ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 8, 2004

MANOVA with singular variance matrix

MUNI S. SRIVASTAVA AND DIETRICH VON ROSEN

ABSTRACT. Classical multivariate analysis of variance for p response variables is extended to cover high-dimensional data. For example, data often comprise many response variables that may be related. Therefore, inference based on all the response variables may be inefficient. However, the relationship between the response variables is usually not known. This leads to the assumption that the p response variables span a linear space of some fixed dimension, say $r < p$; equivalently the $p \times p$ variance matrix is singular of rank r. We will assume that the rank is given. Following the classical approach of doing inference in linear models, parameters are first estimated and thereafter tests are constructed. Estimators and tests are based on the likelihood method. The present model differs from the classical multivariate analysis of variance model and consists of a deterministic part and a random part. It is noticed that the classical approach is a special case of the one which will be considered in this article.

1. Introduction

Often data comprise observations on many response variables of which many may be closely related. Therefore, inference based on all the response variables may be inefficient. In situations like this, usually the principal component method is applied to reduce the dimensionality of the response variables. However, after reducing the dimensionality, it is not known how to carry out the statistical inference on the unknown parameters. For example, in the one-sample problem, it is not known how to estimate the mean vector

Received December 18, 2003.
2000 *Mathematics Subject Classification.* 62G07, 62H12, 62F10.
Key words and phrases. Likelihood ratio test, maximum likelihood estimators, principal component, normal distribution, singular variance matrix, multivariate analysis of variance, confidence intervals, outlier detection.
The support by the Natural Sciences and Engineering Research Council of Canada and the Swedish Natural Science Research Council is gratefully acknowledged.

or how to carry out tests of hypotheses regarding the parameters of the mean vector. In this paper, we assume that the rank of the variance matrix is the same as the number of selected principal components. Thus, we shall assume that the variance matrix of the observations is a singular matrix of a given rank, the rank of the matrix is either determined by the past experience or from a practical consideration such as principal component analysis. We shall consider the multivariate analysis of variance model. Estimators and tests of hypotheses along with their distributions will be given.

Linear models with the singular variance matrix have been studied by many authors including Khatri (1968), Mitra & Rao (1968), Rao (1973), Alalouf (1978), Feuerverger & Fraser (1980) and Nordström (1985). Except Khatri (1968) these authors assume that the variance matrix $\mathbf{\Sigma}$ is known. In this paper this matrix is unknown and therefore has to be estimated. Suppose that the variance matrix of the response variables, $\mathbf{\Sigma} : p \times p$, is of rank r which is a dimensionality reduction if $r < p$, and based on this assumption a MANOVA is performed. The results of this paper will be illustrated by a numerical example. The organization of the paper is as follows. In Section 2 we consider a numerical example to demonstrate the singularity of the variance matrix. Some of its eigenvalues are very small and we suggest that instead of considering initial eight-dimensional data, we will be better off with less than eight. In Section 3 we obtain the maximum likelihood estimators of the parameters and likelihood ratio tests are used to complete the analysis. In Section 4 confidence intervals for mean parameter Ξ are found and Section 5 is dedicated to the detection of outliers.

2. A numerical example

We consider the data of Russell et al. (1967), presented in Srivastava & Carter (1983, p. 128). The data present the clinical analysis of soil characteristics for three contours and four depths of soil. The experiment area was divided into four blocks and samples were taken randomly at various depths and contours. For our purpose we just consider four combinations of depths and contours. Moreover, the four areas are reduced to two by merging the original areas pairwise. The data are presented in Table 2.1.

According to Table 2.1 the multivariate data are obtained from a complete balanced block design. To be even more specific, there are two blocking factors of which one (Area) has two levels and the other (Group) has four levels. Furthermore, there are 16 independent observations of which each comprises 8 response variables.

Table 2.1. Soil data where the following abbreviations are used: $V1$=pH; $V2$=Total nitrogen (%); $V3$=Bulk density (gm/cm^3); $V4$=Total phosphorus (ppm); $V5$=Exchangeable + soluble calcium (me/100 gm); $V6$=Exchangeable + soluble magnesium (me/100 gm); $V7$=Exchangeable + soluble potassium (me/100 gm); $V8$=Exchangeable + soluble sodium (me/100 gm);

V1	5.40	5.65	5.14	5.14	5.14	5.10	4.70	4.46
V2	0.19	0.16	0.26	0.17	0.16	0.09	0.10	0.11
V3	0.92	1.04	0.95	1.10	1.12	1.22	1.52	1.47
V4	215	208	300	248	174	129	117	170
V5	16.4	12.2	13.0	11.9	14.2	8.6	8.7	9.5
V6	7.6	5.2	5.7	7.0	8.1	6.9	8.2	9.2
V7	0.72	0.71	0.68	1.09	0.70	0.81	0.39	0.70
V8	1.14	0.94	0.60	1.01	2.17	2.67	3.32	3.76
Group	1	1	1	1	2	2	2	2
Area	1	1	2	2	1	1	2	2
V1	4.37	4.39	4.17	3.89	3.88	4.07	3.88	3.74
V2	0.11	0.06	0.07	0.08	0.08	0.05	0.06	0.05
V3	1.07	1.54	1.26	1.42	1.25	1.54	1.53	1.40
V4	121	115	112	117	127	91	91	79
V5	8.8	4.7	6.3	6.6	6.4	3.8	5.0	5.9
V6	10.4	6.9	8.0	9.8	11.0	6.6	8.0	10.1
V7	0.74	0.77	0.26	0.41	0.56	0.50	0.23	0.41
V8	5.74	5.85	5.30	8.30	9.67	7.67	8.78	11.04
Group	3	3	3	3	4	4	4	4
Area	1	1	2	2	1	1	2	2

Let the data be collected in $\mathbf{Y} : 8 \times 16$ where the first 4 columns consist of the data from group 1, the next 4 columns consist of data from group 2, etc. The design matrix $\mathbf{X}$ is given by

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{pmatrix}$$

and the following model is assumed to hold:

$$\mathbf{Y} = \Xi\mathbf{X} + \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{E},$$

where the columns of $\mathbf{E}$ are independent identically distributed (i.i.d.) random

vectors with the distribution $N_8(\mathbf{0}, \mathbf{I})$, and $\mathbf{\Sigma}^{\frac{1}{2}}$ is a positive semi-definite square root of $\mathbf{\Sigma}$. The sums-of-squares matrix due to error, i.e. $\mathbf{S}$, is calculated by (3.7):

$$\mathbf{S} = \begin{pmatrix} 0.20 & -0.01 & -0.01 & -14.5 & -0.38 & -1.42 & -0.09 & -0.80 \\ -0.01 & 0.01 & -0.03 & 8.00 & 0.42 & 0.12 & -0.00 & -0.05 \\ -0.01 & -0.03 & 0.26 & -18.1 & -2.34 & -1.33 & 0.02 & -0.03 \\ -14.5 & 8.00 & -18.1 & 8920 & 168.0 & 80.6 & 18.7 & -58.0 \\ -0.38 & 0.42 & -2.34 & 168.0 & 42.4 & 20.4 & -0.73 & 4.01 \\ -1.42 & 0.12 & -1.33 & 80.6 & 20.4 & 26.8 & 0.97 & 10.1 \\ -0.09 & -0.00 & 0.02 & 18.7 & -0.73 & 0.97 & 0.35 & 0.24 \\ -0.80 & -0.05 & -0.03 & -58.0 & 4.01 & 10.1 & 0.24 & 10.8 \end{pmatrix}.$$

An unbiased estimator of the sample variance matrix is $\mathbf{W} = \frac{1}{11}\mathbf{S}$ with eigenvalues

$$811.31 \quad 5.02 \quad 1.50 \quad 0.40 \quad 0.017 \quad 0.013 \quad 0.0027 \quad 0.000039.$$

Since the smallest eigenvalue is very close to zero, it is reasonable to assume that the rank of $\mathbf{S}$ is at most 7. It follows from Lemma 3.1, given in Section 3, that the rank of $\mathbf{W}$ is equal to the rank of $\mathbf{\Sigma}$. The second smallest eigenvalue is 0.003 which is also close to zero. Thus, it may be reasonable to assume that the rank of $\mathbf{\Sigma}$ is 6. The problem of deciding on the rank of $\mathbf{\Sigma}$ is rather difficult. For example, one may argue that since all the four smallest eigenvalues are rather small, the rank of $\mathbf{\Sigma}$ is only four. We may also note that the first eigenvalue contains 99% of the variation in the data, and thus one may argue that the rank of $\mathbf{\Sigma}$ is only one. The fewer components are chosen, the fewer linear combinations of the mean can be tested for significance; the difference can be detected only if it lies in the direction of the chosen components. So, there is a difference between the so called principal component analysis and the problem of drawing inference on its mean components. Any inference based on one component only or on very few components may be very misleading. Here our objective is to select the rank of the variance matrix $\mathbf{\Sigma}$. For this, we should drop from considerations only those components for which the corresponding eigenvalues of the sample variance matrix are zero or very close to zero. In Table 2.2, however, we examine the effect of various variance assumptions on the rank of $\mathbf{\Sigma}$.

We are going to test that there is no area effect in our example, i.e.,

$$H : \mathbf{\Gamma}'\mathbf{\Xi}\mathbf{c} = \mathbf{0} \text{ versus } A : \mathbf{\Gamma}'\mathbf{\Xi}\mathbf{c} \neq \mathbf{0}, \tag{2.1}$$

where

$$\mathbf{c}' = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

and $\mathbf{\Gamma}$ is as in (3.2). Let $\mathcal{C}(\mathbf{Z})$ denote the column vector space generated by

the columns of $\mathbf{Z}$. Clearly $\mathcal{C}(\mathbf{c}) \subseteq \mathcal{C}(\mathbf{X})$ as $r(\mathbf{X}) = 5$ and we will utilize Theorem 3.3. However, note that $r(\mathbf{c}) = 1$ and therefore the exact distribution for λ in the theorem is available, i.e.

$$F = \frac{n - r(\mathbf{X}) + 1 - r}{r} \frac{1 - \lambda}{\lambda} \sim F_{r,n-r(\mathbf{X})+1-r}, \tag{2.2}$$

where $r = r(\mathbf{\Sigma})$. In Table 2.2 the results of this test under various assumptions on $r(\mathbf{\Sigma})$ are presented.

Table 2.2. The hypothesis in (2.1) is tested. Let λ and F be given in Theorem 3.3 and (2.2), respectively and let P denote the probability of obtaining a larger value than F.

$r(\mathbf{\Sigma})$	8	7	6	5	4	3	2	1
λ	0.16	0.16	0.20	0.21	0.63	0.65	0.96	0.98
F	2.64	3.76	4.09	5.16	1.18	1.61	0.19	0.22
P	0.18	0.082	0.055	0.027	0.39	0.25	0.83	0.65

Thus, it follows from Table 2.2 that when we do not put any rank restrictions on $\mathbf{\Sigma}$, i.e. $r(\mathbf{\Sigma}) = 8$, the hypothesis H in (2.1) can not be rejected. However, under assumptions that $\mathbf{\Sigma}$ is singular of rank 6 the hypothesis is rejected at 5.5% level of significance. Our conclusions are that there exist some differences between the areas which we certainly can not draw if just performing a test when $\mathbf{\Sigma}$ is of full rank and which is the standard test in MANOVA. For comparison, each variable has been analyzed separately by performing a univariate ANOVA. The results are presented in Table 2.3.

Table 2.3. Let P denote the significance level for the rejection of the hypothesis of no Area effect in a univariate ANOVA. The variables $V1 - V8$ are defined in Table 2.1.

	V1	V2	V3	V4	V5	V6	V7	V8
P	0.0002	1.00	0.15	0.64	0.32	0.51	0.09	0.14

Hence, from Table 2.3 it follows that there seem to be some differences between the groups although they are not very pronounced. Vaguely speaking, one may say that these differences were found in the multivariate approach under the assumption that the covariance matrix was not of full rank.

3. Tests and estimators of parameters in MANOVA with singular covariance

In this section, we derive the maximum likelihood estimators of the parameters and apply the likelihood ratio test for the general hypothesis.

The general MANOVA model can be written

$$\mathbf{Y} = \Xi\mathbf{X} + \Sigma^{\frac{1}{2}}\mathbf{E}, \tag{3.1}$$

where $\mathbf{Y}$: $p \times n$, Ξ: $p \times k$, $\mathbf{X}$: $k \times n$, the rank of Σ is denoted $r(\Sigma) = r \leq p$, $n - r(\mathbf{X}) \geq r$, the columns of $\mathbf{E}$ are i.i.d. $N_p(\mathbf{0}, \mathbf{I})$ and $\Sigma^{\frac{1}{2}}$ is the positive semi-definite square root of Σ. It is assumed that Ξ and Σ are unknown while $\mathbf{X}$ is a matrix of known constants. As it has been noted by several authors (e.g. Rao 1973, p. 297; Nordström 1985) the singularity in Σ implies certain restrictions on the parameter space as well as the observation matrix. By performing a one-one transformation of (3.1) this can be explored in some detail:

$$\begin{pmatrix} \Gamma' \\ \Gamma_0' \end{pmatrix} \mathbf{Y} = \begin{pmatrix} \Gamma' \\ \Gamma_0' \end{pmatrix} \Xi\mathbf{X} + \begin{pmatrix} \Gamma'\Sigma^{\frac{1}{2}}\mathbf{E} \\ \mathbf{0} \end{pmatrix}, \tag{3.2}$$

where $(\Gamma : \Gamma_0)$ spans the whole space, $\Gamma'\Gamma_0 = \mathbf{0}$, $\Gamma_0'\Sigma = \mathbf{0}$, $\Gamma'\Gamma = \mathbf{I}_r$, $\Gamma : p \times r$, $\Sigma = \Gamma\Lambda\Gamma'$, where $\Lambda = (\lambda_i)$ is the diagonal matrix of positive and ordered eigenvalues of Σ, $\Gamma_0 : p \times (p - r)$ and $\Gamma_0'\Gamma_0 = \mathbf{I}_{p-r}$. Throughout, a matrix $\mathbf{A}_0$ will be any matrix which satisfies $\mathcal{C}(\mathbf{A}_0) = \mathcal{C}(\mathbf{A})^{\perp}$.

From (3.2) it follows with probability 1 that

$$\Gamma_0'\mathbf{Y} = \Gamma_0'\Xi\mathbf{X}. \tag{3.3}$$

Equation (3.3) may be viewed as a linear system of equations in Ξ. From Rao & Mitra (1971, Theorem 2.3.2, p. 24) a necessary and sufficient condition for the equation (3.3) to have a solution is that

$$\Gamma_0'(\Gamma_0')^{+}\Gamma_0'\mathbf{Y}\mathbf{X}^{+}\mathbf{X} = \Gamma_0'\mathbf{Y}, \tag{3.4}$$

where $\mathbf{A}^{+}$ denotes the Moore–Penrose inverse of $\mathbf{A}$. That is $\mathbf{A}^{+}$ satisfies the four conditions (i) $\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}$, (ii) $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+}$, (iii) $\mathbf{A}^{+}\mathbf{A} = (\mathbf{A}^{+}\mathbf{A})'$ and (iv) $\mathbf{A}\mathbf{A}^{+} = (\mathbf{A}\mathbf{A}^{+})'$. It may be noted that in the above equation we may use any generalized inverse of $\mathbf{A}$, say $\mathbf{A}^{-}$, satisfying only the condition $\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$. However, the Moore–Penrose inverse is unique. Since

$$\Gamma_0'\Gamma_0\Gamma_0' = \Gamma_0', \quad \Gamma_0\Gamma_0'\Gamma_0 = \Gamma_0,$$

and $\Gamma_0\Gamma_0'$ and $\Gamma_0'\Gamma_0$ are symmetric matrices, it follows that $\Gamma_0 : (p - r) \times p$ is the Moore–Penrose inverse of Γ_0'. It can also be shown that the Moore–Penrose inverse of $\mathbf{X}$ is given by

$$\mathbf{X}^{+} = \mathbf{X}'(\mathbf{X}\mathbf{X}')^{+},$$

where it is known that $(\mathbf{X}\mathbf{X}')^+ = \mathbf{X}^{+'}\mathbf{X}^+$, $(\mathbf{X}')^+ = (\mathbf{X}^+)'$ and $(\mathbf{X}^+)^+ = \mathbf{X}$. Then, the left-side of (3.4) becomes

$$\mathbf{\Gamma}_0'\mathbf{Y}\mathbf{X}^+\mathbf{X} = (\mathbf{\Gamma}_0'\mathbf{\Xi}\mathbf{X})\mathbf{X}^+\mathbf{X} = \mathbf{\Gamma}_0'\mathbf{\Xi}\mathbf{X},$$

which is equal to $\mathbf{\Gamma}_0'\mathbf{Y}$, from (3.3). Thus, the necessary and sufficient condition is satisfied. From equation (2.3.5) of Rao & Mitra (1971, p. 24), a general solution written in terms of Moore–Penrose inverses is given by

$$\mathbf{\Xi} = \mathbf{\Gamma}_0\mathbf{\Gamma}_0'\mathbf{Y}\mathbf{X}^+ + \mathbf{U} - \mathbf{\Gamma}_0\mathbf{\Gamma}_0'\mathbf{U}\mathbf{X}\mathbf{X}^+,$$

where $\mathbf{U} : p \times k$ is an arbitrary matrix of parameters, which is of the same dimension as $\mathbf{\Xi}$. Since

$$\mathbf{\Gamma}_0\mathbf{\Gamma}_0'\mathbf{\Gamma}a\mathbf{\Gamma}' = \mathbf{I}_p,$$

the general solution can be rewritten as

$$\mathbf{\Xi} = \mathbf{\Gamma}_0\mathbf{\Gamma}_0'\mathbf{Y}\mathbf{X}^+\mathbf{\Gamma}a\mathbf{\Theta} + \mathbf{\Gamma}_0\boldsymbol{\eta}(\mathbf{I} - \mathbf{X}\mathbf{X}^+), \tag{3.5}$$

where $\mathbf{\Theta} = \mathbf{\Gamma}'\mathbf{U} : r \times k$, and $\boldsymbol{\eta} = \mathbf{\Gamma}_0'\mathbf{U} : (p-r) \times k$. It may be noted that if the $k \times n$ matrix $\mathbf{X}$ is of full rank $k \leq n$, then $\mathbf{X}\mathbf{X}^+ = \mathbf{I}_k$ and the third term on the right-side of (3.4) vanishes. Thus, the $(p-r) \times k$ matrix $\boldsymbol{\eta}$ is an arbitrary matrix due to the rank deficiency in $\mathbf{X}$, while the $r \times k$ matrix $\mathbf{\Theta}$ is a matrix of parameters that can be estimated from the random part of the model (3.2).

The solution (3.5) has been obtained from the deterministic part of the model (3.2) and does not affect the random part of (3.2) as can be seen by substituting the solution into (3.2). The random part of the model is given by

$$\mathbf{\Gamma}'\mathbf{Y} = \mathbf{\Theta}\mathbf{X} + \mathbf{\Gamma}'\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{E}. \tag{3.6}$$

In (3.6) the dispersion matrix $\mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma}$ is non-singular as we shall show later. Since $\mathbf{\Gamma}$ is unknown, it is not obvious that estimation can be carried out. However, as will be seen from the following lemma, the known sums-of-squares matrix

$$\mathbf{S} = \mathbf{Y}(\mathbf{I} - \mathbf{X}'(\mathbf{X}\mathbf{X}')^-\mathbf{X})\mathbf{Y}' \tag{3.7}$$

gives some insight in $\mathcal{C}(\mathbf{\Gamma}) = \mathcal{C}(\mathbf{\Sigma})$ which turns out to be crucial for estimation and testing.

In the subsequent $W_p(\bullet, n)$ represents the Wishart distribution with n degrees of freedom.

Lemma 3.1. *Let* $\mathbf{S} \sim W_p(\mathbf{\Sigma}, n - r(\mathbf{X}))$. *Then, with probability 1,* $\mathcal{C}(\mathbf{S}) \subseteq \mathcal{C}(\mathbf{\Sigma})$ *and if* $n - r(\mathbf{X}) \geq r = r(\mathbf{\Sigma})$, $\mathcal{C}(\mathbf{S}) = \mathcal{C}(\mathbf{\Sigma})$.

Proof. We may note that $\mathbf{X}^+\mathbf{X}$ is a symmetric matrix. That is, $\mathbf{X}^+\mathbf{X} = (\mathbf{X}'\mathbf{X}^+)'$. Hence

$$\mathbf{S} = \mathbf{Y}(\mathbf{I}_n - \mathbf{X}^+\mathbf{X})\mathbf{Y}' = \mathbf{Y}(\mathbf{I}_n - \mathbf{X}^+\mathbf{X})(\mathbf{I}_n - \mathbf{X}^+\mathbf{X})'\mathbf{Y}'.$$

Thus,

$$\mathcal{C}(\mathbf{S}) \subseteq \mathcal{C}(\mathbf{Y}(\mathbf{I}_n - \mathbf{X}^+\mathbf{X})) = \mathcal{C}(\Xi\mathbf{X} + \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{E})(\mathbf{I}_n - \mathbf{X}^+\mathbf{X}))$$
$$= \mathcal{C}(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{E}(\mathbf{I}_n - \mathbf{X}^+\mathbf{X})) \subseteq \mathcal{C}(\boldsymbol{\Sigma}^{\frac{1}{2}}) = \mathcal{C}(\boldsymbol{\Sigma}).$$

Hence, the rank of $\mathbf{S}$, $r(\mathbf{S}) \leq r(\boldsymbol{\Sigma}) = r$, that is the rank of the covariance matrix $\mathbf{S}$ is always less or equal to the rank of $\boldsymbol{\Sigma}$. However, to prove that the rank of $\mathbf{S}$ is exactly equal to $r = r(\boldsymbol{\Sigma})$ when $r \leq n$, we use the assumption that $\mathbf{S}$ is Wishart distributed. Since $\boldsymbol{\Sigma}$ is of rank r we can write $\boldsymbol{\Sigma} = \mathbf{BB}'$, where $\mathbf{B}$ is a $p \times r$ matrix of rank r (see Srivastava & Khatri, 1979, Definition 2.1.1, p. 43). Hence, under the assumption of Wishartness, $\mathbf{S}$ can be written $\mathbf{S} = \mathbf{BZZ}'\mathbf{B}'$, where $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ is an $r \times n$ matrix and $\mathbf{z}_i$'s are i.i.d. $N_r(\mathbf{0}, \mathbf{I})$. Since $n \geq r$, the rank of the matrix $\mathbf{Z}$ is r, with probability one (see Srivastava & Khatri 1979, p. 73). Hence,

$$r(\mathbf{S}) = r(\mathbf{BZ}) = r(\mathbf{B}) = r,$$

from Corollary 1.5.3 of Srivastava & Khatri (1979, p. 10). □

Corollary 3.1L. *If* $\mathbf{S} = \mathbf{HLH}'$ *where* $\mathbf{H}'\mathbf{H} = \mathbf{I}_r$, *then* $\boldsymbol{\Gamma} = \mathbf{HQ}$ *where* $\mathbf{Q}$ *is an* $r \times r$ *orthogonal matrix.*

From the above result, it follows that $\boldsymbol{\Gamma}_0'\mathbf{S}\boldsymbol{\Gamma}_0 = \mathbf{0}$ with probability one and thus $\boldsymbol{\Gamma}_0$ may be considered as known. Next we show that the matrix $\boldsymbol{\Gamma}'\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{E}$ occurring in (3.6) is of rank $r = r(\boldsymbol{\Sigma})$. We note that since $\boldsymbol{\Sigma}^{\frac{1}{2}} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Gamma}'$, and $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \mathbf{I}_r$,

$$\boldsymbol{\Gamma}'\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{E} = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Gamma}'\mathbf{E},$$

where the columns of $\mathbf{E}$ are i.i.d. $N_p(\mathbf{0}, \mathbf{I})$. Thus, the n columns of $\boldsymbol{\Gamma}'\mathbf{E}$ are i.i.d. $N_r(\mathbf{0}, \mathbf{I})$ and hence as above the rank is r. Thus

$$r(\boldsymbol{\Gamma}'\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{E}) = r(\boldsymbol{\Gamma}'\mathbf{E}) = r$$

as $\boldsymbol{\Lambda}^{\frac{1}{2}}$ is nonsingular. From this it follows that

$$(\boldsymbol{\Gamma}'\mathbf{Y} - \boldsymbol{\Theta}\mathbf{X})(\boldsymbol{\Gamma}'\mathbf{Y} - \boldsymbol{\Theta}\mathbf{X})'$$

is nonsingular with probability one. In the model (3.6), the matrices of unknown parameters are $\boldsymbol{\Gamma}$, $\boldsymbol{\Theta}$ and Λ as

$$\boldsymbol{\Gamma}'\boldsymbol{\Sigma}^{\frac{1}{2}} = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Gamma}'.$$

Thus, the likelihood function for these parameters is given by

$$L(\boldsymbol{\Theta}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}) = c_0|\boldsymbol{\Lambda}|^{-\frac{1}{2}n}\text{etr}\{-\tfrac{1}{2}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Gamma}'\mathbf{Y} - \boldsymbol{\Theta}\mathbf{X})(\boldsymbol{\Gamma}'\mathbf{Y} - \boldsymbol{\Theta}\mathbf{X})'\}, \; c_0 = (2\pi)^{-\frac{nr}{2}}, \tag{3.8}$$

where $\text{etr}(\mathbf{A})$ stands for the exponential of the trace of the matrix $\mathbf{A}$. From Srivastava & von Rosen (2002) we get the following theorem.

Theorem 3.1. *Assume that* $n-k \geq r(\mathbf{\Sigma})$. *Then the maximum likelihood estimators of the parameters are given by*

$$\begin{aligned}
\widehat{\mathbf{\Gamma}} &= \mathbf{H}, \\
\widehat{\mathbf{\Lambda}} &= \mathbf{L}/n, \\
\widehat{\mathbf{\Sigma}} &= \mathbf{S}/n, \\
\widehat{\mathbf{\Theta}} &= \mathbf{H}'\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{+} + \mathbf{U}_0(\mathbf{I} - \mathbf{X}\mathbf{X}^{+}), \\
\widehat{\mathbf{\Xi}}\mathbf{X} &= \mathbf{Y}\mathbf{X}^{+}\mathbf{X}, \\
\widehat{\mathbf{\Xi}} &= \mathbf{Y}\mathbf{X}^{+} + \mathbf{H}\mathbf{U}_0(\mathbf{I} - \mathbf{X}\mathbf{X}^{+}),
\end{aligned}$$

where $\mathbf{U}_0$ *is an arbitrary matrix.*

Corollary 3.1. *Assume that the* $k \times n$ *matrix* $\mathbf{X}$ *is of full rank. Then the maximum likelihood estimators of* $\mathbf{\Theta}$ *and* $\mathbf{\Xi}$ *are given by*

$$\begin{aligned}
\widehat{\mathbf{\Theta}} &= \mathbf{H}'\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X})^{-1}, \\
\widehat{\mathbf{\Xi}} &= \mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X})^{-1},
\end{aligned}$$

respectively, and all the other estimators remain the same as in Theorem 3.1.

Since $\mathbf{\Gamma}_0'\mathbf{Y} = \mathbf{\Gamma}_0'\mathbf{\Xi}\mathbf{X}$ with probability 1, the unknown mean parameters that may be tested for redundancy are $\mathbf{\Gamma}'\mathbf{\Xi}$. We may wish to test the hypothesis

$$H : \mathbf{\Gamma}'\mathbf{\Xi}\mathbf{C} = \mathbf{0} \text{ versus } A : \mathbf{\Gamma}'\mathbf{\Xi}\mathbf{C} \neq \mathbf{0}, \tag{3.9}$$

where $\mathbf{C} : k \times l$. The restrictions in (3.9) are equivalent to [see (3.5)]

$$\mathbf{\Theta}\mathbf{C} = \mathbf{0}$$

which may be written as

$$\mathbf{\Theta} = \mathbf{\Theta}_1\mathbf{C}_0',$$

where $\mathbf{\Theta}_1$ is a new parameter matrix. Therefore, instead of (3.6) we have the model

$$\mathbf{\Gamma}'\mathbf{Y} = \mathbf{\Theta}_1\mathbf{C}_0'\mathbf{X} + \mathbf{\Gamma}'\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{E}.$$

Thus, the next theorem is a consequence of Theorem 3.1.

Theorem 3.2. *Let*

$$\mathbf{S}_H = \mathbf{Y}(\mathbf{I} - \mathbf{X}'\mathbf{C}_0(\mathbf{C}_0'\mathbf{X}\mathbf{X}'\mathbf{C}_0)^{-}\mathbf{C}_0'\mathbf{X})\mathbf{Y}',$$

$\mathbf{H}_H$*:* $p \times r$ *be a semiorthogonal matrix of eigenvectors which correspond to the* $r = r(\mathbf{\Sigma})$ *largest eigenvalues* l_{Hi} *of* $\mathbf{S}_H$, *i.e.* $\mathbf{H}_H'\mathbf{H}_H = \mathbf{I}_r$, *and* $(\mathbf{L}_H)_{ii} =$

(l_{Hi}). *Assume that* $n - r(\mathbf{X} : \mathbf{C}) + r(\mathbf{C}) \geq r$. *The maximum likelihood estimators of the parameters in* (3.1) *under* H *in* (3.9) *are given by*

$$\begin{aligned}
\widehat{\mathbf{\Lambda}} &= \tfrac{1}{n}\mathbf{L}_H, \\
\widehat{\mathbf{\Gamma}} &= \mathbf{H}_H, \\
\widehat{\mathbf{\Sigma}} &= \tfrac{1}{n}\mathbf{H}_H\mathbf{L}_H\mathbf{H}_H', \\
\widehat{\mathbf{\Xi}} &= \mathbf{Y}\mathbf{X}'\mathbf{C}_0(\mathbf{C}_0'\mathbf{X}\mathbf{X}'\mathbf{C}_0)^- + \mathbf{U}(\mathbf{I} - (\mathbf{C}_0'\mathbf{X})(\mathbf{C}_0'\mathbf{X})^-),
\end{aligned}$$

where $\mathbf{U}$ *is an arbitrary matrix.*

Observe that the test in (3.9) is in fact a between individuals test, e.g. a comparison of different treatment groups. The $\mathbf{\Gamma}$ matrix is only used for reducing the dimension and this reduction is the same for each treatment group. If we are curious about what kind of linear combinations are tested, Corollary 3.1L may be utilized and with probability 1 we are studying $\mathbf{H}'\mathbf{\Xi}\mathbf{C} = \mathbf{0}$.

We end this section by presenting the likelihood ratio test for H in (3.9). Let $U_{p,m,n}$ be the standard test statistic in MANOVA, i.e.

$$U_{p,m,n} = \frac{|\mathbf{SSE}|}{|\mathbf{SSE} + \mathbf{SST}|},$$

where $\mathbf{SSE} \sim W_p(\mathbf{\Sigma}, n)$ is the sums-of-squares matrix due to error and $\mathbf{SST} \sim \mathbf{W}_p(\mathbf{\Sigma}, m)$ is the sums-of-squares matrix due to hypothesis (see e.g. Srivastava & Carter 1983, p. 97).

Theorem 3.3. *Suppose* $\mathcal{C}(\mathbf{C}) \subseteq \mathcal{C}(\mathbf{X})$. *The likelihood ratio test rejects* H, *given by* (3.9), *if*

$$\lambda = U_{r(\mathbf{\Sigma}),r(\mathbf{C}),n-r(\mathbf{X})} \leq c,$$

where

$$\lambda = \frac{|\widehat{\mathbf{\Gamma}}'\mathbf{S}\widehat{\mathbf{\Gamma}}|}{|\widehat{\mathbf{\Gamma}}_H'\mathbf{S}_H\widehat{\mathbf{\Gamma}}_H|} = \frac{\prod_{i=1}^r l_i}{\prod_{i=1}^r l_{Hi}},$$

when $l_1 > \cdots > l_r$ *and* $l_{H1} > \cdots > l_{Hr}$ *are the ordered eigenvalues of* $\mathbf{S}$ *and* $\mathbf{S}_H$, *respectively,* $\widehat{\mathbf{\Gamma}}$ *and* $\mathbf{S}$ *are as in Theorem 3.1, and* $\widehat{\mathbf{\Gamma}}_H$ *and* $\mathbf{S}_H$ *are as in Theorem 3.2. The constant* c *is chosen so that a prespecified significance level is obtained.*

Proof. Now,

$$\begin{aligned}\lambda^{-1} =& \frac{|\widehat{\mathbf{\Gamma}}_H'\mathbf{Y}(\mathbf{I}-\mathbf{X}'\mathbf{C}_0(\mathbf{C}_0'\mathbf{X}\mathbf{X}'\mathbf{C}_0)^-\mathbf{C}_0'\mathbf{X})\mathbf{Y}'\widehat{\mathbf{\Gamma}}|}{|\widehat{\mathbf{\Gamma}}'\mathbf{S}\widehat{\mathbf{\Gamma}}|} \\ =& \frac{\widehat{\mathbf{\Gamma}}_H'\mathbf{Y}(\mathbf{I}-\mathbf{X}'(\mathbf{X}\mathbf{X}')^-\mathbf{X})\mathbf{Y}'\widehat{\mathbf{\Gamma}}_H}{|\widehat{\mathbf{\Gamma}}'\mathbf{S}\widehat{\mathbf{\Gamma}}|} \\ &+ \frac{\widehat{\mathbf{\Gamma}}_H'\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^-\mathbf{C}(\mathbf{C}'(\mathbf{X}\mathbf{X}')^-\mathbf{C})^-\mathbf{C}'(\mathbf{X}\mathbf{X}')^-\mathbf{X}\mathbf{Y}'\widehat{\mathbf{\Gamma}}|}{|\widehat{\mathbf{\Gamma}}'\mathbf{S}\widehat{\mathbf{\Gamma}}|}\end{aligned} \tag{3.10}$$

because $\mathcal{C}(\mathbf{C}) \subseteq \mathcal{C}(\mathbf{X})$. However, since for some orthogonal matrix $\mathbf{Q}$, $\widehat{\mathbf{\Gamma}} = \widehat{\mathbf{\Gamma}}_H\mathbf{Q}$,

$$|\widehat{\mathbf{\Gamma}}_H'\mathbf{S}\widehat{\mathbf{\Gamma}}_H| = |\widehat{\mathbf{\Gamma}}'\mathbf{S}\widehat{\mathbf{\Gamma}}|.$$

Moreover, $\mathcal{C}(\widehat{\mathbf{\Gamma}}_H) = \mathcal{C}(\mathbf{\Gamma})$ implies that

$$\widehat{\mathbf{\Gamma}}_H'(\widehat{\mathbf{\Gamma}}_H'\mathbf{S}\widehat{\mathbf{\Gamma}}_H)^{-1}\widehat{\mathbf{\Gamma}}_H' = \mathbf{\Gamma}'(\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}'.$$

Thus, from (3.10), with probability 1,

$$\lambda^{-1} = \frac{|\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma} + \mathbf{\Gamma}'\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^-\mathbf{C}(\mathbf{C}'(\mathbf{X}\mathbf{X}')^-\mathbf{C})^-\mathbf{C}'(\mathbf{X}\mathbf{X}')^-\mathbf{X}\mathbf{Y}'\mathbf{\Gamma}|}{|\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma}|}.$$

Since

$$\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma}\mathbf{\Lambda}^{-\frac{1}{2}} \sim \mathbf{W}_{r(\mathbf{\Sigma})}(\mathbf{I}, n - r(\mathbf{X}))$$

and under H

$$\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Gamma}'\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^-\mathbf{C} \sim \mathbf{N}_{r(\mathbf{\Sigma}),r(\mathbf{C})}(\mathbf{0}, \mathbf{I}, \mathbf{C}'(\mathbf{X}\mathbf{X}')^-\mathbf{C}),$$

and both expressions are independent of $\mathbf{\Gamma}$ as well as $\mathbf{\Lambda}$, the theorem is established. Here $N_{p,q}(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{\Psi})$ denotes the matrix normal distribution, i.e. $\mathbf{X} \sim N_{p,q}(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{\Psi})$ is equivalent to $\mathrm{vec}\mathbf{X} \sim N_{pq}(\mathrm{vec}\boldsymbol{\mu}, \mathbf{\Psi} \otimes \mathbf{\Sigma})$. □

Remark. If we do not require $\mathcal{C}(\mathbf{C}) \subseteq \mathcal{C}(\mathbf{X})$, it follows with probability 1 that

$$\lambda = \frac{|\widehat{\mathbf{\Gamma}}'\mathbf{S}\widehat{\mathbf{\Gamma}}|}{|\widehat{\mathbf{\Gamma}}_H'\mathbf{S}_H\widehat{\mathbf{\Gamma}}_H|} \sim U_{r(\mathbf{\Sigma}),r(\mathbf{N}),n-r(\mathbf{X})},$$

where $\mathbf{N}$ is a matrix satisfying $\mathcal{C}(\mathbf{N}) = \mathcal{C}(\mathbf{X}) \cap \mathcal{C}(\mathbf{C})$.

4. Confidence intervals for $\mathbf{\Xi}$

In this section we will construct four different types of confidence intervals: a single confidence interval for an arbitrary bilinear combination of the elements in $\mathbf{\Xi}$, i.e. $\mathbf{a}'\mathbf{\Xi}\mathbf{b}$ for any given $\mathbf{a}$ and $\mathbf{b}$; simultaneous within subjects confidence intervals, $\mathbf{a}'\mathbf{\Xi}\mathbf{b}$, i.e. for all $\mathbf{a}$ and any given $\mathbf{b}$; simultaneous between subjects confidence intervals, $\mathbf{a}'\mathbf{\Xi}\mathbf{b}$, i.e. for a given $\mathbf{a}$ and all choices of $\mathbf{b}$; simultaneous confidence intervals, $\mathbf{a}'\mathbf{\Xi}\mathbf{b}$, i.e. for all choices of $\mathbf{a}$ and $\mathbf{b}$.

Throughout, $\mathbf{a} \in \mathcal{C}(\mathbf{\Sigma})$ and $\mathbf{b} \in \mathcal{C}(\mathbf{X})$, because we are dealing with matrices which are not supposed to be of full rank.

For the first case we start with the ratio

$$R = \frac{(\mathbf{a}'(\mathbf{YX}'(\mathbf{XX}')^- - \mathbf{\Xi})\mathbf{b})^2}{\mathbf{a}'\mathbf{Sab}'(\mathbf{XX}')^-\mathbf{b}}.$$

The ratio is meaningful since $\mathbf{a} \in \mathcal{C}(\mathbf{\Sigma})$ and $\mathbf{b} \in \mathcal{C}(\mathbf{X}')$. Then

$$\frac{\mathbf{a}'\mathbf{Sa}}{\mathbf{a}'\mathbf{\Sigma a}} \sim \chi^2(n - r(\mathbf{X}))$$

and

$$\frac{(\mathbf{a}'(\mathbf{YX}'(\mathbf{XX}')^- - \mathbf{\Xi})\mathbf{b})^2}{\mathbf{a}'\mathbf{\Sigma ab}'(\mathbf{XX}')^-\mathbf{b}} \sim \chi^2(1)$$

which are mutually independent random variables. Thus,

$$(n - r(\mathbf{X}))R \sim F(1, n - r(\mathbf{X})).$$

In the next theorem we present a confidence interval based on these quantities.

Theorem 4.1. *Let* $\mathbf{a} \in \mathcal{C}(\mathbf{\Sigma})$ *and* $\mathbf{b} \in \mathcal{C}(\mathbf{X})$. *For the MANOVA model* (3.1),

$$P(\mathbf{a}'\mathbf{\Xi b} \in \mathbf{a}'\mathbf{YX}'(\mathbf{XX}')^-\mathbf{b} \pm \{\frac{1}{n - r(\mathbf{X})} F_{1-\alpha}(1, n - r(\mathbf{X}))\mathbf{a}'\mathbf{Sab}'(\mathbf{XX}')^-\mathbf{b}\}^{\frac{1}{2}}) = 1 - \alpha.$$

To obtain simultaneous within subjects confidence intervals we start again with R but now we rely on the Cauchy-Schwarz inequality because we are interested in

$$\max_{\mathbf{a} \in \mathcal{C}(\mathbf{\Sigma})} R.$$

Since $\mathbf{\Sigma} = \mathbf{\Gamma\Delta\Gamma}'$, we study

$$\max_{\mathbf{a} \in \mathcal{C}(\mathbf{\Sigma})} R = \max_{\mathbf{q}} \frac{(\mathbf{q}'\mathbf{\Gamma}'(\mathbf{YX}'(\mathbf{XX}')^- - \mathbf{\Xi})\mathbf{b})^2}{\mathbf{q}'\mathbf{\Gamma}'\mathbf{S\Gamma q b}'(\mathbf{XX}')^-\mathbf{b}}$$
$$= \frac{\mathbf{b}'(\mathbf{YX}'(\mathbf{XX}')^- - \mathbf{\Xi})'\mathbf{\Gamma}(\mathbf{\Gamma}'\mathbf{S\Gamma})^{-1}\mathbf{\Gamma}'(\mathbf{YX}'(\mathbf{XX}')^- - \mathbf{\Xi})\mathbf{b}}{\mathbf{b}'(\mathbf{XX}')^-\mathbf{b}}$$

which is a Hotelling T^2 statistic since

$$\mathbf{\Gamma}'\mathbf{S\Gamma} \sim W_r(\mathbf{\Lambda}, n - r(\mathbf{X}))$$

and

$$\mathbf{\Gamma}'(\mathbf{YX}'(\mathbf{XX}')^- - \mathbf{\Xi})\mathbf{b}(\mathbf{b}'(\mathbf{XX}')^-\mathbf{b})^{-1/2} \sim \mathbf{N_r}(\mathbf{0}, \mathbf{\Lambda}).$$

Therefore the next theorem has been established.

Theorem 4.2. *Let* $\mathbf{b} \in \mathcal{C}(\mathbf{X})$ *and all* $\mathbf{a} \in \mathcal{C}(\mathbf{\Sigma})$. *Simultaneous within individuals confidence intervals for the MANOVA model* (3.1) *are given by*

$$P(\mathbf{a}'\mathbf{\Xi}\mathbf{b} \in \mathbf{a}'\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{b} \pm \{\tfrac{r}{n-r(\mathbf{X})-r+1}F_{1-\alpha}(r, n-r(\mathbf{X})-r+1)\mathbf{a}'\mathbf{S}\mathbf{a}\mathbf{b}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{b}\}^{\frac{1}{2}} = 1-\alpha.$$

For the simultaneous confidence intervals between the individuals it is noted that

$$\max_{\mathbf{b}\in\mathcal{C}(\mathbf{X})} R = \frac{\mathbf{a}'(\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-} - \mathbf{\Xi})\mathbf{X}\mathbf{X}'(\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-} - \mathbf{\Xi})'\mathbf{a}}{\mathbf{a}'\mathbf{S}\mathbf{a}}$$

which is F-distributed.

Theorem 4.3. *Let* $\mathbf{a} \in \mathcal{C}(\mathbf{\Sigma})$ *and all* $\mathbf{b} \in \mathcal{C}(\mathbf{X})$. *For the MANOVA model* (3.1), *simultaneous between individuals confidence intervals are given by*

$$P(\mathbf{a}'\mathbf{\Xi}\mathbf{b} \in \mathbf{a}'\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{b} \pm \{\tfrac{r(\mathbf{X})}{n-r(\mathbf{X})}F_{1-\alpha}(r(\mathbf{X}), n-r(\mathbf{X}))\mathbf{a}'\mathbf{S}\mathbf{a}\mathbf{b}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{b}\}^{\frac{1}{2}}) = 1-\alpha.$$

Finally for the simultaneous confidence intervals, i.e. when both $\mathbf{a}$ and $\mathbf{b}$ may vary, we note that

$$\max_{\mathbf{a}\in\mathcal{C}(\mathbf{\Sigma}),\mathbf{b}\in\mathcal{C}(\mathbf{X})} R = \max_{\mathbf{b}\in\mathcal{C}(\mathbf{X})} \frac{\mathbf{b}'(\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-} - \mathbf{\Xi})'\mathbf{\Gamma}(\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}'(\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-} - \mathbf{\Xi})\mathbf{b}}{\mathbf{b}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{b}} \equiv e_1, \tag{4.1}$$

where e_1 is the largest eigenvalue of $\mathbf{B}\mathbf{W}^{-1}$, where

$$\mathbf{B} = \mathbf{\Gamma}'(\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-} - \mathbf{\Xi})\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{X}(\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-} - \mathbf{\Xi})'\mathbf{\Gamma} \sim \mathbf{W}_{\mathbf{r}}(\mathbf{\Lambda}, \mathbf{r}(\mathbf{X})).$$

and

$$\mathbf{W} = \mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma} \sim W_r(\mathbf{\Lambda}, n-r(\mathbf{X})).$$

The distribution of the largest root of $\mathbf{W}^{-1}\mathbf{B}$ has been derived by Pillai (1955) and Roy (1957) and tables where given by Heck (1960). It can also be obtained from the relationship $e_{1\alpha} = \frac{c_\alpha}{1-c_\alpha}$, where c_α is tabulated by Srivastava (2002, Table B.7).

Theorem 4.4. *Let* c_α *be the* α*-quantile of the distribution of* e_1 *in* (4.1). *Simultaneous confidence intervals for the MANOVA model* (3.1) *are given by*

$$P(\mathbf{a}'\mathbf{\Xi}\mathbf{b} \in \mathbf{a}'\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{b} \pm \{\mathbf{c}_{1-\alpha}/(1-\mathbf{c}_{1-\alpha})\mathbf{a}'\mathbf{S}\mathbf{a}\mathbf{b}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{b}\}^{\frac{1}{2}}) = 1-\alpha.$$

5. Outliers

The problem of finding outliers has received much attention over the years. For a brief review see Srivastava & von Rosen (1998). Consider (3.1) which can be rewritten as

$$\mathbf{y}_i = \Xi\mathbf{x_i} + \Sigma^{1/2}\epsilon_i, \qquad \mathbf{i} = 1, 2, \ldots, \mathbf{n},$$

where $\epsilon_i \sim N(\mathbf{0}, \Sigma)$. A shift in the mean of the ith observation vector may be expressed as

$$E[\mathbf{y}_i] = \Xi\mathbf{x_i} + \delta. \tag{5.1}$$

We define an observation $\mathbf{y}_i$ to be an outlier if the hypothesis $H_0 : \delta = \mathbf{0}$ is rejected according to a fixed significance level. However, since we will test each observation, many tests are performed and the problem is to find a proper critical value. Let $\mathbf{e}_i$ be the ith column of $\mathbf{I}_n$. Then from (5.1) it follows that

$$E[\mathbf{Y}] = (\delta, \Xi)\begin{pmatrix}\mathbf{e}_i' \\ \mathbf{X}\end{pmatrix} \equiv \Xi^\star\mathbf{X_i^\star}. \tag{5.2}$$

If $\delta = \mathbf{0}$ then we say that we have model H. Otherwise the model in (5.2) is denoted H_i. The likelihood ratio criterion for testing H against H_i is according to Theorem 3.3, in obvious notations, given by

$$\lambda_i = \frac{|\widehat{\mathbf{\Gamma}}_{H_i}'\mathbf{S}_i\widehat{\mathbf{\Gamma}}_{H_i}|}{|\widehat{\mathbf{\Gamma}}_H'\mathbf{S}\widehat{\mathbf{\Gamma}}_H|},$$

where

$$\begin{aligned}\mathbf{S} &= \mathbf{Y}(\mathbf{I} - \mathbf{R})\mathbf{Y}', \\ \mathbf{S}_i &= \mathbf{Y}(\mathbf{I} - \mathbf{R}_i)\mathbf{Y}', \\ \mathbf{R} &= \mathbf{X}'(\mathbf{X}\mathbf{X}')^-\mathbf{X}, \qquad \mathbf{R}_i = \mathbf{X}_i^{\star\prime}(\mathbf{X}_i^\star\mathbf{X}_i^{\star\prime})^-\mathbf{X}_i^\star.\end{aligned}$$

Now, with probability 1

$$\lambda_i = \frac{|\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma}|}{|\mathbf{\Gamma}'\mathbf{S}_i\mathbf{\Gamma}|}.$$

Observe that

$$\mathbf{S}_i = \mathbf{S} - (1 - r_{ii})^{-1}\mathbf{d}_i\mathbf{d}_i',$$

where

$$\begin{aligned}\mathbf{d}_i &= \mathbf{y}_i - \mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^-\mathbf{x}_i, \\ r_{ii} &= \mathbf{x}_i'(\mathbf{X}\mathbf{X}')^-\mathbf{x}_i.\end{aligned}$$

Thus we may consider

$$T_i = \mathbf{d}_i'\mathbf{\Gamma}(\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}'\mathbf{d}_i/(1 - r_{ii}),$$

since

$$\lambda_i = (1 - T_i)^{-1}.$$

If we want to have a numerical value of T_i we can replace $\mathbf{\Gamma}$ by $\widehat{\mathbf{\Gamma}}_H$, for example. This will not change the distribution of T_i. Observe that H is rejected for large values of T_i. An overall test statistic will be based on

$$T = \max_{1 \le i \le n} T_i.$$

A conservative test based on a Bonferroni upper bound is given by

$$P(T \ge t) \le \sum_{i=1}^{n} P(T_i \ge t)$$

and

$$\frac{n - r(\mathbf{X}) - r(\mathbf{\Sigma})}{r(\mathbf{\Sigma})} \frac{T_i}{1 - T_i} \sim F_{r(\mathbf{\Sigma}), n - r(\mathbf{X}) - r(\mathbf{\Sigma})}.$$

Thus,

$$P(T \ge t) \le P(F_{r(\mathbf{\Sigma}), n - r(\mathbf{X}) - r(\mathbf{\Sigma})} \ge \frac{n - r(\mathbf{X}) - r(\mathbf{\Sigma})}{r(\mathbf{\Sigma})} \frac{t}{1 - t}).$$

To sharpen the above result we may use a generalized Bonferroni inequality (Hunter 1976, Worsley 1982, Meng & von Rosen 2003):

$$P(T \ge t) \le nP(T_1 \ge t) - \sum_{i=1}^{n-1} P(T_i \ge t, T_{i+1} \ge t).$$

Hence, the joint distribution of T_i and T_{i+1} is needed. Let

$$\mathbf{l}_i = \mathbf{e}_i(\mathbf{I} - \mathbf{R})(1 - r_{ii})^{1/2},$$
$$\mathbf{L}' = (\mathbf{l}_i, \mathbf{l}_j), \qquad \mathbf{F} = \mathbf{L}\mathbf{L}'$$

and then

$$\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma} = \mathbf{\Gamma}'\mathbf{U}\mathbf{F}^{-1}\mathbf{U}'\mathbf{\Gamma} + \mathbf{\Gamma}'\mathbf{W}\mathbf{\Gamma},$$

where

$$(\mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma})^{-1/2}\mathbf{\Gamma}'\mathbf{U} = (\mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma})^{-1/2}\mathbf{\Gamma}'\mathbf{Y}\mathbf{L}' \sim N_{r(\mathbf{\Sigma}),2}(\mathbf{0}, \mathbf{I}, \mathbf{F}),$$
$$(\mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma})^{-1/2}\mathbf{\Gamma}'\mathbf{W}\mathbf{\Gamma}(\mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma})^{-1/2} \sim W_{r(\mathbf{\Sigma})}(\mathbf{I}, n - r(\mathbf{X}) - 2).$$

Now

$$T_i = \mathbf{u}_i'\mathbf{\Gamma}(\mathbf{\Gamma}'\mathbf{U}\mathbf{F}^{-1}\mathbf{U}'\mathbf{\Gamma} + \mathbf{\Gamma}'\mathbf{W}\mathbf{\Gamma})^{-1}\mathbf{u}_i, \tag{5.3}$$

where

$$\mathbf{U} = (\mathbf{u}_i, \mathbf{u}_j) = (\mathbf{Y}\mathbf{l}_i, \mathbf{Y}\mathbf{l}_j)$$

and

$$T_{ij} = \mathbf{u}_i'\mathbf{\Gamma}(\mathbf{\Gamma}'\mathbf{U}\mathbf{F}^{-1}\mathbf{U}'\mathbf{\Gamma} + \mathbf{\Gamma}'\mathbf{W}\mathbf{\Gamma})^{-1}\mathbf{u}_j,$$
$$\rho_{ij} = \mathbf{l}_i'\mathbf{l}_j.$$

In (5.3) T_i is beta distributed (see Srivastava & Khatri, 1979). By applying Theorem 2.1 in Srivastava & von Rosen (1998) we get the joint density of T_i and T_j.

Theorem 5.1. *Let the statistics T_i, $i = 1, 2, \ldots, n$, be defined by (5.3). Then the joint density of T_i and T_j, $i \neq j$, is given by*

$$(1-\rho_{ij}^2)^{-\frac{1}{2}(n-r(\mathbf{X})-3)} \frac{\Gamma_{r(\mathbf{\Sigma})}(\frac{1}{2}(n-r(\mathbf{X})))}{\Gamma_{r(\mathbf{\Sigma})}(\frac{1}{2}(n-r(\mathbf{X})-2))\Gamma_2(\frac{1}{2}r(\mathbf{\Sigma}))}$$
$$\times \int_a^b (T_iT_j - T_{ij})^{\frac{r(\mathbf{\Sigma})-3}{2}} \{(1-T_i)(1-t_j) - (\rho_{ij}-T_{ij})^2\}^{\frac{1}{2}(n-k-r(\mathbf{\Sigma})-3)} dT_{ij},$$

where

$$a = \max(-\sqrt{T_iT_j}, \rho_{ij} - \sqrt{(1-T_i)(1-t_j)}),$$
$$b = \min(\sqrt{T_iT_j}, \rho_{ij} + \sqrt{(1-T_i)(1-t_j)}).$$

References

Alalouf, I. S. (1978). An explicit treatment of the general linear model with singular covariance matrix. *Sankyā Ser. B* **40**, 65–73.

Feuerverger, A. and Fraser, D. A. S. (1980). Categorical information and the singular linear model. *Canad. J. Statist.* **8**, 41–45.

Heck, D. L. (1960). Charts of some upper percentage points of the distribution of the largest characteristic root. *Ann. Math. Statist.* **31**, 625–642.

Khatri, C. G. (1968). Some results for the singular normal multivariate regression models. *Sankyā Ser. A* **30**, 267–280.

Meng, Q. and von Rosen, D. (2003). Bonferroni-type bounds of degree three. *Comm. Statist. Simulation Comput.* **32**, 857–872.

Mitra, S. K. and Rao, C. R. (1968). Some results in estimation and tests for linear hypothesis under the Gauss-Markoff model. *Sankyā Ser. A* **30**, 281–290.

Nordström, K. (1985). On a decomposition of the singular Gauss-Markov model. *Lecture Notes in Statistics* **35**, 231–245. Springer, New York.

Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *Ann. Math. Statist.* **26**, 117–121.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

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

Russell, J. S., Moore, A. W. and Coaldrake, J. E. (1967). Relationships between subtropical semiarid forest of *Acacia Narpophylla*, micro-relief and chemical properties of gilgai soil. *Austral. J. Botany* **15**, 481–498.

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

Srivastava, M. S. (2002). *Methods of Multivariate Statistics*. Wiley, New York.

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

Srivastava, M. S. and von Rosen, D. (1998). Outliers in multivariate regression models. *J. Multivariate Anal.* **65**, 195–208.

UNIVERSITY OF TORONTO, DEPARTMENT OF STATISTICS, 100 ST GEORGE STREET, TORONTO, ONTARIO, CANADA M5S 1A1
E-mail address: srivasta@utstat.toronto.edu

SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES, DEPARTMENT OF BIOMETRY, BOX 7013, SE-750 07 UPPSALA, SWEDEN
E-mail address: dietrich.von.rosen@bi.slu.se