{D}$ such that

$$\mathbf{\Sigma} = \mathbf{HD}^{-1}\mathbf{H}^T, \quad \mathbf{S} = \mathbf{HH}^T \tag{2.5}$$

where $\mathbf{H}$ is non-singular and $\mathbf{D} = (d_1, d_2, \ldots, d_p)$ is diagonal. Thus

$$\begin{aligned}
|\mathbf{\Sigma}|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}tr(\mathbf{\Sigma}^{-1}\mathbf{S})\right) &= |\mathbf{HD}^{-1}\mathbf{H}^T|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}tr(\mathbf{D})\right) \\
&= |\mathbf{HH}^T|^{-\frac{1}{2}n} \prod_{i=1}^{p} d_i^{\frac{n}{2}} \exp\left(-\frac{1}{2}d_i\right) \\
&\leq |\mathbf{HH}^T|^{-\frac{1}{2}n} n^{\frac{np}{2}} \exp\left(-\frac{1}{2}np\right) \\
&= |\frac{1}{n}\mathbf{S}|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}np\right)
\end{aligned}$$

where equality holds if and only if $d_i = n$ which means that $n\mathbf{\Sigma} = \mathbf{S}$. The inequality holds since it is the maximum value of the product on the right side of the inequality that is represented on the right side of the inequality. $\square$

Now we start deriving the maximum likelihood estimates of the growth curve by direct use of the likelihood. The density of the matrix normal distribution, $N_{p,n}(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{\Psi})$ is given as

$$f_{\mathbf{Y}}(\mathbf{Y}) = (2\pi)^{-\frac{1}{2}pn}|\mathbf{\Sigma}|^{-n/2}|\mathbf{\Psi}|^{-p/2} \exp\left(-\frac{1}{2}tr(\mathbf{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\mathbf{\Psi}^{-1}(\mathbf{Y}-\boldsymbol{\mu})^T)\right) \quad (2.6)$$

From this density we derive the likelihood function as

$$L(\mathbf{B}, \mathbf{\Sigma}) = (2\pi)^{-\frac{1}{2}pn}|\mathbf{\Sigma}|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}tr(\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{XBC})(\mathbf{Y}-\mathbf{XBC})^T)\right). \quad (2.7)$$

by the fact that $\mathbf{\Psi} = \mathbf{I}$ and $\boldsymbol{\mu} = E[\mathbf{Y}] = \mathbf{XBC}$. Then using Lemma 2.2 we get that

$$|\mathbf{\Sigma}|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}tr(\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{XBC})(\mathbf{Y}-\mathbf{XBC})^T)\right) \quad (2.8)$$

$$\leq \left|\frac{1}{n}(\mathbf{Y}-\mathbf{XBC})(\mathbf{Y}-\mathbf{XBC})^T\right|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}np\right) \quad (2.9)$$

where equality is obtained if and only if $n\mathbf{\Sigma} = (\mathbf{Y}-\mathbf{XBC})(\mathbf{Y}-\mathbf{XBC})^T$. Note that $\mathbf{\Sigma}$ has been estimated as a function of the mean and now the mean is estimated. Normally one estimate the mean and thereafter search for an estimate of the covariance matrix. This is achieved if a lower bound of $|(\mathbf{Y}-\mathbf{XBC})(\mathbf{Y}-\mathbf{XBC})^T|$, which is independent of $\mathbf{B}$, is found and obtained for at least one specific choice of $\mathbf{B}$.

In order to find a lower bound two steps are used. First we split the product $(\mathbf{Y} - \mathbf{XBC})(\mathbf{Y}-\mathbf{XBC})^T$ into two parts

$$(\mathbf{Y}-\mathbf{XBC})(\mathbf{Y}-\mathbf{XBC})^T = \mathbf{S} + \mathbf{V}\mathbf{V}^T \quad (2.10)$$

where

$$\mathbf{S} = \mathbf{Y}(\mathbf{I} - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^-\mathbf{C})\mathbf{Y}^T \quad (2.11)$$

and

$$\mathbf{V} = \mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^-\mathbf{C} - \mathbf{XBC}. \quad (2.12)$$

Note that $\mathbf{S}$ is Wishart distributed written as $W_p(\frac{1}{n}\mathbf{\Sigma}, n - r(\mathbf{C}))$ since $\mathbf{Y}$ is normally distributed, and the first element of $\mathbf{S}$ is $\mathbf{Y}\mathbf{Y}^T$ which is the definition of a Wishart distribution, seen in Definition A.11. Furthermore note that $\mathbf{S}$ is not dependent on $\mathbf{B}$. A rewrite of $\mathbf{S} + \mathbf{V}\mathbf{V}^T$ is given as

$$|\mathbf{S} + \mathbf{V}\mathbf{V}^T| = |\mathbf{S}||\mathbf{I} + \mathbf{S}^{-1}\mathbf{V}\mathbf{V}^T| = |\mathbf{S}||\mathbf{I} + \mathbf{V}^T\mathbf{S}^{-1}\mathbf{V}| \quad (2.13)$$

where $\mathbf{S}^{-1}$ exist with probability 1 since $\mathbf{\Sigma}$ is positive definite, and it also holds that $n \geq p$ then $\mathbf{S}$ is positive definite and hence has an inverse. The last equality is true since for $\mathbf{A} : m \times n, \mathbf{B} : n \times m$ it holds that $|\mathbf{I}_m + \mathbf{AB}| = |\mathbf{I}_n + \mathbf{BA}|$.

The second idea is based on Corollary A.2 where we can isolate $\mathbf{S}^{-1}$ to obtain

$$\mathbf{S}^{-1} = \mathbf{S}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1} + \mathbf{X}^o(\mathbf{X}^{oT}\mathbf{S}\mathbf{X}^o)^{-}\mathbf{X}^{oT}. \tag{2.14}$$

Then we can use this estimate in the rewrite and obtain

$$\begin{aligned}|(\mathbf{Y}-\mathbf{XBC})(\mathbf{Y}-\mathbf{XBC})^T| &= |\mathbf{S}||\mathbf{I}+\mathbf{V}^T\mathbf{S}^{-1}\mathbf{V}| \\ &= |\mathbf{S}||\mathbf{I}+\mathbf{V}^T\mathbf{S}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{V}+\mathbf{V}^T\mathbf{X}^o(\mathbf{X}^{oT}\mathbf{S}\mathbf{X}^o)^{-}\mathbf{X}^{oT}\mathbf{V}| \\ &\geq |\mathbf{S}||\mathbf{I}+\mathbf{V}^T\mathbf{X}^o(\mathbf{X}^{oT}\mathbf{S}\mathbf{X}^o)^{-}\mathbf{X}^{oT}\mathbf{V}| \qquad (2.15)\end{aligned}$$

which is independent of $\mathbf{B}$ since $\mathbf{X}^{oT}\mathbf{V} = \mathbf{X}^{oT}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C}$. The inequality holds since $\mathbf{V}^T\mathbf{S}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{V} \geq \mathbf{0}$. Equality in (2.15) holds if and only if

$$\mathbf{V}^T\mathbf{S}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{V} = \mathbf{0} \tag{2.16}$$

which is equivalent to

$$\mathbf{V}^T\mathbf{S}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-} = \mathbf{0} \tag{2.17}$$

Since $\mathbf{V} = \mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{C} - \mathbf{XBC}$ is the only estimate dependent on $\mathbf{B}$ it is a linear equation in $\mathbf{B}$. Using this fact and Theorem A.3 it follows that a general solution is given by

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} + (\mathbf{X}^T)^o\mathbf{Z}_1 + \mathbf{X}^T\mathbf{Z}_2\mathbf{C}^{oT} \tag{2.18}$$

where $\mathbf{Z}_1, \mathbf{Z}_2$ are arbitrary matrices. If it holds that $\mathbf{X}, \mathbf{C}$ are of full rank, i.e. $r(\mathbf{X}) = q, r(\mathbf{C}) = k$, a unique solution exists

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}. \tag{2.19}$$

Furthermore the maximum likelihood estimator of $\mathbf{\Sigma}$ is given by

$$n\hat{\mathbf{\Sigma}} = (\mathbf{Y}-\mathbf{XBC})(\mathbf{Y}-\mathbf{XBC})^T = \mathbf{S} + \hat{\mathbf{V}}\hat{\mathbf{V}}^T \tag{2.20}$$

where $\mathbf{V}$ is given in (2.12) with $\mathbf{B}$ replaced by the estimate. Then it follows that

$$\mathbf{X}\hat{\mathbf{B}}\mathbf{C} = \mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{C} \tag{2.21}$$

is always unique if $\mathbf{X}$ and $\mathbf{C}$ have full rank and hence $\hat{\mathbf{\Sigma}}$ is also uniquely estimated.

2.3 Extended Growth Curve Model

In this section an extension of the model in Definition 2.1 is given. This extension allows for more individual mean structures and can also handle if the estimates of $\hat{\mathbf{B}}_i, i = 1, 2, \ldots, m$ are correlated. As a motivation for the extension of the GMANOVA, we will start off by an example provided by [3].

Example 2.3. This article [3] presents a balanced two way analysis, which is a model with two random effects in the model. Since it is a two way ANOVA model, we have two factors to model the mean structure

$$E[\mathbf{Y}] = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} \mathbf{1}_n + \mathbf{1}_p(\beta_1, \dots, \beta_n) \tag{2.22}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are each a mean for a grouping defined by data. This defines the two way ANOVA without an error term. If we set $\mathbf{X}_1 = \mathbf{1}_p, \mathbf{B}_1 = \boldsymbol{\beta}^T, \mathbf{B}_2 = \boldsymbol{\alpha}$ and $\mathbf{C}_2 = \mathbf{1}_n^T$ then we have that

$$E[\mathbf{Y}] = \mathbf{X}_1\mathbf{B}_1 + \mathbf{B}_2\mathbf{C}_2. \tag{2.23}$$

Furthermore it is assumed that the columns are independent with a multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$. This model is assumed to be replicated r times so one could write $\tilde{\mathbf{C}}_1 = (\mathbf{I}_n, \dots, \mathbf{I}_n)$ and $\tilde{\mathbf{C}}_2 = (\mathbf{C}_2, \dots, \mathbf{C}_2)$ since the $\mathbf{C}$'s are the only known matrices that change between individuals. So the model becomes

$$\mathbf{Y} = \mathbf{X}_1\mathbf{B}_1\tilde{\mathbf{C}}_1 + \mathbf{B}_2\tilde{\mathbf{C}}_2 + \boldsymbol{\Sigma}^{1/2}\mathbf{E} \tag{2.24}$$

Since there are two factors that needs to be estimated, we can not use the same approach as for GMANOVA. Therefore we need to develop a new approach to this problem.

The example motivates the use of an extended GMANOVA, which will now be introduced.

Definition 2.4 (Extended growth curve model)**.** Let $\mathbf{Y} : p \times n, \mathbf{X}_i : p \times q_i, q_i \leq p, \mathbf{B}_i : q_i \times k_i, \mathbf{C}_i : k_i \times n, r(\mathbf{C}_1) + p \leq n, i = 1, 2, \dots, m$ where $C(\mathbf{C}_i^T) \subseteq C(\mathbf{C}_{i-1}^T), i = 2, 3, \dots, m$ and $\boldsymbol{\Sigma} : p \times p$ is a positive definite matrix. Then

$$\mathbf{Y} = \sum_{i=1}^{m} \mathbf{X}_i\mathbf{B}_i\mathbf{C}_i + \boldsymbol{\Sigma}^{1/2}\mathbf{E} \tag{2.25}$$

defines the extended growth curve model, where $\mathbf{E} \sim N_{p,n}(\mathbf{0}, \mathbf{I}_p, \mathbf{I}_n)$, $\mathbf{X}_i, \mathbf{C}_i$ are known matrices and $\mathbf{B}_i, \boldsymbol{\Sigma}$ are unknown parameter matrices.

Note that when $m = 1$ it reduces to the growth curve model in Definition 2.1. The extended growth curve model introduces a more general mean structure, since the variance component $\boldsymbol{\Sigma}^{1/2}\mathbf{E}$ is unchanged.

Writing the likelihood function

$$\begin{aligned} L(\{\mathbf{B}_i\}, \boldsymbol{\Sigma}) &= (2\pi)^{-\frac{1}{2}np}|\boldsymbol{\Sigma}|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \sum_{i=1}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i)(\mathbf{Y} - \sum_{i=1}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i)^T)\right) \\ &\leq (2\pi)^{-\frac{1}{2}np}|\frac{1}{n}(\mathbf{Y} - \sum_{i=1}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i)(\mathbf{Y} - \sum_{i=1}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i)^T|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}np\right) \end{aligned}$$

where equality holds if and only if

$$n\mathbf{\Sigma} = (\mathbf{Y} - \sum_{i=1}^{m} \mathbf{X}_i\mathbf{B}_i\mathbf{C}_i)(\mathbf{Y} - \sum_{i=1}^{m} \mathbf{X}_i\mathbf{B}_i\mathbf{C}_i)^T \tag{2.26}$$

Since it is only $\{\mathbf{B}_i\}$ that is missing, we can minimize the determinant

$$(\mathbf{Y} - \sum_{i=1}^{m} \mathbf{X}_i\mathbf{B}_i\mathbf{C}_i)(\mathbf{Y} - \sum_{i=1}^{m} \mathbf{X}_i\mathbf{B}_i\mathbf{C}_i)^T \tag{2.27}$$

with respect to $\{\mathbf{B}_i\}$. Since it holds that $C(\mathbf{C}_{i+1}^T) \subseteq C(\mathbf{C}_i^T)$ we can present the determinant as

$$|\mathbf{S}_1 + \mathbf{V}_1\mathbf{V}_1^T| \tag{2.28}$$

where

$$\mathbf{S}_1 = \mathbf{Y}(\mathbf{I} - \mathbf{C}_1^T(\mathbf{C}_1\mathbf{C}_1^T)^-\mathbf{C}_1)\mathbf{Y}^T \tag{2.29}$$

and

$$\mathbf{V}_1 = \mathbf{Y}\mathbf{C}_1^T(\mathbf{C}_1\mathbf{C}_1^T)^-\mathbf{C}_1 - \sum_{i=1}^{m} \mathbf{X}_i\mathbf{B}_i\mathbf{C}_i. \tag{2.30}$$

Note that $\mathbf{S}_1$ is identical to (2.11) given the constraint and that $\mathbf{V}_1$ and (2.12) have the same structure. Hence the same approach may be applied. To ease notation the following is introduced

$$\mathbf{P}_{\mathbf{X},\mathbf{S}} = \mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^-\mathbf{X}^T\mathbf{S}^{-1} \tag{2.31}$$

which is a projector and can be rewritten as

$$\mathbf{P}_{\mathbf{X},\mathbf{S}} = \mathbf{I} - \mathbf{P}_{\mathbf{X}^o,\mathbf{S}^{-1}}^T = \mathbf{I} - \mathbf{S}\mathbf{X}^o(\mathbf{X}^{oT}\mathbf{S}\mathbf{X}^o)^-\mathbf{X}^{oT} \tag{2.32}$$

Modifying (2.28) yields

$$\begin{aligned}
|\mathbf{S}_1 + \mathbf{V}_1\mathbf{V}_1^T| &= |\mathbf{S}_1||\mathbf{I} + \mathbf{S}_1^{-1}\mathbf{V}_1\mathbf{V}_1^T| \\
&= |\mathbf{S}_1||\mathbf{I} + \mathbf{V}_1^T\mathbf{S}_1^{-1}\mathbf{V}_1| \\
&= |\mathbf{S}_1||\mathbf{I} + \mathbf{V}_1^T\mathbf{P}_{\mathbf{X}_1,\mathbf{S}_1}^T\mathbf{S}_1^{-1}\mathbf{P}_{\mathbf{X}_1,\mathbf{S}_1}\mathbf{V}_1 + \mathbf{V}_1^T\mathbf{P}_{\mathbf{X}_1^o,\mathbf{S}_1^{-1}}^T\mathbf{S}_1^{-1}\mathbf{P}_{\mathbf{X}_1^o,\mathbf{S}_1^{-1}}^T\mathbf{V}_1| \\
&\geq |\mathbf{S}_1||\mathbf{I} + \mathbf{V}_1^T\mathbf{P}_{\mathbf{X}_1^o,\mathbf{S}_1^{-1}}^T\mathbf{S}_1^{-1}\mathbf{P}_{\mathbf{X}_1^o,\mathbf{S}_1^{-1}}^T\mathbf{V}_1| \\
&= |\mathbf{S}_1||\mathbf{I} + \mathbf{W}_1^T\mathbf{P}_{\mathbf{X}_1^o,\mathbf{S}_1^{-1}}^T\mathbf{S}_1^{-1}\mathbf{P}_{\mathbf{X}_1^o,\mathbf{S}_1^{-1}}^T\mathbf{W}_1|
\end{aligned} \tag{2.33}$$

where

$$\mathbf{W}_1 = \mathbf{Y}\mathbf{C}_1^T(\mathbf{C}_1\mathbf{C}_1^T)^-\mathbf{C}_1 - \sum_{i=2}^{m} \mathbf{X}_i\mathbf{B}_i\mathbf{C}_i. \tag{2.34}$$

So the difference between $\mathbf{W}_1$ and $\mathbf{V}_1$ is that the sum in $\mathbf{W}_1$ does not include the first summand. The reason for exchanging $\mathbf{V}_1$ with $\mathbf{W}_1$ is that the $\mathbf{X}_1$ is zero, since $\mathbf{P}_{\mathbf{X}_1^o,\mathbf{S}_1^{-1}}^T$ is the projection containing $\mathbf{X}_1^o$ and hence when multiplied is zero. This eliminates the first summand, which makes $\mathbf{V}_1 = \mathbf{W}_1$. The inequality in (2.33) holds since $|\mathbf{V}_1^T\mathbf{P}_{\mathbf{X}_1,\mathbf{S}_1}^T\mathbf{S}_1^{-1}\mathbf{P}_{\mathbf{X}_1,\mathbf{S}_1}\mathbf{V}_1| \geq 0$. There is equality if and only if $\mathbf{P}_{\mathbf{X}_1,\mathbf{S}_1}\mathbf{V}_1 = \mathbf{0}$ which is equivalent to

$$\mathbf{X}_1^T\mathbf{S}_1^{-1}\mathbf{V}_1 = \mathbf{0}. \tag{2.35}$$

By Theorem A.3 gives the estimate of $\hat{\mathbf{B}}_1$ as

$$\hat{\mathbf{B}}_1 = (\mathbf{X}_1^T\mathbf{S}_1^{-1}\mathbf{X}_1)^{-}\mathbf{X}_1^T\mathbf{S}_1^{-1}(\mathbf{Y} - \sum_{i=2}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i)\mathbf{C}_1^T(\mathbf{C}_1\mathbf{C}_1^T)^{-} + (\mathbf{X}_1^T)^o\mathbf{Z}_{11} + \mathbf{X}_1^T\mathbf{Z}_{12}\mathbf{C}_1^{oT}. \tag{2.36}$$

and noting that $\mathbf{P}_1 = \mathbf{I}$. We can also estimate $\hat{\mathbf{\Sigma}}$ by inserting the estimate of $\hat{\mathbf{B}}_1$ into (2.26).

Using the same approach as in (2.13), we can rewrite (2.33) as

$$|\mathbf{S}_1 + \mathbf{T}_1\mathbf{W}_1\mathbf{W}_1^T\mathbf{T}_1^T| \tag{2.37}$$

where $\mathbf{T}_1 = \mathbf{I} - \mathbf{P}_{\mathbf{A}_1,\mathbf{S}_1}$. This procedure of finding the determinant can be iteratively done by noting that

$$\begin{aligned}
\mathbf{V}_i &= \mathbf{T}_{i-1}\mathbf{Y}\mathbf{C}_i^T(\mathbf{C}_i\mathbf{C}_i^T)^{-}\mathbf{C}_i - \mathbf{T}_{i-1}\sum_{i}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i\\
\mathbf{W}_i &= \mathbf{Y}\mathbf{C}_i^T(\mathbf{C}_i\mathbf{C}_i^T)^{-}\mathbf{C}_i - \sum_{j=i+1}^{m}\mathbf{X}_j\mathbf{B}_j\mathbf{C}_j\\
\mathbf{T}_i &= \mathbf{I} - \mathbf{P}_{\mathbf{T}_{i-1}\mathbf{X}_i,\mathbf{S}_i}\\
\mathbf{P}_{i+1} &= \prod_{j=1}^{i}\mathbf{T}_i
\end{aligned}$$

where the $\mathbf{P}_{i+1}$ is the projection for the next iteration, i.e. $\mathbf{P}^T_{(\mathbf{P}_i\mathbf{A}_i)^o,\mathbf{S}_i^{-1}}$. Thus we need to find the solution to

$$\mathbf{X}_i^T\mathbf{P}_i^T\mathbf{S}_i^{-1}\mathbf{V}_i = \mathbf{0}. \tag{2.38}$$

Then we can find the maximum likelihood estimator of $\{\mathbf{B}_i\}$ and $\mathbf{\Sigma}$ when $\mathbf{Y} \sim N_{p,n}\left(\sum_{i=1}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i, \mathbf{\Sigma}, \mathbf{I}_n\right)$.

Theorem 2.5. Let

$$\begin{aligned}
\mathbf{P}_{\mathbf{C}_{j-1}^T} &= \mathbf{C}_{j-1}^T(\mathbf{C}_{j-1}\mathbf{C}_{j-1}^T)^{-}\mathbf{C}_{j-1}, \quad j = 1, \dots, m+1\\
\mathbf{K}_j &= \mathbf{P}_j\mathbf{Y}\mathbf{P}_{\mathbf{C}_{j-1}^T}(\mathbf{I} - \mathbf{P}_{\mathbf{C}_j^T})\mathbf{P}_{\mathbf{C}_{j-1}^T}\mathbf{Y}^T\mathbf{P}_j^T, \quad j = 1, \dots, m\\
\mathbf{S}_i &= \sum_{j=1}^{i}\mathbf{K}_j \quad i = 1, 2, \dots, m\\
\mathbf{T}_i &= \mathbf{I} - \mathbf{P}_i\mathbf{X}_i(\mathbf{X}_i^T\mathbf{P}_i^T\mathbf{S}_i^{-1}\mathbf{P}_i\mathbf{X}_i)^{-}\mathbf{X}_i^T\mathbf{P}_i^T\mathbf{S}_i^{-1} \quad i = 1, 2, \dots, m\\
\mathbf{P}_r &= \prod_{j=0}^{r-1}\mathbf{T}_j \quad r = 1, 2, \dots, m+1
\end{aligned}$$

where $\mathbf{T}_0 = \mathbf{I}$ and $\mathbf{C}_0 = \mathbf{I}$. Assuming that $\mathbf{S}_1$ is positive definite then the maximum likelihood estimators are given by

$$\begin{aligned}\hat{\mathbf{B}}_r &= (\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{P}_r\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\left(\mathbf{Y}-\sum_{i=r+1}^m \mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i\right)\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^- \\ &+ (\mathbf{X}_r^T\mathbf{P}_r^T)^o\mathbf{Z}_{r1} + \mathbf{X}_r^T\mathbf{P}_r^T\mathbf{Z}_{r2}\mathbf{C}_r^{oT}, \quad r = 1,2\ldots,m \\ n\hat{\mathbf{\Sigma}} &= (\mathbf{Y}-\sum_{i=1}^m \mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i)(\mathbf{Y}-\sum_{i=1}^m \mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i)^T \\ &= \mathbf{S}_m + \mathbf{P}_{m+1}\mathbf{Y}\mathbf{C}_m^T(\mathbf{C}_m\mathbf{C}_m^T)^-\mathbf{C}_m\mathbf{Y}\mathbf{P}_{m+1}^T\end{aligned}$$

where $\mathbf{Z}_{rj}$ are arbitrary matrices. Here $\sum_{i=m+1}^m \mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i = \mathbf{0}$.

Due to the uniqueness of the projector, the estimate of $\mathbf{P}_r\sum_{i=r}^m \mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i$ is always unique. This is also seen in the next theorem.

Theorem 2.6. For the estimators $\hat{\mathbf{B}}_i$ given in Theorem 2.5 it holds that

$$\mathbf{P}_r\sum_{i=r}^m \mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i = \sum_{i=r}^m(\mathbf{I}-\mathbf{T}_i)\mathbf{Y}\mathbf{C}_i^T(\mathbf{C}_i\mathbf{C}_i^T)^-\mathbf{C}_i. \tag{2.39}$$

A consequence of this theorem is the following

Corollary 2.7. In the extended model, the maximum likelihood estimator of $\mathbf{\Sigma}$ is always unique.

Even though the estimate of $\hat{\mathbf{B}}$ is unique, given the projector, there are still some requirements that need to be fulfilled in order for it to be unique. These are given in the next section. But before we move on we finish the model given in Example 2.3.

Example 2.8. Using the theory as described in the former section, the estimates for $\hat{\mathbf{B}}_1, \hat{\mathbf{B}}_2$ become

$$\begin{aligned}\hat{\mathbf{B}}_1 &= (\mathbf{X}_1^T\mathbf{S}_1^{-1}\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{S}_1^{-1}\left(\mathbf{Y}-\mathbf{X}_2\hat{\mathbf{B}}_2\tilde{\mathbf{C}}_2\right)\tilde{\mathbf{C}}_1^T(\tilde{\mathbf{C}}_1\tilde{\mathbf{C}}_1^T)^- \\ \hat{\mathbf{B}}_2 &= (\mathbf{X}_2^T\mathbf{P}_2^T\mathbf{S}_2^{-1}\mathbf{P}_2\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{P}_2^T\mathbf{S}_2^{-1}\mathbf{Y}\tilde{\mathbf{C}}_2^T(\tilde{\mathbf{C}}_2\tilde{\mathbf{C}}_2^T)^-\end{aligned}$$

where

$$\begin{aligned}\mathbf{P}_{\mathbf{C}_0^T} &= \mathbf{C}_0^T(\mathbf{C}_0\mathbf{C}_0^T)^{-1}\mathbf{C}_0 = \mathbf{I} \\ \mathbf{P}_1 &= \mathbf{T}_0 = \mathbf{I} \\ \mathbf{P}_{\mathbf{C}_1^T} &= \tilde{\mathbf{C}}_1^T(\tilde{\mathbf{C}}_1\tilde{\mathbf{C}}_1^T)^-\tilde{\mathbf{C}}_1 \\ \mathbf{S}_1 &= \mathbf{Y}(\mathbf{I}-\tilde{\mathbf{C}}_1^T(\tilde{\mathbf{C}}_1\tilde{\mathbf{C}}_1^T)^-\tilde{\mathbf{C}}_1)\mathbf{Y}^T \\ \mathbf{P}_2 &= \mathbf{X}_1(\mathbf{X}_1^T\mathbf{S}_1^{-1}\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{S}_1^{-1} \\ \mathbf{S}_2 &= \sum_{j=1}^2 \mathbf{P}_j\mathbf{Y}\mathbf{P}_{\mathbf{C}_{j-1}^T}(\mathbf{I}-\mathbf{P}_{\mathbf{C}_j^T})\mathbf{P}_{\mathbf{C}_{j-1}^T}\mathbf{Y}^T\mathbf{P}_j^T\end{aligned}$$

Here $\hat{\mathbf{B}}_2$ and $\mathbf{S}_2$ relies on the structure of $\tilde{\mathbf{C}}_2$ which is defined by data and hence is impossible to go further with in a non-data example. But since we know that $\tilde{\mathbf{C}}_1$ is the identity matrix replicated r times, we get that $\mathbf{Y}(\mathbf{I}-\tilde{\mathbf{C}}_1^T(\tilde{\mathbf{C}}_1\tilde{\mathbf{C}}_1^T)^-\tilde{\mathbf{C}}_1)\mathbf{Y}^T$ becomes

$$\begin{aligned}\mathbf{Y}(\mathbf{I}-\tilde{\mathbf{C}}_1^T(\tilde{\mathbf{C}}_1\tilde{\mathbf{C}}_1^T)^-\tilde{\mathbf{C}}_1)\mathbf{Y}^T &= \mathbf{Y}(\mathbf{I}-\tilde{\mathbf{C}}_1^T\frac{1}{r}\mathbf{I}\tilde{\mathbf{C}}_1)\mathbf{Y}^T \\ &= \mathbf{Y}(\mathbf{I}-\frac{1}{r}\mathbf{M}_r)\mathbf{Y}^T \\ &= \mathbf{Y}\frac{r-1}{r}\mathbf{M}_r\mathbf{Y}^T \\ &= \frac{r-1}{r}\sum_i \mathbf{Y}_{i\cdot}\mathbf{Y}_{i\cdot}^T\end{aligned}$$

where $\mathbf{M}_r$ is a $r\times r$ block matrix of $\mathbf{I}$ matrices. So $\mathbf{Y}(\mathbf{I}-\tilde{\mathbf{C}}_1^T(\tilde{\mathbf{C}}_1\tilde{\mathbf{C}}_1^T)^-\tilde{\mathbf{C}}_1)\mathbf{Y}^T$ becomes the sum of squares. So we get that

$$\begin{aligned}\hat{\mathbf{B}}_1 &= (\mathbf{X}_1^T\mathbf{S}_1^{-1}\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{S}_1^{-1}\left(\mathbf{Y}-\mathbf{X}_2\hat{\mathbf{B}}_2\tilde{\mathbf{C}}_2\right)\tilde{\mathbf{C}}_1^T(\tilde{\mathbf{C}}_1\tilde{\mathbf{C}}_1^T)^- \\ &= (\mathbf{X}_1^T\mathbf{S}_1^{-1}\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{S}_1^{-1}\left(\mathbf{Y}-\mathbf{X}_2\hat{\mathbf{B}}_2\tilde{\mathbf{C}}_2\right)\frac{r}{n-r}\end{aligned}$$

and

$$\mathbf{S}_1 = (\frac{r-1}{r}\sum_i \mathbf{Y}_{i\cdot}\mathbf{Y}_{i\cdot}^T)$$

for this specific example.

2.4 Uniqueness Of The Maximum Likelihood Estimates

In this section the uniqueness of the maximum likelihood estimates of $\hat{\mathbf{B}}$ is discussed. In the source it is given for $m=3$ but in this section the generalized concept will be derived. This is theoretically heavy and readers who does not care about details can jump to Theorem 2.9. First we derive for the GMANOVA and then for the extended GMANOVA.

For the model in Definition 2.1 the maximum likelihood estimate was given as

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^-\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^- + (\mathbf{X}^T)^o\mathbf{Z}_1 + \mathbf{X}^T\mathbf{Z}_2\mathbf{C}^{oT}. \tag{2.40}$$

Since $\mathbf{Z}_1,\mathbf{Z}_2$ are arbitrary matrices, $\hat{\mathbf{B}}$ becomes unique if and only if $(\mathbf{X}^T)^o=\mathbf{0}$ and $\mathbf{C}^{oT}=\mathbf{0}$. Furthermore since the rank of these are given as $r((\mathbf{X}^T)^o)=q-r((\mathbf{X}^T))$ and $r(\mathbf{C}^{oT})=k-r(\mathbf{C})$ respectively, it must hold that, for $\hat{\mathbf{B}}$ to be unique, $r(\mathbf{X}^T)=q$ and $r(\mathbf{C})=k$ which means that $\mathbf{X},\mathbf{C}$ must span the whole of their respective spaces. This is sufficient requirements for $\hat{\mathbf{B}}$ to be unique for the GMANOVA model.

Now we treat the extended model. In Theorem 2.5 it was found that the estimate for

$\hat{\mathbf{B}}_r$ was given as

$$\hat{\mathbf{B}}_r = (\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{P}_r\mathbf{X}_r)^{-}\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\left(\mathbf{Y} - \sum_{i=r+1}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i\right)\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^{-}$$
$$+ (\mathbf{X}_r^T\mathbf{P}_r^T)^o\mathbf{Z}_{r1} + \mathbf{X}_r^T\mathbf{P}_r^T\mathbf{Z}_{r2}\mathbf{C}_r^{oT}, \quad r = 1, 2\ldots, m$$

From the GMANOVA it can be seen that

$$r(\mathbf{X}_r^T\mathbf{P}_r^T) = r(\mathbf{X}_r) = q_r$$
$$r(\mathbf{C}_r) = k_r$$

must be required, since this eliminates the $\mathbf{Z}_{r1}$ and $\mathbf{Z}_{r2}$ as the arbitrary matrices. But we also need some restrictions on the column space, as the projector is also involved. Since we know that $\mathbf{P}_r = \prod_{i=1}^{r-1}\mathbf{T}_i$ is a function of the observations through $\mathbf{S}_{r-1}$, it must be a projector on a certain space which is independent of the observations, as $\mathbf{S}_{r-1}$ acts as an estimator of the inner product matrix and has nothing to do with the space where $\mathbf{X}_r$ is projected on. Thus the rank conditions should be independent of the observations.

However for $\hat{\mathbf{B}}_r$ to be unique it must hold that

$$(\mathbf{X}_r^T\mathbf{S}_r^{-1}\mathbf{X}_r)^{-}(\mathbf{X}_r^T\mathbf{S}_r^{-1}(\sum_{i=r+1}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i)\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^{-}) \tag{2.41}$$

is unique. But since every $\hat{\mathbf{B}}_i$ contains arbitrary $\mathbf{Z}$ matrices, these must be eliminated for all $i > r$. A necessary condition for this is

$$C(\mathbf{X}_s^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{X}_r(\mathbf{X}_r^T\mathbf{S}_r^{-1}\mathbf{X}_r)^{-}) \subseteq C(\mathbf{X}_s^T\prod_{i=r+1}^{s}\mathbf{T}_i^T) \tag{2.42}$$

where $\mathbf{P} = \mathbf{I} - \mathbf{X}_{r+1}(\mathbf{X}_{r+1}^T\mathbf{T}_r\mathbf{S}_{r+1}^{-1}\mathbf{T}_r\mathbf{X}_{r+1})^{-}\mathbf{X}_{r+1}^T\mathbf{T}_r^T\mathbf{S}_{r+1}^{-1}\mathbf{T}_r$ and $s = r+1, r+2, \ldots, m-r$. Using the definition of $\mathbf{T}_r$ equation (2.42) can be written as

$$C(\mathbf{X}_s^T(\mathbf{I} - \mathbf{T}_{s-1}^T)) \subseteq C(\mathbf{X}_s^T\prod_{i=r+1}^{s}\mathbf{T}_i^T) \tag{2.43}$$

which, by Proposition A.4 is equivalent to

$$C(\mathbf{X}_s^T) \subseteq C(\mathbf{X}_s^T\prod_{i=r+1}^{s}\mathbf{T}_i^T) \tag{2.44}$$

which by Theorem A.6 holds if and only if

$$C(\mathbf{X}_r) \cap C(\mathbf{X}_1 : \mathbf{X}_2 : \ldots : \mathbf{X}_{r-1}) = \{\mathbf{0}\}, \quad r > 1 \tag{2.45}$$

so this must be a sufficient condition for eliminating the $\mathbf{Z}$ matrices. But we still need to have the projector included separately. For this note first that

$$\prod_{i=1}^{m}\mathbf{T}_i^T = \mathbf{P}^T\prod_{i=1}^{m-1}\mathbf{T}_i^T \tag{2.46}$$

and since

$$C(\mathbf{X}_s^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{X}_r(\mathbf{X}_r^T\mathbf{S}_r^{-1}\mathbf{X}_r)^-) = C(\mathbf{X}_s^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{X}_r(\mathbf{X}_r^T\mathbf{S}_r^{-1}\mathbf{X}_r)^-\mathbf{X}_r^T) \tag{2.47}$$

we have that (2.42) is identical to

$$C(\mathbf{X}_s^T\mathbf{P}^T(\mathbf{I}-\mathbf{T}_{s-1}^T)) \subseteq C(\mathbf{X}_s^T\mathbf{P}^T\prod_{i=r+1}^{s}\mathbf{T}_i^T). \tag{2.48}$$

By Proposition A.4 we have that

$$C(\mathbf{X}_s^T\mathbf{P}^T) = C(\mathbf{X}_s^T\prod_{i=r}^{s}\mathbf{T}_i) \tag{2.49}$$

But since $\mathbf{P}$ is a projector it is also idempotent and it holds that $\mathbf{P}^T$ spans the orthogonal complement to $\mathbf{X}_{r+1}$. Hence we can define projectors

$$\begin{aligned}\mathbf{P}_{\mathbf{X}_s^o} &= \mathbf{I}-\mathbf{X}_s(\mathbf{X}_s^T\mathbf{X}_s)^-\mathbf{X}_s^T\\ \mathbf{P}_{(\mathbf{X}_r:\mathbf{X}_s)^o} &= \mathbf{P}_{\mathbf{X}_s^o}-\mathbf{P}_{\mathbf{X}_s^o}\mathbf{X}_r(\mathbf{X}_r^T\mathbf{P}_{\mathbf{X}_s^o}\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{P}_{\mathbf{X}_s^o}.\end{aligned}$$

Since $C(\mathbf{P}^T) = C(\mathbf{P}_{\mathbf{X}_s^o})$ and $C(\prod_{i=1}^m \mathbf{T}_i^T) = C(\mathbf{P}_{(\mathbf{X}_r:\mathbf{X}_s)^o})$ then by Proposition A.4, we have that (2.49) is equivalent to

$$C(\mathbf{X}_s^T\mathbf{P}_{\mathbf{X}_s^o}) = C(\mathbf{X}_s^T\mathbf{P}_{(\mathbf{X}_r:\mathbf{X}_s)^o}). \tag{2.50}$$

Since $C(\mathbf{I}-\mathbf{X}_r(\mathbf{X}_r^T\mathbf{P}_{\mathbf{X}_s^o}\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{P}_{\mathbf{X}_s^o}) = C(\mathbf{P}_{\mathbf{X}_s^o}\mathbf{X}_r)^\perp$, Theorem A.6 and the definition of $\mathbf{P}_{(\mathbf{X}_r:\mathbf{X}_s)^o}$ we have that (2.49) is equivalent to

$$C(\mathbf{P}_{\mathbf{X}_s^o}\mathbf{X}_s) \cup C(\mathbf{P}_{\mathbf{X}_s^o}\mathbf{X}_r) = \{\mathbf{0}\} \tag{2.51}$$

and by using properties for a projector and Theorem A.5 it follows that

$$C(\mathbf{X}_{s,r})^\perp \cap \{C(\mathbf{X}_{s,r})+C(\mathbf{X}_{s+r})\} \cap \{C(\mathbf{X}_{s,r})+C(\mathbf{X}_r)\} = \{\mathbf{0}\}, \quad r>1, s=1,2,\ldots,m-r \tag{2.52}$$

All of the conditions given here such that $\hat{\mathbf{B}}$ is unique is collected and given in the following theorem.

Theorem 2.9. Let

$$\begin{aligned}\mathbf{X}_{s,r} &= (\mathbf{X}_1:\mathbf{X}_2:\ldots:\mathbf{X}_{r-1}:\mathbf{X}_{r+1}:\ldots\mathbf{X}_{r+s+1}), \quad s=2,3,\ldots,m-r, r>1\\ \mathbf{X}_{s,r} &= (\mathbf{X}_1:\mathbf{X}_2:\ldots:\mathbf{X}_s), \quad s=2,3,\ldots,m-r, r=1\\ \mathbf{X}_{s,r} &= (\mathbf{X}_1:\mathbf{X}_2:\ldots:\mathbf{X}_{r-1}), \quad s=1, r>1\\ \mathbf{X}_{s,r} &= \mathbf{0} \quad s=1, r=1.\end{aligned}$$

Then $\hat{\mathbf{B}}$ given in Theorem 2.5 is unique if and only if

$$
\begin{aligned}
r(\mathbf{X}_r) &= m_r \\
r(\mathbf{C}_r) &= k_r \\
C(\mathbf{X}_r) \cap C(\mathbf{X}_1 : \mathbf{X}_2 : \ldots : \mathbf{X}_{r-1}) &= \{\mathbf{0}\}, \quad r > 1 \\
C(\mathbf{X}_{s,r})^{\perp} \cap \{C(\mathbf{X}_{s,r}) + C(\mathbf{X}_{s+r})\} \cap \{C(\mathbf{X}_{s,r}) + C(\mathbf{X}_r)\} &= \{\mathbf{0}\}, \quad r > 1, s = 1, 2, \ldots, m-r
\end{aligned}
$$

Since we now have the unique estimates, we need to find the moment estimates in order to be able to estimate the distributions of the estimators.

2.5 Moments Of Estimates

In this section first order moments of estimates of the maximum likelihood estimates are found. These results play a vital role in approximating the distributions of the estimators.

2.5.1 Moments Of The GMANOVA

For the GMANOVA the estimate of $\hat{\mathbf{B}}$ was found to be

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \tag{2.53}$$

where it was assumed that $r(\mathbf{X}) = q$ and $r(\mathbf{C}) = k$ for $\hat{\mathbf{B}}$ to be unique. We know that this estimate is non-linear in $\mathbf{Y}$ since $\mathbf{S}$ contains $\mathbf{Y}$ and the inverse of $\mathbf{S}$ is used in the estimate for $\hat{\mathbf{B}}$. The estimate $\hat{\mathbf{B}}$ consists of two parts $(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}$, which is a non-linear random expression in $\mathbf{Y}$, and $\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}$ which is linear in $\mathbf{Y}$. Due to the first part of $\hat{\mathbf{B}}$ is not available in simple form, one resorts to comparing $\hat{\mathbf{B}}$ with a more simple distribution. For this simple distribution one could choose

$$\hat{\mathbf{B}}_{\mathbf{G}} = (\mathbf{X}^T\mathbf{G}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{G}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \tag{2.54}$$

where $\mathbf{G}$ is assumed to be a non-random positive definite matrix. One obvious choice for $\mathbf{G}$ in this setting is $\mathbf{G} = \mathbf{I}$. According to Theorem A.7 the distribution of $\hat{\mathbf{B}}_{\mathbf{G}}$ is matrix normal and hence it would be valuable to compare moments of $\hat{\mathbf{B}}$ with $\hat{\mathbf{B}}_{\mathbf{G}}$. Furthermore it is tempting to condition on $\mathbf{S}$ since the distribution of $\mathbf{S}$ does not involve the parameter $\mathbf{B}$. Hence it is of interest how an omission of the variation in $\mathbf{S}$ affects the moments in $\hat{\mathbf{B}}$.

Theorem 2.10. Let

$$
\begin{aligned}
\hat{\mathbf{B}} &= (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \\
\mathbf{X}\hat{\mathbf{B}}\mathbf{C} &= \mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C}
\end{aligned}
$$

Then the following statements hold

1. $E[\hat{\mathbf{B}}] = \mathbf{B}$

2. if $n-k-p+q-1>0$ then

$$D[\hat{\mathbf{B}}] = \frac{n-k-1}{n-k-p+q-1}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-} \otimes (\mathbf{C}\mathbf{C}^T)^{-} \tag{2.55}$$

3. $E[\mathbf{X}\hat{\mathbf{B}}\mathbf{C}] = \mathbf{X}\mathbf{B}\mathbf{C}$

4. if $n-r(\mathbf{C})-p+r(\mathbf{X})-1>0$ then

$$D[\mathbf{X}\hat{\mathbf{B}}\mathbf{C}] = \frac{n-r(\mathbf{C})-1}{n-r(\mathbf{C})-p+r(\mathbf{X})-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T \otimes \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C} \tag{2.56}$$

Note from 2 to 4 that in 2 it is assumed that the matrices $\mathbf{X},\mathbf{C}$ are of full rank, q,k respectively.

Proof. We first prove 3. Since by Theorem A.8, $\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}$ and $\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C}$ are independent we can split $E[\mathbf{X}\hat{\mathbf{B}}\mathbf{C}]$ into

$$E[\mathbf{X}\hat{\mathbf{B}}\mathbf{C}] = E[\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}]E[\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C}] \tag{2.57}$$

But since $E[\mathbf{Y}] = \mathbf{X}\mathbf{B}\mathbf{C}$ implies $E[\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C}] = \mathbf{X}\mathbf{B}\mathbf{C}$ we get that

$$\begin{aligned} E[\mathbf{X}\hat{\mathbf{B}}\mathbf{C}] &= E[\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}]\mathbf{X}\mathbf{B}\mathbf{C} \\ &= E[\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X}]\mathbf{B}\mathbf{C} \\ &= \mathbf{X}\mathbf{B}\mathbf{C} \end{aligned}$$

where the last equality comes from $(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X} = \mathbf{I}$. In the same manner 1 can be proven.

To prove the dispersion we consider

$$D[\mathbf{X}\hat{\mathbf{B}}\mathbf{C}] = E[vec(\mathbf{X}(\hat{\mathbf{B}}-\mathbf{B})\mathbf{C})vec^T(\mathbf{X}(\hat{\mathbf{B}}-\mathbf{B})\mathbf{C})] \tag{2.58}$$

and note that

$$\mathbf{X}(\hat{\mathbf{B}}-\mathbf{B})\mathbf{C} = \mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}(\mathbf{Y}-\mathbf{X}\mathbf{B}\mathbf{C})\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C}. \tag{2.59}$$

It will be utilised that

$$\mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1} = \underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^{-}\underline{\mathbf{X}}^T\mathbf{S}^{-1} \tag{2.60}$$

where $\underline{\mathbf{X}}$ is any matrix of full rank such that $r(\mathbf{X}) = r(\underline{\mathbf{X}})$ which follows from the uniqueness property of projectors given in Proposition A.9. What this means is that $\underline{\mathbf{X}}$ is $\mathbf{X}$ where all linear dependent rows and corresponding columns are removed, and hence will have the same rank.

Put

$$\mathbf{Q} = (\mathbf{Y}-\mathbf{X}\mathbf{B}\mathbf{C})\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C} \tag{2.61}$$

which is independent of $\mathbf{S}$ by Theorem A.8 and the dispersion of $\mathbf{Q}$ equals

$$D[\mathbf{Q}] = \boldsymbol{\Sigma} \otimes \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C}.$$

From (2.58) and (2.60) it follows that

$$\begin{aligned}
D[\mathbf{X}\hat{\mathbf{B}}\mathbf{C}] &= E[(\mathbf{I}\otimes\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T\mathbf{S}^{-1})E[vec(\mathbf{Q})vec^T(\mathbf{Q})](\mathbf{I}\otimes\mathbf{S}^{-1}\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T)]\\
&= E[(\mathbf{I}\otimes\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T\mathbf{S}^{-1})D[\mathbf{Q}](\mathbf{I}\otimes\mathbf{S}^{-1}\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T)]\\
&= E[\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T\mathbf{S}^{-1}\mathbf{\Sigma}\mathbf{S}^{-1}\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T]\otimes\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^-\mathbf{C} \qquad (2.62)
\end{aligned}$$

For the rest of this proof we will utilise the canonical representation of $\underline{\mathbf{X}}^T\mathbf{\Sigma}^{-1/2}$ where $\mathbf{\Sigma}^{-1/2}$ is a symmetric square root of $\mathbf{\Sigma}^{-1}$. Proposition A.10 implies that there exists a non-singular matrix $\mathbf{H}$ and an orthogonal matrix $\mathbf{\Gamma}$ such that

$$\underline{\mathbf{X}}^T\mathbf{\Sigma}^{-1/2} = \mathbf{H}(\mathbf{I}_{r(\mathbf{X})} : \mathbf{0})\mathbf{\Gamma} = \mathbf{H}\mathbf{\Gamma}_1 \qquad (2.63)$$

where $\mathbf{\Gamma}^T = (\mathbf{\Gamma}_1^T : \mathbf{\Gamma}_2^T), (p\times r(\mathbf{X}) : p\times(p-r(\mathbf{X}))$. Let

$$\mathbf{V} = \mathbf{\Sigma}^{-1/2}\mathbf{S}\mathbf{\Sigma}^{-1/2} \qquad (2.64)$$

where we have from Theorem A.12 that $\mathbf{V}\sim W_p(\mathbf{I}, n-r(\mathbf{C}))$ since $\mathbf{S}$ is Wishart distributed and furthermore the matrices $\mathbf{V}, \mathbf{V}^{-1}$ is partitioned into

$$\mathbf{V} = \begin{bmatrix}\mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{bmatrix}, \quad \begin{matrix} r(\mathbf{X})\times r(\mathbf{X}) & r(\mathbf{X})\times(p-r(\mathbf{X}))\\ (p-r(\mathbf{X}))\times r(\mathbf{X}) & (p-r(\mathbf{X}))\times(p-r(\mathbf{X}))\end{matrix} \qquad (2.65)$$

$$\mathbf{V}^{-1} = \begin{bmatrix}\mathbf{V}^{11} & \mathbf{V}^{12}\\ \mathbf{V}^{21} & \mathbf{V}^{22}\end{bmatrix}, \quad \begin{matrix} r(\mathbf{X})\times r(\mathbf{X}) & r(\mathbf{X})\times(p-r(\mathbf{X}))\\ (p-r(\mathbf{X}))\times r(\mathbf{X}) & (p-r(\mathbf{X}))\times(p-r(\mathbf{X}))\end{matrix} \qquad (2.66)$$

Using these facts it follows that

$$\begin{aligned}
&E[\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T\mathbf{S}^{-1}\mathbf{\Sigma}\mathbf{S}^{-1}\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T] \qquad (2.67)\\
&\quad= E[\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}_1^T(\mathbf{V}^{11})^{-1}(\mathbf{V}^{11}:\mathbf{V}^{12})(\mathbf{V}^{11}:\mathbf{V}^{12})^T(\mathbf{V}^{11})^{-1}\mathbf{\Gamma}_1\mathbf{\Sigma}^{1/2}]\\
&\quad= E[\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}_1^T(\mathbf{I}:(\mathbf{V}^{11})^{-1}\mathbf{V}^{12})(\mathbf{I}:(\mathbf{V}^{11})^{-1}\mathbf{V}^{12})^T\mathbf{\Gamma}_1\mathbf{\Sigma}^{1/2}]\\
&\quad= E[\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}_1^T(\mathbf{I}+(\mathbf{V}^{11})^{-1}\mathbf{V}^{12}\mathbf{V}^{21}(\mathbf{V}^{11})^{-1})\mathbf{\Gamma}_1\mathbf{\Sigma}^{1/2}]\\
&\quad= E[\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}_1^T\mathbf{\Gamma}_1\mathbf{\Sigma}^{1/2}+\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}_1^T\mathbf{V}_{12}(\mathbf{V}_{22})^{-1}(\mathbf{V}_{22})^{-1}\mathbf{V}_{21}\mathbf{\Gamma}_1\mathbf{\Sigma}^{1/2}] \qquad (2.68)
\end{aligned}$$

by using Proposition A.13 to get $(\mathbf{V}^{11})^{-1}\mathbf{V}^{12} = -\mathbf{V}_{12}\mathbf{V}_{22}^{-1}$. The first equality is gotten since

$$E[\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T\mathbf{S}^{-1}\mathbf{\Sigma}\mathbf{S}^{-1}\underline{\mathbf{X}}(\underline{\mathbf{X}}^T\mathbf{S}^{-1}\underline{\mathbf{X}})^-\underline{\mathbf{X}}^T] = \frac{n-1}{n-p+r(\mathbf{X})-1}\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T \qquad (2.69)$$

if $n-p+r(\mathbf{X})-1>0$, which is already assumed in the Theorem. Now we focus on the $\mathbf{V}_{12}(\mathbf{V}_{22})^{-1}(\mathbf{V}_{22})^{-1}\mathbf{V}_{21}$ part. Since $\mathbf{V}\sim W_p(\mathbf{I}, n-r(\mathbf{C}))$ there exists, according to Definition A.11, a matrix $\mathbf{U}\sim N_{p,n-r(\mathbf{C})}(\mathbf{0},\mathbf{I},\mathbf{I})$ such that $\mathbf{V}=\mathbf{U}\mathbf{U}^T$. If we partition $\mathbf{U} = (\mathbf{U}_1^T : \mathbf{U}_2^T)^T$ so that $\mathbf{V}_{21} = \mathbf{U}_2\mathbf{U}_1^T$ and $\mathbf{V}_{11} = \mathbf{U}_1\mathbf{U}_1^T$ then we get that

$$E[\mathbf{V}_{12}(\mathbf{V}_{22})^{-1}(\mathbf{V}_{22})^{-1}\mathbf{V}_{21}] = E[\mathbf{U}_1\mathbf{U}_2^T(\mathbf{U}_2\mathbf{U}_2^T)^{-1}(\mathbf{U}_2\mathbf{U}_2^T)^{-1}\mathbf{U}_2\mathbf{U}_1^T]. \qquad (2.70)$$

By Theorem A.14 and the independence between $\mathbf{U}_1, \mathbf{U}_2$ we get

$$\begin{aligned} E[\mathbf{V}_{12}(\mathbf{V}_{22})^{-1}(\mathbf{V}_{22})^{-1}\mathbf{V}_{21}] &= E[tr\left(\mathbf{U}_2^T(\mathbf{U}_2\mathbf{U}_2^T)^{-1}(\mathbf{U}_2\mathbf{U}_2^T)^{-1}\mathbf{U}_2\right)]\mathbf{I} \\ &= E[tr(\mathbf{U}_2\mathbf{U}_2^T)^{-1}]\mathbf{I} \end{aligned}$$

Since $\mathbf{U}_2\mathbf{U}_2^T \sim W_{p-r(\mathbf{X})}(\mathbf{I}, n-r(\mathbf{C}))$ it follows from Theorem A.15 that

$$E[\mathbf{V}_{12}(\mathbf{V}_{22})^{-1}(\mathbf{V}_{22})^{-1}\mathbf{V}_{21}] = \frac{n-r(\mathbf{C})-1}{n-r(\mathbf{C})-p+r(\mathbf{X})-1}\mathbf{I} \tag{2.71}$$

Finally we note that $\mathbf{\Gamma}_1\mathbf{\Gamma}_1^T = \mathbf{I}$ and $\mathbf{H}$ is non-singular, and we get that

$$\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}_1^T\mathbf{\Gamma}_1\mathbf{\Sigma}^{1/2} = \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T. \tag{2.72}$$

We prove 4 by combining (2.62), (2.67), (2.71) and (2.72) to get the desired statement. 2 follows immediately from 4 by multiplying $(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-} \otimes (\mathbf{C}\mathbf{C}^T)^{-}$ by $\mathbf{X}, \mathbf{C}$ respectively and also assuming full rank of $\mathbf{X}, \mathbf{C}$. □

Now we actually have all elements necessary for approximating the distribution of $\hat{\mathbf{B}}$ and hence we only need to approximate the distribution of $\hat{\mathbf{\Sigma}}$. This estimator is given as

$$n\hat{\mathbf{\Sigma}} = \mathbf{S} + (\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C} - \mathbf{X}\hat{\mathbf{B}}\mathbf{C})(\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C} - \mathbf{X}\hat{\mathbf{B}}\mathbf{C})^T \tag{2.73}$$

where

$$\begin{aligned} \mathbf{S} &= \mathbf{Y}(\mathbf{I} - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C})\mathbf{Y}^T \\ \mathbf{X}\hat{\mathbf{B}}\mathbf{C} &= \mathbf{X}(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C}. \end{aligned}$$

Then we get the following Theorem, which will be given without proof. This project refer to [2] for a proof of this.

Theorem 2.11. Let $\hat{\mathbf{\Sigma}}$ be given as in (2.73).

1. If $n - r(\mathbf{C}) - p + r(\mathbf{X}) - 1 > 0$ then

$$E[\hat{\mathbf{\Sigma}}] = \mathbf{\Sigma} - r(\mathbf{C})\frac{1}{n}\frac{n-r(\mathbf{C})-2(p-r(\mathbf{X}))-1}{n-r(\mathbf{C})-p-r(\mathbf{X})-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T \tag{2.74}$$

2. If $n - r(\mathbf{C}) - p + r(\mathbf{X}) - 3 > 0$ then

$$\begin{aligned} D[\hat{\mathbf{\Sigma}}] =& d_1(\mathbf{I}+\mathbf{K}_{p,p})\left((\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T) \otimes (\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T)\right) \\ &+ d_2(\mathbf{I}+\mathbf{K}_{p,p})\left((\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T) \otimes (\mathbf{\Sigma} - \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T)\right) \\ &+ d_2(\mathbf{I}+\mathbf{K}_{p,p})\left((\mathbf{\Sigma} - \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T) \otimes (\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T)\right) \\ &+ \frac{1}{n}(\mathbf{I}+\mathbf{K}_{p,p})\left((\mathbf{\Sigma} - \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T) \otimes (\mathbf{\Sigma} - \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T)\right) \\ &+ d_3 vec(\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T)vec^T(\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T) \end{aligned}$$

where

$$d_1 = \frac{n - r(\mathbf{C})}{n^2} + 2r(\mathbf{C})\frac{p - r(\mathbf{X})}{n^2(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 1)} + r(\mathbf{C})\frac{2c_1 + c_2 + c_3}{n^2} + r(\mathbf{C})^2\frac{c_3}{n^2} \tag{2.75}$$

and

$$\begin{aligned} d_2 &= \frac{n - p + r(\mathbf{X}) - 1}{n(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 1)} \\ d_3 &= \frac{2r(\mathbf{C})(n - r(\mathbf{C}) - 1)(n - p + r(\mathbf{X}) - 1)(p - r(\mathbf{X}))}{n^2(n - r(\mathbf{C}) - p + r(\mathbf{X}))(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 1)^2(n - r(\mathbf{C} - p + r(\mathbf{X}) - 3)} \end{aligned}$$

where further

$$\begin{aligned} c_1 &= \frac{p - r(\mathbf{X})}{n - r(\mathbf{C}) - p + r(\mathbf{X}) - 1} \\ c_2 &= \frac{2(p - r(\mathbf{X}))(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 1) + (2 + (n - r(\mathbf{C}) - p + r(\mathbf{X}))}{(n - r(\mathbf{C}) - p + r(\mathbf{X}))(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 1)^2(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 3)} \\ &\cdot \frac{(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 3)(p - r(\mathbf{X}))^2}{(n - r(\mathbf{C}) - p + r(\mathbf{X}))(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 1)^2(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 3)} \\ c_3 &= \frac{p - r(\mathbf{X})}{(n - r(\mathbf{C}) - p + r(\mathbf{X}))(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 3)} \\ &+ \frac{p - r(\mathbf{X})^2}{(n - r(\mathbf{C}) - p + r(\mathbf{X}))(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 1)(n - r(\mathbf{C}) - p + r(\mathbf{X}) - 3)} \end{aligned}$$

As can be seen in the theorem, $\hat{\mathbf{\Sigma}}$ is a biased estimator of $\mathbf{\Sigma}$. Normally one can overcome this by multiplying with a constant. Such a constant exists and is given by

$$\frac{1}{n - r(\mathbf{C})}\mathbf{S}. \tag{2.76}$$

This is an unbiased estimate since using the definition of $\mathbf{S}$ in (2.73) and the fact that $\mathbf{Y}$ can be written as $\mathbf{Y} = \mathbf{X}\hat{\mathbf{B}}\mathbf{C} + \mathbf{u}$ where $\mathbf{u} \sim N(0, \sigma^2)$ is the error, we get

$$\begin{aligned} \hat{\mathbf{S}} &= \mathbf{Y}(\mathbf{I} - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^-\mathbf{C})\mathbf{Y}^T \\ &= \mathbf{Y}(\mathbf{C}^T\hat{\mathbf{B}}^T\mathbf{X}^T + \mathbf{u}^T - \mathbf{C}^T\hat{\mathbf{B}}^T\mathbf{X}^T - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^-\mathbf{C}\mathbf{u}) \\ &= (\mathbf{X}\hat{\mathbf{B}}\mathbf{C} + \mathbf{u})(\mathbf{u}^T - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^-\mathbf{C}\mathbf{u}) \\ &= (\mathbf{u}\mathbf{u}^T + \mathbf{X}\hat{\mathbf{B}}\mathbf{C}\mathbf{u}^T - \mathbf{u}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^-\mathbf{C}\mathbf{u}^T - \mathbf{X}\hat{\mathbf{B}}\mathbf{C}\mathbf{u}^T) \\ &= (\mathbf{u}\mathbf{u}^T - \mathbf{u}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^-\mathbf{C}\mathbf{u}^T) \end{aligned}$$

Then it can be seen that the expected value is 0 since $E[\mathbf{u}] = 0$. Then multiplying this on the estimate of $\hat{\mathbf{\Sigma}}$ will give an unbiased estimate.

Now we have derived the mean and dispersion for the standard GMANOVA model, so now we will try to extract them for the extended model.

2.5.2 Moments Of The Extended GMANOVA

The derivation of the moments are quite complicated, as it was already very complicated for the simple GMANOVA model. Therefore we will not be deriving the distributions to its full extent, but give pointers towards how to derive this. It is basically the same procedure as for the simple GMANOVA model.

First we are going to see that $\{\hat{\mathbf{B}}\}$ is an unbiased estimator under the uniqueness assumptions and hence $\sum_{i=1}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i$ is also unbiased. In Theorem 2.5 the estimators $\hat{\mathbf{B}}_r, r = 1, 2, \ldots, m$ were presented and since $C(\mathbf{C}_j^T) \subseteq C(\mathbf{C}_k^T)$ for $j \geq k$ it follows that $\mathbf{P}_r^T\mathbf{S}_r^{-1}$ is independent of $\mathbf{Y}\mathbf{C}_r^T$. Hence

$$
\begin{aligned}
E[\hat{\mathbf{B}}_r] = E&\left[(\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{P}_r\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\left(E[\mathbf{Y}] - \sum_{i=r+1}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i\right)\right]\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^- \\
&= \mathbf{B}_r - E\left[(\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{P}_r\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\times\left(\sum_{i=r+1}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i - \sum_{i=r+1}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i\right)\right]\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^- \\
&= \mathbf{B}_r - E\Big[(\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{P}_r\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\times \\
&\left((\mathbf{I}-\mathbf{T}_{r+1})E[\mathbf{Y}\mathbf{C}_{r+1}^T(\mathbf{C}_{r+1}\mathbf{C}_{r+1}^T)^-\mathbf{C}_{r+1}] + \mathbf{T}_{r+1}\sum_{i=r+2}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i - \sum_{i=r+1}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i\right)\Bigg]\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^- \\
&= \mathbf{B}_r - E\Big[(\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{P}_r\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{T}_{r+1}\times \\
&\left(\sum_{i=r+2}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i - \sum_{i=r+2}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i\right)\Bigg]\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^- \\
&= \mathbf{B}_r - E\Big[(\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{P}_r\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{T}_{r+1}\times \\
&\Big((\mathbf{I}-\mathbf{T}_{r+2})E[\mathbf{Y}\mathbf{C}_{r+2}^T(\mathbf{C}_{r+2}\mathbf{C}_{r+2}^T)^-\mathbf{C}_{r+2}] \\
&+ \sum_{i=r+2}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i - \mathbf{T}_{r+2}\sum_{i=r+2}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i - \sum_{i=r+2}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i\Bigg)\Bigg]\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^- \\
&= \mathbf{B}_r - E\Big[(\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{P}_r\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{P}_r^T\mathbf{S}_r^{-1}\mathbf{T}_{r+1}\mathbf{T}_{r+2}\times \\
&\left(\sum_{i=r+3}^{m}\mathbf{X}_i\hat{\mathbf{B}}_i\mathbf{C}_i - \sum_{i=r+3}^{m}\mathbf{X}_i\mathbf{B}_i\mathbf{C}_i\right)\Bigg]\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^- \\
&= \ldots = \mathbf{B}_r
\end{aligned}
$$

which gives the following theorem.

Theorem 2.12. The estimator $\hat{\mathbf{B}}_r$ in Theorem 2.5 is an unbiased estimator under the uniqueness assumptions.

The dispersion matrix is a bit more complicated to derive. The idea is to look at the linear combinations

$$
\mathbf{G}_{r-1}^T\mathbf{X}_r\hat{\mathbf{B}}_r\mathbf{C}_r \tag{2.77}
$$

that we can do since

$$(\mathbf{X}_r^T\mathbf{G}_{r-1}\mathbf{G}_{r-1}^T\mathbf{X}_r)^-\mathbf{X}_r^T\mathbf{G}_{r-1}\mathbf{G}_{r-1}^T\mathbf{X}_r\hat{\mathbf{B}}_r\mathbf{C}_r\mathbf{C}_r^T(\mathbf{C}_r\mathbf{C}_r^T)^- = \hat{\mathbf{B}}_r \tag{2.78}$$

where $\mathbf{G}_{r+1} = \mathbf{G}_r(\mathbf{G}_r^T\mathbf{X}_{r+1})^o$. Since $\hat{\mathbf{B}}_r$ is an unbiased estimator we consider

$$\mathbf{G}_{r-1}\mathbf{X}_r(\hat{\mathbf{B}}_r - \mathbf{B}_r)\mathbf{C}_r. \tag{2.79}$$

Then one can transform this expression and decompose the dispersion in the following

$$\begin{aligned}
D[\mathbf{G}_{r-1}\mathbf{X}_r(\hat{\mathbf{B}}_r - \mathbf{B}_r)\mathbf{C}_r] &= D[\mathbf{G}_{r-1}^T\mathbf{R}_{r-1}(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{P}_{\mathbf{C}_r^T} - \mathbf{P}_{\mathbf{C}_{r+1}^T})] \\
&+ D[\mathbf{G}_{r-1}\mathbf{R}_{r-1}(\mathbf{I} - \mathbf{R}_r)(\mathbf{Y} - E[\mathbf{Y}])\mathbf{P}_{\mathbf{C}_{r+1}^T}] \\
&+ D[\mathbf{G}_{r-1}\mathbf{R}_{r-1}(\mathbf{I} - \mathbf{R}_r)(\mathbf{I} - \mathbf{R}_{r+1})(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{P}_{\mathbf{C}_{r+2}^T} - \mathbf{P}_{\mathbf{C}_{r+3}^T})] \\
&+ D[\mathbf{G}_{r-1}\mathbf{R}_{r-1}(\mathbf{I} - \mathbf{R}_r)(\mathbf{I} - \mathbf{R}_{r+1})(\mathbf{I} - \mathbf{R}_{r+2})(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{P}_{\mathbf{C}_{r+3}^T} - \mathbf{P}_{\mathbf{C}_{r+4}^T})] \\
&\vdots
\end{aligned}$$

where $\mathbf{R}_{r-1} = \mathbf{X}_r(\mathbf{X}_r^T\mathbf{G}_{r-1}(\mathbf{G}_{r-1}^T\mathbf{W}_r\mathbf{G}_{r-1})^-\mathbf{G}_{r-1}^T\mathbf{X}_r^T)^-\mathbf{X}_r^T\mathbf{G}_{r-1}(\mathbf{G}_{r-1}^T\mathbf{W}_r\mathbf{G}_{r-1})^-\mathbf{G}_{r-1}^T$. Then one can treat each expression differently and derive the dispersion matrix expression.

Now we have given ideas for the distribution of $\hat{\mathbf{B}}$ and hence we need to derive the distribution of $\hat{\mathbf{\Sigma}}$. This is stated in the following Theorem.

Theorem 2.13. Let $\mathbf{K}_i, i = 1, 2, \ldots, m$ be defined as in Theorem 2.5 and let

$$\begin{aligned}
c_{r-1} &= \frac{n - r(\mathbf{C}_r) - m_{r-1} - 1}{n - r(\mathbf{C}_{r-1}) - m_{r-1} - 1} \\
m_r &= p - r(\mathbf{X}_1 : \ldots : \mathbf{X}_r) + r(\mathbf{X}_1 : \ldots : \mathbf{X}_{r-1}).
\end{aligned}$$

Let $\mathbf{C}_0 = \mathbf{0}$ and put

$$\begin{aligned}
g_{i,j} &= \frac{c_i c_{i+1} \times \cdots \times c_{j-1} m_j}{n - r(\mathbf{C}_j) - m_j - 1} \quad i < j \\
g_{j,j} &= \frac{m_j}{n - r(\mathbf{C}_j) - m_j - 1}
\end{aligned}$$

which are supposed to be finite and positive. Then for $\hat{\mathbf{\Sigma}}$ given in Theorem 2.5 it holds that

$$\begin{aligned}
E[n\hat{\mathbf{\Sigma}}] = \sum_{j=1}^{m}(r(\mathbf{C}_{j-1}) - r(\mathbf{C}_j)) \times \left(\sum_{i=1}^{j-1} g_{i,j-1}\mathbf{K}_i + \mathbf{\Sigma}\mathbf{G}_{j-1}(\mathbf{G}_{j-1}^T\mathbf{\Sigma}\mathbf{G}_{j-1})^{-1}\mathbf{G}_{j-1}^T\mathbf{\Sigma}\right) \\
+ r(\mathbf{C}_m)\left(\sum_{i=1}^{m} g_{i,m}\mathbf{K}_i + \mathbf{\Sigma}\mathbf{G}_{r-1}(\mathbf{G}_{r-1}^T\mathbf{\Sigma}\mathbf{G}_{r-1})^{-1}\mathbf{G}_{r-1}^T\mathbf{\Sigma}\right)
\end{aligned}$$

where $\sum_{i=1}^{0} g_{i,j-1}\mathbf{K}_i = \mathbf{0}$.

Proof. From Theorem 2.5 it follows that the expectation of $\mathbf{P}_j\mathbf{Y}\mathbf{F}_j\mathbf{F}_j^T\mathbf{Y}^T\mathbf{P}_j^T$, $j = 1, 2, \ldots, m$ and $\mathbf{P}_{m+1}\mathbf{Y}\mathbf{P}_{\mathbf{C}_m^T}\mathbf{Y}^T\mathbf{P}_{m+1}^T$ are needed where $\mathbf{F}_j = \mathbf{P}_{\mathbf{C}_{j-1}^T}(\mathbf{I} - \mathbf{P}_{\mathbf{C}_j^T})$. Since $\mathbf{Y}\mathbf{F}_j\mathbf{F}_j^T\mathbf{Y}^T$ and $\mathbf{Y}\mathbf{P}_{\mathbf{C}_m^T}\mathbf{Y}^T$ are independent of $\mathbf{P}_j$ and $\mathbf{P}_{m+1}$ respectively and

$$E[\mathbf{Y}\mathbf{F}_j\mathbf{F}_j^T\mathbf{Y}^T] = (r(\mathbf{C}_{j-1}) - r(\mathbf{C}_j))\mathbf{\Sigma}$$
$$E[\mathbf{Y}\mathbf{P}_{\mathbf{C}_m^T}\mathbf{Y}^T] = r(\mathbf{C}_m)\mathbf{\Sigma}$$

since

$$E[\mathbf{P}_r\mathbf{\Sigma}\mathbf{P}_r^T] = \sum_{i=1}^{r-1} g_{i,r-1}\mathbf{K}_i + \mathbf{\Sigma}\mathbf{G}_{r-1}(\mathbf{G}_{r-1}^T\mathbf{\Sigma}\mathbf{G}_{r-1})^{-1}\mathbf{G}_{r-1}^T\mathbf{\Sigma}, \quad r = 1, 2, \ldots, m+1. \tag{2.80}$$

□

Like for the GMANOVA this is not an unbiased estimate. An unbiased estimate of $\hat{\mathbf{\Sigma}}$ is given as

$$\hat{\mathbf{\Sigma}} = \tilde{\hat{\mathbf{\Sigma}}} + \frac{1}{n}\sum_{j=1}^{m}(r(\mathbf{C}_{j-1}) - r(\mathbf{C}_j))\sum_{i=1}^{j-1} k_{i,j}\mathbf{X}_i(\mathbf{X}_i^T\mathbf{P}_i^T\hat{\mathbf{\Sigma}}^{-1}\mathbf{P}_i\mathbf{X}_i)^{-}\mathbf{X}_i^T$$
$$+ \frac{1}{n}r(\mathbf{C}_m)\sum_{i=1}^{m} k_{i,m+1}\mathbf{X}_i(\mathbf{X}_i^T\mathbf{P}_i^T\hat{\mathbf{\Sigma}}^{-1}\mathbf{P}_i\mathbf{X}_i)^{-}\mathbf{X}_i^T$$

where $0 < k_{i,j} < \infty$ is given as

$$k_{i,j} = n(g_{i+1,j-1} - g_{i,j-1})/(n - r(\mathbf{C}_i) - m_i) \tag{2.81}$$

and where $\sum_{i=1}^{0} k_{i,j}\mathbf{X}_i(\mathbf{X}_i^T\mathbf{P}_i^T\hat{\mathbf{\Sigma}}^{-1}\mathbf{P}_i\mathbf{X}_i)^{-}\mathbf{X}_i^T = \mathbf{0}$. The $\tilde{\hat{\mathbf{\Sigma}}}$ is the estimate from Theorem 2.5.

So now we have the necessary building blocks for approximating the distributions of the estimators. These will be investigated in the next section.

2.6 Approximations Of The Distributions of Estimators

In this section we are going to derive the distributions of several estimators of the GMANOVA. We will not be deriving the distributions of the extended GMANOVA as this is complicated. We refer to [2] for a derivation of this. Let $f_{\mathbf{X}}(\mathbf{X}_0)$ denote the density of $\mathbf{X}$ evaluated in the point $\mathbf{X}_0$ and let $f_{\mathbf{X}}^{(k)}(\mathbf{X}_0)$ denote the k'th derivative of $f_{\mathbf{X}}(\mathbf{X})$ evaluated at $\mathbf{X}_0$.

2.6.1 Approximation Of Distribution Of Estimators For GMANOVA

We remember that the estimates for the GMANOVA is given as

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}$$
$$\mathbf{S} = \mathbf{Y}(\mathbf{I} - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C})\mathbf{Y}^T$$
$$n\hat{\mathbf{\Sigma}} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\mathbf{C})(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\mathbf{C})^T = \mathbf{S} + \hat{\mathbf{V}}\hat{\mathbf{V}}^T$$
$$\mathbf{V} = \mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C} - \mathbf{X}\hat{\mathbf{B}}\mathbf{C}$$

under the uniqueness conditions. The approach that we are going to be taking is approximating the distribution by asymptotic considerations.

The first thing to note is that

$$\frac{1}{n-k}\mathbf{S} \xrightarrow{\text{P}} \boldsymbol{\Sigma} \quad \text{as} \quad n \to \infty \tag{2.82}$$

since all bias in Theorem 2.11 will go to zero. Hence we can approximate $\hat{\mathbf{B}}$ by

$$\hat{\mathbf{B}}_N = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \tag{2.83}$$

Since (2.83) is a linear equation in $\mathbf{Y}$, the distribution of $\hat{\mathbf{B}}$ will be approximated by a normal distribution $\hat{\mathbf{B}}_N \sim N_{q,k}(\mathbf{B}, (\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X})^{-}, (\mathbf{C}\mathbf{C}^T)^{-})$. By using that $\hat{\mathbf{B}}_N$ is matrix normal distributed and the fact that $D[\mathbf{X}] = \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}$ when $\mathbf{X} \sim N_{q,k}(\mu, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ we get that $D[\hat{\mathbf{B}}_N] = (\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X})^{-} \otimes (\mathbf{C}\mathbf{C}^T)^{-}$. Comparing $\hat{\mathbf{B}}$ and $\hat{\mathbf{B}}_N$ we get from Theorem 2.10 that

$$\begin{aligned}
E[\hat{\mathbf{B}}] &= E[\hat{\mathbf{B}}_N] = \mathbf{B} \\
D[\hat{\mathbf{B}}] - D[\hat{\mathbf{B}}_N] &= \frac{(n-k-1)-(n-k-p+q-1)}{n-k-p+q-1}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-} \otimes (\mathbf{C}\mathbf{C}^T)^{-} \\
&= \frac{p-q}{n-k-p+q-1}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-} \otimes (\mathbf{C}\mathbf{C}^T)^{-}
\end{aligned}$$

where the last difference is positive definite. Hence $D[\hat{\mathbf{B}}_N]$ underestimate $D[\hat{\mathbf{B}}]$, which was expected as the random matrix $\mathbf{S}$ is replaced by the non-random matrix $\boldsymbol{\Sigma}$. Now we are going to use the normal distribution with a correction term. Observe that

$$\begin{aligned}
\hat{\mathbf{B}} &= (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \\
&+ (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1})\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}
\end{aligned}$$

Theorem A.8 bullet 2 yields that

$$(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \tag{2.84}$$

and

$$(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1})\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \tag{2.85}$$

are independently distributed since $\boldsymbol{\Psi} = \mathbf{I}_n$. Moreover the same Theorem bullet 4 also gives that $\mathbf{S}$ and $\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}$ are independent. Define

$$\mathbf{U} = (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1})\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \tag{2.86}$$

which is independent of $\hat{\mathbf{B}}_N$ since (2.84) and (2.85) were independently distributed which are the core components of $\hat{\mathbf{B}}_N$ and $\mathbf{U}$ respectively. Furthermore they also project on orthogonal subspaces. Then we get that $\hat{\mathbf{B}} = \hat{\mathbf{B}}_N + \mathbf{U}$ and $E[\mathbf{U}] = \mathbf{0}$. This can be seen as $E[\mathbf{Y}] = \mathbf{X}\hat{\mathbf{B}}\mathbf{C}$ then

$$\begin{aligned}
E[\mathbf{U}] &= E[(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1})\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}] \\
&= (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1})\mathbf{X}\hat{\mathbf{B}}\mathbf{C}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \\
&= (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1})\mathbf{X}\hat{\mathbf{B}} \\
&= (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}(\mathbf{X}\hat{\mathbf{B}}-\mathbf{X}\hat{\mathbf{B}}) \\
&= (\hat{\mathbf{B}}-\hat{\mathbf{B}}) \\
&= \mathbf{0}
\end{aligned}$$

Hence it makes sense to approximate $\hat{\mathbf{B}}$ by $\hat{\mathbf{B}}_N$.

Theorem 2.14. Let $\hat{\mathbf{B}}, \hat{\mathbf{B}}_N, \mathbf{U}$ be defined as above, then an Edgeworth-type expansion of the density of $\hat{\mathbf{B}}$ equals

$$f_{\hat{\mathbf{B}}}(\mathbf{B}_0) = f_{\mathbf{B}_E}(\mathbf{B}_0) + \cdots \tag{2.87}$$

where

$$\begin{aligned}
f_{\mathbf{B}_E}(\mathbf{B}_0) &= \left(1+\frac{1}{2}s(tr(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{B}_0-\mathbf{B})\mathbf{C}\mathbf{C}^T(\mathbf{B}_0-\mathbf{B}))-kq)\right)f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0) \\
s &= \frac{p-q}{n-k-p+q-1}
\end{aligned}$$

Proof. The form of the approximation

$$f_{\mathbf{B}_E}(\mathbf{B}_0) = f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0) + (-1)^2\frac{1}{2!}E[(vec^T\mathbf{U})^{\otimes 2}]vecf^{(2)}_{\hat{\mathbf{B}}_N}(\mathbf{B}_0) - \cdots \tag{2.88}$$

follows from Theorem A.16 according to [2]. As noted before

$$\hat{\mathbf{B}} = \hat{\mathbf{B}}_N + \mathbf{U} \tag{2.89}$$

which is identical to $\hat{\mathbf{B}} - \mathbf{B} = \hat{\mathbf{B}}_N - \mathbf{B} + \mathbf{U}$ and because of the independence between

$$\hat{\mathbf{B}}_N = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \tag{2.90}$$

and

$$\mathbf{U} = (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}^T\boldsymbol{\Sigma}^{-1})\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-} \tag{2.91}$$

as it was shown above the theorem, it follows that

$$E[(vec\mathbf{U})^{\otimes 2}] = vec(D[\hat{\mathbf{B}}] - D[\hat{\mathbf{B}}_N]) \tag{2.92}$$

since $vec(E[vec\mathbf{U}vec^T\mathbf{U}]) = E[(vec\mathbf{U})^{\otimes 2}]$. We have from Theorem 2.10 how $D[\hat{\mathbf{B}}]$ was obtained and below (2.83) how $D[\hat{\mathbf{B}}_N]$ was obtained. The last term is the differentiated density according to $\hat{\mathbf{B}}_N$ written as $f^{(2)}_{\hat{\mathbf{B}}_N}(\mathbf{B}_0)$ which is equal to

$$\begin{aligned}
\frac{d^{(2)} f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0)}{d\mathbf{B}_0^{(2)}} &= -\frac{d(f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0)\boldsymbol{\Sigma}^{-1}(\mathbf{B}_0-\boldsymbol{\mu}))}{d\mathbf{B}_0} \\
&= -\frac{d(\mathbf{B}_0-\boldsymbol{\mu})^T}{d\mathbf{B}_0}\boldsymbol{\Sigma}^{-1} f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0) - \frac{d f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0)}{d\mathbf{B}_0}(\mathbf{B}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1} \\
&= -\boldsymbol{\Sigma}^{-1} f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0) + \boldsymbol{\Sigma}^{-1}(\mathbf{B}_0-\boldsymbol{\mu})(\mathbf{B}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1} f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0) \\
&= f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0)(\boldsymbol{\Sigma}^{-1}(\mathbf{B}_0-\boldsymbol{\mu})(\mathbf{B}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}) \\
&= f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0)(\boldsymbol{\Sigma}^{-1}(\mathbf{B}_0-\mathbf{B})(\mathbf{B}_0-\mathbf{B})^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1})
\end{aligned}$$

since $E[\hat{\mathbf{B}}_N] = \mathbf{B}$. Then we get that

$$\begin{aligned}
f_{\mathbf{B}_E}(\mathbf{B}_0) &= f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0) + (-1)^2\frac{1}{2!}E[(vec^T\mathbf{U})^{\otimes 2}]vec f^{(2)}_{\hat{\mathbf{B}}_N}(\mathbf{B}_0) \\
&= f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0) + \frac{1}{2}vec^T\left(\frac{p-q}{n-k-p+q-1}(\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X})^- \otimes (\mathbf{C}\mathbf{C}^T)^-\right) \\
&\cdot vec\left(f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0)(\boldsymbol{\Sigma}^{-1}(\mathbf{B}_0-\boldsymbol{\mu})(\mathbf{B}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1})\right) \\
&= f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0)\left(1 + \frac{1}{2}s\left(vec^T((\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X})^- \otimes (\mathbf{C}\mathbf{C}^T)^-)\right.\right. \\
&\left.\left.\cdot vec(\boldsymbol{\Sigma}^{-1}(\mathbf{B}_0-\boldsymbol{\mu})(\mathbf{B}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1})\right)\right) \\
&= \left(1 + \frac{1}{2}s(tr(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{B}_0-\mathbf{B})\mathbf{C}\mathbf{C}^T(\mathbf{B}_0-\mathbf{B})) - kq)\right) f_{\hat{\mathbf{B}}_N}(\mathbf{B}_0)
\end{aligned}$$

since the vec operator only cares about the dimensions and hence the trace. The kq comes from the dimensions of $(\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X})^- \otimes (\mathbf{C}\mathbf{C}^T)^- \cdot \boldsymbol{\Sigma}^{-1}$. □

We have now proved that we can approximate $\hat{\mathbf{B}}$ by $\hat{\mathbf{B}}_N$, and hence use this distribution as the true distribution. So now we approximated a distribution which we can make a Wald test to see if any of the coefficients are significant.

2.7 Wald Test

This section will derive the Wald-test for the GMANOVA. The Wald-test is a test for if a predictor could be more or less influential. For this, we can use the Mahalanobis distance. First we will derive that the squared Mahalanobis distance is chi squared distributed for a Gaussian distribution. This is based on [4]. The Mahalanobis distance is defined as

$$d(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{y})} \tag{2.93}$$

then replacing $\mathbf{x}$ by $\mathbf{X}$ and $\mathbf{y}$ by the mean of the distribution $\boldsymbol{\mu}$ we get

$$D = d(\mathbf{X},\boldsymbol{\mu}) = (\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \tag{2.94}$$

where $\mathbf{\Sigma}$ is a $p \times p$ covariance matrix. To obtain a different representation of $\mathbf{\Sigma}$ a eigenvalue decomposition is performed

$$\begin{aligned}\mathbf{\Sigma} &= \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \\ &= \sum_{k=1}^{p} \lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^T.\end{aligned}$$

Then using this, we get

$$\begin{aligned}D &= (\mathbf{X}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \\ &= (\mathbf{X}-\boldsymbol{\mu})^T(\sum_{k=1}^{p} \lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^T)(\mathbf{X}-\boldsymbol{\mu}) \\ &= \sum_{k=1}^{p} \lambda_k^{-1}(\mathbf{X}-\boldsymbol{\mu})^T\mathbf{u}_k\mathbf{u}_k^T(\mathbf{X}-\boldsymbol{\mu}) \\ &= \sum_{k=1}^{p}(\lambda_k^{-1/2}(\mathbf{X}-\boldsymbol{\mu})^T\mathbf{u}_k\mathbf{u}_k^T(\mathbf{X}-\boldsymbol{\mu}))^2 \\ &= \sum_{k=1}^{p} \mathbf{Y}_k^2\end{aligned}$$

where $\mathbf{Y}_k$ is a new random variable based on an affine linear transformation of $\mathbf{X}$. If we set $\mathbf{a}_k^T = \lambda_k^{-1/2}\mathbf{u}_k$ and $\mathbf{Z} = (\mathbf{X}-\boldsymbol{\mu})$ we get that $\mathbf{Y}_k = \mathbf{a}_k^T\mathbf{Z}$. Note that $\mathbf{Y}_k \sim N(0, \sigma_k^2)$ where σ_k^2 is equal to

$$\begin{aligned}\sigma_k^2 &= \mathbf{a}_k^T\mathbf{\Sigma}\mathbf{a}_k \\ &= \lambda_k^{-1}\mathbf{u}_k^T\mathbf{\Sigma}\mathbf{u}_k \\ &= \lambda_k^{-1}\mathbf{u}_k^T(\sum_{j=1}^{p} \lambda_j^{-1}\mathbf{u}_j\mathbf{u}_j^T)\mathbf{u}_k \\ &= \sum_{j=1}^{p} \lambda_k^{-1}\lambda_j\mathbf{u}_k^T\mathbf{u}_j\mathbf{u}_j^T\mathbf{u}_k\end{aligned}$$

where $\mathbf{u}_i$ are eigenvectors, the dotted products will be 0 for $j \neq k$. When $j = k$ we get that $\sigma_k^2 = 1$ and hence $\mathbf{Y}_k \sim N(0,1)$. By the same arguments as above we get that

$$\sigma_{jk}^2 = \mathbf{a}_j^T\mathbf{\Sigma}\mathbf{a}_k = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases} \tag{2.95}$$

which makes the $\mathbf{Y}_k$ independent of $\mathbf{Y}_j$. Hence we get that

$$D = \sum_{k=1}^{p} \mathbf{Y}_k^2 \tag{2.96}$$

which is a χ^2 distribution with p degrees of freedom.

So now we know that the squared Mahalanobis distance is χ^2 distributed so we need to plug in what values that we have for the test. So setting the $\mathbf{X}$ in (2.94) to be the i'th column in $\hat{\mathbf{B}}$ and we have derived the $\mathbf{\Sigma}$ in (2.55) as the dispersion of $\hat{\mathbf{B}}$. The $\boldsymbol{\mu}$ can be replaced by what we are testing for; it could be if the i'th column of $\hat{\mathbf{B}}$ has no significant effect i.e. if it is zero. It could also be if it should have a bigger impact, hence giving it a greater value than estimated.

Chapter 3

Implementation Of The GMANOVA

This chapter will be focusing on implementing the GMANOVA into *R*. It will be using the theory described in the former chapter. We remember that the estimates were given as

$$\begin{aligned}\hat{\mathbf{B}} &= (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\\ \mathbf{S} &= \mathbf{Y}(\mathbf{I}-\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C})\mathbf{Y}^T\\ n\hat{\mathbf{\Sigma}} &= (\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\mathbf{C})(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\mathbf{C})^T = \mathbf{S}+\hat{\mathbf{V}}\hat{\mathbf{V}}^T\\ \mathbf{V} &= \mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-}\mathbf{C}-\mathbf{X}\hat{\mathbf{B}}\mathbf{C}\end{aligned}$$

We will both implement the GMANOVA and a likelihood ratio test.

3.1 Implementation

This section will be focusing on implementing the GMANOVA into *R*. Note that we do assume full rank for $\mathbf{X}$ and $\mathbf{C}$ hence the estimate of $\hat{\mathbf{B}}$ is unique. The first thing to do is making a design matrix, which is done by the command **model.matrix** for a given formula. Then we need the $\mathbf{C}$ matrix which is a between-individuals matrix. So for an example of $\mathbf{C}$ let us divide the n measurements into three groups where the number of observations in these groups are n_1, n_2, n_3. Then the $\mathbf{C}$ matrix could look like this for this example

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0\\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0\\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 \end{bmatrix} \tag{3.1}$$

where the first 1's are replicated n_1 times, and the second row has n_2 1's and so on, for $\mathbf{C}$ of size $3 \times (n_1 + n_2 + n_3)$ because there are 3 groups. If there is only one intercept, the $\mathbf{C}$ matrix reduces to a $\mathbf{1}_n^T$ vector, which was the same as in our example. So each row represents a site in our data. That means that the $\mathbf{C}$ matrix is a **model.matrix** with the sites, and the $\mathbf{X}$ matrix is a **model.matrix** with our SNPs.

Next thing is the estimation of $\mathbf{S}$ since this is used in the estimation of $\hat{\mathbf{B}}$. So this is written as

$$\mathbf{S} = \mathbf{Y}(\mathbf{I}-\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{C})\mathbf{Y}^T/n \tag{3.2}$$

It has also been divided by n which is done for the $\hat{\mathbf{\Sigma}}$ to be easier to estimate later on. Then we can estimate $\hat{\mathbf{B}}$ as

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{S}^{-1}\mathbf{Y}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1} \tag{3.3}$$

which again is the exact same as in (2.19). Then we can estimate the data by

$$\mathbf{X}\hat{\mathbf{B}}\mathbf{C}. \tag{3.4}$$

The code looks like this

```
gmanova <- function(f1,f2,y,x1,x2){
    x1$y <- 1
    C <- model.matrix(f1,data=x1)
    x2$y <- 1
    X <- model.matrix(f2,data=x2)
    tC <- t.default(C)
    D1 <- tC%*%ginv(C%*%tC)
    n <- dim(y)[2]
    Sinv <- solve(y%*%(diag(1,n)-D1%*%C)%*%t.default(y)/n)
    tX <- t.default(X)
    B <- ginv(tX%*%Sinv%*%X)%*%tX%*%Sinv%*%y%*%D1
    colnames(B) <- rownames(C);rownames(B) <- colnames(X)
    pred <- X%*%B%*%C
}
```

Then we need to estimate the $\mathbf{\Sigma}$ matrix, but that is just the error term, so it can be estimated from data by the variance of the residuals. But this is unbiased as we saw in Chapter 2, so we have two different estimates; one that is unbiased and one that is biased. The unbiased estimate is given in Section 2.5.1, (2.76). So the code looks like the following

```
res <- y -pred
    if(unbiased.estimate == T){
      sigma <- (1/(n-qr(C)$rank))*y%*%(diag(1,n)-D1%*%C)%*%t.default(y)
    }else{
      sigma <- (y - pred)%*%t.default(y-pred)/n
    }
```

which estimates the $\hat{\mathbf{\Sigma}}$. Then combining these two, we get the full GMANOVA function

```
gmanova <- function(f1,f2,y,x1,x2, unbiased.estimate = T){
    x1$y <- 1
    C <- model.matrix(f1,data=x1)
    x2$y <- 1
    X <- model.matrix(f2,data=x2)
    tC <- t.default(C)
    D1 <- tC%*%ginv(C%*%tC)
```

```
    n <- dim(y)[2]
    Sinv <- solve(y%*%(diag(1,n)-D1%*%C)%*%t.default(y)/n)
    tX <- t.default(X)
    B <- ginv(tX%*%Sinv%*%X)%*%tX%*%Sinv%*%y%*%D1
    colnames(B) <- rownames(C);rownames(B) <- colnames(X)
    pred <- X%*%B%*%C
    res <- y -pred
    if(unbiased.estimate == T){
      sigma <- (1/(n-qr(C)$rank))*y%*%(diag(1,n)-D1%*%C)%*%t.default(y)
    }else{
      sigma <- (y - pred)%*%t.default(y-pred)/n
    }
    list(beta=B,sigma=sigma,res=res,y=y,model=list(f1=f1,f2=f2))
}
```

We can also check that the nested models $m_0 \subset m_1$ are not significantly different by a likelihood ratio test. This is straight forward and implemented as follows

```
test.gma <- function(m0,m1){
    df <- prod(dim(m1$beta))-prod(dim(m0$beta))
    n <- dim(m1$y)[1]
    stat <- n*(determinant(m0$sigma)$modulus[1]
    -determinant(m1$sigma)$modulus[1])
    pval <- 1-pchisq(stat,df)
    list(test.stat=stat,df=df,pval=pval)
}
```

The reason for the p-value to be calculated as that, is because the *stat* part is approximately χ^2 distributed. This is due to the following theorem from [5].

Theorem 3.1 (Wilks Theorem)**.** If a population with a variate x is distributed according to the probability function $f(x, \theta_1, \theta_2, \dots, \theta_h)$, such that optimum estimates θ_i of the θ, exist which are distributed in large samples according to $\frac{|c_{ij}|^{\frac{1}{2}}}{(2\pi)^{h/2}} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{h} c_{ij} z_i z_j\right)$, where $z_i = (\hat{\theta}_i - \theta_i)\sqrt{n}, c_{ij} = -E[\frac{\partial^2 \log f}{\partial\theta_i \partial\theta_j}]$ then when the hypothesis $\mathcal{H}$ is true that $\theta_i = \theta_{0i}, i = m+1, m+2, \cdots, h$ where $\hat{\theta}_0 = (\theta_{01}, \dots, \theta_{0m}, \hat{\hat{\theta}}_{m+1}, \dots, \hat{\hat{\theta}}_n)$, the distribution of $-2\log\lambda$, where λ is given by $\frac{l(x;\theta_0)}{l(x;\theta)}$ is, except for terms of order $1/\sqrt{n}$, distributed like χ^2 with $h - m$ degrees of freedom.

This is a asymptotic result which states that the Likelihood Ratio test is asymptotically χ^2 distributed with $dim(m_1) - dim(m_0)$ degrees of freedom for nested models $m_0 \subset m_1$. Then since this handles the less extreme measurements, one could do as in the code to get the $\alpha\%$ most extreme cases, where α is the significance level.

The reason for the quotient $\frac{l(x;\theta_0)}{l(x;\theta)}$ to be written in the code as $n \cdot (\det(\mathbf{\Sigma}) - \det(\mathbf{\Sigma}_0))$

is due to the fact that the likelihood is given as

$$L(\mathbf{B}, \mathbf{\Sigma}) = |\mathbf{\Sigma}|^{n/2} \exp\left(-\frac{1}{2}tr((\mathbf{Y} - \mathbf{XBC})^T\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{XBC}))\right) \tag{3.5}$$

and when replacing $\mathbf{B}$ by the estimate $\hat{\mathbf{B}}$ the likelihood becomes

$$L(\hat{\mathbf{B}}, \mathbf{\Sigma}) = |\mathbf{\Sigma}|^{n/2} \exp\left(-\frac{1}{2}tr(\mathbf{\Sigma}^{-1}n\mathbf{\Sigma})\right) \tag{3.6}$$

so when replacing $\mathbf{\Sigma}$ by its estimate $\hat{\mathbf{\Sigma}}$ we get

$$L(\hat{\mathbf{B}}, \hat{\mathbf{\Sigma}}) = |\mathbf{\Sigma}|^{n/2} \exp\left(-\frac{np}{2}\right). \tag{3.7}$$

Then one can see that the quotient is

$$Q = \frac{L(x;\theta_0)}{L(x;\theta)} = \frac{|\hat{\mathbf{\Sigma}}_0|^{-n/2}}{|\hat{\mathbf{\Sigma}}|^{-n/2}}. \tag{3.8}$$

So the likelihood ratio test becomes

$$-2\log Q = n(\log(\hat{\mathbf{\Sigma}}_0) - \hat{\mathbf{\Sigma}}) \tag{3.9}$$

which is approximately χ^2 distributed with the number of predictors in $\mathbf{B}$ minus the number of predictors in $\mathbf{B}_0$ degrees of freedom.

Before proceeding we need to check that the code actually works and do as it is supposed to. This is done in the next section for a simple example.

3.2 Example

In this section we will be giving a simple example of 2 sites with 1 repetition and 1 SNP. There will be three subjects with 3 different combinations of phenotypes. We will in this example choose the SNP *rs1015362*. So we have that the $\mathbf{X}$ becomes

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \tag{3.10}$$

and that the $\mathbf{C}$ becomes

$$\mathbf{C} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \tag{3.11}$$

The matrix $\hat{\mathbf{B}}$ is the one that we need to estimate, so this can be written as

$$\hat{\mathbf{B}} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}. \tag{3.12}$$

If we take the product **BC** we see that

$$\mathbf{BC} = \begin{bmatrix} b_{11} & b_{11}+b_{12} \\ b_{21} & b_{21}+b_{22} \\ b_{31} & b_{31}+b_{32} \end{bmatrix} \tag{3.13}$$

which is the site effect on each estimate. So b_{11} is the estimate for the standard level site, which is arm, and then $b_{11} + b_{12}$ is the estimate of buttocks. So b_{12} is the effect of going from arm to buttock. Then taking the **X** on this product, gives the estimates for each person, which is the 3 people in this example.

The **y** vector in this given example for three different people chosen at random is given by

$$\mathbf{y} = \begin{bmatrix} 7.20 & 6.50 \\ 8.70 & 6.90 \\ 7.40 & 5.30 \end{bmatrix} \tag{3.14}$$

which is a observation at both places for each person. After doing the formula in Chapter 3.1 we get that the $\hat{\mathbf{B}}$ matrix looks like

	(Intercept)	sitebuttock
(Intercept)	7.40	-2.10
rs1015362CT	1.30	0.30
rs1015362T	-0.20	1.40

which is read as the pigmentation is lower on the buttock than the arm for the standard level, since the standard level intercept on buttock is $7.40 - 2.10 = 5.30$. Furthermore you are getting darker on both sites when you have the CT combination than the CC combination which is the standard level. Then on the arm you are getting lower pigmentation if you have the TT combination compared to standard level. Both places it raises the pigmentation on the buttocks compared to the standard level.

This estimate of $\hat{\mathbf{B}}$ is the same that we get for our example, which makes the function valid and ready to use for our data.

Chapter 4

Analysis

This chapter will be using the GMANOVA command described in Chapter 3 to analyse the dataset described in Section 1.1. First we will be doing some elementary analysis and work our way from there. It should be noted that in section 3.2 it was described how the $\mathbf{X}, \mathbf{B}$ and $\mathbf{C}$ matrices looked like, so this will not be given at any point.

4.1 Preliminary Analysis

In this section we will be doing the initial analysis. The first thought is just doing the analysis with no assumptions. This section will go into detail about every aspect of the analysis, whereas the next sections will not be going into detail, unless an exciting finding is done. So doing the analysis without assumptions results in the following table. That is all SNPs are used and all sites are used without any restrictions.

	(Intercept)	sitebuttock	siteface
(Intercept)	12.88	1.89	2.72
rs1015362CT	0.75	-0.17	-0.31
rs1015362T	-0.02	0.23	-0.16
rs10777129CT	0.13	0.14	0.30
rs10777129T	0.08	0.79	0.49
rs10831496CT	-0.16	-0.35	-0.38
rs10831496T	-0.32	-0.09	-0.24
rs11238349G	-0.66	-0.74	-0.84
rs11238349GA	-0.57	-0.67	-0.47
rs12203592CT	-0.40	-0.37	-0.06
rs12203592T	0.39	-0.25	0.61
rs12350739G	0.34	0.32	0.09
rs12350739GA	0.45	0.30	0.07
rs12668421AT	-0.36	-0.42	-0.28
rs12668421T	-1.09	-0.50	-0.64
rs12896399C	0.34	-0.09	0.04
rs12896399CA	0.70	-0.18	-0.08
rs12913832G	-0.55	-0.44	-0.20
rs12913832GA	-0.66	-0.22	-0.02
rs13289G	0.22	0.09	0.11
rs13289GC	-0.16	0.22	0.16
rs13933350CT	-0.23	0.05	0.32
rs13933350T	-0.05	0.22	-0.35
rs1408799G	-0.43	-0.54	-0.40
rs1408799GA	0.07	-0.48	-0.20
rs1426654G	2.59	-0.86	-1.09
rs1426654GA	-0.11	-0.00	-0.20
rs16891982G	0.83	0.64	-0.18
rs16891982GC	1.01	1.09	-0.13
rs1800407CT	-1.10	-0.50	-0.19
rs1800407T	-2.28	1.25	0.77
rs2031526CT	-0.18	0.10	0.12
rs2031526T	-0.28	0.38	-0.15
rs2424984CT	-1.41	-0.44	0.16
rs2424984T	-1.45	-0.02	0.05
rs2470102G	-1.71	1.22	1.08
rs2470102GA	0.88	-0.08	0.25
rs26722G	-3.13	-1.95	-1.12
rs26722GA	-3.00	-1.64	-1.00
rs4424881CT	0.24	0.18	0.24
rs4424881T	0.71	0.48	0.34
rs4911414GT	-0.48	-0.10	0.21
rs4911414T	1.01	-0.68	-0.45
rs6119471G	0.41	-0.08	-0.23
rs6119471GC	0.44	-0.26	-0.27
rs6742078GT	0.07	0.12	-0.01
rs6742078T	-0.22	0.50	0.04

Table 4.1: Estimates for the model with no assumptions made

In Table 4.1 it can be seen that there are 3 different columns, besides the naming column. The first is an intercept which is the value of the standard level which is arm in this case. So if a person has genotype *rs1015362CT*, the persons pigmentation is increased by 0.75, whereas if the person has *rs1015362T* it is decreased by 0.02 for arm pigmentation. The basic arm pigmentation is 12.88, which is a person that has only the standard level alleles. Since there are only 2 or 3 different genotypes, the standard level is the first letter that comes. So for *rs1015362* it is *rs1015362C* that is the standard level.

The next two columns are the effects of genotypes on the remaining sites(buttock,face), which reveals what impact the given genotype has on the site pigmentation compared to standard level. As an example it gives darker pigmentation on the buttocks if a person has *rs1015362T* while it gives a lighter pigmentation on the face. The arm pigmentation has the lowest amount of pigmentation compared with the others, which is quite surprising comparing with Table 1.1 where buttock has the lowest. So one would expect going from arm to buttock to have a negative effect i.e. a negative number. One more thing to note is that the standard levels is expected to be around the mean, but as can be seen it is higher than the mean.

Now we know how to interpret the estimates and what they represent. Now we need to check that these estimates actually makes sense, which is done by making sure that the error is normally distributed since this is assumed in Definition 2.1. This is done by examining a QQ-plot, which shows the distribution against a normal distribution. Hence, if it is normally distributed, it will lie on a straight line plotted against the generated normal distribution. But since there are 9 measurements for each subject, we need to check each measurement separately in order to find if one measurement is not normally distributed. The QQ-plot can be seen on Figure 4.1 for the residuals, which is the predictions subtracted from the data.

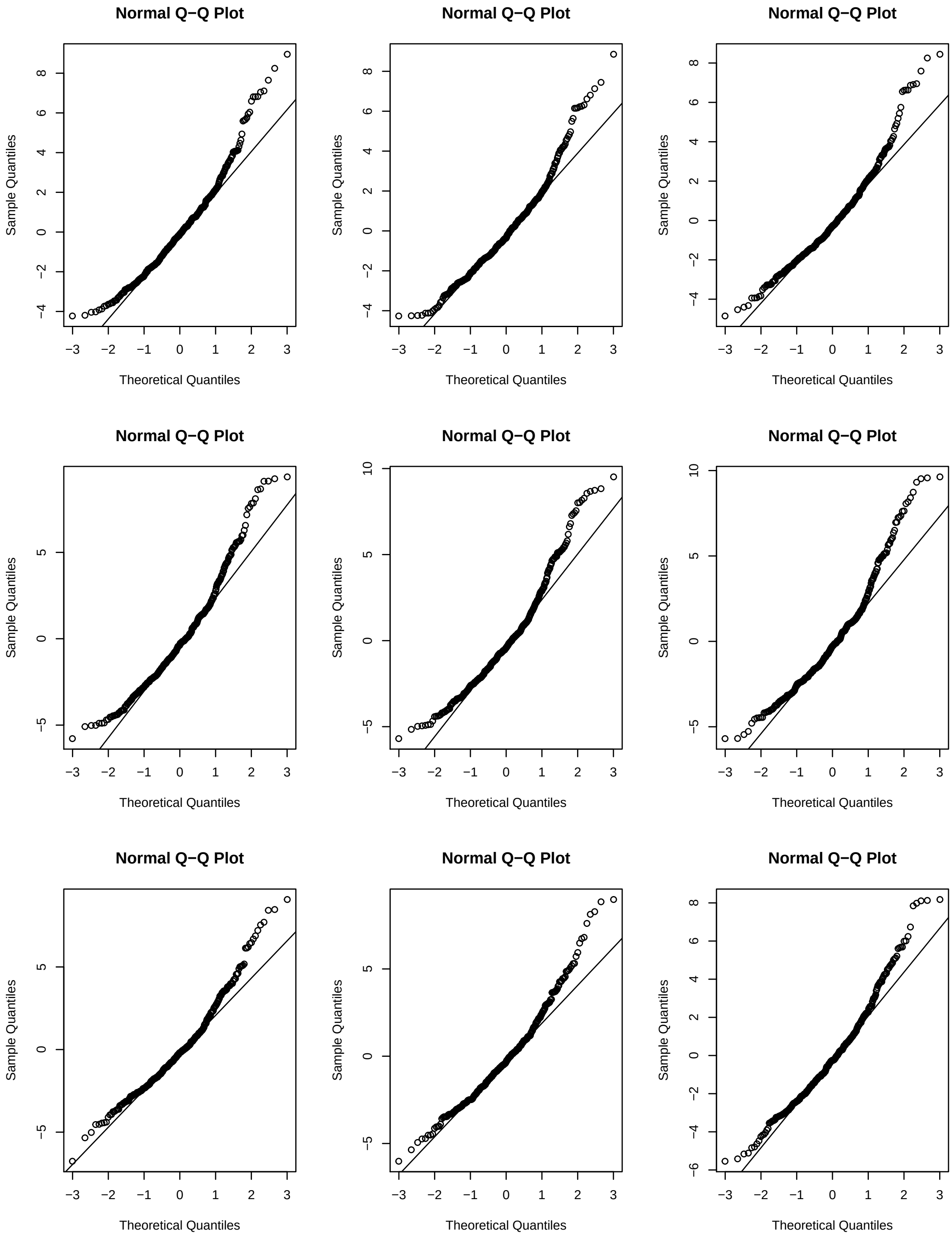

Figure 4.1: QQ-plot for the model with no assumptions

As can be seen non of the measurements seem to violate the assumption regarding normal distribution of the error term. They do however seem to be right skewed, which will be taken care of in Section 4.3. But for now, we will be going forward with this.

Now we know that all assumptions are met, so we need to have a measure of how well the estimated pigmentation matches up against the actual measurements. This is done by a visual representation, where the estimates are plotted against the actuals. This is seen on Figure 4.2.

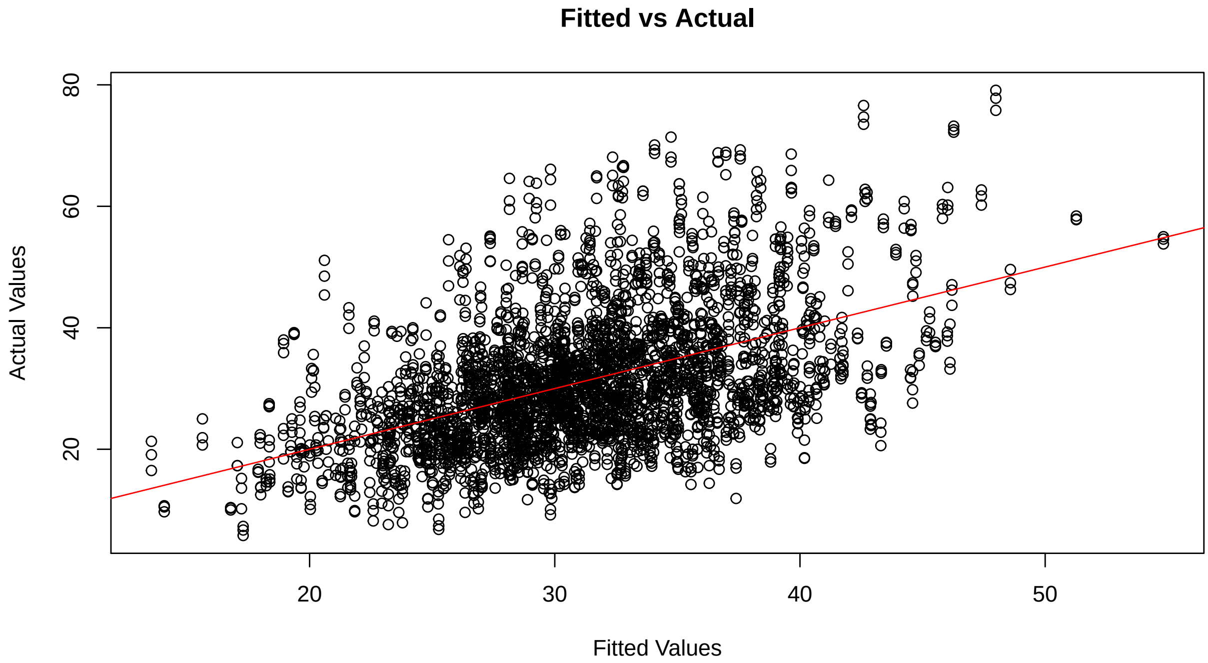

Figure 4.2: Fitted values versus the actual observations

As can be seen on Figure 4.2 there is a great variance, which means that the model does not incorporate the variance properly in the data, but that is also expected as this is just a naive model of data. But as can be seen when zooming in on the big black lump of observations is that the model seems to fit quite well, as there are quite a lot of measurements that lie just around the line. So this means that we can continue with this model in good faith. The mean square error is 6.812 which we want to lower in the remaining of this chapter. We will try to do this in various ways throughout, but this section serves as an introduction to understanding the results, hence the results will not be explained in depth for the rest of this chapter. One new approach to data, is to see what happens when we eliminate the effect of which site the measurement is taken at. This is done in the next section.

4.2 No Site Effect

This section will try to make a model where there is no site effect i.e. $\mathbf{C} = \mathbf{I}$. This will make the estimates be all the same, and hence each subject gets one value for all 9 measurements. This could be thought to violate the normality assumption, but doing the same as in Section 4.1, we see that it does not violate the normal distribution assumption. Hence we can see whether the estimates are better than for the model with no assumptions given in Section 4.1. This is seen on Figure 4.3.

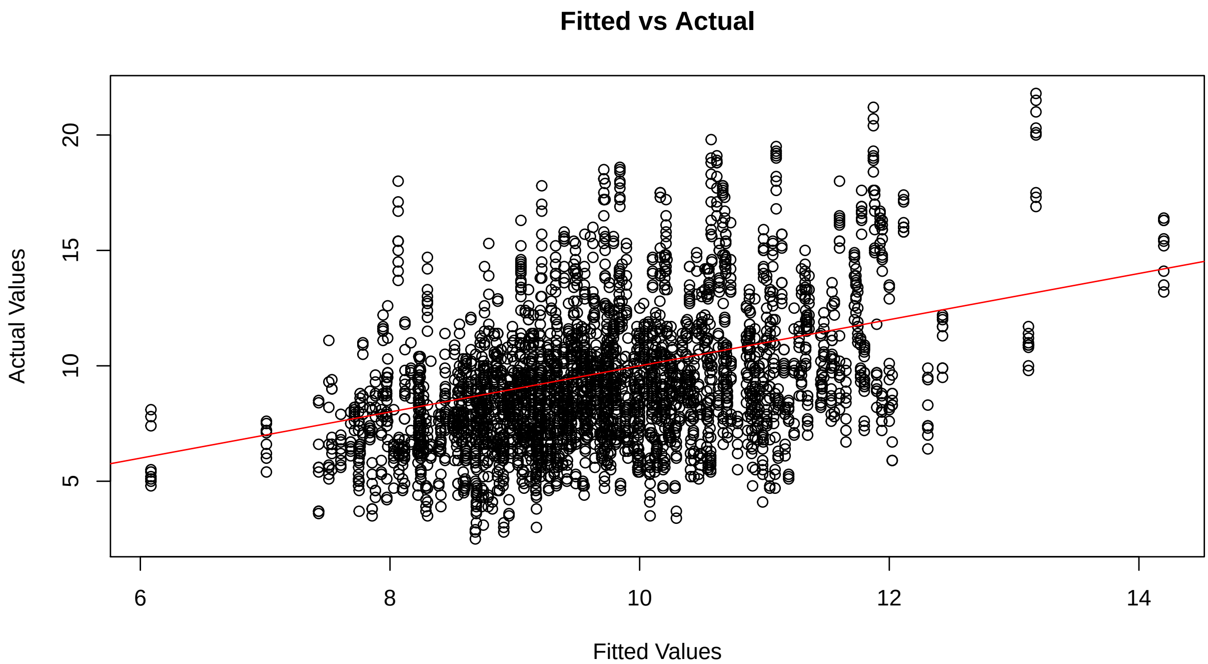

Figure 4.3: Fitted values versus the actual observations with no site effect

Comparing Figure 4.3 and Figure 4.2 it can be seen that Figure 4.2 seems to fit the data much better than Figure 4.3. This is also supported by the mean square error of 7.504 which is bigger than the model with no assumptions. So the site does have an effect on the prediction of pigmentation.

We can also test this hypothesis using our likelihood ratio test. By doing this the p-value is $8.3 \cdot 10^{-14}$ which must mean that we can reject the hypothesis of no site effect. So the site can not be disregarded. So another approach is to log transform the measurements, as we saw in the no assumptions model that it had some difficulties when calculating the more "extreme" measurements.

4.3 Transformed Data

In this section we will try to transform the data such that we may find some better estimates. We are doing this since on Figure 4.1 we saw that the data could benefit from a transformation. Since this data seems to be right-skewed, we might apply a log transformation on the data. The problem was the QQ-plots in Figure 4.1, where we saw a right skewedness, that we want to eliminate. This is seen on Figure 4.4.

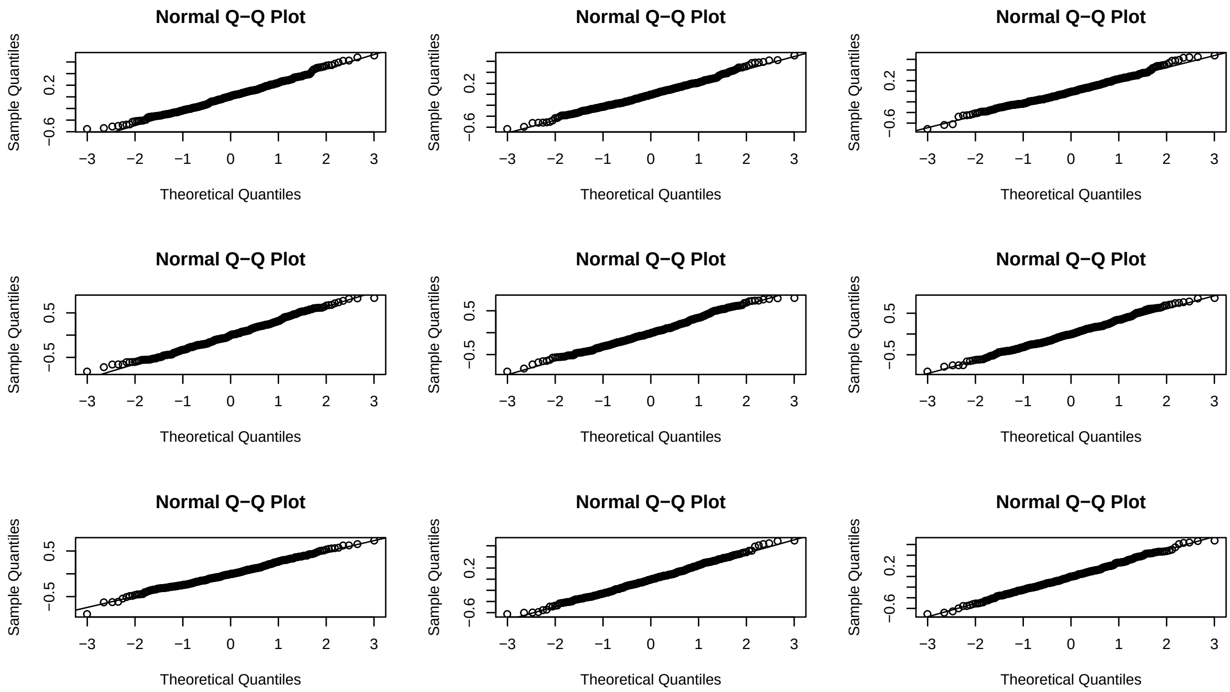

Figure 4.4: Fitted values versus the actual observations with no site effect

The transformed data follow a normal distribution for the errors, and it improved a lot compared to Figure 4.1. So it does make sense to try and analyse our data using this transformation in order to get better estimates.

After investigating it gives a worse estimate of the pigmentation, with a mean square error of 1.064 after transforming back[1], which is higher than the model without any assumptions so this is not the way to go.

Another transformation one might perform is a square root transformation, which is also common for right-skewed data. Such a transformation also yields normality as seen on Figure 4.5.

[1]The mean of the lognormal distribution is $exp(\mu + \sigma^2/2)$

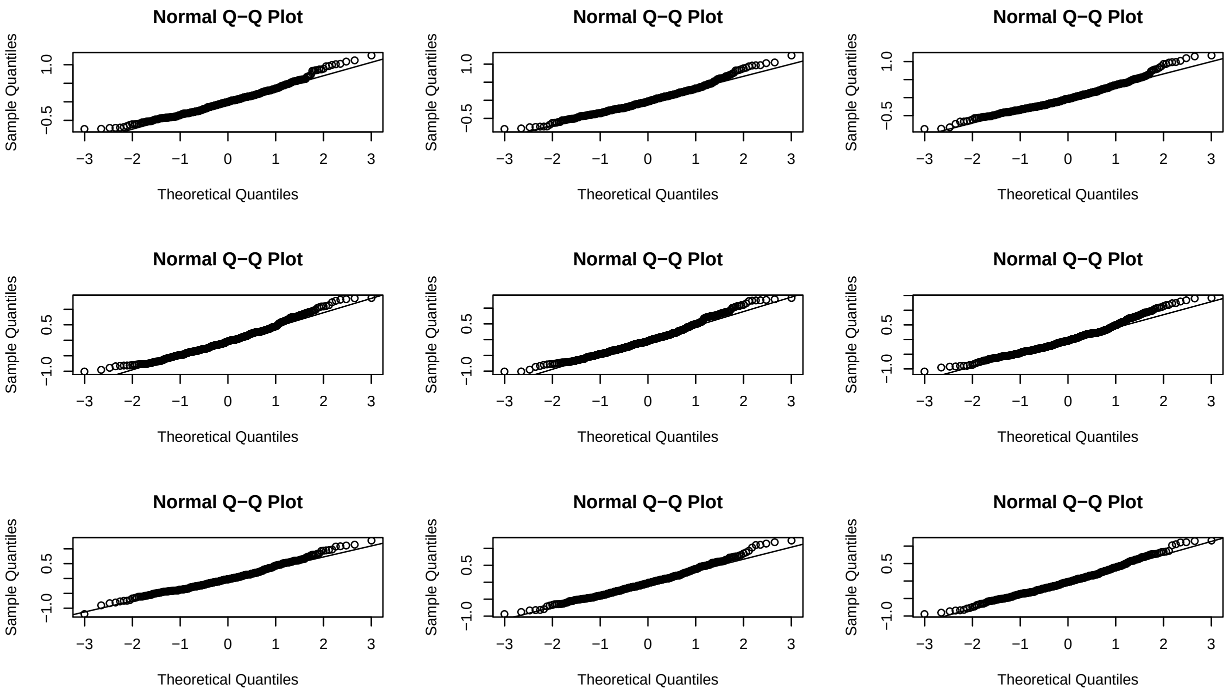

Figure 4.5: Fitted values versus the actual observations with no site effect

These quantiles also look great and normal distributed, so we can use this to model the pigmentation. This yields a mean square error of 0.170 where we have not transformed it back, so it is not comparable.

Now we have tried altering both the effect of the sites where it is measured and transforming the data. So now we are missing one obvious thing; altering the SNPs in a thoughtful way.

4.4 Altering SNPs

In this section we will be trying to alter the layout of the SNPs by trying to eliminate some of them as some might just bring noise to the model and not helping in estimating the pigmentation. So we will try to cut some SNPs using the likelihood ratio test that we implemented in Chapter 3. We will be doing this by cutting those values that have a p-value above 0.05 which is the significance level that is chosen. But just to try different things, we will also be trying with a significance level of 0.01. This gives the following SNPs to be used given in Table 4.2.

0.05	0.01
rs12913832	rs2470102
rs1408799	
rs1800407	
rs2470102	

Table 4.2: Significant SNPs

This means that these 4 and 1 SNPs respectively predicts the pigmentation. So let us see how well it performs compared to the model with no assumptions.

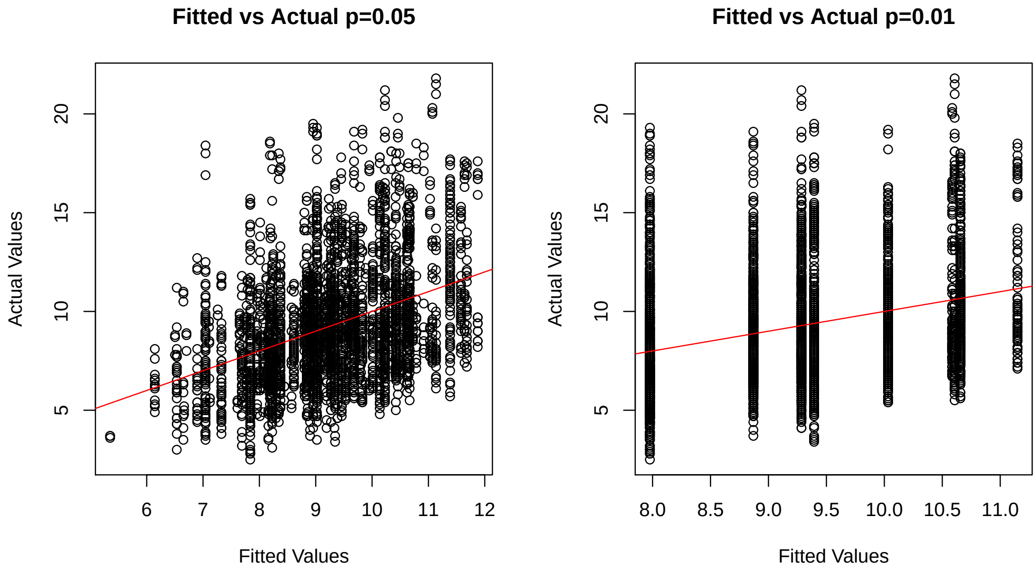

Figure 4.6: Fitted values versus the actual observations with no site effect

These model have a mean square error of 8.122 for the p-value of 0.01 and a mean square error of 7.619 for the p-value of 0.05. So this means that none of the reduced models fit better than the model with no assumptions. This was also expected as comparing Figure 4.6 and Figure 4.2 we see that Figure 4.2 lie more closely around the line.

So four different approaches to data has not resulted in a better estimation than the model with no assumptions. So this thesis will not investigate this any further.

Chapter 5

Discussion and Conclusions

This thesis has analysed the pigmentation of 376 different Brazilians using the GMANOVA approach. This approach was chosen as it could handle both the difference in SNPs between subjects and the difference between the sites. The GMANOVA could also handle the repeated measurements within the subjects very nicely. Furthermore not one observation was deemed to exemplify that it did not meet the assumptions of the model, which makes it valid.

The analysis in this thesis could have been more thorough, but due to time constraints it was not possible to go more in-depth in the analysis. But still it was found that there were 4 SNPs that were used in predicting pigmentation for a significance level of 0.05. This can be seen in Table 4.2. Many different approaches were taken to data including transforming the data and checking significance of all SNPs and of the sites. One thing that could be an extension is doing the same as in our example for our data, which would make make the SNP effect clear as well as the site specific effect. It could also be eliminating the difference between the sites if done for them selves, that is the mean of buttock is lower than for forehead. The only problem with this place specific analysis is that we are interested in modelling a person's general pigmentation, and this is not done for a place specific analysis.

All in all the GMANOVA performs well with low errors and fits the data great. It has also gotten some low variation when comparing the actual values with the fitted values. It performed especially well for the data where we had not done anything, so it is easy to use the model for this data.

This thesis concludes that the GMANOVA uses 4 SNPs for a significance level of 0.05; rs12913832, rs1408799, rs1800407, rs2470102. These four SNPs is the most important but comparing with the former reports, none of these has been deemed important. So this is a quite surprising result.

This thesis concludes that the GMANOVA model with no assumptions is the best suited for this data, as it fits the data the best and has the lowest error.

Bibliography

[1] Single-nucleotide polymorphism analysis by pyrosequencing. https://www-sciencedirect-com.zorac.aub.aau.dk/science/article/pii/S0003269700944932.

[2] D. von Rosen T. Kollo. *Advanced Multivariate Statistics with Matrices.* Springer, 2005.

[3] M. Vaeth I. Olkin. *Maximum Likelihood Estimation in a Two-Way Analysis of Variance with Correlated Errors in One Classification.* 1981.

[4] The relationship between the mahalanobis distance and the chi-squared distribution. https://markusthill.github.io/mahalanbis-chi-squared/.

[5] The large-sample distribution of the likelihood ratio for testing composite hypotheses. https://www.jstor.org/stable/2957648?seq=1/subjects#page_scan_tab_contents.

Appendix A

Useful Results

All these results come from [2] and no proofs will be given. These are given in the source.

Theorem A.1. Let $\mathbf{A} > 0$ and $\mathbf{B}$ symmetric. Then, there exist a non-singular matrix $\mathbf{T}$ and diagonal matrix $\mathbf{\Lambda} = (\lambda_1, \ldots, \lambda_m)_d$ such that

$$\mathbf{A} = \mathbf{T}\mathbf{T}^T \quad \mathbf{B} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^T \tag{A.1}$$

where $\lambda_i, i = 1, 2, \ldots, m$ are eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$.

This theorem also holds if $\mathbf{B}$ is a positive semi definite matrix.

Corollary A.2. For $\mathbf{S} > 0$ and an arbitrary matrix $\mathbf{B}$ of proper size it holds that

$$\mathbf{S}^{-1} - \mathbf{B}^o(\mathbf{B}^{oT}\mathbf{S}\mathbf{B}^o)^{-1}\mathbf{B}^{oT} = \mathbf{S}^{-1}\mathbf{B}(\mathbf{B}^T\mathbf{S}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{S}^{-1} \tag{A.2}$$

Theorem A.3. A representation of the general solution of the consistent equation in $\mathbf{X} : \mathbf{AXB} = \mathbf{C}$ is given by any of the following formulas

$$\begin{aligned}
\mathbf{X} &= \mathbf{X}_0 + (\mathbf{A}^T)^o\mathbf{Z}_1\mathbf{B}^T + \mathbf{A}^T\mathbf{Z}_2\mathbf{B}^{oT} + (\mathbf{A}^T)^o\mathbf{Z}_3\mathbf{B}^{oT} \\
\mathbf{X} &= \mathbf{X}_0 + (\mathbf{A}^T)^o\mathbf{Z}_1 + \mathbf{A}^T\mathbf{Z}_2\mathbf{B}^{oT} \\
\mathbf{X} &= \mathbf{X}_0 + \mathbf{Z}_1\mathbf{B}^{oT} + (\mathbf{A}^T)^o\mathbf{Z}_2\mathbf{B}^T
\end{aligned}$$

where $\mathbf{X}_0$ is a particular solution and $\mathbf{Z}_i, i = 1, 2, 3$ are properly sized, arbitrary matrices.

Proposition A.4. For column vector spaces the following relations hold

- $C(\mathbf{A}) \subseteq C(\mathbf{B})$ if and only if $\mathbf{A} = \mathbf{BQ}$ for some matrix $\mathbf{Q}$
- If $C(\mathbf{A} + \mathbf{BE}) \subseteq C(\mathbf{B})$, for some proper sized matrix $\mathbf{E}$ then $C(\mathbf{A}) \subseteq C(\mathbf{B})$. If $C(\mathbf{A}) \subseteq C(\mathbf{B})$ then $C(\mathbf{A} + \mathbf{BE}) \subseteq C(\mathbf{B})$

- $$C(\mathbf{A}^T\mathbf{B}_1) \subseteq C(\mathbf{A}^T\mathbf{B}_2) \ \text{ if } \ C(\mathbf{B}_1) \subseteq C(\mathbf{B}_2)$$
$$C(\mathbf{A}^T\mathbf{B}_1) = C(\mathbf{A}^T\mathbf{B}_2) \ \text{ if } \ C(\mathbf{B}_1) = C(\mathbf{B}_2)$$

- $C(\mathbf{A}^T\mathbf{B}) = C(\mathbf{A}^T)$ if $C(\mathbf{A}) \subseteq C(\mathbf{B})$

- $C(\mathbf{A}) \cap C(\mathbf{B}) = C((\mathbf{A}^o : \mathbf{B}^o)^o)$

- For any $\mathbf{A}^{-1}$ it holds that $\mathbf{C}\mathbf{A}^{-1}\mathbf{A} = \mathbf{C}$ if and only if $C(\mathbf{C}^T) \subseteq C(\mathbf{A}^T)$

- $C(\mathbf{A}^T) = C(\mathbf{A}^T\mathbf{B})$ if and only if $r(\mathbf{A}^T\mathbf{B}) = r(\mathbf{A}^T)$

- Let $\mathbf{A} \in \mathbb{R}^{p\times q}, \mathbf{S} > 0, r(\mathbf{H}) = p$. Then $C(\mathbf{A}^T) = C(\mathbf{A}^T\mathbf{H}) = C(\mathbf{A}^T\mathbf{S}\mathbf{A})$

- Let $\mathbf{A} \in \mathbb{R}^{p\times q}, \mathbf{S} > 0$ then
$$\mathbf{C}(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S}\mathbf{A} = \mathbf{C} \ \text{ if and only if } \ C(\mathbf{C}^T) \subseteq C(\mathbf{A}^T)$$
$$\mathbf{A}(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S}\mathbf{B} = \mathbf{B} \ \text{ if and only if } \ C(\mathbf{B}) \subseteq C(\mathbf{B})$$
$$\mathbf{C}\mathbf{A}\mathbf{B}(\mathbf{C}\mathbf{A}\mathbf{B})^{-1}\mathbf{C} = \mathbf{C} \ \text{ if} r(\mathbf{C}\mathbf{A}\mathbf{B}) = r(\mathbf{C})$$

- $\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ is invariant under choice of g-inverse if and only if $C(\mathbf{C}^T) \subseteq C(\mathbf{A}^T)$ and $C(\mathbf{B}) \subseteq C(\mathbf{A})$

- Let $\mathbf{S} > 0$ then $\mathbf{C}_1(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1}\mathbf{C}_2$ is invariant under any choice of $(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1}$ if and only if $C(\mathbf{C}_1^T) \subseteq C(\mathbf{A}^T)$ and $C(\mathbf{C}_2) \subseteq C(\mathbf{A}^T)$

- If $C(\mathbf{C}^T) \subseteq C(\mathbf{A}^T)$ and $\mathbf{S} > 0$ then
$$C(\mathbf{C}) = C(\mathbf{C}(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1}) = C(\mathbf{C}(\mathbf{A}^T\mathbf{S}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S}) \tag{A.3}$$

- $C(\mathbf{A}\mathbf{B}) \subseteq C(\mathbf{A}\mathbf{B}^o)$ if and only if $C(\mathbf{A}) \subseteq C(\mathbf{A}\mathbf{B}^o)$

Theorem A.5. Let P be an arbitrary projector and A a linear transformation such that PA is defined. Then
$$\mathcal{R}(PA) = \mathcal{R}(P) \cap (\mathcal{N}(P) + \mathcal{R}(A)) \tag{A.4}$$

Theorem A.6. For any transformations A, B assume that A^TB^o is well defined. Then the following statements are equivalent

- $\mathcal{R}(A^T) \subseteq \mathcal{R}(A^TB^o)$

- $\mathcal{R}(A^T) = \mathcal{R}(A^TB^o)$

- $\mathcal{R}(A) \cap \mathcal{R}(B) = \{\mathbf{0}\}$

Theorem A.7. Let $\mathbf{X} \sim N_{p,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ then for any $\mathbf{A} : q \times p$ and $B : m \times n$ then

$$\mathbf{AXB}^T \sim N_{q,m}(\mathbf{A}\boldsymbol{\mu}\mathbf{B}^T, \mathbf{A\Sigma A}^T, \mathbf{B\Psi B}^T) \tag{A.5}$$

Theorem A.8. Let $\mathbf{X} \sim N_{p,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}), \mathbf{Y} \sim N_{p,n}(0, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ and $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{K}, \mathbf{L}$ are non-random matrices of proper sizes. Then

- $\mathbf{AXK}$ is independent of $\mathbf{CXL}$ for all constant matrices $\mathbf{K}, \mathbf{L}$ if and only if $\mathbf{A\Sigma C}^T = \mathbf{0}$.
- $\mathbf{KXB}^T$ is independent of $\mathbf{LXD}^T$ for all constant matrices $\mathbf{K}, \mathbf{L}$ if $\mathbf{B\Psi D}^T = \mathbf{0}$
- $\mathbf{YAY}^T$ is independent of $\mathbf{YBY}^T$ if and only if

$$\begin{aligned} \mathbf{\Psi A \Psi B}^T\mathbf{\Psi} = \mathbf{0}, \ \ \mathbf{\Psi A}^{\mathbf{T}}\mathbf{\Psi B \Psi} = \mathbf{0} \\ \mathbf{\Psi A \Psi B \Psi} = \mathbf{0}, \ \ \mathbf{\Psi A}^T\mathbf{\Psi B}^T\mathbf{\Psi} = \mathbf{0} \end{aligned}$$

- $\mathbf{YAY}^T$ is independent of $\mathbf{YB}$ if and only if

$$\begin{aligned} \mathbf{B}^T\mathbf{\Psi A}^T\mathbf{\Psi} = \mathbf{0} \\ \mathbf{B}^T\mathbf{\Psi A \Psi} = \mathbf{0} \end{aligned}$$

Proposition A.9. Let P be a projector on $\mathbb{V}_1$ along $\mathbb{V}_2$. Then

- P is a linear transformation
- $PP = P$, i.e. P is idempotent
- $I - P$ is projector on $\mathbb{V}_2$ along $\mathbb{V}_1$ where I is the identity mapping
- The range space $\mathcal{R}(P)$ is identical to $\mathbb{V}_1$ the null space $\mathcal{N}(P)$ equals $\mathcal{R}(I - P)$
- If P is idempotent then P is a projector
- P is unique

Proposition A.10. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ of rank r. Then there exists a triangular matrix $\mathbf{T} \in \mathbb{R}^{m\times m}$ and an orthogonal matrix $\mathbf{H} \in \mathbb{R}^{n\times n}$ such that

$$\mathbf{A} = \mathbf{T}\begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{H} \tag{A.6}$$

and

$$\mathbf{A} = \mathbf{KL} \tag{A.7}$$

where $\mathbf{K} \in \mathbb{R}^{m \times r}, \mathbf{L} \in \mathbb{R}^{r \times n}$ and $\mathbf{K}$ consists of the first r columns of $\mathbf{T}$ and $\mathbf{L}$ consists of the first r rows of $\mathbf{H}$.

Definition A.11. The matrix $\mathbf{W} : p \times p$ is said to be Wishart distributed if and only if $\mathbf{W} = \mathbf{X}\mathbf{X}^T$ for some matrix $\mathbf{X} \sim N_{p,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{I}), \boldsymbol{\Sigma} \geq 0$. If $\boldsymbol{\mu} = \mathbf{0}$ we have a central Wishart distribution which is denoted as $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n)$, and if $\boldsymbol{\mu} \neq \mathbf{0}$ we have a non-central Wishart distribution which will be denoted $W_p(\boldsymbol{\Sigma}, n, \boldsymbol{\Delta})$ where $\boldsymbol{\Delta} = \boldsymbol{\mu}\boldsymbol{\mu}^T$.

Theorem A.12. Let $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n, \boldsymbol{\Delta})$ and $\mathbf{A} \in \mathbb{R}^{q \times p}$. Then

$$\mathbf{A}\mathbf{W}\mathbf{A}^T \sim W_q(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T, n, \mathbf{A}\boldsymbol{\Delta}\mathbf{A}^T) \tag{A.8}$$

Proposition A.13. Let

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \quad \mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{pmatrix} \tag{A.9}$$

so the dimensions of $\mathbf{A}_{11}, \mathbf{A}_{12}, \mathbf{A}_{21}, \mathbf{A}_{22}$ corresponds to those of $\mathbf{A}^{11}, \mathbf{A}^{12}, \mathbf{A}^{21}, \mathbf{A}^{22}$. Then

- $(\mathbf{A}^{11})^{-1}\mathbf{A}^{12} = -\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$

Theorem A.14. Let $\mathbf{X} \sim N_{p,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$. Then

- $E[\mathbf{X}\mathbf{A}\mathbf{X}^T] = tr(\boldsymbol{\Psi}\mathbf{A})\boldsymbol{\Sigma} + \boldsymbol{\mu}\mathbf{A}\boldsymbol{\mu}^T$

Theorem A.15. Let $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n)$ then

- $E[\mathbf{W}^{-1}] = \frac{1}{n-p-1}\boldsymbol{\Sigma}^{-1}, \quad n-p-1 > 0$
- $E[tr(\mathbf{W}^{-1})\mathbf{W}] = \frac{1}{n-p-1}(n\boldsymbol{\Sigma}tr(\boldsymbol{\Sigma}^{-1}) - 2\mathbf{I}), \quad n-p-1 > 0$

Theorem A.16. Let $\mathbf{y}$ and $\mathbf{x}$ be random p-vector and r-vector, $p \leq r$ and let $\mathbf{P} : p \times r$ be a matrix of rank $r(\mathbf{P}) = p$ and $\mathbf{A} : r \times r$ positive definite. Then

$$f_{\mathbf{y}}(\mathbf{y}_0) = |\mathbf{A}|^{\frac{1}{2}}|\mathbf{P}\mathbf{A}^{-1}\mathbf{P}|^{\frac{1}{2}} \left((2\pi)^{\frac{1}{2}(r-p)} f_{\mathbf{x}}(\mathbf{x}_0) + \sum_{k=1}^{m} (-1)^k \frac{1}{k!} vec^T \mathbf{h}_k(\mathbf{0} vec f_{\mathbf{x}}^{(k)}(\mathbf{x}_0) + r_m^* \right) \tag{A.10}$$

and

$$r_m^* = (2\pi)^{-r} \int_{\mathbf{R}^r} r_m(\mathbf{t}) \exp(-i\mathbf{t}^T \mathbf{x}_0) \phi_{\mathbf{x}}(\mathbf{t}) d\mathbf{t}. \tag{A.11}$$

where

$$\mathbf{h}_k(\mathbf{0}) = i^{-k}\mathbf{l}^k(\mathbf{P0}, \mathbf{0})$$
$$\mathbf{l}^k(\mathbf{0}, \mathbf{0}) = (2\pi)^{\frac{1}{2}(r-p)} \frac{\phi_{\mathbf{y}}(\mathbf{0})}{\phi_{\mathbf{x}}(\mathbf{0})}$$
$$r_m(\mathbf{t}) = \frac{1}{(m+1)!} \mathbf{t}\mathbf{l}^{m+1}(\mathbf{P}(\mathbf{\Theta} \circ \mathbf{t}), \mathbf{\Theta} \circ \mathbf{t})\mathbf{t}^{\otimes m}$$

where $\mathbf{\Theta} \circ \mathbf{t}$ is the Hadamard product of $\mathbf{\Theta}, \mathbf{t}$ and $\mathbf{\Theta}$ is a r-vector with elements between 0 and 1.

Definition A.17. Let $\mathbf{\Sigma} = \tau\tau^T$ and $\Psi = \gamma\gamma^T$ where $\tau : p \times r$ and $\gamma : n \times s$. A matrix $\mathbf{X} \sim N_{p,n}(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{\Psi})$ if it has the same distribution as

$$\boldsymbol{\mu} + \tau \mathbf{U} \gamma^T \tag{A.12}$$

where $\boldsymbol{\mu} : p \times n$ is non-random and $\mathbf{U} : r \times s$ consists of s I.I.D $N_r(\mathbf{0}, \mathbf{I})$ vectors $\mathbf{u}_i, i = 1, 2, \ldots, s$.

Theorem A.18. Let $\mathbf{W}$,$\mathbf{Y}$ and $\mathbf{V}$ be $p \times p$ random symmetric matrices with $\mathbf{W} \sim W_p(\mathbf{\Sigma}, n)$ and $\mathbf{V} = \mathbf{W} - n\mathbf{\Sigma}$. Then for the density $f_{\mathbf{Y}}(\mathbf{X})$ the following expansion holds

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{X}) = f_{\mathbf{V}}(\mathbf{X}) \Big(& 1 + E[\mathbf{V}^{(2)}(\mathbf{Y})]^T \mathbf{L}_1^*(\mathbf{X}, \mathbf{\Sigma}) \\ & + \frac{1}{2} vec^T (D[\mathbf{V}^2(\mathbf{Y})] - D[\mathbf{V}^2(\mathbf{V})] + E[\mathbf{V}^2(\mathbf{Y})] E[\mathbf{V}^2(\mathbf{Y})]^T) vec \mathbf{L}_2^*(\mathbf{X}, \mathbf{\Sigma}) \\ & + \frac{1}{6} (vec^T (c_3[\mathbf{V}^2(\mathbf{Y})] - c_3[\mathbf{V}^2(\mathbf{V})])^T \\ & + 3 vec^T (D[\mathbf{V}^2(\mathbf{Y})] - D[\mathbf{V}^2(\mathbf{Y})]) \otimes E[\mathbf{V}^2(\mathbf{Y})]^T + E[\mathbf{V}^2(\mathbf{Y})]^{T \otimes 3}) \\ & \times vec \mathbf{L}_3^*(\mathbf{X}, \mathbf{\Sigma}) + \cdots) \end{aligned}$$

where

$$V^2(\mathbf{W}) = (w_{11}, w_{12}, w_{22}, \ldots, w_{2n})$$
$$\mathbf{L}_k^*(\mathbf{X}, \mathbf{\Sigma}) = \mathbf{L}_k(\mathbf{X} + n\mathbf{\Sigma}, \mathbf{\Sigma})$$

where $\mathbf{L}_k$ is defined on page 270 in [2] for $k = 0, 1, 2, 3$.

Corollary A.19. Let $\mathbf{V} \sim W_p(\mathbf{I}, n)$ and apply the partition

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}, \quad \begin{matrix} r(\mathbf{X}) \times r(\mathbf{X}) & r(\mathbf{X}) \times (p - r(\mathbf{X})) \\ (p - r(\mathbf{X})) \times r(\mathbf{X}) & (p - r(\mathbf{X})) \times (p - r(\mathbf{X})) \end{matrix} \tag{A.13}$$

then

$$\mathbf{V}_{12}\mathbf{V}_{22}^{-1/2} \sim N_{r,p-r}(\mathbf{0}, \mathbf{I}, \mathbf{I}) \tag{A.14}$$

is independent of $\mathbf{V}_{22}$.