Journal of Statistical Theory and Applications
Volume 11, Number 1, 2012, pp. 23-45 ISSN 1538-7887

A Test for Order Restriction of Several Multivariate Normal Mean Vectors against all Alternatives when the Covariance Matrices are Unknown but Common

Abouzar Bazyari *and Rahim Chinipardaz
Department of Statistics, Shahid Chamran University, Ahvaz, Iran

Abstract

In the present paper, we are interested in testing the order restriction of mean vectors against all alternatives based on a sample from several p dimensional normal distributions. Here we consider the case when the covariance matrices are completely unknown but common. We propose a test statistic and obtain the supremum of its upper tail probability under the null hypothesis. A reformulation of the test statistic is also provided based on the orthogonal projections on the closed convex cones to study its null distribution. Finally, a Monte Carlo simulation is presented to estimate its critical values.

Key words and phrases: Closed convex cone; Likelihood ratio test; Monte Carlo simulation; Multivariate isotonic regression; Multivariate normal distribution.

AMS (2000) Subject classification: Primary 62F30; Secondary 62F03, 62H15.

*Address correspondence to Abouzar Bazyari, Department of Statistics, Persian Gulf University, Bushehr 75168, Iran; $E-mail$: *ab_bazyari@yahoo.com*

1 Introduction

Suppose that $\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{in_i}$ are random vectors from a p - dimensional normal distribution $N_p(\boldsymbol{\mu}_i, \Sigma)$ with unknown mean vector $\boldsymbol{\mu}_i$ and nonsingular covariance matrix $\Sigma, i = 1, 2, \ldots, k$. We assume that Σ is unknown. Consider the problem of testing

$$H_0 : \boldsymbol{\mu}_1 \leq \boldsymbol{\mu}_2 \leq \cdots \leq \boldsymbol{\mu}_k,$$

against all alternatives, where $\boldsymbol{\mu}_i \leq \boldsymbol{\mu}_j$ means that all the elements of $\boldsymbol{\mu}_j - \boldsymbol{\mu}_i$ are non-negative. Let H_1 be the hypothesis which has no restriction on $\boldsymbol{\mu}'$s. Then we are interested in testing the hypothesis H_0 against the alternative hypothesis $H_1 - H_0$.

To illustrate the usefulness of order restriction we have taken the following example from Silvapulle and Sen (2005).

Example (Silvapulle and Sen, 2005). An experiment was conducted to evaluate the effect of exercise on the age at which a child starts to walk. Let Y denote the age (in months) at which a child starts to walk; the data on Y are given in Table 1. (The original experiment considered of another treatment, however, here we consider only three treatments for simplicity.) The first treatment group received a special walking exercise for 12 minutes per day beginning

Table 1: The age at which a child walks

Treatment (i)	Age (in months)						n_i	$\bar{y}_i$	μ_i
1	9.00	9.50	9.75	10.00	13.00	9.50	6	10.125	μ_1
2	11.00	10.00	10.00	11.75	10.50	15.00	6	11.375	μ_2
3	13.25	11.50	12.00	13.50	11.50		5	12.35	μ_3

at age 1 week and lasting 7 weeks. The second group received daily exercises but not the special walking exercises. The third group is the control; they did not receive any exercises or other treatments. For treatment $i (i = 1, 2, 3)$, let

$$\mu_i = \text{ Mean age (in months) at which a child starts to walk.}$$

In the traditional analysis of variance (ANOVA), one would usually test

$$H_0' : \mu_1 = \mu_2 = \mu_3$$

against

$$H_1' : \mu_1, \mu_2 \quad \text{and} \quad \mu_3 \quad \text{are not all equal.}$$

However, suppose that the researcher was prepared to assume that the walking exercises would not have the negative effect of increasing the mean age at which a child starts to walk, and it was desired that this additional information be incorporated to improve on the statistical analysis. For illustrative purpose, let us suppose that the researcher wishes to incorporate the information $\mu_1 \leq \mu_2 \leq \mu_3$. In this case, the test changes to

$$H_0' : \mu_1 = \mu_2 = \mu_3$$

against

$$H_1'' : \mu_1 \leq \mu_2 \leq \mu_3 \text{ and } \mu_1, \mu_2 \text{ and } \mu_3 \text{ are not all equal.}$$

Therefore, one would expect that we should be able to do better than the traditional F - test. Let us make a few remarks about this example. If the objective of the experiment was to establish that the special walking exercise results in a reduction in the mean age at which a child starts to walk, the testing problem needs to be formulated differently; for example, it may be formulated as test of

$$H_a : \mu_1 < \mu_2 < \mu_3 \text{ does not hold against } H_b : \mu_1 < \mu_2 < \mu_3.$$

The problem of testing the homogeneity of k univariate normal means against an order restricted alternative hypothesis was given by Bartholomew (1959a). The most well known and extensively studied approach is the likelihood ratio method. He derived the test statistics as well as their null distributions under these assumptions that the variances are known and unknown. Bartholomew (1959b) studied the problem of testing the homogeneity of k univariate normal means against two sided ordered hypothesis. Moreover, he obtained the test criterion for this problem of testing and also derived a good approximation to its distribution in the general case, for $k = 5$. The problems with ordered parameters have been studied to some extents by Bartholomew (1961a, b), Chacko (1963), Shorack (1967) and Kudo and Yao (1982).

Much of the development and theory on this subject was assembled in Barlow et al. (1972) and in Robertson et al. (1988). The null distribution of the likelihood ratio test (LRT) statistic depends heavily on both the specific form of the order restriction and the sample sizes. For the case of equal sample sizes, the null distributions are well tabulated for the simple order case and the simple tree order case (Barlow et al., 1972 and Robertson et al., 1988). For unequal sample sizes, approximations of the null distributions under these constraints are available (Chase, 1974; Siskind, 1976; Robertson and Wright; 1983, 1985 and Wright and Tran, 1985). For most other orderings, however, the null distribution is not readily available.

Kudo (1963) considered a p - dimensional normal distribution with unknown mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)$ and known covariance matrix Σ. The problem of testing was $H_0'' : \boldsymbol{\mu} = \mathbf{0}$ against the restricted alternative $H_1' : \mu_i \geq 0 (i = 1, \ldots, p)$, where the inequality is strict for at least one value of i. He obtained the test statistic based on the likelihood ratio criterion and discussed its existence and geometric nature and also gave a scheme for its computation. Perlman (1969) studied this problem assuming that Σ is completely unknown. Tang et al. (1989) proposed a new multivariate statistic for this problem of testing with known covariance matrix and investigated its null distribution. The null distribution is a special case of the chi-bar-squared distribution, which allows critical values to be obtained. Also this statistic is shown to approximate the likelihood ratio test statistic for such alternatives (Kudo, 1963). Robertson and Wegman (1978) obtained the likelihood ratio test statistic for testing the isotonicness of several univariate normal means against all alternative hypotheses. They calculated its exact critical values at different significance levels for some of the normal distributions and simulated the power by Monte Carlo experiment. Also they considered the test of trend for an exponential class of distributions. Sasabuchi (1980) considered the problem that the multivariate normal mean lies on the boundary of a convex polyhedral cone against the mean vector corresponds in the interior. It was a complete generalization of Inada (1978) who studied it for the bivariate case (see also Sasabuchi, 1988a, b). Anderson (1984) considered the problem of testing the homogeneity of several multivariate normal means against the unrestricted alternative hypothesis.

The extension of the ordered means has considered by some authors. Sasabuchi et al. (1983) extended Bartholomew's (1959a) problem to multivariate normal mean vectors with known covariance matrices. They computed the likelihood ratio test statistic and proposed an iterative algorithm for computing the bivariate isotonic regression. Kulatunga and Sasabuchi (1984) studied its null distribution only in some special situations on the covariance matrices. Kulatunga et al. (1990) considered this problem when the covariance matrices are not diagonal. They proposed some test procedures and studied them by simulation. Sasabuchi et al. (1992) generalized the iterative algorithm to multivariate isotonic regression. Sasabuchi et al. (1998) made some power comparisons by simulation in the bivariate case and showed that under the ordered hypothesis, Sasabuchi et al.'s (1983) test is more powerful than the usual chi-square test. Sasabuchi et al. (2003) considered this problem of testing the case that the covariance matrices are common but unknown. He proposed a test statistic, studied its upper tail probability under the null hypothesis and estimated its critical values. Sasabuchi (2007) provided some tests, which are more powerful than Sasabuchi et al. (2003).

In the present paper we extend Robertson and Wegman's (1978) problem in multivariate normal distribution. The aim of this paper is to derive the null distribution of the test statistic when the covariance matrices are completely unknown but common.

The paper is organized as follows. In Section 2, we introduce some nomenclatures, review the results given by Barlow et al. (1972), Sasabuchi et al. (1983) and also give some definitions and lemmas on the closed convex cone. In Section 3, we derive the likelihood ratio test criterion for our problem of testing when the covariance matrices are known. Also we propose a test statistic when they are completely unknown but common and obtain the supremum of the upper tail probability of the statistic under the null hypothesis. Some theorems which are concerned with the null distribution of the test statistic when the covariance matrices are unknown discussed perfectly in this section. In Section 4, we give some lemmas with their proofs. The critical values of the test statistic are computed by Monte Carlo simulation in Section 5.

2 Nomenclature, review the main results and some definitions

In this section, we give some nomenclatures, main definitions about multivariate isotonic regression and convex cone and also prepare some lemmas. Proofs of the lemmas are omitted.

A_n	closed convex cone
c	convex cone
c_∞	infinite union of the convex cones
$\bar{c}_\infty$	closure of the c_∞
H	hypothesis symbol
I_p	unit matrix
n_i	sample size of the i th population
$N_p(\boldsymbol{\mu}_i, \Sigma)$	multivariate normal distribution with parameters $\boldsymbol{\mu}_i, \Sigma$
R^p	p - dimensional real Euclidean space
S	sample covariance matrix
s_{11}	$(1,1)$ th element of S
t	critical value
$W_p(a,b)$	Wishart distribution with parameters a, b
$\mathbf{X}$	$p-$ dimensional real vector
$\bar{\mathbf{X}}_i$	sample mean vector of the i th population
$\bar{X}_{i1}$	first component of sample mean vector of the i th population
$\bar{X}$	total sample mean vector
α	significance level
$\boldsymbol{\mu}_i$	multivariate mean vector
$\hat{\boldsymbol{\mu}}_i$	multivariate isotonic regression
μ_{i1}	first component of $\boldsymbol{\mu}_i$
Λ	positive definite matrix
Σ	covariance matrix
$\otimes$	Kronecker product
$\langle \cdot, \cdot \rangle_\Lambda$	inner product in R^{pk}
$\pi_\Lambda(\mathbf{x}, \mathbf{c})$	orthogonal projection of $\mathbf{x}$ onto $\mathbf{c}$ with respect to $\langle \cdot, \cdot \rangle_\Lambda$

Definition 1. (Sasabuchi et al., 1983) Given p - variate real vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$ and $p \times p$ positive definite matrices $\Sigma_1, \Sigma_2, \ldots, \Sigma_k$ a $p \times k$ real matrix $(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \ldots, \hat{\boldsymbol{\mu}}_k)$ is said to be the multivariate isotonic regression (MIR) of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$ with weights $\Sigma_1^{-1}, \Sigma_2^{-1}, \ldots, \Sigma_k^{-1}$ if

$(\hat{\boldsymbol{\mu}}_1 \leq \hat{\boldsymbol{\mu}}_2 \leq \cdots \leq \hat{\boldsymbol{\mu}}_k)$ and $(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \ldots, \hat{\boldsymbol{\mu}}_k)$ satisfies

$$\min_{\boldsymbol{\mu}_1 \leq \boldsymbol{\mu}_2 \leq \cdots \leq \boldsymbol{\mu}_k} \sum_{i=1}^{k} (\mathbf{X}_i - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_i) = \sum_{i=1}^{k} (\mathbf{X}_i - \hat{\boldsymbol{\mu}}_i)' \Sigma_i^{-1} (\mathbf{X}_i - \hat{\boldsymbol{\mu}}_i),$$

where $\hat{\boldsymbol{\mu}}_i'$s can be computed by the iterative algorithm proposed by Sasabuchi et al. (1983). This definition includes the definition given by Barlow et al. (1972) for univariate variables. The univariate isotonic regression can be computed easily by the well-known method, Pool Adjacent Violators algorithm (see Barlow et al. 1972). Now, we give some definitions and lemmas on the closed convex cone.

Definition 2. $\mathbf{c}$ is called a convex cone if $x, y \in \mathbf{c}, \lambda \geq 0, \gamma \geq 0$ then $\lambda x + \gamma y \in \mathbf{c}$. Also $\mathbf{c}$ is called a closed convex cone if it is convex cone and close set.

Lemma 1. Let

$$A_n = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \middle| \begin{array}{c} x \in R^1, \ y \in R^1 \\ x \geq 0, \ y \geq -nx \end{array} \right\}, \quad n = 1, 2, \ldots.$$

Then we have

1) A_n is a closed convex cone in $R^2, n = 1, 2, \ldots$

2) $A_n \subset A_{n+1}, n = 1, 2, \ldots$

3) $\bigcup_{n=1}^{\infty} \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \middle| \begin{array}{c} x \geq 0 \\ y \geq 0 \end{array} \right\} \bigcup \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \middle| \begin{array}{c} x > 0 \\ y \leq 0 \end{array} \right\}$.

Now, let $A \otimes B$ be the Kronecker product of matrix $A_{r \times m} = (a_{ij})$ and $B_{h \times s} = (b_{kl})$ and defined as

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & & \vdots \\ a_{r1}B & \cdots & a_{rm}B \end{pmatrix}.$$

Also the square norm in R^{pk} is

$$||\mathbf{x}||^2 = \langle \mathbf{x}, \mathbf{x} \rangle_\Lambda.$$

For p - dimensional real vectors $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)'$ and $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k)'$ their inner product in R^{pk} is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle_\Lambda = \sum_{i=1}^{k} \mathbf{x}_i' \Lambda^{-1} \mathbf{y}_i$$

$$= (\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_k') \begin{pmatrix} \Lambda^{-1} & & 0 \\ & \ddots & \\ 0 & & \Lambda^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix}.$$

Suppose that for $\mathbf{x} \in R^{pk}$, $\pi_\Lambda(\mathbf{x}, \mathbf{c})$ be the point which minimizes $||\mathbf{x} - \mathbf{w}||_\Lambda$, where $\mathbf{w} \in \mathbf{c}$. We note that, since is a closed convex cone, so the uniqueness of $\pi_\Lambda(\mathbf{x}, \mathbf{c})$ is clear.

Lemma 2. Let $\mathbf{x} \in R^{pk}$, then for any $p \times p$ nonsingular matrix B,

$$||\pi_\Lambda(\mathbf{x}, \mathbf{c}) - \mathbf{x}||_\Lambda = ||\pi_{B\Lambda B'} \left((I_k \otimes B)\mathbf{x}, (I_k \otimes B)\mathbf{c} - (I_k \otimes B)\mathbf{x}\right)||_{B\Lambda B'}.$$

Lemma 3. Let $\{\mathbf{c}_n\}_{1,2},\ldots$ be sequence of closed convex cones in R^{pk}. If $\mathbf{c}_1 \subset \mathbf{c}_2 \subset \cdots \mathbf{c}_n \subset \cdots$, then

$$\lim_{n\to\infty} ||\pi_\Lambda(\mathbf{x}, \mathbf{c}_n) - \mathbf{x}||_\Lambda = ||\pi_\lambda(\mathbf{x}, \bar{\mathbf{c}}_\infty) - \mathbf{x}||_\Lambda,$$

where $\bar{\mathbf{c}}_\infty = \bigcup_{n=1}^\infty \mathbf{c}_n$.

3 Likelihood ratio test statistic

Consider the problem of testing

$$H_0 : \boldsymbol{\mu}_1 \leq \boldsymbol{\mu}_2 \leq \ldots \leq \boldsymbol{\mu}_k \text{ versus } H_1 - H_0,$$

where H_1 is the hypothesis has no restriction on $\boldsymbol{\mu}'$s and $\boldsymbol{\mu}_i \leq \boldsymbol{\mu}_j$ means that all the elements of $\boldsymbol{\mu}_j - \boldsymbol{\mu}_i$ are non-negative.

First suppose that the covariance matrix Σ is known, then the likelihood ratio test for testing H_0 against $H_1 - H_0$ is

$$\lambda = \frac{\sup\limits_{H_0} L(\boldsymbol{\mu})}{\sup\limits_{H_1 - H_0} L(\boldsymbol{\mu})} \tag{1}$$

$$= \frac{\sup\limits_{H_0} \prod_{i=1}^k \frac{1}{(2\pi)^{p/2}} |\Sigma|^{-1/2} \left\{-\frac{1}{2}\sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)'\Sigma^{-1}(\mathbf{X}_{ij} - \boldsymbol{\mu}_i)\right\}}{\sup\limits_{H_1 - H_0} \prod_{i=1}^k \frac{1}{(2\pi)^{p/2}} |\Sigma|^{-1/2} \left\{-\frac{1}{2}\sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)'\Sigma^{-1}(\mathbf{X}_{ij} - \boldsymbol{\mu}_i)\right\}}.$$

A likelihood ratio test rejects H_0 for small values of λ.

Theorem 1. Suppose that Σ is known. The likelihood ratio test for H_0 against $H_1 - H_0$ rejects H_0 for large values of the following statistic

$$-2\ln\lambda = \sum_{i=1}^{k} n_i(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i)'\Sigma^{-1}(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i),$$

where $\bar{\mathbf{X}}_i$ is the maximum likelihood estimate of $\boldsymbol{\mu}_i$ under the alternative hypothesis and $(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \ldots, \hat{\boldsymbol{\mu}}_k)$ is the multivariate isotonic regression (MIR) of $\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \ldots, \bar{\mathbf{X}}_k$ with weights $n_1\Sigma_1^{-1}, n_2\Sigma_2^{-1}, \ldots, n_k\Sigma_k^{-1}$ and $\bar{\mathbf{X}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{X}_{ij},\ i = 1, 2, \ldots, k$.

Proof. With some easy manipulations on formula (1) we have

$$\lambda = \exp\left\{\frac{1}{2}\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)'\Sigma^{-1}(\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i) - \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)'\Sigma^{-1}(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)\right]\right\},$$

where $\bar{\mathbf{X}}_i$ and $\boldsymbol{\mu}_i$ are as before.

$$\begin{aligned}
-2\ln\lambda &= \left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)'\Sigma^{-1}(\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i) - \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{X}_{ij} - \mathbf{X}_i)'\Sigma^{-1}(\mathbf{X}_{ij} - \mathbf{X}_i)\right] \quad (2)\\
&= \sum_{i=1}^{k}\left[-n_i(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i)'\Sigma^{-1}(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i) - 2\sum_{j=1}^{n_i}(\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)'\Sigma^{-1}(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i)\right].
\end{aligned}$$

By simplifying the second term of this formula, we have

$$\begin{aligned}
&2\sum_{j=1}^{n_i}(\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)'\Sigma^{-1}(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i)\\
&= 2\left[\sum_{j=1}^{n_i}\mathbf{X}'_{ij}\Sigma^{-1}\hat{\boldsymbol{\mu}}_i - \sum_{j=1}^{n_i}\mathbf{X}'_{ij}\Sigma^{-1}\bar{\mathbf{X}}_i - \sum_{j=1}^{n_i}\hat{\boldsymbol{\mu}}_i\Sigma^{-1}\hat{\boldsymbol{\mu}}_i + \sum_{j=1}^{n_i}\hat{\boldsymbol{\mu}}'_i\Sigma^{-1}\bar{\mathbf{X}}_i\right]\\
&= -2n_i(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i)'\Sigma^{-1}(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i).
\end{aligned}$$

Now, by substituting the obtained term in the second term of formula (2), we derive the following statistic

$$-2\ln\lambda = \sum_{i=1}^{k} n_i(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i)'\Sigma^{-1}(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i).$$

The main attention to this paper is considered to the case that covariance matrix Σ is unknown and common. The likelihood function is

$$\begin{aligned}
L(\boldsymbol{\mu}) &= \prod_{i=1}^{k}\frac{1}{(2\pi)^{p/2}}|\Sigma|^{-1/2}.\exp\left\{-\frac{1}{2}\sum_{j=1}^{n_i}(\mathbf{X}_{ij} - \boldsymbol{\mu}_i)'\Sigma^{-1}(\mathbf{X}_{ij} - \boldsymbol{\mu}_i)\right\}\\
&= \left(\frac{1}{(2\pi)^{kp/2}}\right)|\Sigma|^{-k/2}.\exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{X}_{ij} - \boldsymbol{\mu}_i)'\Sigma^{-1}(\mathbf{X}_{ij} - \boldsymbol{\mu}_i)\right\}.
\end{aligned}$$

Now, the likelihood ratio test statistic is

$$\lambda' = \frac{\sup_{\Sigma} \sup_{\boldsymbol{\mu} \in H_0} L(\boldsymbol{\mu})}{\sup_{\Sigma} \sup_{\boldsymbol{\mu} \in H_1 - H_0} L(\boldsymbol{\mu})},$$

where $\sup_{\boldsymbol{\mu} \in H_0}$ is the supremum for $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ respect to H_0 and $\sup_{\Sigma}$ is the supremum for all the $p \times p$ positive definite matrices. Then

$$\begin{aligned} \lambda' &= \frac{\sup_{\Sigma} \sup_{\boldsymbol{\mu} \in H_0} |\Sigma|^{-k/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)' \Sigma^{-1} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)\right\}}{\sup_{\Sigma} \sup_{\boldsymbol{\mu} \in H_1 - H_0} |\Sigma|^{-k/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)' \Sigma^{-1} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)\right\}} \\ &= \frac{|\hat{\Sigma}_0|^{-k/2} \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)' \hat{\Sigma}_0^{-1} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)\right]\right\}}{|\hat{\Sigma}_1|^{-k/2} \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)' \hat{\Sigma}_1^{-1} (\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)\right]\right\}}, \end{aligned}$$

where $\hat{\Sigma}_0$ and $\hat{\Sigma}_1$ are the estimators of the unknown Σ under the hypotheses H_0 and H_1 respectively. So that

$$\begin{aligned} -2 \ln \lambda' &= \left[k \ln |\hat{\Sigma}_0| + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)' \hat{\Sigma}_0^{-1} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)\right] \\ &- \left[k \ln |\hat{\Sigma}_1| + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)' \hat{\Sigma}_1^{-1} (\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)\right] \\ &= \left[k \ln |\hat{\Sigma}_0| + tr \sum_{i=1}^{k} \sum_{j=1}^{n_i} \hat{\Sigma}_0^{-1} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)' (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)\right] \\ &- \left[k \ln |\hat{\Sigma}_1| + tr \sum_{i=1}^{k} \hat{\Sigma}_1^{-1} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)' (\mathbf{X}_{ij} - \hat{\boldsymbol{\mu}}_i)\right], \end{aligned}$$

where the symbol tr denotes the trace of matrix. Now, using lemma 3.2.2 of Anderson (1984) in the way similar to that of Anderson [(1984), Section 8.8], in order to get the likelihood ratio test for our problem we need to minimize the determinant

$$\left| S + \sum_{i=1}^{k} n_i (\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)' \right|,$$

under the order hypothesis $\boldsymbol{\mu}_1 \leq \boldsymbol{\mu}_2 \leq \cdots \leq \boldsymbol{\mu}_k$ and $S = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)'$ is distributed with Wishart distribution $W_p(N - k, \Sigma)$.

On the other hand, we have

$$\left| S \;+\; \sum_{i=1}^{k} n_i (\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)' \right|$$

$$= |S|\Big|I_p + S^{-1/2}\sum_{i=1}^{k} n_i(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)'S^{-1/2}\Big|.$$

Thus it is enough to minimize the term

$$\Big|I_p + S^{-1/2}\sum_{i=1}^{k} n_i(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)'S^{-1/2}\Big|.$$

Now suppose that A is a $p \times p$ non-negative definite real (symmetric) matrix, $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the characteristic roots of A and ϵ is a positive number. Then

$$\Big|I_p + \epsilon A\Big| = \prod_{i=1}^{p}(1 + \epsilon\lambda_i) = 1 + \sum_{i=1}^{p}\epsilon\lambda_i + o(\epsilon^2) = 1 + tr(\epsilon A) + o(\epsilon^2).$$

Then our problem is to minimize the following term

$$\begin{aligned} 1 \quad + \quad & tr\left[S^{-1/2}\sum_{i=1}^{k} n_i(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)'S^{-1/2}\right] \\ = \quad & 1 + \sum_{i=1}^{k} n_i(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i)'S^{-1/2}(\boldsymbol{\mu}_i - \bar{\mathbf{X}}_i), \end{aligned}$$

under the order hypothesis $\boldsymbol{\mu}_1 \leq \boldsymbol{\mu}_2 \leq \cdots \leq \boldsymbol{\mu}_k$. Therefore, it is useful if we get the statistic

$$T = \sum_{i=1}^{k} n_i(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i)'S^{-1}(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i),$$

where $(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \ldots, \hat{\boldsymbol{\mu}}_k)$ is the multivariate isotonic regression of $\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \ldots, \bar{\mathbf{X}}_k$ with weights $n_1S^{-1}, n_2S^{-1}, \ldots, n_kS^{-1}$. For testing H_0 against $H_1 - H_0$ the large values of this statistic rejects H_0, then for given the significance level α and any constant t, we have under H_0

$$\alpha = \sup_{\Sigma}\sup_{H_0} P_{\boldsymbol{\mu},\Sigma}(T \geq t),$$

where $\sup_{\Sigma}$ is the supremum value for all $p \times p$ positive definite matrices and $\sup_{H_0}$ is the supremum value for all $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots \boldsymbol{\mu}_k$ under the hypothesis $\boldsymbol{\mu}_1 \leq \boldsymbol{\mu}_2 \leq \ldots \leq \boldsymbol{\mu}_k$. Consider another hypothesis

$$H_2 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k.$$

Then it is clear that $H_2 \subset H_0 \subset H_1$. In fact, H_2 is the least favorable among hypotheses satisfying H_0 with the largest type I error probability (Silvaputlle and Sen, 2005).

Hence we can write

$$\begin{aligned} \alpha \quad = \quad & \sup_{\Sigma}\sup_{H_0} P_{\boldsymbol{\mu},\Sigma}(T \geq t) = \sup_{\Sigma} P_{\boldsymbol{\mu}\in H_2,\Sigma}(T \geq t) \\ = \quad & \sup_{\Sigma} P_{\boldsymbol{\mu}_0,\Sigma}(T \geq t), \end{aligned}$$

where $\boldsymbol{\mu}_0$ is the common value of $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ under H_2.

3.1 Test statistic based on the orthogonal projections

We saw the test statistic is

$$T = \sum_{i=1}^{k} n_i(\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i)' S^{-1} (\hat{\boldsymbol{\mu}}_i - \bar{\mathbf{X}}_i).$$

In this subsection, we try to reformulate this statistic based on the orthogonal projections on the closed convex cones. Let us define the vectors

$$\mathbf{X} = \begin{pmatrix} \bar{\mathbf{X}}_1 \\ \vdots \\ \bar{\mathbf{X}}_k \end{pmatrix}, \qquad \hat{\boldsymbol{\mu}} = \begin{pmatrix} \hat{\boldsymbol{\mu}}_1 \\ \vdots \\ \hat{\boldsymbol{\mu}}_k \end{pmatrix}.$$

Then the new form of the T statistic is

$$T = ||\hat{\boldsymbol{\mu}} - \mathbf{X}||_S^2.$$

Now define the closed convex cones in R^{pk} as following

$$\begin{aligned} \mathbf{c}_0 &= \left\{ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_k \end{pmatrix} \Big| \boldsymbol{\mu}_1 \leq \cdots \leq \boldsymbol{\mu}_k, \boldsymbol{\mu}_i \in R^{pk}, i = 1, \ldots, k \right\}, \\ \mathbf{c}_1 &= \left\{ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_k \end{pmatrix} \Big| \boldsymbol{\mu}_i \in R^{pk}, i = 1, \ldots, k \right\}, \end{aligned}$$

where under the closed convex cone $\mathbf{c}_1$ there is no any restriction on $\boldsymbol{\mu}_i'$ s. Since $\pi_S(\mathbf{X}, \mathbf{c}_0)$ is the point which minimizes $||\mathbf{X} - \boldsymbol{\mu}||_S^2$, where $\boldsymbol{\mu} \in \mathbf{c}_0$, and we note that $(\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_k)$ is the multivariate isotonic regression under the hypothesis H_0, we have that

$$T = ||\pi_S(\mathbf{X}, \mathbf{c}_0) - \mathbf{X}||_S^2.$$

We have reformulated the T statistic based on the orthogonal projection to give the following theorem.

Theorem 2. Under the hypothesis H_2, the distribution of the T statistic is independent of $\boldsymbol{\mu}_0$.

Proof. Define the random vector $\mathbf{Y}$ by

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_k \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{X}}_1 - \boldsymbol{\mu}_0 \\ \vdots \\ \bar{\mathbf{X}}_k - \boldsymbol{\mu}_0 \end{pmatrix} = \mathbf{X} - \begin{pmatrix} \boldsymbol{\mu}_0 \\ \vdots \\ \boldsymbol{\mu}_0 \end{pmatrix}.$$

Then it is clear that the distribution of $\mathbf{Y}$ in independent of $\boldsymbol{\mu}_0$ and is distributed with $N_p\left(\mathbf{0}, \frac{\Sigma}{n_i}\right)$. On the other hand

$$\begin{aligned} T &= ||\pi_S(\mathbf{X}, \mathbf{c}_0) - \mathbf{X})||_S^2 \\ &= ||\pi_S(\mathbf{Y}, \mathbf{c}_0) - \mathbf{Y}||_S^2. \end{aligned}$$

Since the distribution of $|\pi_S(\mathbf{Y}, \mathbf{c}_0) - \mathbf{Y}||_S^2$ is independent of $\boldsymbol{\mu}_0$, so the distribution of the T statistic is independent of $\boldsymbol{\mu}_0$.

So, we assume that $\boldsymbol{\mu}_0 = \mathbf{0}$. But the distribution of T depends on the unknown Σ.

To solve this problem, we introduce the following statistic:

$$\bar{T} = \frac{1}{s_{11}} \sum_{i=1}^{k} n_i (\bar{X}_{i1} - \hat{\mu}_{i1})^2,$$

where $\bar{X}_{i1}$ and s_{11} are defined as section 2 and $(\hat{\mu}_{11}, \ldots, \hat{\mu}_{k1})$ is the univariate isotonic regression of $\bar{X}_{11}, \ldots, \bar{X}_{k1}$ with weights $n_1, \ldots, n_k$. Now, consider the convex cone $\mathbf{c}_2$ in R^{pk} by

$$\mathbf{c}_2 = \left\{ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_k \end{pmatrix} \mid \mu_{11} \leq \mu_{21} \leq \cdots \leq \mu_{k1}, \boldsymbol{\mu}_i \in R^{pk}, i = 1, \ldots, k \right\},$$

where μ_{i1} is defined as section 2. Also we define the T^* statistic by

$$T^* = ||\pi_S(\mathbf{X}, \mathbf{c}_2) - \mathbf{X}||_S^2.$$

lemma 4. Under the hypothesis H_2, the distribution of the T^* statistic is independent of $\boldsymbol{\mu}_0$ and Σ.

Proof. Suppose that $\mathbf{Y}$ be the random vector as before, then

$$T^* = ||\pi_S(\mathbf{Y}, \mathbf{c}_2) - \mathbf{Y}||_S^2,$$

and hence it is clear that the distribution of the T^* statistic is independent of $\boldsymbol{\mu}_0$. Let W be the $p \times p$ orthogonal matrix, then by lemma 5.1 of Sasabuchi et al. (2003), for the convex cone

$\mathbf{c}_2$,

$$\left(I_k \otimes (W\Sigma^{-1/2})\mathbf{c}_2\right) = \mathbf{c}_2.$$

and by lemma 2, T^* has the following form,

$$\begin{aligned} T^* &= ||\left(I_k \otimes (W\Sigma^{-1/2})\right)\mathbf{Y} \\ &- \pi_{W\Sigma^{-1/2}S\Sigma^{-1/2}W'}\left(I_k \otimes (W\Sigma^{-1/2})\right)\mathbf{Y}, \left(I_k \otimes (W\Sigma^{-1/2})\right)\mathbf{c}_2||^2_{W\Sigma^{-1/2}S\Sigma^{-1/2}W'}. \end{aligned}$$

Now, let

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_k \end{pmatrix} = \begin{pmatrix} W\Sigma^{-1/2}\mathbf{Y}_1 \\ \vdots \\ W\Sigma^{-1/2}\mathbf{Y}_k \end{pmatrix} = \left(I_k \otimes W\Sigma^{-1/2}\right)\mathbf{Y},$$

and

$$S^* = W\Sigma^{-1/2}S\Sigma^{-1/2}W'.$$

Then, by lemma 5.1 of Sasabuchi et al. (2003), we get

$$T^* = ||\pi_S(\mathbf{Z}, \mathbf{c}_2) - \mathbf{Z}||^2_S.$$

Since, the variables $\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{in_i}$ are independent, so, and $\mathbf{Z}_1, \ldots, \mathbf{Z}_k$ are mutually independent. On the other hand S^* is distributed with $W_p(n-k, I)$ and $\mathbf{Z}_i$ has the p -dimensional normal distribution $N_p\left(\mathbf{0}, \frac{I_p}{n_i}\right), i = 1, \ldots, k$. Thus, the T^* statistic does not depend on the unknown Σ.

Theorem 3. For testing H_0 against $H_1 - H_0$, for given the significance level α and any constant t, we have

$$\begin{aligned} \alpha &= \sup_{\Sigma}\sup_{H_0} P_{\boldsymbol{\mu},\Sigma}(T \geq t) = \sup_{\Sigma} P_{\boldsymbol{\mu}\in H_2,\Sigma}(T \geq t) \\ &= \sup_{\Sigma} P_{\boldsymbol{\mu}_0,\Sigma}(T \geq t) = \sup_{\Sigma} P_{\mathbf{0},\Sigma}(T \geq t) \\ &= \sup_{\mathbf{0},I_p}(\bar{T} \geq t). \end{aligned}$$

To prove this theorem, we need to give the following lemmas in section 4.

4 Other main lemmas

In this section, to complete the presented argument in section 3, some lemmas are given.

Lemma 5. For any constant t,

$$\alpha = \sup_{\Sigma} \sup_{H_0} P_{\boldsymbol{\mu},\Sigma}(T \geq t) = \sup_{\mathbf{0},\Sigma}(T \geq t) \leq P_{\mathbf{0},I_p}(T^* \geq t).$$

Proof. By defined the convex cones $\mathbf{c}_0, \mathbf{c}_1$ and $\mathbf{c}_2$, we note that $\mathbf{c}_0 \subset \mathbf{c}_1 \subset \mathbf{c}_2$. Thus

$$||\pi_S(\mathbf{X}, \mathbf{c}_0) - \mathbf{X}||_S^2 \leq ||\pi_S(\mathbf{X}, \mathbf{c}_2 - \mathbf{X})||_S^2,$$

and hence $T \leq T^*$. We get

$$\begin{aligned} \sup_{\Sigma} \sup_{H_0} P_{\boldsymbol{\mu}_0,\Sigma}(T \geq t) &= \sup_{\Sigma} P_{\boldsymbol{\mu}_0,\Sigma}(T \geq t) \\ &\leq \sup_{\boldsymbol{\mu}_0,\Sigma} (T^* \geq t) \\ &= P_{\mathbf{0},I_p}(T^* \geq t). \end{aligned}$$

Lemma 6. Consider the closed convex cones $\mathbf{c}_0, \mathbf{c}_1$ and $\mathbf{c}_2$. For a sequence of $p \times p$ nonsingular real matrices $\{F_n\},\ n = 1, 2, \ldots$ and a sequence of $p \times p$ positive real matrices $\{\Sigma_n\}$ such that

$$F_n = \begin{pmatrix} 1 & & & 0 \\ -n & 1 & & \\ \vdots & & \ddots & \\ -n & 0 & & 1 \end{pmatrix}, \qquad \Sigma_n = (F_n' F_n)^{-1}.$$

Then

1) $\mathbf{c}_1 \subset (I_k \otimes F_1)\mathbf{c}_0 \subset (I_k \otimes F_2)\mathbf{c}_0 \subset \cdots \subset (I_k \otimes F_n)\mathbf{c}_0 \subset \cdots$

2) $\overline{\bigcup_{n=1}^{\infty} (I_k \otimes F_n)\mathbf{c}_0} = \mathbf{c}_2$.

The proof of this lemma is similar to proof of lemma 5.4 which given by Sasabuchi et al. (2003).

Lemma 7. For any constant t and $n = 1, 2, \ldots$, we have

$$P_{\mathbf{0},\Sigma_n}(T \geq t) = P_{\mathbf{0},I_p}\left(||\pi_S(\mathbf{X}, (I_k \otimes F_n)\mathbf{c}_0) - \mathbf{X}||_S^2 \geq t\right).$$

Proof. Let

$$\mathbf{X}^* = \begin{pmatrix} \mathbf{X}_1^* \\ \vdots \\ \mathbf{X}_k^* \end{pmatrix} = \begin{pmatrix} F_n\bar{\mathbf{X}}_1 \\ \vdots \\ F_n\bar{\mathbf{X}}_k \end{pmatrix} = (I_k \otimes F_n)\mathbf{X}, \qquad S^* = F_n S F_n'.$$

Then by lemmas 2 and 6, we get that

$$\begin{aligned} ||\pi_S(\mathbf{X},\mathbf{c}_0)-\mathbf{X}||_S^2 &= ||\pi_{F_nSF_n'}((I_k\otimes F_n)\mathbf{X},(I_k\otimes F_n)\mathbf{c}_0)-(I_k\otimes F_n)\mathbf{X}||^2_{F_nSF_n'} \\ &= ||\pi_S(\mathbf{X}^*,\mathbf{c}_0)-\mathbf{X}^*||_S^2. \end{aligned}$$

and we have

$$P_{\mathbf{0},\Sigma_n}(T\geq t)=P_{\mathbf{0},\Sigma_n}\left(||\pi_S(\mathbf{X},\mathbf{c}_0-\mathbf{X})||_S^2\geq t\right),$$

where we know that $\bar{X}_i$ is distributed as $N_p\left(\mathbf{0},\frac{\Sigma_n}{n_i}\right)$ and S has Wishart distribution $W_p(n-k,\Sigma_n), i=1,2,\ldots,k$.
By using $\Sigma_n=(F_n'F_n)^{-1}$, it is clear that $\mathbf{X}_i^*\sim N_p\left(\mathbf{0},\frac{I_p}{n_i}\right)$ and $S^*\sim W_p(n-k,I_p)$. Thus we have the result that

$$\begin{aligned} P_{\mathbf{0},\Sigma_n}(T\geq t) &= P_{\mathbf{0},\Sigma_n}\left(||\pi_S(\mathbf{X},\mathbf{c}_0-\mathbf{X})||_S^2\geq t\right) \\ &= P_{\mathbf{0},\Sigma_n}\left(||\pi_{F_nSF_n'}((I_k\otimes F_n)\mathbf{X},(I_k\otimes F_n)\mathbf{c}_0)-\mathbf{X}^*||_S^2\geq t\right), \end{aligned}$$

and it is worthful if we write

$$P_{\mathbf{0},\Sigma_n}(T\geq t)=P_{\mathbf{0},I_p}\left(||\pi_S\left(\mathbf{X},(I_k\otimes F_n)\mathbf{c}_0\right)-\mathbf{X}||_S^2\geq t\right),$$

where $\bar{\mathbf{X}}_i\sim N_p\left(\mathbf{0},\frac{I_p}{n_i}\right)$ and $S\sim W_p(n-k,I_p),\ i=1,2,\ldots,k.$

Lemma 8. For any constant t,

$$\alpha=\sup_{\Sigma}\sup_{H_0}P_{\boldsymbol{\mu},\Sigma}(T\geq t)=\sup_{\Sigma}P_{\mathbf{0},\Sigma}(T\geq t)\geq P_{\mathbf{0},I_p}(T^*\geq t).$$

Proof. By using the lemmas 3, 6 and 7,

$$\begin{aligned} \lim_{n\to\infty}P_{\mathbf{0},\Sigma_n}(T\geq t) &= \lim_{n\to\infty}P_{\mathbf{0},I_p}(||\pi_S(\mathbf{X},\mathbf{c}_0)-\mathbf{X}||_S^2\geq t) \\ &= \lim_{n\to\infty}P_{\mathbf{0},I_p}(||\pi_S(\mathbf{X},I_k\otimes F_n)\mathbf{c}_0-\mathbf{X}||_S^2\geq t) \\ &= P_{\mathbf{0},I_p}\left[\left|\left|\pi_S\left(\mathbf{X},\overline{\bigcup_{n=1}^{\infty}(I_k\otimes F_n)\mathbf{c}_0}\right)-\mathbf{X}\right|\right|_S^2\geq t)\right] \\ &= P_{\mathbf{0},I_p}(||\pi_S(\mathbf{X},\mathbf{c}_2)-\mathbf{X}||_S^2\geq t) \\ &= P_{\mathbf{0},I_p}(T^*\geq t). \end{aligned}$$

Now, we have

$$\begin{aligned} \alpha &= \sup_{\Sigma}\sup_{H_0}P_{\boldsymbol{\mu},\Sigma}(T\geq t)=\sup_{\Sigma}P_{\mathbf{0},\Sigma}(T\geq t) \\ &\geq \lim_{n\to\infty}P_{\mathbf{0},\Sigma_n}(T\geq t)=P_{\mathbf{0},I_p}(\bar{T}\geq t). \end{aligned}$$

To prove the theorem 3 in subsection 3.1, it is enough to show that $\bar{T} = T^*$. Consider that the T^* statistic has the following form

$$\begin{aligned} T^* &= ||\pi_S(\mathbf{X}, \mathbf{c}_2) - \mathbf{X}||_S^2 \\ &= \min_{\boldsymbol{\mu} \in c_2} ||\mathbf{X} - \boldsymbol{\mu}||_S^2 = \min_{\boldsymbol{\mu} \in c_2} \sum_{i=1}^{k} n_i (\bar{\mathbf{X}}_i - \boldsymbol{\mu}_i)' S^{-1} (\bar{\mathbf{X}}_i - \boldsymbol{\mu}_i). \end{aligned}$$

Now, according to Anderson (1984), the partition on $\boldsymbol{\mu}_i, \mathbf{X}_i$ and S is considered,

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_{i1} \\ \boldsymbol{\mu}_{i2} \end{pmatrix}, \quad \bar{\mathbf{X}}_i = \begin{pmatrix} \bar{X}_{i1} \\ \bar{\mathbf{X}}_{i2} \end{pmatrix}, \quad S = \begin{pmatrix} s_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}.$$

With this partition, we have

$$\begin{aligned} & \sum_{i=1}^{k} n_i (\bar{\mathbf{X}}_i - \boldsymbol{\mu}_i)' S^{-1} (\bar{\mathbf{X}}_i - \boldsymbol{\mu}_i) \\ = & \begin{pmatrix} \bar{X}_{i1} - \mu_{i1} \\ \bar{\mathbf{X}}_{i2} - \boldsymbol{\mu}_{i2} \end{pmatrix} \begin{pmatrix} s_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}' \begin{pmatrix} \bar{X}_{i1} - \mu_{i1} \\ \bar{\mathbf{X}}_{i2} - \boldsymbol{\mu}_{i2} \end{pmatrix} \\ = & \frac{1}{s_{11}} (\bar{X}_{i1} - \mu_{i1})^2 + \left\{ \bar{\mathbf{X}}_{i2} - \frac{S_{21}}{s_{11}} (\bar{X}_{i1} - \mu_{i1}) - \boldsymbol{\mu}_{i2} \right\}' \\ \times & \left(S_{22} - \frac{S_{22}^2}{s_{11}} \right)^{-1} \left\{ \bar{\mathbf{X}}_{i2} - \frac{S_{21}}{s_{11}} (\bar{X}_{i1} - \boldsymbol{\mu}_{i1}) - \boldsymbol{\mu}_{i2} \right\}. \end{aligned}$$

So, we get

$$\begin{aligned} & ||\pi_S(\mathbf{X}, \mathbf{c}_2) - \mathbf{X}||_S^2 \\ = & \min_{\boldsymbol{\mu} \in c_2} \sum_{i=1}^{k} n_i \left[\frac{1}{s_{11}} (\bar{X}_{i1} - \mu_{i1}) + \left\{ \bar{\mathbf{X}}_{i2} - \frac{S_{21}}{s_{11}} (\bar{X}_{i1} - \mu_{i1}) - \boldsymbol{\mu}_{i2} \right\}' \right. \\ \times & \left. \left(S_{22} - \frac{S_{22}^2}{s_{11}} \right)^{-1} \left\{ \bar{\mathbf{X}}_{i2} - \frac{S_{21}}{s_{11}} (\bar{X}_{i1} - \boldsymbol{\mu}_{i1}) - \boldsymbol{\mu}_{i2} \right\} \right] \\ = & \min_{\mu_{11} \leq \mu_{21} \leq \cdots \leq \mu_{k1}} \min_{\boldsymbol{\mu}_{12}, \boldsymbol{\mu}_{22}, \ldots, \boldsymbol{\mu}_{k2}} \sum_{i=1}^{k} \left[\frac{1}{s_{11}} (\bar{X}_{i1} - \mu_{i1})^2 \right. \\ + & \sum_{i=1}^{k} n_i \left\{ \bar{\mathbf{X}}_{i2} - \frac{S_{21}}{s_{11}} (\bar{X}_{i1} - \mu_{i1}) - \boldsymbol{\mu}_{i2} \right\}' \\ \times & \left. \left(S_{22} - \frac{S_{22}^2}{s_{11}} \right)^{-1} \left\{ \bar{\mathbf{X}}_{i2} - \frac{S_{21}}{s_{11}} (\bar{X}_{i1} - \boldsymbol{\mu}_{i1}) - \boldsymbol{\mu}_{i2} \right\} \right] \\ = & \min_{\mu_{11} \leq \mu_{21} \leq \cdots \leq \mu_{k1}} \sum_{i=1}^{k} \left[\frac{1}{s_{11}} (\bar{X}_{i1} - \mu_{i1}) \right. \end{aligned}$$

$$
\begin{aligned}
&+ \min_{\boldsymbol{\mu}_{12},\boldsymbol{\mu}_{22},\ldots,\boldsymbol{\mu}_{k2}} \sum_{i=1}^{k} n_i \left\{ \bar{\mathbf{X}}_{i2} - \frac{S_{21}}{s_{11}}(\bar{X}_{i1} - \mu_{i1}) - \boldsymbol{\mu}_{i2} \right\}' \\
&\times \left(S_{22} - \frac{S_{22}^2}{s_{11}} \right)^{-1} \left\{ \bar{\mathbf{X}}_{i2} - \frac{S_{21}}{s_{11}}(\bar{X}_{i1} - \boldsymbol{\mu}_{i1}) - \boldsymbol{\mu}_{i2} \right\} \Bigg] \\
&= \min_{\mu_{11}\le\mu_{21}\le\cdots\le\mu_{k1}} \frac{1}{s_{11}} \sum_{i=1}^{k} n_i(\bar{X}_{i1} - \mu_{i1})^2 = \frac{1}{s_{11}} \sum_{i=1}^{k} n_i(\bar{X}_{i1} - \hat{\mu}_{i1})^2.
\end{aligned}
$$

Since under the hypothesis which has no restriction on $\mu_{12}, \mu_{22}, \ldots, \mu_{k2}$, the estimator of parameter μ_{i1} to minimize the second term is $\bar{X}_{i1}$, but under the hypothesis $\mu_{11} \le \mu_{21} \le \cdots \le \mu_{k1,}$ by using the Pool Adjacent Violators algorithm, we use $\hat{\mu}_{i1}$ as the estimator μ_{i1}. Hence, we write that

$$T^* = \min_{\boldsymbol{\mu}\in c_2} \sum_{i=1}^{k} n_i(\bar{\mathbf{X}}_i - \boldsymbol{\mu}_i)' S^{-1} (\bar{\mathbf{X}}_i - \boldsymbol{\mu}_i) = \bar{T}.$$

Now, we can present the proof of theorem 3. By lemmas 5 and 8, we write that

$$
\begin{aligned}
\alpha &= \sup_{\Sigma} \sup_{H_0} P_{\boldsymbol{\mu},\Sigma}(T \ge t) = \sup_{\Sigma} P_{\boldsymbol{\mu}\in H_2,\Sigma}(T \ge t) \\
&= \sup_{\Sigma} P_{\boldsymbol{\mu}_0,\Sigma}(T \ge t) = \sup_{\Sigma} P_{\mathbf{0},\Sigma}(T \ge t) \\
&= \sup_{\mathbf{0},I_p}(T^* \ge t) = P_{\mathbf{0},I_p}(\bar{T} \ge t).
\end{aligned}
$$

and this completes the proof.

5 Critical values for the given significance levels

In this section, we shall obtain the critical values of the test statistic. To get these values we need to derive the distribution of $\bar{T}$ statistic. As we have seen the form of this statistic is

$$\bar{T} = \frac{1}{s_{11}} \sum_{i=1}^{k} n_i(\bar{X}_{i1} - \hat{\mu}_{i1})^2,$$

where $(\hat{\mu}_{11}, \ldots, \hat{\mu}_{k1})$ is the univariate isotonic regression of $\bar{X}_{11}, \ldots, \bar{X}_{k1}$ with weights $n_1, \ldots, n_k$. In order to obtain the critical values, it is necessary to know the sample distribution of $\bar{T}$. Unfortunately, it is not straightforward to determine the distribution of $\bar{T}$. To illustrate this, first, suppose that $k = 2$. Then, under the hypothesis $\mu_{11} \le \mu_{21} \le \cdots \le \mu_{k1}$, we have

$$\bar{T} = \begin{cases} \frac{1}{s_{11}} \sum_{i=1}^{2} n_i(\bar{X}_{i1} - \bar{\bar{X}})^2 & \bar{X}_{11} \ge \bar{X}_{21} \\ 0 & \bar{X}_{11} < \bar{X}_{21}, \end{cases}$$

where $\bar{\bar{X}}$ is the mean of $\bar{X}_{11}$ and $\bar{X}_{21}$.

For tests of significance and for any constant , under the null hypothesis be true, the probability is

$$\begin{aligned} P(\bar{T} \geq t) &= P\left\{\bar{X}_{11} \geq \bar{X}_{21}, \frac{1}{s_{11}} \sum_{i=1}^{2} n_i(\bar{X}_{i1} - \bar{\bar{X}})^2 \geq t\right\} \\ &= P(\bar{X}_{11} \geq \bar{X}_{21}) P\left(\frac{1}{s_{11}} \sum_{i=1}^{2} n_i(\bar{X}_{i1} - \bar{\bar{X}})^2 \geq t\right) \\ &= P(\bar{X}_{11} \geq \bar{X}_{21}) P\left(\sum_{i=1}^{2} n_i(\bar{X}_{i1} - \bar{\bar{X}})^2 \geq t'\right). \end{aligned}$$

Under the hypothesis $H_0, P(\bar{X}_{11} \geq \bar{X}_{21}) = \frac{1}{2}$. But, On the other hand, since the random variable $\bar{X}_{i1}, i = 1, 2$, is distributed as $N\left(0, \frac{1}{n_i}\right)$, we get that

$$\bar{X}_{11} - \bar{\bar{X}} \sim N\left(0, \frac{1}{4}\left(\frac{3}{n_1} - \frac{1}{n_2}\right)\right), \quad \bar{X}_{21} - \bar{\bar{X}} \sim N\left(0, \frac{1}{4}\left(\frac{3}{n_2} - \frac{1}{n_1}\right)\right),$$

Thus, it is cumbersome to determine the null distribution of $\sum_{i=1}^{2} n_i(\bar{X}_{i1} - \bar{\bar{X}})^2$. Also, determination of the null distribution of the related statistics when are so difficult. In this situation, we present a Monte Carlo study to estimate the critical values. In this computer simulation, in order to obtain the critical values of T under H_0, we need to derive these values for $\bar{T}$ when $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_p$. For the computation of univariate isotonic regression of $\bar{T}$ statistic under the restriction $\mu_{11} \leq \mu_{21} \leq \cdots \leq \mu_{k1}$, the S-PLUS software is used. A computer program is written for estimation the univariate isotonic regression based on the Pool Adjacent Violators algorithm. We supposed that $n_1 = n_2 = \cdots = n_k$. Here are the steps:

1) Generate a p - dimensional random vectors of size n from $N_p(\mathbf{0}, I_p)$.

2) Compute a value for $\bar{T}$, denote it by $\bar{t}$.

3) Repeat the previous two steps 1000 times, and obtain $\bar{t}_1, \bar{t}_2, \ldots, \bar{t}_{1000}$.

4) For a given significance level α, derive the sample $100(1-\alpha)\%$ percentile of $\bar{t}_1, \bar{t}_2, \ldots, \bar{t}_{1000}$ and take it as the critical value. These critical values are listed in Table 2.

Acknowledgement: The authors would like to express their sincere thanks to the Editor and referees for their constructive comments and suggestions, which led to an improved paper.

References

[1] Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis, 2nd edition. Wiley, New York.

[2] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression.* Wiley, New York.

[3] Bartholomew, D.J. (1959a). A test of homogeneity for ordered alternatives. *Biometrika,* 46, 36-48.

[4] Bartholomew, D.J. (1959b). A test of homogeneity for ordered alternatives II. *Biometrika,* 46, 328-335.

[5] Bartholomew, D.J. (1961a). A test of homogeneity of means under restricted alternatives. *J. Roy. Statist. Soc.* B, 23, 239-281.

[6] Bartholomew, D.J. (1961b). Ordered tests in the analysis of variance. *Biometrika,* 48, 325-332.

[7] Chacko, V.J. (1963). Testing homogeneity against ordered alternatives. *The Annals of Mathematical Statistics,* 34, 945-956.

[8] Chase, G.R. (1974). On testing for ordered alternatives with an increased sample size for a control.*Biometrika,* 61, 569-578.

[9] Inada, K. (1978). Some bivariate tests of composite hypotheses with restricted alternatives. *Reports of the Faculty of Science, Kagoshima University. Mathematics, Physics and Chemistry,* 11, 25-31.

[10] Kudo, A. (1963). A multivariate analogue of the one-sided test. *Biometrika,* 50, 403-418.

[11] Kulatunga, D.D.S. and Sasabuchi, S. (1984). A test of homogeneity of mean vectors against multivariate isotonic alternatives. *Memoirs of the Faculty of Science,* Kyushu University. Seies A, Mathematics, 38, 151-161.

[12] Kulatunga, D.D.S., Inutsuka, M. and Sasabuchi, S. (1990). *A Simulation Study of Some Test Procedures for Testing Homogeneity of Mean Vectors against Multivariate Isotonic Alternatives.* Technical Report, Kyushu University.

[13] Perlman, M.D. (1969). One-sided testing problems in multivariate analysis. *The Annals of Mathematical Statistics,* 40(2), 549-567.

[14] Robertson, T. and Wegman, E.T. (1978). Likelihood ratio tests for order restrictions in exponential families. *The Annals of Statistics,* 6(3), 485-505.

[15] Robertson, T. and Wright, F.T. (1983). On approximation of the level probabilities and associated distributions in order restricted inference. *Biometrika,* 70, 597-606.

[16] Robertson, T. and Wright, F.T. (1985). One-sided comparisons for treatments with control. *Canadian Journal of Statistics,* 13, 109-120.

[17] Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference.* Wiley, New York.

[18] Sasabuchi, S. (1980). A test of a multivariate normal mean with composite hypotheses determined by linear inequalities. *Biometrika,* 67, 429-439.

[19] Sasabuchi, S. (1988a). *A multivariate test with composite hypotheses determined by linear inequalities when the covariance matrix is completely unknown.* Memoirs of the Faculty of Science, Kyushu University. Seies A, Mathematics, 42(1), 37-46.

[20] Sasabuchi, S. (1988b). *A multivariate test with composite hypotheses determined by linear inequalities when the covariance matrix has an unknown scale factor.* Memoirs of the Faculty of Science, Kyushu University. Seies A, Mathematics, 42(1), 9-19.

[21] Sasabuchi, S. (2007). More powerful tests for homogeneity of multivariate normal mean vectors under an order restriction. *Sankhya,* 69(4), 700-716.

[22] Sasabuchi, S., Inutsuka, M. and Kulatunga, D.D.S. (1983). A multivariate version of isotonic regression. *Biometrika,* 70, 465-472.

[23] Sasabuchi, S., Inutsuka, M. and Kulatunga, D.D.S. (1992). An algorithm for computing multivariate isotonic regression. *Hiroshima Mathematical Journal,* 22, 551-560.

[24] Sasabuchi, S. and Kulatunga, D.D.S. and Saito, H. (1998) Comparison of powers some tests for testing homogeneity under order restrictions in multivariate normal means. *American Journal of Mathematical and Management Sciences,* 18, 131-158.

[25] Sasabuchi, S., Tanaka, K. and Takeshi, T. (2003). Testing homogeneity of multivariate normal mean vectors under an order restriction when the covariance matrices are common but unknown. *The Annals of Statistics,* 31(5), 1517-1536.

[26] Shorack, G.R. (1967). Testing against ordered alternative in model I analysis of variance: normal theory and nonparametric. *The Annals of Mathematical Statistics,* 38, 1740-1752.

[27] Silvapulle, M.J. and Sen, P.K. (2005). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions.* Wiley, New York.

[28] Siskind, V. (1976). Approximate probability integrals and critical values for Bartholomew's test for ordered means. *Biometrika,* 63, 647-654.

[29] Tang, D.I., Gnecco, C. and Geller, N. (1989). An approximate likelihood ratio test for a normal mean vector with nonnegative components with application to clinical trials. *Biometrika,* 76, 577-583.

[30] Wright, F.T. and Tran, T.A. (1985). Approximating the level probabilities in order restricted inference: the simple tree ordering. *Biometrika,* 72, 429-439.

Appendix

Table 2: Critical values of the test statistic, $\bar{T}$, significance levels $0.025, 0.01, 0.05$, $n_1 = \cdots = n_k$

α	p	k	5	10	15	20
0.025	2	2	0.206	0.256	0.220	0.126
		3	0.343	0.390	0.281	0.110
		4	0.360	0.113	0.094	0.037
	3	2	0.471	0.496	0.438	0.255
		3	0.522	0.312	0.257	0.180
		4	0.610	0.439	0.455	0.341
	4	2	0.472	0.381	0.237	0.182
		3	0.615	0.605	0.405	0.219
		4	0.665	0.422	0.271	0.160
0.01	2	2	0.352	0.328	0.411	0.228
		3	0.415	0.503	0.326	0.175
		4	0.476	0.291	0.286	0.072
	3	2	0.536	0.550	0.457	0.382
		3	0.755	0.591	0.310	0.228
		4	0.688	0.625	0.523	0.402
	4	2	0.620	0.554	0.441	0.232
		3	0.725	0.724	0.542	0.352
		4	0.865	0.583	0.325	0.273
0.05	2	2	0.432	0.433	0.450	0.267
		3	0.540	0.563	0.423	0.207
		4	0.695	0.466	0.387	0.218
	3	2	0.620	0.618	0.503	0.411
		3	0.892	0.735	0.363	0.326
		4	0.695	0.853	0.577	0.472
	4	2	0.730	0.615	0.512	0.285
		3	0.814	0.226	0.630	0.404
		4	0.892	0.340	0.364	0.362