Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 97 (2006) 2162–2176
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Two-sample inference for normal mean vectors based on monotone missing data

Jianqi Yu[a], K. Krishnamoorthy[a,*], Maruthy K. Pannala[b]

[a]*Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA*
[b]*Actuarial Department, United Guarantee Corporation, Greensboro, NC 27420, USA*

Received 5 January 2005
Available online 28 August 2006

Abstract

Inferential procedures for the difference between two multivariate normal mean vectors based on incomplete data matrices with different monotone patterns are developed. Assuming that the population covariance matrices are equal, a pivotal quantity, similar to the Hotelling T^2 statistic, is proposed, and its approximate distribution is derived. Hypothesis testing and confidence estimation of the difference between the mean vectors based on the approximate distribution are outlined. The validity of the approximation is investigated using Monte Carlo simulation. Monte Carlo studies indicate that the approximate method is very satisfactory even for small samples. A multiple comparison procedure is outlined and the proposed methods are illustrated using an example.

AMS 1991 subject classification: 62H12; 62H15

Keywords: Coverage probability; Incomplete data; Maximum likelihood estimators; Moment approximation; Powers; Sizes

1. Introduction

The problem of incomplete data arises commonly in many practical situations, especially in public survey. Missing data arises, for example, during data gathering and recording, when the experiment is involved a group of individuals over a period of time like in clinical trials or in a planned experiment where the variables that are expensive to measure are collected only from a subset of a sample. The causes for missing data are not our concern but to ignore the process that causes missing data it is assumed that the data are missing at random (MAR). Recently, Lu

* Corresponding author. Fax: +1 337 482 4356.
E-mail address: krishna@louisiana.edu (K. Krishnamoorthy).

0047-259X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2006.07.002

and Copas [11] pointed out that inference from the likelihood method ignoring the missing data mechanism is valid if and only if the missing data mechanism is MAR. For formal definition and exposition of MAR or missing completely at random we refer to Little and Rubin [9] or Little [8].

There are a few missing patterns considered in the literature, but the incomplete data with monotone pattern (see display 1) not only occurs frequently in practice but also it is convenient for making inference. In particular, if multivariate normality is assumed then the monotone pattern allows the exact calculation of the maximum likelihood estimators (MLEs), the likelihood ratio statistics and relevant distributions. Several authors have considered the monotone missing pattern under normality assumption, and provided asymptotic as well as approximate test procedures about the normal mean vector. Anderson [1], one of the earliest papers in this area, gives a simple approach to derive the MLEs and present them for a special case of monotone pattern and some other patterns. Kanda and Fujkoshi [4] studied some basic properties of the MLEs based on monotone data. Many authors developed asymptotic inferential procedures based on the likelihood ratio approach for multivariate normal distribution. We note, among many other papers, Bhargava [2], Morrison and Bhoj [12] and Naik [14]. Many of these papers considered primarily hypothesis testing problem, and only recently Krishnamoorthy and Pannala [6] provided an accurate simple approach to construct confidence region for a normal mean vector.

In this article, we consider the problems of hypothesis testing and confidence estimation of the difference between two normal mean vectors based on sample data matrices that are of monotone pattern. Our approach is essentially based on the ones given in Krishnamoorthy and Pannala [5,6] and Hao and Krishnamoorthy [3]. Specifically, we develop a pivotal quantity based on the MLEs (similar to the Hotelling T^2 statistic for the one sample case), and derive its approximate distribution to make inferential procedures.

To formulate the problem, let $\mathbf{x}$ follow a p-variate normal distribution with mean vector μ and covariance matrix Σ. We write this as $\mathbf{x} \sim N_p(\mu, \Sigma)$. Let $\mathbf{y} \sim N_p(\beta, \Sigma)$ independently of $\mathbf{x}$. Suppose that we have a sample of N_1 observations available on $\mathbf{x}$, and a sample of M_1 observations available on $\mathbf{x}$. Assume that the samples have the following monotone pattern:

$$\begin{array}{cc} x_{11}, \ldots, x_{1N_k}, \ldots, x_{1N_2}, \ldots, x_{1N_1}, & y_{11}, \ldots, y_{1M_k}, \ldots, y_{1M_2}, \ldots, y_{1M_1}, \\ x_{21}, \ldots, x_{2N_k}, \ldots, x_{2N_2}, & y_{21}, \ldots, y_{2M_k}, \ldots, y_{2M_2}, \\ \vdots & \vdots \\ x_{k1}, \ldots, x_{kN_k}, & y_{k1}, \ldots, y_{kM_k}, \end{array} \tag{1}$$

where x_{ij} is a $p_i \times 1$ vector, $j = 1, \ldots, N_i$, while y_{ij} is a $q_i \times 1$ vector, $j = 1, \ldots, M_i$, $i = 1, \ldots, k$. In other words, in the x-sample, there are N_1 observations available on the first p_1 components, N_2 observations available on the first $p_1 + p_2$ components, and so on. Notice that $N_1 \geqslant N_2 \geqslant \cdots \geqslant N_k$, $M_1 \geqslant M_2 \geqslant \cdots \geqslant M_k$, and $p_1 + \cdots + p_k = q_1 + \cdots + q_k = p$.

Krishnamoorthy and Pannala [6] considered the one-sample case, and provided approximate methods for constructing confidence region and hypothesis testing for the mean vector. Using their idea, we develop inferential procedures to the present two-sample problem. However, unlike the complete data case, extending the solution of the one-sample problem to the two-sample case is not easy. Indeed, the problem is much more complex than the one-sample problem, and so methods for the two-sample case are really warranted for easy reference.

In the following section, we present first some preliminaries in the notations of Krishnamoorthy and Pannala [5] for the data matrices in (1) with $k = 3$ and $p_i = q_i$, $i = 1, 2, 3$. We present the MLEs of the relevant parameters in terms of these notations. Using these MLEs, we propose a

pivotal statistic similar to the Hotelling T^2 statistic, and derive an approximation to its distribution. We outline procedures for hypothesis testing and constructing confidence region for $\mu-\beta$ based on the approximate distribution. Required results for approximating the null distribution of the pivotal quantity are also given for the general case. We describe a method of constructing simultaneous confidence intervals for the components of $\mu-\beta$. We also point out the results for equal monotone pattern (that is, $p_i = q_i, i = 1, \ldots, k$) can be extended to the case of unequal monotone patterns (i.e., in (1), $p_i \neq q_i$ for some i). The accuracies of the approximation are appraised by Monte Carlo simulation in Section 3. Simulation studies show that the approximation is very satisfactory even for small samples. Our limited power studies in Section 4 indicate that the proposed test has some natural power properties. The methods are illustrated using an example in Section 5, and some concluding remarks are given in Section 6.

2. Inference on $\mu - \beta$

To develop inferential procedures about $\mu - \beta$, we first need to obtain the MLEs of the parameters μ, β and the common covariance matrix Σ. In the following section, we present some preliminaries in the notations of Krishnamoorthy and Pannala [5], and present the MLEs of the relevant parameters for the samples of two-block monotone pattern (that is $k = 2$ in (1)). The MLEs can be easily expressed for the general case.

2.1. The maximum likelihood estimators

Consider the data matrices in (1) with $k = 3$ and assume that $p_i = q_i = r_i, i = 1, \ldots, k$, and partition the data matrices as follows:

$$\mathbf{x}_1 = \left(x_{11}, \ldots, x_{1N_3}, \ldots, x_{1N_2}, \ldots, x_{1N_1}\right)_{r_1 \times N_1},$$

$$\mathbf{x}_2 = \begin{pmatrix} x_{11}, \ldots, x_{1N_2} \\ x_{21}, \ldots, x_{2N_2} \end{pmatrix}_{(r_1+r_2)\times N_2},$$

$$\mathbf{x_3} = \begin{pmatrix} x_{11}, \ldots, x_{1N_3} \\ x_{21}, \ldots, x_{2N_3} \\ x_{21}, \ldots, x_{2N_3} \end{pmatrix}_{(r_1+r_2+r_3)\times N_3}. \tag{2}$$

That is, $\mathbf{x}_l$ is the submatrix of $\mathbf{x}$ in (1) formed by the first N_l columns and the first $p_1 + \cdots + p_l$ rows, $l = 1, \ldots, 3$. Partition the matrix $\mathbf{y}$ similarly. That is,

$$\mathbf{y}_1 = \left(y_{11}, \ldots, y_{1M_3}, \ldots, y_{1M_2}, \ldots, y_{1M_1}\right)_{r_1 \times M_1},$$

$$\mathbf{y}_2 = \begin{pmatrix} y_{11}, \ldots, y_{1M_2} \\ y_{21}, \ldots, y_{2M_2} \end{pmatrix}_{(r_1+r_2)\times M_2},$$

$$\mathbf{y_3} = \begin{pmatrix} y_{11}, \ldots, y_{1M_3} \\ y_{21}, \ldots, y_{2M_3} \\ y_{21}, \ldots, y_{2M_3} \end{pmatrix}_{(r_1+r_2+r_3)\times M_3}. \tag{3}$$

Let $\bar{\mathbf{x}}_l$ and $\mathbf{S}_l$ denote, respectively, the sample mean vector and the sums of squares and products matrix based on $\mathbf{x}_l$, $l = 1, 2, 3$. Similarly, let $\bar{\mathbf{y}}_l$ and $\mathbf{V}_l$ denote, respectively, the sample mean

vector and the sums of squares and products matrix based on $\mathbf{y}_l$, $l = 1, 2, 3$. We partition these means and matrices accordingly as following:

$$\bar{\mathbf{x}}_1 = \bar{x}_1^{(1)}, \quad \bar{\mathbf{x}}_2 = \begin{pmatrix} \bar{x}_2^{(1)} : r_1 \times 1 \\ \bar{x}_2^{(2)} : r_2 \times 1 \end{pmatrix}, \quad \bar{\mathbf{x}}_3 = \begin{pmatrix} \bar{x}_3^{(1)} : r_1 \times 1 \\ \bar{x}_3^{(2)} : r_2 \times 1 \\ \bar{x}_3^{(3)} : r_3 \times 1 \end{pmatrix},$$

$$\mathbf{S}_1 = S_1^{(1,1)}, \quad \mathbf{S}_2 = \begin{pmatrix} S_2^{(1,1)} : r_1 \times r_1 & S_2^{(1,2)} : r_1 \times r_2 \\ S_2^{(2,1)} : r_2 \times r_1 & S_2^{(2,2)} : r_2 \times r_2 \end{pmatrix}$$

and

$$\mathbf{S}_3 = \begin{pmatrix} S_3^{(1,1)} : r_1 \times r_1 & S_3^{(1,2)} : r_1 \times r_2 & S_3^{(1,3)} : r_1 \times r_3 \\ S_3^{(2,1)} : r_2 \times r_1 & S_3^{(2,2)} : r_2 \times r_2 & S_3^{(2,3)} : r_2 \times r_3 \\ S_3^{(3,1)} : r_3 \times r_1 & S_3^{(3,2)} : r_3 \times r_2 & S_3^{(3,3)} : r_3 \times r_3 \end{pmatrix}. \tag{4}$$

Notice that $\mathbf{x}_l^{(i)} : r_i \times 1$ is the mean of the ith block of the data matrix $\mathbf{x}_l$, $i = 1, \ldots, l$ and $l = 1, 2, 3$. We also read $\mathbf{S}_l^{(i,j)} : r_i \times r_j$ as the (i, j)th submatrix of $\mathbf{S}_l$ based on the data matrix $\mathbf{x}_l$, $l = 1, 2, 3$.

The statistics $\bar{\mathbf{y}}$ and $\mathbf{V}$ based on the data matrix $\mathbf{y}$ in (3) are also partitioned like $\bar{\mathbf{x}}_l$ and $\mathbf{S}_l$. That is, $\bar{y}_l^{(i)} : r_i \times 1$ is the mean of the ith block of data matrix $\mathbf{y}$, $i = 1, \ldots, l$ and $l = 1, 2, 3$, and $V_l^{(i,j)} : r_i \times r_j$ is the (i, j)th submatrix of $\mathbf{V}_l$, $i, j = 1, \ldots, l$ and $l = 1, 2, 3$.

Finally, we partition the parameters as follows:

$$\mu = \begin{pmatrix} \mu_1 : r_1 \times 1 \\ \mu_2 : r_2 \times 1 \\ \mu_3 : r_3 \times 1 \end{pmatrix}_{p \times 1}, \quad \beta = \begin{pmatrix} \beta_1 : r_1 \times 1 \\ \beta_2 : r_2 \times 1 \\ \beta_3 : r_3 \times 1 \end{pmatrix}_{p \times 1}$$

and

$$\Sigma = \begin{pmatrix} \Sigma_{11} : r_1 \times r_1 & \Sigma_{12} : r_1 \times r_2 & \Sigma_{13} : r_1 \times r_3 \\ \Sigma_{21} : r_2 \times r_1 & \Sigma_{22} : r_2 \times r_2 & \Sigma_{23} : r_2 \times r_3 \\ \Sigma_{31} : r_3 \times r_1 & \Sigma_{32} : r_3 \times r_2 & \Sigma_{33} : r_3 \times r_3 \end{pmatrix}_{p \times p}.$$

It should be noted that the way the data matrices and summary statistics are partitioned is different from the one given in Krishnamoorthy and Pannala [6] for the one-sample case. We found the MLEs can be expressed in simple forms in terms of the above partitioned sample mean vectors, $\mathbf{S}$ and $\mathbf{V}$. We now give the MLEs of the partitioned mean vectors and sub-matrices of Σ. Let

$$B_{21} = (S_2^{(2,1)} + V_2^{(2,1)})(S_2^{(1,1)} + V_2^{(1,1)})^{-1},$$

$$(B_{31}, B_{32}) = (S_3^{(3,1)} + V_3^{(3,1)}, S_3^{(3,2)} + V_3^{(3,2)}) \begin{pmatrix} S_3^{(1,1)} + V_3^{(1,1)}, S_3^{(1,2)} + V_3^{(1,2)} \\ S_3^{(2,1)} + V_3^{(2,1)}, S_3^{(2,2)} + V_3^{(2,2)} \end{pmatrix}^{-1},$$

$$\widehat{\Sigma}_{2.1} = \frac{1}{N_2 + M_2}[(S_2^{(2,2)} + V_2^{(2,2)}) - (S_2^{(2,1)} + V_2^{(2,1)})(S_2^{(1,1)} + V_2^{(1,1)})^{-1}(S_2^{(1,2)} + V_2^{(1,2)})]$$

and

$$\widehat{\Sigma}_{3.21} = \frac{1}{N_3 + M_3}\left[(S_3^{(3,3)} + V_3^{(3,3)}) - (S_3^{(3,1)} + V_3^{(3,1)}, S_3^{(3,2)} + V_3^{(3,2)}) \right.$$
$$\left. \times \begin{pmatrix} S_3^{(1,1)} + V_3^{(1,1)} & S_3^{(1,2)} + V_3^{(1,2)} \\ S_3^{(2,1)} + V_3^{(2,1)} & S_3^{(2,2)} + V_3^{(2,2)} \end{pmatrix}^{-1} \begin{pmatrix} S_3^{(1,3)} + V_3^{(1,3)} \\ S_3^{(2,3)} + V_3^{(2,3)} \end{pmatrix}\right].$$

The MLEs are given by

$$\widehat{\mu}_1 = \overline{x}_1, \quad \widehat{\mu}_2 = \overline{x}_2^{(2)} - B_{21}(\overline{x}_2^{(1)} - \widehat{\mu}_1), \quad \widehat{\mu}_3 = \overline{x}_3^{(3)} - B_{31}(\overline{x}_3^{(1)} - \widehat{\mu}_1) - B_{32}(\overline{x}_3^{(2)} - \widehat{\mu}_2),$$
$$\widehat{\beta}_1 = \overline{y}_1, \quad \widehat{\beta}_2 = \overline{y}_2^{(2)} - B_{21}(\overline{y}_2^{(1)} - \widehat{\beta}_1), \quad \widehat{\beta}_3 = \overline{y}_3^{(3)} - B_{31}(\overline{y}_3^{(1)} - \widehat{\beta}_1) - B_{32}((\overline{y}_3^{(2)} - \widehat{\beta}_2),$$
$$\widehat{\Sigma}_{11} = (S_1^{(1,1)} + V_1^{(1,1)})/(N_1 + M_1), \quad \widehat{\Sigma}_{21} = B_{21}\widehat{\Sigma}_{11}, \quad \widehat{\Sigma}_{22} = \widehat{\Sigma}_{2.1} + B_{21}\widehat{\Sigma}_{12},$$
$$\widehat{\Sigma}_{31} = B_{31}\widehat{\Sigma}_{11} + B_{32}\widehat{\Sigma}_{21}, \quad \widehat{\Sigma}_{32} = B_{31}\widehat{\Sigma}_{12} + B_{32}\widehat{\Sigma}_{22} \quad \text{and} \quad \widehat{\Sigma}_{33} = \widehat{\Sigma}_{3.21} + B_{31}\widehat{\Sigma}_{13} + B_{32}\widehat{\Sigma}_{23}.$$

Let $\widehat{\mu} = (\widehat{\mu}_1', \widehat{\mu}_2', \widehat{\mu}_3')'$ and $\widehat{\beta} = (\widehat{\beta}_1', \widehat{\beta}_2', \widehat{\beta}_3')'$, and let Ω denote the covariance matrix of $(\widehat{\mu} - \widehat{\beta})$. Then, the MLE of Ω is given by

$$\widehat{\Omega} = \begin{pmatrix} \widehat{\Sigma}_{11}/W_1 & \widehat{\Sigma}_{12}/W_1 & \widehat{\Sigma}_{13}/W_1 \\ \widehat{\Sigma}_{21}/W_1 & \widehat{\Sigma}_{22}/W_2 - \frac{W_1 - W_2}{W_1 W_2} B_{21}\widehat{\Sigma}_{12} & \widehat{\Sigma}_{23}/W_2 - \frac{W_1 - W_2}{W_1 W_2} B_{21}\widehat{\Sigma}_{13} \\ \widehat{\Sigma}_{31}/W_1 & \widehat{\Sigma}_{32}/W_2 - \frac{W_1 - W_2}{W_1 W_2} B_{21}'\widehat{\Sigma}_{31} & \widehat{\Sigma}_{33}/W_3 - \frac{W_1 - W_3}{W_1 W_3}(B_{31}\widehat{\Sigma}_{13} \\ & & + B_{32}\widehat{\Sigma}_{23}) + \frac{W_1 - W_2}{W_1 W_2} B_{32}\widehat{\Sigma}_{22.1} B_{23} \end{pmatrix}, \tag{5}$$

where

$$W_1^{-1} = N_1^{-1} + M_1^{-1}, \quad W_2^{-1} = N_2^{-1} + M_2^{-1} \quad \text{and} \quad W_3^{-1} = N_3^{-1} + M_3^{-1}.$$

2.2. Hypothesis test and confidence region for $\mu - \beta$

The pivotal quantity that we consider to construct confidence region for $\mu - \beta$ or to test about $\mu - \beta$ is given by

$$\begin{aligned} Q &= [(\widehat{\mu} - \widehat{\beta}) - (\mu - \beta)]'[\widehat{\Omega}]^{-1}[(\widehat{\mu} - \widehat{\beta}) - (\mu - \beta)] \\ &= Q_1 + Q_2 + Q_3, \end{aligned} \tag{6}$$

where

$$Q_1 = W_1(\widehat{\mu}_1 - \widehat{\beta}_1 - (\mu_1 - \beta_1))'[\widehat{\Sigma}_{11}]^{-1}(\widehat{\mu}_1 - \widehat{\beta}_1 - (\mu_1 - \beta_1)),$$

$$\begin{aligned} Q_2 = W_2[(\widehat{\mu}_{2.1} - \widehat{\beta}_{2.1}) - (\mu_2 - \beta_2 - B_{21}(\mu_1 - \beta_1))]'[\widehat{\Sigma}_{2.1}]^{-1} \\ \times[(\widehat{\mu}_{2.1} - \widehat{\beta}_{2.1}) - (\mu_2 - \beta_2 - B_{21}(\mu_1 - \beta_1))], \end{aligned}$$

$$\begin{aligned} Q_3 = W_3[(\widehat{\mu}_{3.21} - \widehat{\beta}_{3.21}) - (\mu_3 - \beta_3 - B_{31}(\mu_1 - \beta_1) - B_{32}(\mu_2 - \beta_2))]'[\widehat{\Sigma}_{3.21}]^{-1} \\ \times[(\widehat{\mu}_{3.21} - \widehat{\beta}_{3.21}) - (\mu_3 - \beta_3 - B_{31}(\mu_1 - \beta_1) - B_{32}(\mu_2 - \beta_2))], \end{aligned}$$

$$\widehat{\mu}_{2.1} = \widehat{\mu}_2 - B_{21}\widehat{\mu}_1,$$
$$\widehat{\beta}_{2.1} = \widehat{\beta}_2 - B_{21}\widehat{\beta}_1,$$
$$\widehat{\mu}_{3.21} = \widehat{\mu}_3 - B_{31}\widehat{\mu}_1 - B_{32}\widehat{\mu}_2$$

and

$$\widehat{\beta}_{3.21} = \widehat{\beta}_3 - B_{31}\widehat{\beta}_1 - B_{32}\widehat{\beta}_2.$$

The expression for Q clearly suggests that it is difficult to derive the exact distribution of Q. Because Q is resembling the Hotelling-T^2 statistic, and its distribution is free of any parameters (see Appendix A), it is reasonable to approximate its distribution by the distribution of $dF_{p,v}$, where d is a positive constant, and $F_{a,b}$ denotes the F random variable with numerator degrees of freedom a and the denominator degrees of freedom b. The unknown constants d and v can be determined so that the first two moments of Q are equal to those of $dF_{p,v}$. In the appendix, following the lines of Krishnamoorthy and Pannala [6], we evaluated an exact expression G_1 for $E(Q)$ and an approximation G_2 for $E(Q^2)$. Using these G_1 and G_2 (see Appendix A), we see that

$$Q \sim dF_{p,v} \text{ approximately}, \quad \text{where } v = \frac{4pG_2 - 2(p+2)G_1^2}{pG_2 - (p+2)G_1^2} \quad \text{and} \quad d = G_1\frac{v-2}{v}. \tag{7}$$

We again note that d and v were determined so that $E(Q) = E(dF_{p,v})$ and $E(Q^2) = E[(dF_{p,v})^2]$. Letting $\delta = \mu - \beta$, an approximate $1-\alpha$ confidence set for $\mu - \beta$ is the set of values of δ that satisfy

$$[(\widehat{\mu} - \widehat{\beta}) - \delta]'[\widehat{\Omega}]^{-1}[(\widehat{\mu} - \widehat{\beta}) - \delta] \leqslant dF_{p,v}(1-\alpha), \tag{8}$$

where $F_{p,v}(1-\alpha)$ is the $(1-\alpha)$th quantile of the $F_{p,v}$ distribution.

An approximate α-level test rejects the null hypothesis $\mathrm{H}_0 : \mu = \beta$ when

$$(\widehat{\mu} - \widehat{\beta})'[\widehat{\Omega}]^{-1}(\widehat{\mu} - \widehat{\beta}) > dF_{p,v}(1-\alpha).$$

The results for a general monotone patterns are given in Appendix B. In particular, we write

$$\begin{aligned} Q &= [(\widehat{\mu} - \widehat{\beta}) - (\mu - \beta)]'[\widehat{\Omega}]^{-1}[(\widehat{\mu} - \widehat{\beta}) - (\mu - \beta)] \\ &= Q_1 + \cdots + Q_k \end{aligned}$$

and give expression G_1 for $E(Q)$ and an approximation G_2 for $E(Q^2)$.

2.3. Simultaneous confidence intervals

Approximate simultaneous confidence intervals for $\delta_i = \mu_i - \beta_i$, $i = 1, \ldots, p$, can be constructed using Scheffé's S-method. Towards this, we note that the inequality in (8) holds if and only if

$$\frac{[\mathbf{a}'(\hat{\delta} - \delta)]^{2'}}{\mathbf{a}}\widehat{\Omega}\mathbf{a} \leqslant dF_{p,v} \quad \text{for all } \mathbf{a} \in R^p.$$

Therefore, we have

$$P\left(\mathbf{a}'\hat{\delta} - \mathbf{c}_\alpha\sqrt{\mathbf{a}'\widehat{\Omega}\mathbf{a}} \leqslant \mathbf{a}'\delta \leqslant \mathbf{a}'\hat{\delta} + c_\alpha\sqrt{\mathbf{a}'\widehat{\Omega}\mathbf{a}} \text{ for all } \mathbf{a}\right) \simeq 1 - \alpha,$$

where $\hat{\delta} = \widehat{\mu} - \widehat{\beta}$ and $c_\alpha = \sqrt{dF_{p,\nu}(1-\alpha)}$. It follows from the above equation that

$$\hat{\delta}_i - c_\alpha\sqrt{\hat{\omega}_{ii}} \leqslant \delta_i \leqslant \hat{\delta}_i + c_\alpha\sqrt{\hat{\omega}_{ii}} \quad \text{for } i = 1, \ldots, p, \tag{9}$$

where $\hat{\omega}_{ii}$ is the (i, i)th element of $\widehat{\Omega}$ in (5), with probability at least $1 - \alpha$.

2.4. The case of unequal monotone patterns

We shall now consider the data matrices in (1) with different monotone pattern. For convenience and simplicity, let us assume that $k = 2$, and without loss of generality, $p_1 > q_1$. That is, we have the following data matrices.

$$\begin{aligned} &(x_{11}, \ldots, x_{1L_2}, \ldots, x_{1L_1})_{p_1\times L_1}, \quad (y_{11}, \ldots, y_{1T_2}, \ldots, y_{1T_1})_{q_1\times T_1}, \\ &(x_{21}, \ldots, x_{2L_2})_{p_2\times L_2}, \quad (y_{21}, \ldots, y_{2T_2})_{q_2\times T_2}. \end{aligned} \tag{10}$$

Now, let $\mathbf{x}_1$ denote the data matrix formed by the first q_1 rows of $(x_{11}, \ldots, x_{1L_2}, \ldots, x_{1L_1})_{p_1\times L_1}$, $\mathbf{x}_2$ denote the data matrix $(x_{11}, \ldots, x_{1L_2}, \ldots, x_{1L_1})_{p_1\times L_1}$ and

$$\mathbf{x}_3 = \begin{pmatrix} x_{11}, \ldots, x_{1L_2} \\ x_{21}, \ldots, x_{2L_2} \end{pmatrix}_{(p_1+p_2)\times L_2}.$$

Notice that $q_2 = (p_1 - q_1) + p_2$ because $p_1 + p_2 = q_1 + q_2 = p$. Let $\mathbf{y}_1 = (y_{11}, \ldots, y_{1T_2}, \ldots, y_{1T_1})_{q_1\times T_1}$, $\mathbf{y}_2$ denote the data matrix formed by $(y_{11}, \ldots, y_{1T_2})_{q_1\times T_1}$ as the first block of rows and $(p_1 - q_1)$ rows of $(y_{21}, \ldots, y_{2T_2})_{q_2\times T_2}$ as second block of rows and

$$\mathbf{y}_3 = \begin{pmatrix} y_{11}, \ldots, y_{1T_2} \\ y_{21}, \ldots, y_{2T_2} \end{pmatrix}_{(q_1+q_2)\times T_2}.$$

Thus, we can partition the data matrices of unequal monotone patterns to make equal monotone pattern. Specifically, we have $\mathbf{x}_1 : q_1 \times L_1$, $\mathbf{x}_2 : p_1 \times L_1$, $\mathbf{x}_3 : (p_2 + p_1) \times L_2$, $\mathbf{y}_1 : q_1 \times T_1$, $\mathbf{y}_2 : p_1 \times T_2$, and $\mathbf{y}_3 : (q_1+q_2)\times T_2$. So, by setting $(r_1, r_2, r_3) = (q_1, p_1-q_1, p_2)$, $(N_1, N_2, N_3) = (L_1, L_1, L_2)$ and $(M_1, M_2, M_3) = (T_1, T_2, T_2)$, we can apply the method for equal monotone pattern case to the present unequal monotone patterns case. We also note that any type of unequal monotone patterns data can be rearranged to form equal monotone pattern.

3. Validity of the approximation

To appraise the accuracy of the F approximation to the distribution of the pivotal quantity Q in (6), we estimated the coverage probabilities of the 95% confidence region based on (8) for various sample size configurations using Monte Carlo simulation. Each simulation result is based on 100,000 runs. The multivariate normal random vectors were generated using IMSL subroutine RNMVN. In Table 1a, we present the estimated coverage probabilities for the equal monotone pattern with $(r_1, r_2, r_3) = (1, 1, 1)$, and various values of $(N_1, N_2, N_3, M_1, M_2, M_3)$. The estimated coverage probabilities are given in Table 1b for 2-block unequal monotone patterns with $(p_1, p_2) = (3, 1)$ and $(q_1, q_2) = (1, 3)$, and in Table 1c for $(p_1, p_2) = (2, 2)$ and

Table 1
Critical values $dF_{p,\nu}(0.95)$ and Monte Carlo estimates of the coverage probabilities of the 95% confidence region in (8)

	(M_1, M_2, M_3)		
(N_1, N_2, N_3)	(12, 6, 6)	(21, 15, 7)	(19, 16, 16)
	$(r_1, r_2, r_3) = (1, 1, 1)$		
(12,6,6)	17.03(.950)	14.19(.951)	11.10(.948)
(12,12,6)	15.66(.952)	13.89(.952)	10.80(.948)
(19,11,11)	12.57(.949)	11.48(.950)	10.29(.948)
(11,11,11)	12.75(.948)	11.57(.949)	10.40(.949)
(16,12,7)	14.53(.950)	13.09(.952)	10.62(.947)
(18,14,7)	14.31(.951)	13.00(.952)	10.53(.948)
(21,15,7)	14.19(.950)	12.95(.952)	10.48(.948)
(32,22,11)	11.65(.951)	11.10(.949)	9.90(.950)
(32,21,18)	10.48(.948)	10.08(.949)	9.57(.948)
(40,30,10)	11.75(.951)	11.25(.951)	9.83(.949)
(60,40,30)	9.29(.950)	9.16(.949)	8.96(.947)

(N_1, N_2)	(M_1, M_2)					
	(20, 15)	(20, 10)	(30, 20)	(30, 15)	(15, 10)	(12, 10)
	$(p_1, p_2) = (2, 1)$, $(q_1, q_2) = (1, 2)$					
(15,10)	10.37(.948)	11.16(.949)	9.82(.949)	10.27(.948)	11.23(.949)	11.81(.948)
(15,9)	10.45(.949)	11.35(.948)	9.87(.948)	10.36(.948)	11.42(.948)	12.08(.949)
(15,7)	10.67(.949)	11.85(.948)	9.99(.948)	10.58(.949)	11.91(.949)	12.84(.950)
(18,15)	9.92(.948)	10.40(.948)	9.53(.948)	9.84(.949)	10.46(.948)	10.78(.949)
(18,12)	10.10(.948)	10.71(.947)	9.65(.949)	10.02(.949)	10.77(.949)	11.19(.946)
(18,9)	10.34(.948)	11.19(.949)	9.79(.949)	10.26(.948)	11.25(.948)	11.87(.949)
(25,20)	9.52(.948)	9.82(.948)	9.26(.947)	9.46(.948)	9.86(.948)	10.04(.948)
(25,15)	9.73(.949)	10.14(.949)	9.40(.948)	9.67(.948)	10.18(.950)	10.45(.947)
(25,7)	10.37(.950)	11.46(.950)	9.78(.950)	10.32(.949)	11.50(.949)	12.35(.950)

(N_1, N_2)	(M_1, M_2)					
	(20, 15)	(20, 10)	(20, 7)	(30, 15)	(30, 10)	(30, 7)
	$(p_1, p_2) = (2, 2)$, $(q_1, q_2) = (1, 3)$					
(30,20)	11.76(.948)	12.12(.948)	12.43(.948)	11.65(.949)	12.01(.950)	12.32(.949)
(30,15)	12.03(.948)	12.58(.948)	13.10(.950)	11.92(.949)	12.47(.950)	12.99(.950)
(30,10)	12.49(.949)	13.49(.950)	14.69(.951)	12.38(.950)	13.39(.950)	14.59(.951)
(18,15)	12.53(.946)	13.18(.947)	13.81(.947)	12.33(.947)	12.99(.948)	13.63(.949)
(18,12)	12.77(.947)	13.63(.947)	14.54(.947)	12.57(.947)	13.44(.950)	14.36(.948)
(18, 9)	13.10(.947)	14.36(.949)	15.93(.952)	12.91(.948)	14.18(.949)	15.76(.950)
(40,20)	11.53(.948)	11.85(.949)	12.13(.948)	11.45(.948)	11.78(.949)	12.06(.950)
(40,15)	11.80(.949)	12.32(.949)	12.82(.949)	11.72(.947)	12.24(.949)	12.75(.949)
(40,10)	12.27(.950)	13.52(.949)	14.43(.951)	12.19(.949)	13.18(.951)	14.36(.952)

$(q_1, q_2) = (1, 3)$. We observe from all these three table values that the estimated coverage probabilities are very close to the nominal level 0.95 for all the cases considered. Even for small samples, our approximate procedure is very accurate (for example, see $(N_1, N_2, N_3) = (M_1, M_2, M_3) = (12, 6, 6)$ in Table 1a, $(N_1, N_2, M_1, M_2) = (15, 10, 12, 8)$ in Table 1b). We

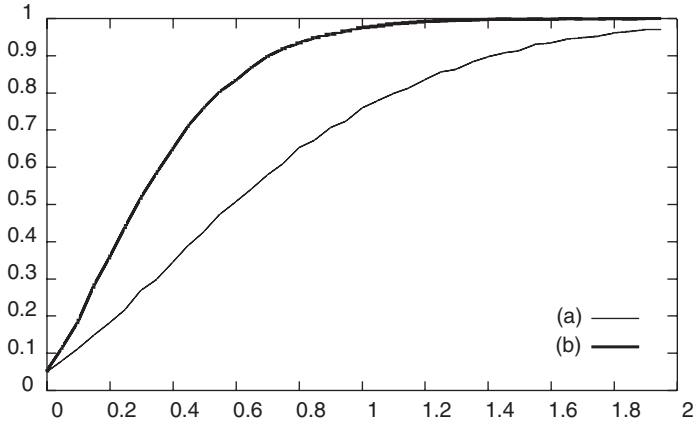

Fig. 1. Monte Carlo estimates of the powers of the test: (a) $(N_1, N_2, N_3) = (12, 8, 6)$; $(M_1, M_2, M_3) = (15, 11, 7)$; (b) $(N_1, N_2, N_3) = (20, 18, 12)$, $(M_1, M_2, M_3) = (19, 16, 10)$.

also estimated the coverage probabilities of the confidence region in (8) at confidence level 0.99 for sample sizes given in Tables 1. Because the results are similar to the case of 95% confidence level, they are not reported here.

4. Power studies

To understand the nature of the power function of the test in Section 2.2, we estimated the powers via Monte Carlo simulation consisting of 100,000 runs. The powers are estimated as a function of $\delta = (\mu - \beta)'\Sigma^{-1}(\mu - \beta)$. For fixed sample sizes, the powers are estimated using $\Sigma = I_p$ and $\mu - \beta = \sqrt{\delta}\mathbf{1}$, where $\mathbf{1}$ denotes the vector of ones. We observe from power plots in Fig. 1 that for fixed sample sizes, the power is an increasing function of δ. Also, for fixed δ, the power is an increasing function of sample sizes because the power curve for $(N_1, N_2, N_3) = (20, 18, 12)$, $(M_1, M_2, M_3) = (19, 16, 10)$ fall above the power curve for $(N_1, N_2, N_3) = (12, 8, 6)$, $(M_1, M_2, M_3) = (15, 11, 7)$. The power studies for other sample sizes and parameter configurations exhibited similar properties and so they are not reported here. Thus, our proposed test possesses some natural power properties.

5. An illustrative example

We shall now illustrate the methods using the "Fisher's Iris Data" which represent measurements of the sepal length and width, and pedal length and width in centimeters of fifty plants for each of three types of iris; Iris setosa, Iris versicolor and Iris virginica. The data sets are posted in many websites, and we downloaded them from http://javeeh.net/sasintro/intro151.html. For illustration purpose, we use the data on virginica ($\mathbf{x}$) and versicolor ($\mathbf{y}$). Also, we use only sepal length, width and pedal length as three components. We applied the modified likelihood ratio test (e.g., Muirhead [13, p. 309]) to check the equality of covariance matrices. The test produced a p-value of 0.412, and so the assumption of equality of covariance matrices is tenable.

We created monotone patterns by discarding the last 10 measurements on x_2 (sepal length of virginica), the last 20 measurements on x_3 (pedal length of virginica), the last 18 measurements on

y_2 (sepal length of versicolor), and the last 25 measurements on y_3 (pedal length of versicolor). That is, we have $(N_1, N_2, N_3) = (50, 40, 30)$ and $(M_1, M_2, M_3) = (50, 42, 25)$. Let $\mu' = (\mu_1, \mu_2, \mu_3) =$ (average sepal length, average sepal width, average pedal length) of virginica, and $\beta' = (\beta_1, \beta_2, \beta_3) =$ (average sepal length, average sepal width, average pedal length) of versicolor. We want to test

$$\mathrm{H}_0 : \mu' - \beta' = \mathbf{d}_0' \quad \text{vs.} \quad \mathrm{H}_0 : \mu' - \beta' \neq \mathbf{d}_0', \tag{11}$$

where $\mathbf{d}_0' = (0.4, 0.0, 1.1)$ and construct simultaneous confidence intervals for $\mu_1 - \beta_1$, $\mu_2 - \beta_2$ and $\mu_3 - \beta_3$.

We present the results for three different cases: (i) complete data sets containing 50 observations from each group; (ii) incomplete monotone pattern data, and (iii) partially complete data (that is, a vector observation is discarded if any of its components are missing; in the present case, $N_1 = N_2 = N_3 = 30$ and $M_1 = M_2 = M_3 = 25$).

5.1. Results based on complete data

$$\bar{\mathbf{x}} - \bar{\mathbf{y}} = (0.65200, 0.20400, 1.29200), \quad \widehat{\mathrm{Cov}(\bar{\mathbf{x}} - \bar{\mathbf{y}})} = \begin{pmatrix} 0.01342 & 0.00358 & 0.00972 \\ 0.00358 & 0.00405 & 0.00308 \\ 0.00972 & 0.00308 & 0.01051 \end{pmatrix} \tag{12}$$

and

$$[\widehat{\mathrm{Cov}(\bar{\mathbf{x}} - \bar{\mathbf{y}})}]^{-1} = \begin{pmatrix} 235.350 & -54.473 & -201.814 \\ -54.473 & 330.440 & -46.468 \\ -201.814 & -46.468 & 295.538 \end{pmatrix},$$

where $\widehat{\mathrm{Cov}(\bar{\mathbf{x}} - \bar{\mathbf{y}})} = (1/N_1 + 1/M_1) S_p$ and S_p is the pooled covariance matrix so that $E(S_p) = \Sigma$. The Hotelling T^2-statistic is computed as

$$T^2 = (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \mathbf{d}_0)' [\widehat{\mathrm{Cov}(\bar{\mathbf{x}} - \bar{\mathbf{y}})}]^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \mathbf{d}_0) = 10.8220.$$

Noticing that T^2 statistic is distributed as $\frac{(N_1+M_1-2)p}{N_1+N_2-p-1} F_{p, N_1+N_2-p-1}$, the p-value for testing (11) is given by

$$P\left(\frac{(N_1 + M_1 - 2)p}{N_1 + M_1 - p - 1} F_{p, N_1+M_1-p-1} > T^2 \right) = P\left(\frac{294}{96} F_{3,96} > 10.8220 \right) = 0.0177.$$

To get simultaneous confidence intervals, we computed $c = \frac{294}{96} F_{3,96}(0.95) = 8.2669$. Using Scheffé's method,

$$\bar{x}_i - \bar{y}_i \pm \sqrt{\frac{(N_1 + M_1 - 2)p}{N_1 + M_1 - p - 1} F_{p, N_1+M_1-p-1}(1 - \alpha) v_{ii}}, \quad i = 1, 2, 3,$$

where v_{ii} is the (i, i)th element of $\widehat{\mathrm{Cov}(\bar{\mathbf{x}} - \bar{\mathbf{y}})}$ given in (12), we computed 95% simultaneous confidence intervals for $\mu_1 - \beta_1$, $\mu_2 - \beta_2$ and $\mu_3 - \beta_3$ as

$$0.652 \pm 0.333, \quad 0.204 \pm 0.183 \quad \text{and} \quad 1.292 \pm 0.295,$$

respectively.

5.2. Results based on incomplete data

As pointed out earlier in the section, here we consider the monotone data with $(N_1, N_2, N_3) = (50, 40, 30)$ and $(M_1, M_2, M_3) = (50, 42, 25)$. The MLE of $\mu - \beta$ is given by

$$(\hat{\mu} - \hat{\beta})' = (0.65200, 0.18693, 1.35280)$$

and the estimate $\hat{\Omega}$ of the covariance matrix of $\hat{\mu} - \hat{\beta}$ in (5) is computed as

$$\hat{\Omega} = \begin{pmatrix} 0.01315 & 0.00351 & 0.00994 \\ 0.00351 & 0.00495 & 0.00291 \\ 0.00994 & 0.00291 & 0.01300 \end{pmatrix} \quad \text{and} \quad \hat{\Omega}^{-1} = \begin{pmatrix} 193.608 & -58.073 & -135.020 \\ -58.073 & 250.036 & -11.521 \\ -135.020 & -11.521 & 182.737 \end{pmatrix}.$$

The value of the statistic Q in (6) is given by 8.9468. The required values to compute the critical value are $G_1 = E(Q) = 3.2876$, $G_2 = E(Q^2) = 18.3749$, $d = 3.2243$ and $v = 103.7985$. The critical value $dF_{p,v}(0.95) = 8.6802$. The p-value for testing (11) is given by

$$P\left(dF_{p,v} > Q\right) = P\left(3.2243F_{3,103.7985} > 8.9468\right) = 0.0451.$$

The 95% simultaneous confidence intervals for $\mu_1 - \beta_1$, $\mu_2 - \beta_2$ and $\mu_3 - \beta_3$ based on (9) are

$$0.652 \pm 0.337, \quad 0.187 \pm 0.207 \quad \text{and} \quad 1.353 \pm 0.336,$$

respectively.

5.3. The results based on partially complete data

As we already mentioned, here we form complete data sets by dropping vector observations with missing components. In this case, $N_1 = N_2 = N_3 = 30$ and $M_1 = M_2 = M_3 = 25$. Using these complete vector observations, we found

$$\bar{\mathbf{x}} - \bar{\mathbf{y}} = (0.57133, 0.15733, 1.29133), \quad \widehat{\mathrm{Cov}(\bar{\mathbf{x}} - \bar{\mathbf{y}})} = \begin{pmatrix} 0.02901 & 0.00802 & 0.02194 \\ 0.00802 & 0.00862 & 0.00647 \\ 0.02194 & 0.00647 & 0.02230 \end{pmatrix} \tag{13}$$

and

$$[\widehat{\mathrm{Cov}(\bar{\mathbf{x}} - \bar{\mathbf{y}})}]^{-1} = \begin{pmatrix} 142.518 & -34.917 & -130.110 \\ -34.917 & 156.752 & -11.102 \\ -130.110 & -11.102 & 176.092 \end{pmatrix}.$$

The p-value and other critical values can be computed using the formulas in Section 4.1 with $N_1 = 30$ and $M_1 = 25$. The Hotelling T^2 statistic is computed as 3.4289 with p-value 0.3578. The 95% simultaneous confidence intervals for $\mu_1 - \beta_1$, $\mu_2 - \beta_2$ and $\mu_3 - \beta_3$ are given by

$$0.571 \pm 0.502, \quad 0.157 \pm 0.274 \quad \text{and} \quad 1.291 \pm 0.440,$$

respectively.

We observe from the above results that the conclusions of the tests based on complete data (Section 5.1) and on incomplete data are the same. Also, as expected, the simultaneous confidence

intervals based on incomplete data are wider than the corresponding ones based on complete data, and shorter than those based on partially complete data.

6. Concluding remarks

In this article, we proposed a Hotteling T^2 type test for testing the equality of two normal mean vectors when the covariance matrices are equal. The test is simple to use and the monotone patterns of the samples are not necessarily similar. We also note that in many practical situations the covariance matrices need not be equal. It is plausible that we can extend the present approach for the case of unequal covariance matrices along the lines Krishnamoorthy and Yu [7] who gave a simple test procedures when there is no missing data. We are currently working on this multivariate Behrens–Fisher problem with missing data.

As pointed out by a reviewer, the setup for the present problem is a special case of the setup for a multivariate linear regression. In particular, they are special cases of the models considered in Liu [10] who provides the MLEs for the parameters in multivariate linear regression model with missing data. Even though the MLEs are readily available, it is not straightforward to get the moment approximation for the distribution of the pivotal quantity of the form in (6). We plan to investigate the applicability of our approach to this general setup, and publish the results elsewhere.

Acknowledgment

The authors are grateful to a reviewer for providing valuable comments and suggestions.

Appendix A.

We here evaluate the first two moments of Q in (6). Define

$$Q_{2d} = W_2[(\bar{x}_2^{(1)} - \bar{y}_2^{(1)}) - (\mu_1 - \beta_1)]'[S_2^{(1,1)} + V_2^{(1,1)}]^{-1}[(\bar{x}_2^{(1)} - \bar{y}_2^{(1)}) - (\mu_1 - \beta_1)],$$

$$\begin{aligned} Q_{3d} = {} & W_3([(\bar{x}_3^{(1)} - \bar{y}_3^{(1)}) - (\mu_1 - \beta_1)]', [(\bar{x}_3^{(2)} - \bar{y}_3^{(2)}) - (\mu_2 - \beta_2)]') \\ & \times \begin{pmatrix} S_3^{(1,1)} + V_3^{(1,1)} & S_3^{(1,2)} + V_3^{(1,2)} \\ S_3^{(2,1)} + V_3^{(2,1)} & S_3^{(2,2)} + V_3^{(2,2)} \end{pmatrix}^{-1} \begin{pmatrix} (\bar{x}_3^{(1)} - \bar{y}_3^{(1)}) - (\mu_1 - \beta_1) \\ (\bar{x}_3^{(2)} - \bar{y}_3^{(2)}) - (\mu_2 - \beta_2) \end{pmatrix}, \end{aligned}$$

$$R_2 = \frac{Q_2}{(1 + Q_{2d})} \quad \text{and} \quad R_3 = \frac{Q_3}{(1 + Q_{3d})}.$$

The following results can be easily deduced from the result in Seber [15, p. 52]. The variables Q_{2d} and Q_{3d} are independent with

$$Q_{2d} \sim \frac{r_1}{N_2 + M_2 - r_1 - 1} F_{r_1, N_2 + M_2 - r_1 - 1} \tag{A.1}$$

and

$$Q_{3d} \sim \frac{r_1 + r_2}{N_3 + M_3 - (r_1 + r_2) - 1} F_{(r_1 + r_2), N_3 + M_3 - (r_1 + r_2) - 1}.$$

Also, Q_1 in (6), R_2 and R_3 are independent with

$$Q_1 \sim \frac{(N_1+M_1)r_1}{N_1+M_1-r_1-1} F_{r_1,N_1+M_1-r_1-1} \tag{A.2}$$

and

$$R_2 \sim \frac{(N_2+M_2)r_2}{N_2+M_2-(r_1+r_2)-1} F_{r_2,N_2+M_2-(r_1+r_2)-1},$$

$$R_3 \sim \frac{(N_3+M_3)r_3}{N_3+M_3-p-1} F_{r_3,N_3+M_3-p-1}.$$

Furthermore, Q_{2d} and Q_{3d} are distributed independently of R_2 and R_3. However, Q_1 and (Q_{2d}, Q_{3d}) are not independent. Notice that the pivotal quantity Q in (6) can be written as

$$\begin{aligned} Q &= Q_1 + Q_2 + Q_3 \\ &= Q_1 + R_2(1+Q_{2d}) + R_3(1+Q_{3d}) \end{aligned}$$

and hence, it follows from (A.1) and (A.2) that, the distribution of Q is free of any parameters.

We shall now evaluate the first moment and an approximation to the second moment of $Q = Q_1 + Q_2 + Q_3$. Using the above distributional results, we have

$$E(Q_1) = \frac{(N_1+M_1)r_1}{N_1+M_1-r_1-3},$$

$$\begin{aligned} E(Q_2) &= E(R_2(1+Q_{2d})) = E(R_2)E(1+Q_{2d}) \\ &= \frac{(N_2+M_2)(r_2)(N_2+M_2-3)}{(N_2+M_2-r_1-3)(N_2+M_2-r_1-r_2-3)}, \end{aligned}$$

$$\begin{aligned} E(Q_3) &= E(R_3(1+Q_{3d})) = E(R_3)E(1+Q_{3d}) \\ &= \frac{(N_3+M_3)r_3(N_3+M_3-3)}{(N_3+M_3-r_1-r_2-3)(N_3+M_3-p-3)}. \end{aligned}$$

The second moments of Q_i's are given by

$$E(Q_1^2) = \frac{(N_1+M_1)^2 r_1(r_1+2)}{(N_1+M_1-r_1-3)(N_1+M_1-r_1-5)},$$

$$\begin{aligned} &E(Q_2^2) \\ &\quad = \frac{(N_2+M_2)^2(N_2+M_2-3)(N_2+M_2-5)(r_2)(r_2+2)}{(N_2+M_2-r_1-3)(N_2+M_2-r_1-r_2-3)(N_2+M_2-r_1-5)(N_2+M_2-r_1-r_2-5)}, \end{aligned}$$

$$\begin{aligned} &E(Q_3^2) \\ &\quad = \frac{(N_3+M_3)^2(N_3+M_3-3)(N_3+M_3-5)r_3(r_3+2)}{(N_3+M_3-r_1-r_2-3)(N_3+M_3-p-3)(N_3+M_3-r_1-r_2-5)(N_3+M_3-p-5)}. \end{aligned}$$

Using the arguments of Krishnamoorthy and Pannala [6], it can be shown that $E(Q_1Q_2) \simeq E(Q_1)E(Q_2)$ and $E(Q_1Q_3) \simeq E(Q_1)E(Q_3)$. Thus, we have

$$E(Q) = E(Q_1) + E(Q_2) + E(Q_3) = G_1$$

and

$$E(Q^2) \simeq E(Q_1^2) + E(Q_2^2) + E(Q_3^2) + 2E(Q_1)E(Q_2) + 2E(Q_3)E(Q_2) + 2E(Q_1)E(Q_3) = G_2.$$

Appendix B. Generalization

Assume that the samples have the monotone pattern in (1). In this general case, the quadratic form Q in (6) can be expressed as

$$Q = [(\widehat{\mu} - \widehat{\beta}) - (\mu - \beta)]'[\widehat{\Omega}]^{-1}[(\widehat{\mu} - \widehat{\beta}) - (\mu - \beta)] = Q_1 + Q_2 + \cdots + Q_k,$$

where

$$Q_1 = W_1(\widehat{\mu}_1 - \widehat{\beta}_1 - (\mu_1 - \beta_1))'[\widehat{\Sigma}_{11}]^{-1}(\widehat{\mu}_1 - \widehat{\beta}_1 - (\mu_1 - \beta_1)),$$

and, for $l = 2, \ldots, k$,

$$Q_l = W_l \left[(\widehat{\mu}_{l,\overline{l-1}\ldots 1} - \widehat{\beta}_{l,\overline{l-1}\ldots 1}) - \left(\mu_l - \beta_l - \sum_{j=1}^{l-1} B_{lj}(\mu_j - \beta_j) \right) \right]' [\widehat{\Sigma}_{l,\overline{l-1}\ldots 1}]^{-1} \times [(\widehat{\mu}_{l,\overline{l-1}\ldots 1} - \widehat{\beta}_{l,\overline{l-1}\ldots 1})],$$

with

$$(B_{l1}, \ldots, B_{\overline{l-1}}) = \left(S_l^{(l,1)} + V_l^{(l,1)}, \ldots, S_l^{(l,\overline{l-1})} + V_l^{(l,\overline{l-1})} \right) \times \begin{pmatrix} S_l^{(1,1)} + V_l^{(1,1)} & \cdots & S_l^{(1,\overline{l-1})} + V_l^{(1,\overline{l-1})} \\ \vdots & \vdots & \vdots \\ S_l^{(\overline{l-1},1)} + V_l^{(\overline{l-1},1)} & \cdots & S_l^{(\overline{l-1},\overline{l-1})} + V_l^{(\overline{l-1},\overline{l-1})} \end{pmatrix}^{-1},$$

$$\widehat{\Sigma}_{l,\overline{l-1}\ldots 1} = \frac{1}{N_l + M_l} \left[(S_l^{(l,l)} + V_l^{(l,l)}) - (B_{l1}, \ldots, B_{\overline{l-1}}) \begin{pmatrix} S_l^{(1,l)} + V_l^{(1,l)} \\ \vdots \\ S_l^{(l-1,l)} + V_l^{(l-1,l)} \end{pmatrix} \right],$$

and

$$\widehat{\mu}_{l,\overline{l-1}\ldots 1} = \overline{X}_l^l - \sum_{j=1}^{l-1} B_{lj}\overline{X}_l^j, \widehat{\beta}_{l,\overline{l-1}\ldots 1} = \overline{Y}_l^l - \sum_{j=1}^{l-1} B_{lj}\overline{Y}_l^j.$$

Furthermore, for $l = 1, \ldots, k$, we have

$$E(Q_l) = \frac{(N_l + M_l)p_l(N_l + M_l - 3)}{(N_l + M_l - p_{(l-1)} - 3)(N_l + M_l - p_{(l)} - 3)}$$

and

$$E(Q_l^2) = \frac{(N_l+M_l)^2(N_l+M_l-3)(N_l+M_l-5)(p_l)(p_l+2)}{(N_l+M_l-p_l-3)(N_l+M_l-p_{(l-1)}-3)(N_l+M_l-p_l-5)(N_l+M_l-p_{(l)}-5)},$$

where $p_{(l)} = \sum_{i=1}^{l} p_i$ and $p_{(0)} = 0$. Also, using the approximation that $E(R_l Q_s) \simeq E(Q_l)E(Q_s)$ for $l \neq s$, we can get $G_1 = E(Q)$ and an approximation G_2 for $E(Q^2)$. Thus,

$$Q \sim dF_{p,v} \text{ approximately}, \quad \text{where } v = \frac{4pG_2 - 2(p+2)G_1^2}{pG_2 - (p+2)G_1^2} \quad \text{and} \quad d = G_1 \frac{v-2}{v}.$$

References

[1] T.W. Anderson, Maximum likelihood estimates for a multivariate normal distribution when some observations are missing, J. Amer. Statist. Assoc. 52 (1957) 200–203.

[2] R.P. Bhargava, Multivariate tests of hypotheses with incomplete data, Ph.D. Dissertation, Stanford University, Stanford, CA, 1962.

[3] J. Hao, K. Krishnamoorthy, Inferences on normal covariance matrix and generalized variance with incomplete data, J. Multivariate Anal. 78 (2001) 62–82.

[4] T. Kanda, Y. Fujikoshi, Some basic properties of the MLEs for a multivariate normal distribution with monotone missing data, J. Math. Manage. Sci. 18 (1998) 161–190.

[5] K. Krishnamoorthy, M. Pannala, Some simple test procedures for normal mean vector with incomplete data, Ann. Inst. Statist. Math. 50 (1998) 531–542.

[6] K. Krishnamoorthy, M. Pannala, Confidence estimation of normal mean vector with incomplete data, Canad. J. Statist. 27 (1999) 395–407.

[7] K. Krishnamoorthy, J. Yu, Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem, Statist. Probab. Lett. 66 (2004) 161–169.

[8] R.J.A. Little, A test of missing completely at random for multivariate data with missing values, J. Amer. Statist. Assoc. 83 (1988) 1198–1202.

[9] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, Wiley, New York, 1987.

[10] C. Liu, Bayesian robust multivariate linear regression with incomplete data, J. Amer. Statist. Assoc. 91 (1996) 1219–1227.

[11] G.B. Lu, J.B. Copas, Missing at random, likelihood ignorability and model completeness, Ann. Statist. 32 (2004) 754–765.

[12] D.F. Morrison, D. Bhoj, Power of the likelihood ratio test on the mean vector of the multivariate normal distribution with missing observations, Biometrika 60 (1973) 365–368.

[13] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.

[14] U.D. Naik, On testing equality of means of correlated variables with incomplete data, Biometrika 62 (1975) 615–622.

[15] G.A.F. Seber, Multivariate Observations, Wiley, New York, 1984.