Tests for Normal Mean Vectors with Monotone Incomplete Data

Ayaka Yagi and Takashi Seo

Department of Mathematical Information Science
Tokyo University of Science
1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

Abstract

In this article, we consider tests for mean vectors when each data set has a monotone missing data pattern. We obtain the simplified Hotelling's T^2-type statistics and their approximate upper percentiles in the case of data with general k-step monotone missing data patterns. We also consider multivariate multiple comparisons for mean vectors with general k-step monotone missing data. Approximate simultaneous confidence intervals for pairwise comparisons among mean vectors and comparisons with a control are obtained using Bonferroni's approximation procedure. Finally, the accuracy and asymptotic behavior of the approximations are investigated by Monte Carlo simulation.

Key Words and Phrases: Comparisons with a control, Hotelling's T^2-type statistic, Maximum likelihood estimator, One-sample problem, Pairwise comparisons, Two-sample problem.

1 Introduction

The one-sample and two-sample problems of testing for mean vectors with monotone missing data are considered in this study. The case in which the missing observations are of the monotone type has been considered by several authors, including Rao (1956), Anderson (1957), and Bhargava (1962). The closed form expressions for the maximum likelihood estimators (MLEs) of the mean vector and the covariance matrix in the case of k-step monotone missing data under multivariate normality were derived by Jinadasa and Tracy (1992). Kanda and Fujikoshi (1998) discussed the distribution of the MLEs in the case of k-step monotone missing data. These results were derived for the one-sample problem. In the case of a two-step monotone missing data pattern, the usual Hotelling's T^2 statistic and various properties were derived by Chang and Richards (2009) and Seko, Yamazaki and Seo (2012), among others. Further, for the case of a three-step monotone missing data pattern, Krishnamoorthy and Pannala (1999) derived the Hotelling's

T^2 statistic and F approximation, and Yagi and Seo (2014) gave a simplified Hotelling's T^2-type statistic and its approximation to the upper percentiles under the one-sample problem. In the two-sample problem in a case of a two-step monotone missing data pattern, Seko, Kawasaki and Seo (2011) derived a Hotelling's T^2 statistic, the likelihood ratio test statistic, and their approximate upper percentiles. In addition, Yu, Krishnamoorthy and Pannala (2006) derived the Hotelling's T^2 statistic and its approximate distribution using another approach. Seko (2012) discussed tests for mean vectors with two-step monotone missing data for the m-sample problem. In the case of three-step monotone missing data, Yagi and Seo (2015b) used the concepts of Yagi and Seo (2014) to determine the approximate upper percentiles of the simplified Hotelling's T^2-type statistic for the two-sample problem. In this article, for the two-sample and m-sample problems, we propose a simplified Hotelling's T^2-type statistic and its approximate upper percentile in the case of general k-step monotone missing data. This result is an extension of Yagi and Seo (2014, 2015b).

The remainder of this article is organized as follows. In Section 2, some preliminary notations, the MLEs of the mean vectors, and the common covariance matrix for the m-sample problem are given in the case of k-step monotone missing data. In Section 3, for the one-sample problem, we discuss the test for the mean vector in the case of k-step monotone missing data. Further, we give the Hotelling's T^2-type statistic to test the equality of two mean vectors and their approximate upper percentiles in the case of k-step monotone missing data. In addition, we discuss the Hotelling's T^2-type statistics when two data sets have unequal monotone missing data patterns. We also present simultaneous confidence intervals for multiple comparisons among mean vectors under the two-sample and m-sample problems. In order to obtain the simultaneous confidence intervals, we derive the approximate upper percentiles of the $T^2_{\max}$-type statistics by Bonferroni's approximation. Finally, in Section 4, we give the simulation results and state our conclusions.

2 Monotone missing data and MLEs

In this section, we consider the MLEs of the mean vectors and the covariance matrix for the m-sample problem in the case of k-step monotone missing data. We assume that m

covariance matrices are equal and unknown. We first present some notations, definitions, and the setting in this aricle. Then, we derive the MLEs using the derivation of Yagi and Seo (2015b).

2.1 k-step monotone missing data

Let $\boldsymbol{x}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $\boldsymbol{x}_i = (\boldsymbol{x})_i$ be the subvector of $\boldsymbol{x}$ containing the first p_i components of $\boldsymbol{x}$. Then, $\boldsymbol{x}_i (= (x_1, x_2, \ldots, x_{p_i})')$ is distributed as $N_{p_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2, \ldots, k$, with $p = p_1 > p_2 > \cdots > p_k > 0$, where $\boldsymbol{\mu}_i = (\boldsymbol{\mu})_i = (\mu_1, \mu_2, \ldots, \mu_{p_i})'$ and $\boldsymbol{\Sigma}_i$ is the $p_i \times p_i$ principal submatrix of $\boldsymbol{\Sigma} (= \boldsymbol{\Sigma}_1)$. Suppose we have n_1 observations on $\boldsymbol{x}_1$, n_2 observations on $\boldsymbol{x}_2$, $\cdots$, n_k observations on $\boldsymbol{x}_k$. If $\boldsymbol{x}_{ij}$ denotes the jth observation on $\boldsymbol{x}_i$, then the k-step monotone missing data set is of the form

$$\left(\begin{array}{ccc}
 & \boldsymbol{x}'_{11} & \\
 & \vdots & \\
 & \boldsymbol{x}'_{1n_1} & \\
 & \boldsymbol{x}'_{21} & * \cdots * \\
 & \vdots & \vdots \quad \vdots \\
 & \boldsymbol{x}'_{2n_2} & * \cdots * \\
 & \cdot & \\
 & \cdot & \\
 & \cdot & \\
\boldsymbol{x}'_{k1} & * \cdots * & * \cdots * \\
\vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
\boldsymbol{x}'_{kn_k} & * \cdots * & * \cdots *
\end{array}\right),$$

where "$*$" indicates a missing observation. We now define $N_1 = 0$ and $N_{i+1} = \sum_{j=1}^{i} n_j$, $i = 1, 2, \ldots, k$.

Further, for the covariance matrix, let $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}$ and, for $1 \leq i < j \leq k$, let $(\boldsymbol{\Sigma}_i)_j$ be the principal submatrix of $\boldsymbol{\Sigma}_i$ of order $p_j \times p_j$; we define

$$\boldsymbol{\Sigma}_{i+1} = (\boldsymbol{\Sigma}_1)_{i+1}, \ \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{i+1} & \boldsymbol{\Sigma}_{i+1,2} \\ \boldsymbol{\Sigma}'_{i+1,2} & \boldsymbol{\Sigma}_{i+1,3} \end{pmatrix}$$

and

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{i+1} & \boldsymbol{\Sigma}_{(i,2)} \\ \boldsymbol{\Sigma}'_{(i,2)} & \boldsymbol{\Sigma}_{(i,3)} \end{pmatrix}, \ i = 1, 2, \ldots, k-1.$$

We use these notations, which are based on Jinadasa and Tracy (1992) and Yagi and Seo (2015b), throughout this article.

2.2 The MLEs of the mean vectors and the covariance matrix

Using the notations in Subsection 2.1, we consider the MLEs of the mean vectors and the common covariance matrix for the m-sample problem.

Let $\boldsymbol{x}_{i1}^{(\ell)}, \boldsymbol{x}_{i2}^{(\ell)}, \ldots, \boldsymbol{x}_{in_i^{(\ell)}}^{(\ell)}$ be distributed as $N_{p_i}(\boldsymbol{\mu}_i^{(\ell)}, \boldsymbol{\Sigma}_i)$ for $i = 1, 2, \ldots, k$ and $\ell = 1, 2, \ldots, m$, where $\boldsymbol{\mu}_i^{(\ell)} = (\mu_1^{(\ell)}, \mu_2^{(\ell)}, \ldots, \mu_{p_i}^{(\ell)})'$ and $\boldsymbol{\Sigma}_i$ is the $p_i \times p_i$ covariance matrix, where $p = p_1 > p_2 > \cdots > p_k > 0$, $\sum_{\ell=1}^m n_1^{(\ell)} - m \geq p \geq k$. Let

$$\overline{\boldsymbol{x}}_i^{(\ell)} = \frac{1}{n_i^{(\ell)}} \sum_{j=1}^{n_i^{(\ell)}} \boldsymbol{x}_{ij}^{(\ell)},$$

$$\boldsymbol{E}_i^{(\ell)} = \sum_{j=1}^{n_i^{(\ell)}} (\boldsymbol{x}_{ij}^{(\ell)} - \overline{\boldsymbol{x}}_i^{(\ell)})(\boldsymbol{x}_{ij}^{(\ell)} - \overline{\boldsymbol{x}}_i^{(\ell)})', \ i = 1, 2, \ldots, k.$$

Further, we define

$$N_1^{(\ell)} = 0, \ N_{i+1}^{(\ell)} = \sum_{j=1}^{i} n_j^{(\ell)}, \ i = 1, 2, \ldots, k,$$

$$\nu_{i \cdot m} = \sum_{\ell=1}^{m} n_i^{(\ell)}, \ i = 1, 2, \ldots, k, \quad M_{i \cdot m} = \sum_{\ell=1}^{m} N_i^{(\ell)}, i = 1, 2, \ldots, k+1,$$

$$\boldsymbol{d}_1^{(\ell)} = \overline{\boldsymbol{x}}_1^{(\ell)}, \ \boldsymbol{d}_i^{(\ell)} = \frac{n_i^{(\ell)}}{N_{i+1}^{(\ell)}} \left[\overline{\boldsymbol{x}}_i^{(\ell)} - \frac{1}{N_i^{(\ell)}} \sum_{j=1}^{i-1} n_j^{(\ell)} (\overline{\boldsymbol{x}}_j^{(\ell)})_i \right], \ i = 2, 3, \ldots, k,$$

$$\boldsymbol{f}_1^{(\ell)} = \boldsymbol{d}_1^{(\ell)}, \quad \boldsymbol{f}_i^{(\ell)} = \boldsymbol{U}_i \boldsymbol{d}_i^{(\ell)}, \quad i = 2, 3, \ldots, k,$$

$$\boldsymbol{U}_1 = \boldsymbol{T}_1, \quad \boldsymbol{U}_i = \boldsymbol{U}_{i-1} \boldsymbol{T}_i, \quad i = 2, 3, \ldots, k,$$

$$\boldsymbol{T}_1 = \boldsymbol{I}_{p_1}, \quad \boldsymbol{T}_{i+1} = \begin{pmatrix} \boldsymbol{I}_{p_{i+1}} \\ \boldsymbol{\Sigma}'_{(i,2)} \boldsymbol{\Sigma}_{i+1}^{-1} \end{pmatrix}, \quad i = 1, 2, \ldots, k-1.$$

The MLEs of $\boldsymbol{\mu}^{(\ell)}$ and $\boldsymbol{\Sigma}$ are given in the following theorem.

Theorem 1. *Let $\boldsymbol{x}_{ij}^{(\ell)}$ $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, n_i^{(\ell)}$, $\ell = 1, 2, \ldots, m$ be the j-th random vector of the i-th step from the ℓ-th population distributed as $N_{p_i}(\boldsymbol{\mu}_i^{(\ell)}, \boldsymbol{\Sigma}_i)$. Then, the MLEs of $\boldsymbol{\mu}^{(\ell)}$, $\ell = 1, 2, \ldots, m$ are given by*

$$\widehat{\boldsymbol{\mu}}^{(\ell)} = \sum_{i=1}^{k} \widehat{\boldsymbol{f}}_i^{(\ell)},$$

where

$$\widehat{\boldsymbol{f}}_1^{(\ell)} = \boldsymbol{d}_1^{(\ell)}, \ \widehat{\boldsymbol{f}}_i^{(\ell)} = \widehat{\boldsymbol{U}}_i^{[pl]} \boldsymbol{d}_i^{(\ell)}, \ i = 2, 3, \ldots, k,$$

$$\widehat{\boldsymbol{U}}_1^{[pl]} = \boldsymbol{T}_1, \quad \widehat{\boldsymbol{U}}_i^{[pl]} = \widehat{\boldsymbol{U}}_{i-1}^{[pl]} \widehat{\boldsymbol{T}}_i^{[pl]}, \quad i = 2, 3, \ldots, k,$$

$$\boldsymbol{T}_1 = \boldsymbol{I}_{p_1}, \quad \widehat{\boldsymbol{T}}_{i+1}^{[pl]} = \begin{pmatrix} \boldsymbol{I}_{p_{i+1}} \\ \widehat{\boldsymbol{\Sigma}}_{(i,2)}^{[pl]\prime} \widehat{\boldsymbol{\Sigma}}_{i+1}^{[pl]-1} \end{pmatrix}, \quad i = 1, 2, \ldots, k-1;$$

then, the MLE of the covariance matrix is given by

$$\widehat{\boldsymbol{\Sigma}}^{[pl]} = \frac{1}{M_{2\cdot m}} \sum_{\ell=1}^{m} \boldsymbol{H}_1^{(\ell)} + \sum_{\ell=1}^{m} \sum_{i=2}^{k} \frac{1}{M_{i+1\cdot m}} \boldsymbol{F}_i^{[pl]} \left[\boldsymbol{H}_i^{(\ell)} - \frac{\nu_{i\cdot m}}{M_{i\cdot m}} \boldsymbol{L}_{i-1,1}^{(\ell)} \right] \boldsymbol{F}_i^{[pl]\prime},$$

where

$$\boldsymbol{H}_1^{(\ell)} = \boldsymbol{E}_1^{(\ell)}, \ \boldsymbol{H}_i^{(\ell)} = \boldsymbol{E}_i^{(\ell)} + \frac{N_i^{(\ell)} N_{i+1}^{(\ell)}}{n_i^{(\ell)}} \boldsymbol{d}_i^{(\ell)} \boldsymbol{d}_i^{(\ell)\prime}, \ i = 2, 3, \ldots, k,$$

$$\boldsymbol{L}_1^{(\ell)} = \boldsymbol{H}_1^{(\ell)}, \ \boldsymbol{L}_i^{(\ell)} = (\boldsymbol{L}_{i-1}^{(\ell)})_i + \boldsymbol{H}_i^{(\ell)}, \ i = 2, 3, \ldots, k,$$

$$\boldsymbol{L}_{i1}^{(\ell)} = (\boldsymbol{L}_i^{(\ell)})_{i+1}, \ \boldsymbol{L}_i^{(\ell)} = \begin{pmatrix} \boldsymbol{L}_{i1}^{(\ell)} & \boldsymbol{L}_{i2}^{(\ell)} \\ \boldsymbol{L}_{i2}^{(\ell)\prime} & \boldsymbol{L}_{i3}^{(\ell)} \end{pmatrix}, \ i = 1, 2, \ldots, k-1$$

and

$$\boldsymbol{F}_1^{[pl]} = \boldsymbol{G}_1, \quad \boldsymbol{F}_i^{[pl]} = \boldsymbol{F}_{i-1}^{[pl]} \boldsymbol{G}_i^{[pl]}, \quad i = 2, 3, \ldots, k,$$

$$\boldsymbol{G}_1 = \boldsymbol{I}_{p_1}, \quad \boldsymbol{G}_{i+1}^{[pl]} = \begin{pmatrix} \boldsymbol{I}_{p_{i+1}} \\ \left(\sum_{\ell=1}^{m} \boldsymbol{L}_{i2}^{(\ell)} \right)' \left(\sum_{\ell=1}^{m} \boldsymbol{L}_{i1}^{(\ell)} \right)^{-1} \end{pmatrix}, \quad i = 1, 2, \ldots, k-1.$$

The above result of Theorem 1 is an extension of Theorem 1 in Yagi and Seo (2015b). Further, we note that this result can be applied to the case in which the data sets have unequal monotone missing data patterns.

3 A simplified Hotelling's T^2-type statistic

In this section, we first consider the one-sample problem of the test for the mean vector with k-step monotone missing data. We present the simplified Hotelling's T^2-type statistic and its approximate upper percentiles using the MLEs in the previous section. As in the case of the one-sample problem, we also consider the tests for the equality of two mean vectors and the simultaneous confidence intervals for any and all linear compounds of the mean. Further, we consider the simultaneous confidence intervals for the pairwise comparisons and the comparisons with a control under the m-sample problem with k-step monotone missing data.

3.1 One-sample problem

In this subsection, we consider the following hypothesis test with a k-step monotone missing data pattern:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \text{ vs. } H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0,$$

where $\boldsymbol{\mu}_0$ is known. Without loss of generality, we can assume that $\boldsymbol{\mu}_0 = \mathbf{0}$. To test the hypothesis H_0, we consider the simplified Hotelling's T^2-type statistic given by

$$\widetilde{T}_1^2 = \widehat{\boldsymbol{\mu}}' \, \widetilde{\boldsymbol{\Gamma}}^{-1} \widehat{\boldsymbol{\mu}},$$

where $\widehat{\boldsymbol{\mu}} = \sum_{i=1}^{k} \widehat{\boldsymbol{f}}_i$, $\widetilde{\boldsymbol{\Gamma}} = \widehat{\mathrm{Cov}}[\widetilde{\boldsymbol{\mu}}]$, and $\widetilde{\boldsymbol{\mu}} = \sum_{i=1}^{k} \boldsymbol{f}_i$. In this article, we use the Hotelling's T^2-type statistic with $\widetilde{\boldsymbol{\Gamma}}$ instead of $\widehat{\boldsymbol{\Gamma}}(= \widehat{\mathrm{Cov}}[\widehat{\boldsymbol{\mu}}])$ since $\mathrm{Cov}[\widehat{\boldsymbol{\mu}}]$ is complicated and $\widetilde{\boldsymbol{\Gamma}}$ and $\widehat{\boldsymbol{\Gamma}}$ are asymptotically equivalent. Then, we have the following theorem.

Theorem 2. *If the data have a k-step monotone pattern of missing observations, then the covariance matrix of $\widetilde{\boldsymbol{\mu}}$ is given by*

$$\mathrm{Cov}[\widetilde{\boldsymbol{\mu}}] = \frac{1}{N_2}\boldsymbol{\Sigma}_1 - \sum_{i=2}^{k} \frac{n_i}{N_i N_{i+1}} \boldsymbol{U}_i \boldsymbol{\Sigma}_i \boldsymbol{U}_i'.$$

Proof. First, since $\mathrm{Cov}[\widetilde{\boldsymbol{\mu}}] = \mathrm{E}[\widetilde{\boldsymbol{\mu}}\widetilde{\boldsymbol{\mu}}'] - \boldsymbol{\mu}\boldsymbol{\mu}'$ and $\widetilde{\boldsymbol{\mu}} = \sum_{i=1}^{k} \boldsymbol{f}_i$, we have

$$\mathrm{E}[\widetilde{\boldsymbol{\mu}}\widetilde{\boldsymbol{\mu}}'] = \mathrm{E}[\boldsymbol{f}_1\boldsymbol{f}_1'] + \sum_{r=2}^{k} \mathrm{E}[\boldsymbol{f}_r\boldsymbol{f}_r'] + 2\left[\sum_{s=2}^{k} \mathrm{E}[\boldsymbol{f}_1\boldsymbol{f}_s'] + \sum_{\substack{r=2 \\ r<s}}^{k}\sum_{s=3}^{k} \mathrm{E}[\boldsymbol{f}_r\boldsymbol{f}_s']\right].$$

Further, using the following results,

$$\mathrm{E}[\boldsymbol{f}_1\boldsymbol{f}_1'] = \frac{1}{n_1}\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1',$$

$$\mathrm{E}[\boldsymbol{f}_r\boldsymbol{f}_r'] = \frac{n_r}{N_r N_{r+1}} \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{U}_r', \quad r = 2, 3, \ldots, k,$$

$$\mathrm{E}[\boldsymbol{f}_1\boldsymbol{f}_s'] = -\frac{n_s}{N_s N_{s+1}} \begin{pmatrix} \boldsymbol{\Sigma}_s \\ \boldsymbol{\Sigma}_{s2}' \end{pmatrix} \boldsymbol{U}_s', \quad s = 2, 3, \ldots, k,$$

and

$$\mathrm{E}[\boldsymbol{f}_r\boldsymbol{f}_s'] = \boldsymbol{O}, \quad 2 \leq r < s \leq k,$$

we obtain

$$\mathrm{Cov}[\widetilde{\boldsymbol{\mu}}] = \frac{1}{N_2}\boldsymbol{\Sigma}_1 + \sum_{r=2}^{k} \frac{n_r}{N_r N_{r+1}} \left[\boldsymbol{U}_r \boldsymbol{\Sigma}_r - 2 \begin{pmatrix} \boldsymbol{\Sigma}_r \\ \boldsymbol{\Sigma}'_{r2} \end{pmatrix} \right] \boldsymbol{U}'_r.$$

Therefore,

$$\mathrm{Cov}[\widetilde{\boldsymbol{\mu}}] = \frac{1}{N_2}\boldsymbol{\Sigma}_1 - \sum_{r=2}^{k} \frac{n_r}{N_r N_{r+1}} \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{U}'_r$$

since $\boldsymbol{U}_r \boldsymbol{\Sigma}_r = \begin{pmatrix} \boldsymbol{\Sigma}_r \\ \boldsymbol{\Sigma}'_{r2} \end{pmatrix}$. □

For a two-step monotone missing data pattern, Yagi and Seo (2015a) gave $\mathrm{Cov}(\widehat{\boldsymbol{\mu}})$ as well as $\mathrm{Cov}(\widetilde{\boldsymbol{\mu}})$, and Seko et al. (2012) discussed the usual Hotelling's T^2 statistic, $T_1^2 = \widehat{\boldsymbol{\mu}}' \, \widehat{\boldsymbol{\Gamma}}^{-1} \widehat{\boldsymbol{\mu}}$, and its null distribution using other definitions.

We note that under H_0, the T^2-type statistic is asymptotically distributed as a χ^2 distribution with p degrees of freedom when $n_1, N_{k+1} \to \infty$ with $n_1/N_{k+1} \to \delta \in (0, 1]$. However, it has been noted that the χ^2 approximation is not a good approximation to the upper percentile of the T^2-type statistic when the sample is not large. Using the same concept for three-step monotone missing data used by Yagi and Seo (2014), we propose the approximate upper percentile of the $\widetilde{T}_1^2$ statistic since it is difficult to find the exact upper percentiles of the $\widetilde{T}_1^2$ statistic.

Theorem 3. *If the data have a k-step monotone pattern of missing observations, then the two kinds of approximate upper 100α percentiles of the $\widetilde{T}_1^2$ statistic are given by*

$$t^2_{\mathrm{YS}\cdot\mathrm{L1}}(\alpha) = (1-\omega_1) T^2_{n_1,\alpha} + \omega_1 T^2_{N_{k+1},\alpha},$$
$$t^2_{\mathrm{YS}\cdot F1}(\alpha) = \frac{n_1^* p_1}{n_1^* - p_1} F_{p_1, n_1^* - p_1, \alpha},$$

where

$$T^2_{n_1,\alpha} = \frac{n_1 p_1}{n_1 - p_1} F_{p_1, n_1 - p_1, \alpha}, \ T^2_{N_{k+1},\alpha} = \frac{N_{k+1} p_1}{N_{k+1} - p_1} F_{p_1, N_{k+1} - p_1, \alpha},$$

$$\omega_1 = \frac{\sum_{i=2}^{k} n_i p_i}{p_1 \sum_{i=2}^{k} n_i}, \ n_1^* = \frac{1}{p_1} \sum_{i=1}^{k} n_i p_i,$$

and $F_{p,q,\alpha}$ is the upper 100α percentile of the F distribution with p and q degrees of freedom.

Further, we consider the simultaneous confidence intervals for any and all linear compounds of the mean when the data have k-step monotone missing observations. Using the approximate upper percentiles of $\widetilde{T}_1^2$, for any nonnull vector $\boldsymbol{c} = (c_1, c_2, \ldots, c_p)'$, the approximate simultaneous confidence intervals for $\boldsymbol{c}'\boldsymbol{\mu}$ are given by

$$\boldsymbol{c}'\boldsymbol{\mu} \in \left[\boldsymbol{c}'\widehat{\boldsymbol{\mu}} \pm t_{\text{app}\cdot 1}(\alpha)\{\boldsymbol{c}'\widetilde{\boldsymbol{\Gamma}}\boldsymbol{c}\}^{\frac{1}{2}}\right], \quad \forall \boldsymbol{c} \in \boldsymbol{R}^p - \{\boldsymbol{0}\},$$

where $t^2_{\text{app}\cdot 1}(\alpha)$ is the value of $t^2_{\text{YS}\cdot\text{L}1}(\alpha)$ or $t^2_{\text{YS}\cdot F1}(\alpha)$. For three-step monotone missing data, see Yagi and Seo (2014).

3.2 Two-sample problem

In this section, we test the equality of two mean vectors with k-step monotone missing data. We give the simplified T^2-type statistic and its approximate upper percentiles in the case of unequal monotone missing data. Further, we consider multiple comparisons among mean vectors with k-step monotone missing data.

To test the hypothesis $H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$ vs. $H_1 : \boldsymbol{\mu}^{(1)} \neq \boldsymbol{\mu}^{(2)}$ when two data sets have the same k-step monotone missing data pattern, we adopt the Hotelling's T^2-type statistic given by

$$\widetilde{T}_2^2 = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})' \, \widetilde{\boldsymbol{\Gamma}}^{[pl]^{-1}} (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}),$$

where $\widehat{\boldsymbol{\mu}}^{(\ell)} = \sum_{i=1}^k \widehat{\boldsymbol{f}}_i^{(\ell)}$, $\ell = 1, 2$, and $\widetilde{\boldsymbol{\Gamma}}^{[pl]}$ is an estimator of $\mathrm{Cov}[\widetilde{\boldsymbol{\mu}}^{(1)} - \widetilde{\boldsymbol{\mu}}^{(2)}]$, $\widetilde{\boldsymbol{\mu}}^{(\ell)} = \sum_{i=1}^k \boldsymbol{f}_i^{(\ell)}$, $\ell = 1, 2$. Then, using the result for the one-sample problem in the previous subsection, we have

$$\widetilde{\boldsymbol{\Gamma}}^{[pl]} = \frac{n_1^{(1)} + n_1^{(2)}}{n_1^{(1)} n_1^{(2)}} \widehat{\boldsymbol{\Sigma}}_1^{[pl]} - \sum_{\ell=1}^{2} \sum_{i=2}^{k} \frac{n_i^{(\ell)}}{N_i^{(\ell)} N_{i+1}^{(\ell)}} \widehat{\boldsymbol{U}}_i^{[pl]} \widehat{\boldsymbol{\Sigma}}_i^{[pl]} \widehat{\boldsymbol{U}}_i^{[pl]'},$$

where $\widehat{\boldsymbol{\Sigma}}^{[pl]}$ is the MLE when $m = 2$ in Theorem 1.

We note that under H_0, $\widetilde{T}_2^2$ is asymptotically distributed as a χ^2 distribution with p degrees of freedom when $n_1^{(\ell)}, N_{k+1}^{(\ell)} \to \infty$ with $n_1^{(\ell)}/N_{k+1}^{(\ell)} \to \delta^{(\ell)} \in (0, 1]$, $\ell = 1, 2$. However, as with the one-sample problem, we note that the χ^2 approximation is not a good approximate upper percentile of the $\widetilde{T}_2^2$ statistic when the sample size is not large. To obtain an approximation that is accurate even for a small sample, we use the following theorem.

Theorem 4. *Suppose that two data sets have the same k-step monotone missing data pattern. Then, the two approximate upper 100α percentiles of the $\widetilde{T}_2^2$ statistic are given by*

$$t^2_{\mathrm{YS\cdot L}2}(\alpha) = (1-\omega_2)T^2_{\nu_{1\cdot 2},\alpha} + \omega_2 T^2_{M_{k+1\cdot 2},\alpha},$$
$$t^2_{\mathrm{YS\cdot}F2}(\alpha) = \frac{n_2^* p_1}{n_2^* - p_1 - 1} F_{p_1, n_2^*-p_1-1,\alpha},$$

where

$$\omega_2 = \frac{\sum_{i=2}^k \nu_{i\cdot 2} p_i}{p_1 \sum_{i=2}^k \nu_{i\cdot 2}},\ n_2^* = \frac{1}{p_1}\sum_{i=1}^k \nu_{i\cdot 2} p_i,$$

$$T^2_{\nu_{1\cdot 2},\alpha} = \frac{\nu_{1\cdot 2} p_1}{\nu_{1\cdot 2} - p_1 - 1} F_{p_1,\nu_{1\cdot 2}-p_1-1,\alpha},\ T^2_{M_{k+1\cdot 2},\alpha} = \frac{M_{k+1\cdot 2} p_1}{M_{k+1\cdot 2} - p_1 - 1} F_{p_1, M_{k+1\cdot 2}-p_1-1,\alpha},$$

and $F_{p,q,\alpha}$ is the upper 100α percentile of the F distribution with p and q degrees of freedom.

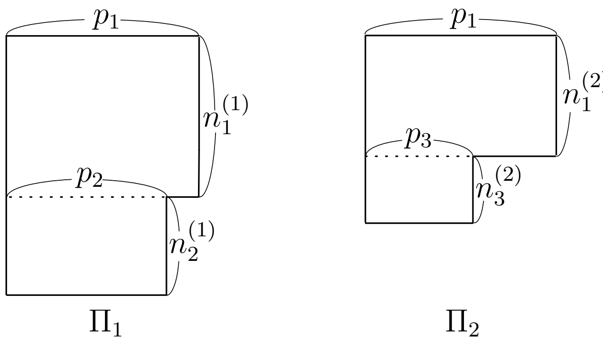

Figure 1: Unequal two-step monotone missing data patterns

Further, we test the equality of two mean vectors when two data sets have unequal general step monotone missing data patterns. For example, the two data sets Π_1 and Π_2 are of the forms given in Figure 1. Then, we can apply the results of Theorem 1 if we put $n_3^{(1)} = 0$ and $n_2^{(2)} = 0$. For details, see Yagi and Seo (2015b).

Next, under the two-sample problem, we consider the simultaneous confidence intervals when each data set has k-step monotone missing observations.

For any nonnull vector $\boldsymbol{c} = (c_1, c_2, \ldots, c_p)'$, the simultaneous confidence intervals for $\boldsymbol{c}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$ with the confidence level $(1-\alpha)$ are given by

$$\boldsymbol{c}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) \in \left[\boldsymbol{c}'(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}) \pm t_2(\alpha)\{\boldsymbol{c}'\widehat{\boldsymbol{\Gamma}}^{[pl]}\boldsymbol{c}\}^{\frac{1}{2}},\right],\ \forall \boldsymbol{c} \in \boldsymbol{R}^p - \{\boldsymbol{0}\},$$

where $t_2^2(\alpha)$ is the upper 100α percentile of the $T_2^2(= (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\boldsymbol{\Gamma}}^{[pl]^{-1}}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}))$ statistic and $\widehat{\boldsymbol{\Gamma}}^{[pl]}$ is an estimator of $\mathrm{Cov}[\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}]$. However, it is not easy to obtain $t_2^2(\alpha)$. Therefore, using the approximate upper percentiles of the $\widetilde{T}_2^2$ statistic, $t^2_{\mathrm{YS}\cdot\mathrm{L}2}(\alpha)$ or $t^2_{\mathrm{YS}\cdot F2}(\alpha)$, for any nonnull vector $\boldsymbol{c} = (c_1, c_2, \ldots, c_p)'$, the approximate simultaneous confidence intervals for $\boldsymbol{c}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$ can be obtained by

$$\boldsymbol{c}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) \in \left[\boldsymbol{c}'(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}) \pm t_{\mathrm{app}\cdot 2}(\alpha)\{\boldsymbol{c}'\widetilde{\boldsymbol{\Gamma}}^{[pl]}\boldsymbol{c}\}^{\frac{1}{2}}\right], \ \forall \boldsymbol{c} \in \boldsymbol{R}^p - \{\boldsymbol{0}\},$$

where the value of $t^2_{\mathrm{app}\cdot 2}(\alpha)$ is $t^2_{\mathrm{YS}\cdot\mathrm{L}2}(\alpha)$ or $t^2_{\mathrm{YS}\cdot F2}(\alpha)$.

3.3 Simultaneous confidence intervals for multiple comparisons among mean vectors

Under the m-sample problem, we consider the simultaneous confidence intervals for pairwise multiple comparisons among mean vectors when each data set has k-step monotone missing observations. We also consider and construct the simultaneous confidence intervals for comparisons with a control. Let $\boldsymbol{x}_{i1}^{(\ell)}, \boldsymbol{x}_{i2}^{(\ell)}, \ldots, \boldsymbol{x}_{in_i^{(\ell)}}^{(\ell)}$ be distributed as $N_{p_i}(\boldsymbol{\mu}_i^{(\ell)}, \boldsymbol{\Sigma}_i)$ for $i = 1, 2, \ldots, k$ and $\ell = 1, 2, \ldots, m$. Further, we define the $T^2_{\mathrm{max}\cdot\mathrm{p}}$ statistic as

$$T^2_{\mathrm{max}\cdot\mathrm{p}} = \max_{1\le a<b\le m} T^2_{ab},$$

where $T^2_{ab} = (\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)})'\widehat{\boldsymbol{\Gamma}}_{ab}^{[pl]^{-1}}(\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)})$ and $\widehat{\boldsymbol{\Gamma}}_{ab}^{[pl]}$ is an estimator of $\mathrm{Cov}[\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)}]$. Then, for the case of pairwise multiple comparisons, the simultaneous confidence intervals for $\boldsymbol{c}'(\boldsymbol{\mu}^{(a)} - \boldsymbol{\mu}^{(b)})$, $1 \le a < b \le m$ are given by

$$\boldsymbol{c}'(\boldsymbol{\mu}^{(a)} - \boldsymbol{\mu}^{(b)}) \in \left[\boldsymbol{c}'(\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)}) \pm t_{\mathrm{max}\cdot\mathrm{p}}(\alpha)\{\boldsymbol{c}'\widehat{\boldsymbol{\Gamma}}_{ab}^{[pl]}\boldsymbol{c}\}^{\frac{1}{2}}\right],$$
$$1 \le a < b \le m, \ \forall \boldsymbol{c} \in \boldsymbol{R}^p - \{\boldsymbol{0}\},$$

where $t^2_{\mathrm{max}\cdot\mathrm{p}}(\alpha)$ is the upper percentile of the $T^2_{\mathrm{max}\cdot\mathrm{p}}$ statistic.

Similarly, for the case of comparisons with a control, let $\boldsymbol{\mu}^{(1)}$ be a control and define the $T^2_{\mathrm{max}\cdot\mathrm{c}}$ statistic as

$$T^2_{\mathrm{max}\cdot\mathrm{c}} = \max_{2\le b\le m} T^2_{1b},$$

where $T^2_{1b} = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(a)})'\widehat{\boldsymbol{\Gamma}}_{1b}^{[pl]^{-1}}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(b)})$ and $\widehat{\boldsymbol{\Gamma}}_{1b}^{[pl]}$ is an estimator of $\mathrm{Cov}[\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(b)}]$.

Then, the simultaneous confidence intervals for $\boldsymbol{c}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(b)})$, $2 \leq b \leq m$ are given by

$$\boldsymbol{c}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(b)}) \in \big[\boldsymbol{c}'(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(b)}) \pm t_{\mathrm{max\cdot c}}(\alpha)\{\boldsymbol{c}'\widehat{\boldsymbol{\Gamma}}^{[pl]}_{1b}\boldsymbol{c}\}^{\frac{1}{2}}\big],$$
$$2 \leq b \leq m,\ \forall \boldsymbol{c} \in \boldsymbol{R}^p - \{\boldsymbol{0}\},$$

where $t^2_{\mathrm{max\cdot c}}(\alpha)$ is the upper percentile of the $T^2_{\mathrm{max\cdot c}}$ statistic.

However, it is not easy to obtain $t^2_{\mathrm{max\cdot p}}(\alpha)$ and $t^2_{\mathrm{max\cdot c}}(\alpha)$ even under non-missing multivariate normality (see Seo and Siotani (1992), Seo, Mano and Fujikoshi (1994)). Therefore, in this article, we adopt Bonferroni's approximation, which is one of the solutions to this problem. Let $n_i^{(1)} = n_i^{(2)} = \cdots = n_i^{(m)}$, $i = 1, 2, \ldots, k$; then, the null distributions of T^2_{ab} and T^2_{1b} are identical. Their approximate simultaneous confidence intervals for pairwise comparisons and comparisons with a control are given by

$$\boldsymbol{c}'(\boldsymbol{\mu}^{(a)} - \boldsymbol{\mu}^{(b)}) \in \big[\boldsymbol{c}'(\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)}) \pm t_{\mathrm{Bon}}(\alpha_{\mathrm{p}})\{\boldsymbol{c}'\widetilde{\boldsymbol{\Gamma}}^{[pl]}_{ab}\boldsymbol{c}\}^{\frac{1}{2}}\big],$$
$$1 \leq a < b \leq m,\ \forall \boldsymbol{c} \in \boldsymbol{R}^p - \{\boldsymbol{0}\},$$

and

$$\boldsymbol{c}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(b)}) \in \big[\boldsymbol{c}'(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(b)}) \pm t_{\mathrm{Bon}}(\alpha_{\mathrm{c}})\{\boldsymbol{c}'\widetilde{\boldsymbol{\Gamma}}^{[pl]}_{1b}\boldsymbol{c}\}^{\frac{1}{2}}\big],$$
$$2 \leq b \leq m,\ \forall \boldsymbol{c} \in \boldsymbol{R}^p - \{\boldsymbol{0}\},$$

respectively, where the value of $t^2_{\mathrm{Bon}}(\alpha_{\mathrm{p}})$ is $t^2_{\mathrm{YS\cdot L}m}(\alpha_{\mathrm{p}})$ or $t^2_{\mathrm{YS\cdot}Fm}(\alpha_{\mathrm{p}})$ and the value of $t^2_{\mathrm{Bon}}(\alpha_{\mathrm{c}})$ is $t^2_{\mathrm{YS\cdot L}m}(\alpha_{\mathrm{c}})$ or $t^2_{\mathrm{YS\cdot}Fm}(\alpha_{\mathrm{c}})$, which are given in the following Theorem 5, and

$$\alpha_{\mathrm{p}} = \frac{2\alpha}{m(m-1)},\ \ \alpha_{\mathrm{c}} = \frac{\alpha}{m-1}.$$

We note that $\widetilde{\boldsymbol{\Gamma}}^{[pl]}_{ab}$ and $\widetilde{\boldsymbol{\Gamma}}^{[pl]}_{1b}$ are estimated by the use of $\widehat{\boldsymbol{\Sigma}}^{[pl]}$ in Theorem 1.

Theorem 5. *Suppose that m data sets have the same k-step monotone missing data pattern. Then, the two approximate upper 100α percentiles of the $T^2_{\mathrm{max\cdot p}}$ and $T^2_{\mathrm{max\cdot c}}$ statistics are given by*

$$t^2_{\mathrm{YS\cdot L}m}(\alpha_{\mathrm{p}}) = (1-\omega_m)T^2_{\nu_{1\cdot m},\alpha_{\mathrm{p}}} + \omega_m T^2_{M_{k+1\cdot m},\alpha_{\mathrm{p}}},$$
$$t^2_{\mathrm{YS\cdot}Fm}(\alpha_{\mathrm{p}}) = \frac{n^*_m p_1}{n^*_m - p_1 - (m-1)} F_{p_1, n^*_m - p_1 - (m-1), \alpha_{\mathrm{p}}},$$

and

$$t^2_{\mathrm{YS}\cdot \mathrm{L}m}(\alpha_\mathrm{c}) = (1-\omega_m)T^2_{\nu_{1\cdot m},\alpha_\mathrm{c}} + \omega_m T^2_{M_{k+1\cdot m},\alpha_\mathrm{c}},$$
$$t^2_{\mathrm{YS}\cdot Fm}(\alpha_\mathrm{c}) = \frac{n^*_m p_1}{n^*_m - p_1 - (m-1)} F_{p_1, n^*_m - p_1 - (m-1), \alpha_\mathrm{c}},$$

respectively, where

$$\alpha_\mathrm{p} = \frac{2\alpha}{m(m-1)},\ \alpha_\mathrm{c} = \frac{\alpha}{m-1},\ \omega_m = \frac{\sum_{i=2}^k \nu_{i\cdot m} p_i}{p_1 \sum_{i=2}^k \nu_{i\cdot m}},\ n^*_m = \frac{1}{p_1}\sum_{i=1}^k \nu_{i\cdot m} p_i,$$

$$T^2_{\nu_{1\cdot m},\alpha} = \frac{\nu_{1\cdot m} p_1}{\nu_{1\cdot m} - p_1 - (m-1)} F_{p_1, \nu_{1\cdot m} - p_1 - (m-1), \alpha},$$

$$T^2_{M_{k+1\cdot m},\alpha} = \frac{M_{k+1\cdot m} p_1}{M_{k+1\cdot m} - p_1 - (m-1)} F_{p_1, M_{k+1\cdot m} - p_1 - (m-1), \alpha},$$

and $F_{p,q,\alpha}$ *is the upper 100*α *percentile of the* F *distribution with* p *and* q *degrees of freedom.*

4 Simulation studies

In this section, we investigate the accuracy and asymptotic behavior of the approximations for the upper percentiles of Hotelling's T^2-type statistic for one-sample, two-sample, and m-sample problems by Monte Carlo simulation. We provide the simulated upper percentiles and their approximations for selected parameters.

4.1 One-sample problem

For the one-sample problem, we compute the upper percentiles of the Hotelling's T^2-type statistic with k-step monotone missing data using Monte Carlo simulation (10^6 runs). That is, the $\widetilde{T}^2_1$ statistic is computed 10^6 times based on the normal random vectors generated from $N_p(\mathbf{0}, \boldsymbol{I}_p)$. Note that the Hotelling's T^2-type statistics with monotone missing data are asymptotically invariant under the nonsingular transformation. In particular, we evaluate the accuracy of the proposed approximations in Theorem 3 for the one-sample problem.

Tables 1 and 2 give the simulated upper 100α percentiles of the $\widetilde{T}^2_1$ statistic with five-step and ten-step monotone missing data patterns. That is, we provide $\tilde{t}^2_{\mathrm{simu}\cdot 1} (= \tilde{t}^2_{\mathrm{simu}\cdot 1}(\alpha))$

for the following cases:

Five-step Case: $(p_1, p_2, p_3, p_4, p_5) = (15, 12, 9, 6, 3)$,

$n_1 = 25(5)50, 100, 200, 400, 800,\ n_2 = n_3 = \cdots = n_5 = 5, 10,$

$\alpha = 0.05, 0.01.$

Ten-step Case: $(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}) = (20, 18, 16, 14, 12, 10, 8, 6, 4, 2)$,

$n_1 = 25(5)50, 100, 200, 400, 800,\ n_2 = n_3 = \cdots = n_{10} = 5, 10,$

$\alpha = 0.05, 0.01.$

These tables also give the approximations to the upper percentiles of the $\widetilde{T}_1^2$ statistic, that is, $t^2_{\mathrm{YS}\cdot\mathrm{L1}}(= t^2_{\mathrm{YS}\cdot\mathrm{L1}}(\alpha))$ and $t^2_{\mathrm{YS}\cdot F1}(= t^2_{\mathrm{YS}\cdot F1}(\alpha))$ in Theorem 3. In addition, we provide the simulated coverage probabilities for the approximate upper percentiles in Tables 1 and 2, which are given by

$$\widetilde{\mathrm{CP}}(t^2_{\mathrm{YS}\cdot\mathrm{L1}}(\alpha)) = 1 - \Pr\{\widetilde{T}_1^2 > t^2_{\mathrm{YS}\cdot\mathrm{L1}}(\alpha)\},$$

$$\widetilde{\mathrm{CP}}(t^2_{\mathrm{YS}\cdot F1}(\alpha)) = 1 - \Pr\{\widetilde{T}_1^2 > t^2_{\mathrm{YS}\cdot F1}(\alpha)\},$$

$$\widetilde{\mathrm{CP}}(\chi^2_{p,\alpha}) = 1 - \Pr\{\widetilde{T}_1^2 > \chi^2_{p,\alpha}\}.$$

It may be noted from Tables 1 and 2 that the simulated values, $\widetilde{t}^2_{\mathrm{simu}\cdot 1}(\alpha)$, are closer to the upper percentiles of the χ^2 distribution when the sample size n_1 becomes large. However, the upper percentiles of the χ^2 distribution, $\chi^2_{p,\alpha}$, are not good approximations to those of the $\widetilde{T}_1^2$ statistic for small sample sizes. At the same time, the proposed approximate upper percentiles $t^2_{\mathrm{YS}\cdot\mathrm{L1}}$ and $t^2_{\mathrm{YS}\cdot F1}$ are good even for small sample sizes; in particular, $t^2_{\mathrm{YS}\cdot\mathrm{L1}}$ is considerably good for all cases.

4.2 Two-sample problem

To investigate the accuracy of some of the approximations under the two-sample and m-sample problems, we compute the upper percentiles of the $\widetilde{T}_2^2$, $\widetilde{T}^2_{\max\cdot\mathrm{p}}$ and $\widetilde{T}^2_{\max\cdot\mathrm{c}}$ statistics by Monte Carlo simulation (10^6 runs). As with the one-sample problem in Subsection 4.1, the $\widetilde{T}_2^2$, $\widetilde{T}^2_{\max\cdot\mathrm{p}}$, and $\widetilde{T}^2_{\max\cdot\mathrm{c}}$ statistics are computed 10^6 times for each set $(\alpha, p_i, n_i^{(\ell)})$ of parameters based on the normal random vectors $\boldsymbol{x}_{ij}^{(\ell)}$ generated from $N_{p_i}(\mathbf{0}, \boldsymbol{I}_{p_i})$.

The simulation results related to the upper percentiles of the $\widetilde{T}_2^2$ statistic and their approximations in the cases of five-step and ten-step monotone missing data are summarized in Tables 3 and 4. Computations are carried out for the following two cases:

$$\begin{aligned}
\text{Five-step Case:}\quad & (p_1, p_2, p_3, p_4, p_5) = (15, 12, 9, 6, 3),\\
& n_1^{(1)} = n_1^{(2)} = 25(5)50, 100, 200, 400, 800,\\
& n_2^{(\ell)} = n_3^{(\ell)} = \cdots = n_5^{(\ell)} = 5, 10,\ \ \ell = 1, 2,\\
& \alpha = 0.05, 0.01.
\end{aligned}$$

$$\begin{aligned}
\text{Ten-step Case:}\quad & (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}) = (20, 18, 16, 14, 12, 10, 8, 6, 4, 2),\\
& n_1^{(1)} = n_1^{(2)} = 25(5)50, 100, 200, 400, 800,\\
& n_2^{(\ell)} = n_3^{(\ell)} = \cdots = n_{10}^{(\ell)} = 5, 10,\ \ \ell = 1, 2,\\
& \alpha = 0.05, 0.01.
\end{aligned}$$

Tables 3 and 4 give the simulated upper 100α percentiles of the $\widetilde{T}_2^2$ statistic $(\widetilde{t}^2_{\text{simu}\cdot 2}(\alpha))$, the approximate upper 100α percentiles of $\widetilde{T}_2^2$ $(t^2_{\text{YS}\cdot\text{L}2}(\alpha),\ t^2_{\text{YS}\cdot F2}(\alpha))$, and the upper 100α percentiles of the χ^2 distribution with p degrees of freedom $(\chi^2_{p,\alpha})$. In the tables, we denote $\widetilde{t}^2_{\text{simu}\cdot 2}(\alpha)$, $t^2_{\text{YS}\cdot\text{L}2}(\alpha)$, and $t^2_{\text{YS}\cdot F2}(\alpha)$ as $\widetilde{t}^2_{\text{simu}\cdot 2}$, $t^2_{\text{YS}\cdot\text{L}2}$, and $t^2_{\text{YS}\cdot F2}$, respectively. In addition, we provide the simulated coverage probabilities for the approximate upper 100α percentiles given by

$$\widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L}2}(\alpha)) = 1 - \Pr\{\widetilde{T}_2^2 > t^2_{\text{YS}\cdot\text{L}2}(\alpha)\},$$
$$\widetilde{\text{CP}}(t^2_{\text{YS}\cdot F2}(\alpha)) = 1 - \Pr\{\widetilde{T}_2^2 > t^2_{\text{YS}\cdot F2}(\alpha)\},$$
$$\widetilde{\text{CP}}(\chi^2_{p,\alpha}) = 1 - \Pr\{\widetilde{T}_2^2 > \chi^2_{p,\alpha}\}.$$

It may be noted from Tables 3 and 4 that the simulated values are not close to the upper percentiles of the χ^2 distribution even when the sample size $n_1^{(\ell)}$ is moderately large. However, the proposed approximations are accurate even for cases in which $n_1^{(\ell)}$ is not large. In particular, the values of $t^2_{\text{YS}\cdot\text{L}2}(\alpha)$ are highly accurate for all cases. In other words, the simulated coverage probabilities for $t^2_{\text{YS}\cdot\text{L}2}(\alpha)$ are considerably close to the nominal level $1-\alpha$. For simulation results in the case of three-step monotone missing data, see Yagi and Seo (2015b). Thus, it can be concluded that the approximation $t^2_{\text{YS}\cdot\text{L2}}(\alpha)$ is highly accurate even for small samples and unbalanced cases when the data have a k-step monotone pattern of missing observations.

Next, in order to compare the approximate values with the simulated values in the cases of pairwise comparisons and comparisons with a control, we compute for the following case:

$$\begin{aligned}
\text{Five-step Case:}\quad & m = 6, 10, \\
& (p_1, p_2, p_3, p_4, p_5) = (15, 12, 9, 6, 3), \\
& n_1^{(\ell)} = 25(5)50, 100, 200, 400, 800,\ \ \ell = 1, 2, \ldots, m, \\
& n_2^{(\ell)} = n_3^{(\ell)} = \cdots = n_5^{(\ell)} = 5, 10,\ \ \ell = 1, 2, \ldots, m, \\
& \alpha = 0.05, 0.01.
\end{aligned}$$

Tables 5 and 6 give the simulated upper 100α percentiles of the $\widetilde{T}^2_{\text{max}\cdot\text{p}}$ statistic ($\widetilde{t}^2_{\text{simu}\cdot\text{p}}(\alpha)$), the simulated upper $100\alpha_{\text{p}}$ percentiles of the $\widetilde{T}^2_{ab}$ statistic ($\widetilde{t}^2_{\text{simu}\cdot\text{Bon}}(\alpha_{\text{p}})$), the approximate upper $100\alpha_{\text{p}}$ percentiles of the $\widetilde{T}^2_{ab}$ statistic ($t^2_{\text{YS}\cdot\text{L}m}(\alpha_{\text{p}})$, $t^2_{\text{YS}\cdot Fm}(\alpha_{\text{p}})$), and the upper $100\alpha_{\text{p}}$ percentiles of the χ^2 distribution with p degrees of freedom ($\chi^2_{p,\alpha_{\text{p}}}$). The values of $\widetilde{t}^2_{\text{simu}\cdot\text{Bon}}(\alpha_{\text{p}})$ are simulated values obtained via Monte Carlo simulation. In the tables, we denote $\widetilde{t}^2_{\text{simu}\cdot\text{p}}(\alpha)$, $\widetilde{t}^2_{\text{simu}\cdot\text{Bon}}(\alpha_{\text{p}})$, $t^2_{\text{YS}\cdot\text{L}m}(\alpha_{\text{p}})$, and $t^2_{\text{YS}\cdot Fm}(\alpha_{\text{p}})$ as $\widetilde{t}^2_{\text{simu}\cdot\text{p}}$, $\widetilde{t}^2_{\text{simu}\cdot\text{Bon}}$, $t^2_{\text{YS}\cdot\text{L}m}$, and $t^2_{\text{YS}\cdot Fm}$, respectively. In addition, we provide the simulated coverage probabilities given by

$$\begin{aligned}
\widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L}m}(\alpha_{\text{p}})) &= 1 - \Pr\{\widetilde{T}^2_{\text{max}\cdot\text{p}} > t^2_{\text{YS}\cdot\text{L}m}(\alpha_{\text{p}})\}, \\
\widetilde{\text{CP}}(t^2_{\text{YS}\cdot Fm}(\alpha_{\text{p}})) &= 1 - \Pr\{\widetilde{T}^2_{\text{max}\cdot\text{p}} > t^2_{\text{YS}\cdot Fm}(\alpha_{\text{p}})\}, \\
\widetilde{\text{CP}}(\chi^2_{p,\alpha_{\text{p}}}) &= 1 - \Pr\{\widetilde{T}^2_{\text{max}\cdot\text{p}} > \chi^2_{p,\alpha_{\text{p}}}\}.
\end{aligned}$$

As Tables 5 and 6 show, the simulated values for $\widetilde{t}^2_{\text{simu}\cdot\text{Bon}}(\alpha_{\text{p}})$ are larger than the simulated values for $\widetilde{t}^2_{\text{simu}\cdot\text{p}}(\alpha)$. It may be seen from the tables that the approximate values of $t^2_{\text{YS}\cdot\text{L}m}(\alpha_{\text{p}})$ and $t^2_{\text{YS}\cdot Fm}(\alpha_{\text{p}})$ are closer to the simulated values of $\widetilde{t}^2_{\text{simu}\cdot\text{p}}(\alpha)$ when the sample size becomes large. The simulation studies show that $t^2_{\text{YS}\cdot\text{L}m}(\alpha_{\text{p}})$ is close to $\widetilde{t}^2_{\text{simu}\cdot\text{p}}(\alpha)$ and is a conservative approximation.

Tables 7 and 8 list the results for the case of comparisons with a control. We provide $\widetilde{t}^2_{\text{simu}\cdot\text{c}}(\alpha)$, $\widetilde{t}^2_{\text{simu}\cdot\text{Bon}}(\alpha_{\text{c}})$, $t^2_{\text{YS}\cdot\text{L}m}(\alpha_{\text{c}})$, $t^2_{\text{YS}\cdot Fm}(\alpha_{\text{c}})$, and $\chi^2_{p,\alpha_{\text{c}}}$ as well as $\widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L}m}(\alpha_{\text{c}}))$,

$\widetilde{\mathrm{CP}}(t^2_{\mathrm{YS}\cdot Fm}(\alpha_{\mathrm{c}}))$, and $\widetilde{\mathrm{CP}}(\chi^2_{p,\alpha_{\mathrm{c}}})$. The accuracy of the approximations is similar to that in the case of pairwise comparisons. Additional simulation results are given in Yagi and Seo (2015b).

In conclusion, we developed the approximate upper percentiles of the Hotelling's T^2-type statistic for tests of mean vectors with k-step monotone missing data under one-sample and two-sample problems. Further, we presented the approximate simultaneous confidence intervals for pairwise comparisons among mean vectors and comparisons with a control using Bonferroni's approximation. The proposed approximate values can be easily calculated, and the accuracy of the approximations is considerably higher than that of the χ^2 approximations in almost all cases.

Table 1: Simulated and approximate values and coverage probabilities when $(p_1, p_2, p_3, p_4, p_5) = (15, 12, 9, 6, 3)$

Sample Size		Upper Percentile			Coverage Probability		
n_1	$n_2 = n_3 = \cdots = n_5$	$\widetilde{t}^2_{\text{simu}\cdot 1}$	$t^2_{\text{YS}\cdot \text{L1}}$	$t^2_{\text{YS}\cdot F1}$	$\widetilde{\text{CP}}_{\text{YS}\cdot \text{L1}}$	$\widetilde{\text{CP}}_{\text{YS}\cdot F1}$	$\widetilde{\text{CP}}_{\chi^2}$
			$\alpha = 0.05$				
25	5	90.68	76.01	57.84	.915	.825	.295
30	5	63.22	57.08	50.13	.927	.885	.429
35	5	52.04	48.76	45.33	.934	.911	.527
40	5	46.20	44.02	42.06	.938	.924	.598
45	5	42.46	40.91	39.69	.940	.931	.652
50	5	39.74	38.71	37.90	.943	.937	.693
100	5	31.03	30.86	30.79	.948	.948	.850
200	5	27.77	27.76	27.75	.950	.950	.908
400	5	26.34	26.34	26.34	.950	.950	.931
800	5	25.71	25.66	25.66	.949	.949	.940
25	10	84.33	71.59	45.33	.919	.743	.336
30	10	59.05	53.73	42.06	.929	.837	.470
35	10	49.08	46.13	39.69	.934	.881	.564
40	10	43.95	41.90	37.90	.937	.903	.630
45	10	40.66	39.17	36.49	.940	.916	.680
50	10	38.35	37.25	35.36	.942	.925	.716
100	10	30.69	30.43	30.22	.947	.945	.856
200	10	27.72	27.64	27.61	.949	.949	.908
400	10	26.36	26.31	26.31	.949	.949	.931
800	10	25.65	25.65	25.65	.950	.950	.941
			$\alpha = 0.01$				
25	5	145.85	115.84	81.06	.978	.930	.432
30	5	91.99	80.66	68.40	.982	.964	.588
35	5	72.12	66.53	60.75	.985	.976	.692
40	5	62.48	58.84	55.65	.986	.981	.759
45	5	56.34	53.96	52.01	.987	.984	.806
50	5	52.17	50.57	49.28	.988	.985	.840
100	5	39.19	38.89	38.78	.989	.989	.946
200	5	34.52	34.45	34.44	.990	.990	.974
400	5	32.45	32.45	32.45	.990	.990	.983
800	5	31.54	31.50	31.50	.990	.990	.987
25	10	137.95	109.05	60.75	.978	.872	.480
30	10	85.96	75.57	55.65	.982	.937	.633
35	10	67.84	62.58	52.01	.985	.962	.727
40	10	59.06	55.68	49.28	.986	.972	.788
45	10	53.65	51.37	47.17	.987	.978	.829
50	10	49.96	48.41	45.48	.988	.981	.858
100	10	38.48	38.27	37.95	.990	.989	.949
200	10	34.34	34.28	34.24	.990	.990	.974
400	10	32.48	32.41	32.40	.990	.990	.983
800	10	31.50	31.49	31.49	.990	.990	.987

Note. $\widetilde{\text{CP}}_{\text{YS}\cdot\text{L1}} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L1}}(\alpha))$, $\widetilde{\text{CP}}_{\text{YS}\cdot F1} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot F1}(\alpha))$, $\widetilde{\text{CP}}_{\chi^2} = \widetilde{\text{CP}}(\chi^2_{p,\alpha})$, $\chi^2_{15,0.05} = 25.00$, $\chi^2_{15,0.01} = 30.58$.

Table 2: Simulated and approximate values and coverage probabilities when $(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}) = (20, 18, 16, 14, 12, 10, 8, 6, 4, 2)$

Sample Size		Upper Percentile			Coverage Probability		
n_1	$n_2 = n_3 = \cdots = n_{10}$	$\widetilde{t}^2_{\text{simu}\cdot 1}$	$t^2_{\text{YS}\cdot\text{L1}}$	$t^2_{\text{YS}\cdot F1}$	$\widetilde{\text{CP}}_{\text{YS}\cdot\text{L1}}$	$\widetilde{\text{CP}}_{\text{YS}\cdot F1}$	$\widetilde{\text{CP}}_{\chi^2}$
			$\alpha = 0.05$				
25	5	333.46	252.88	67.91	.918	.447	.082
30	5	119.90	107.28	61.48	.929	.697	.209
35	5	83.24	77.62	56.96	.934	.805	.323
40	5	68.57	65.16	53.61	.936	.858	.418
45	5	60.78	58.28	51.03	.938	.886	.493
50	5	55.71	53.88	48.99	.940	.904	.554
100	5	40.94	40.49	40.01	.946	.942	.795
200	5	35.79	35.71	35.66	.949	.949	.886
400	5	33.52	33.53	33.52	.950	.950	.921
800	5	32.48	32.46	32.46	.950	.950	.936
25	10	316.56	248.27	49.96	.924	.355	.126
30	10	108.26	103.34	48.12	.943	.609	.277
35	10	74.53	74.21	46.61	.949	.743	.400
40	10	62.19	62.18	45.35	.950	.812	.491
45	10	55.62	55.65	44.29	.950	.853	.561
50	10	51.68	51.55	43.37	.949	.878	.612
100	10	39.78	39.54	38.40	.948	.937	.815
200	10	35.55	35.39	35.22	.948	.947	.890
400	10	33.47	33.43	33.41	.950	.949	.922
800	10	32.46	32.44	32.43	.950	.950	.936
			$\alpha = 0.01$				
25	5	761.78	509.35	90.38	.978	.602	.141
30	5	194.43	162.55	80.52	.982	.840	.329
35	5	119.24	107.98	73.73	.984	.919	.477
40	5	93.25	87.16	68.77	.985	.951	.588
45	5	80.10	76.24	64.99	.986	.965	.667
50	5	72.07	69.48	62.03	.987	.973	.726
100	5	50.47	49.97	49.29	.989	.988	.915
200	5	43.45	43.37	43.29	.990	.990	.965
400	5	40.40	40.41	40.40	.990	.990	.980
800	5	38.96	38.97	38.97	.990	.990	.986
25	10	744.34	502.77	63.43	.979	.488	.202
30	10	179.29	156.96	60.77	.985	.762	.414
35	10	107.25	103.17	58.61	.988	.875	.565
40	10	84.32	82.97	56.81	.989	.924	.663
45	10	73.04	72.55	55.30	.990	.948	.732
50	10	66.36	66.21	54.00	.990	.961	.779
100	10	48.87	48.66	47.05	.990	.986	.927
200	10	43.10	42.93	42.70	.990	.989	.967
400	10	40.33	40.28	40.24	.990	.990	.981
800	10	38.97	38.94	38.93	.990	.990	.986

Note. $\widetilde{\text{CP}}_{\text{YS}\cdot\text{L1}} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L1}}(\alpha))$, $\widetilde{\text{CP}}_{\text{YS}\cdot F1} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot F1}(\alpha))$, $\widetilde{\text{CP}}_{\chi^2} = \widetilde{\text{CP}}(\chi^2_{p,\alpha})$, $\chi^2_{20,0.05} = 31.41$, $\chi^2_{20,0.01} = 37.57$.

Table 3: Simulated and approximate values and coverage probabilities when $(p_1, p_2, p_3, p_4, p_5) = (15, 12, 9, 6, 3)$

Sample Size		Upper Percentile			Coverage Probability		
$n_1^{(\ell)}$	$n_2^{(\ell)} = n_3^{(\ell)} = \cdots = n_5^{(\ell)}$	$\widetilde{t}^2_{\text{simu}\cdot 2}$	$t^2_{\text{YS}\cdot\text{L2}}$	$t^2_{\text{YS}\cdot F2}$	$\widetilde{\text{CP}}_{\text{YS}\cdot\text{L2}}$	$\widetilde{\text{CP}}_{\text{YS}\cdot F2}$	$\widetilde{\text{CP}}_{\chi^2}$
			$\alpha = 0.05$				
25	5	39.37	38.20	36.09	.942	.923	.697
30	5	36.16	35.39	34.23	.944	.933	.754
35	5	34.14	33.61	32.91	.945	.939	.790
40	5	32.78	32.38	31.91	.946	.941	.815
45	5	31.77	31.46	31.14	.947	.944	.835
50	5	30.98	30.76	30.52	.948	.945	.850
100	5	27.83	27.78	27.75	.949	.949	.907
200	5	26.39	26.38	26.37	.950	.950	.930
400	5	25.66	25.68	25.68	.950	.950	.941
800	5	25.33	25.34	25.34	.950	.950	.945
25	10	37.50	36.76	32.91	.944	.905	.730
30	10	34.77	34.23	31.91	.945	.921	.778
35	10	33.03	32.66	31.14	.947	.931	.810
40	10	31.90	31.58	30.52	.947	.936	.831
45	10	31.07	30.79	30.02	.947	.939	.848
50	10	30.44	30.18	29.60	.947	.941	.859
100	10	27.67	27.59	27.50	.949	.948	.909
200	10	26.33	26.32	26.31	.950	.950	.931
400	10	25.67	25.67	25.67	.950	.950	.941
800	10	25.36	25.34	25.34	.950	.950	.945
			$\alpha = 0.01$				
25	5	51.50	49.73	46.46	.987	.981	.844
30	5	46.64	45.48	43.72	.988	.984	.886
35	5	43.50	42.85	41.79	.989	.987	.910
40	5	41.57	41.03	40.35	.989	.987	.926
45	5	40.15	39.71	39.23	.989	.988	.938
50	5	38.96	38.69	38.35	.989	.989	.946
100	5	34.48	34.46	34.42	.990	.990	.974
200	5	32.54	32.49	32.49	.990	.990	.983
400	5	31.54	31.53	31.53	.990	.990	.987
800	5	31.05	31.05	31.05	.990	.990	.989
25	10	48.73	47.66	41.79	.988	.973	.869
30	10	44.52	43.82	40.35	.989	.980	.903
35	10	41.93	41.49	39.23	.989	.984	.924
40	10	40.23	39.91	38.35	.989	.986	.936
45	10	39.03	38.75	37.63	.989	.987	.945
50	10	38.09	37.87	37.03	.990	.987	.951
100	10	34.26	34.19	34.06	.990	.989	.975
200	10	32.42	32.41	32.39	.990	.990	.984
400	10	31.46	31.51	31.51	.990	.990	.987
800	10	31.11	31.05	31.05	.990	.990	.988

Note. $\widetilde{\text{CP}}_{\text{YS}\cdot\text{L2}} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L2}}(\alpha))$, $\widetilde{\text{CP}}_{\text{YS}\cdot F2} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot F2}(\alpha))$, $\widetilde{\text{CP}}_{\chi^2} = \widetilde{\text{CP}}(\chi^2_{p,\alpha})$, $\chi^2_{15,0.05} = 25.00$, $\chi^2_{15,0.01} = 30.58$.

Table 4: Simulated and approximate values and coverage probabilities when $(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}) = (20, 18, 16, 14, 12, 10, 8, 6, 4, 2)$

Sample Size		Upper Percentile			Coverage Probability		
$n_1^{(\ell)}$	$n_2^{(\ell)} = n_3^{(\ell)} = \cdots = n_{10}^{(\ell)}$	$\widetilde{t}^2_{\text{simu}\cdot 2}$	$t^2_{\text{YS}\cdot\text{L}2}$	$t^2_{\text{YS}\cdot F2}$	$\widetilde{\text{CP}}_{\text{YS}\cdot\text{L}2}$	$\widetilde{\text{CP}}_{\text{YS}\cdot F2}$	$\widetilde{\text{CP}}_{\chi^2}$
			$\alpha = 0.05$				
25	5	53.21	53.05	44.01	.949	.872	.589
30	5	47.92	47.61	42.42	.948	.902	.670
35	5	44.74	44.50	41.19	.948	.918	.721
40	5	42.78	42.47	40.20	.948	.926	.758
45	5	41.34	41.03	39.40	.947	.932	.786
50	5	40.19	39.94	38.72	.948	.936	.805
100	5	35.69	35.56	35.39	.949	.947	.887
200	5	33.56	33.51	33.49	.950	.949	.921
400	5	32.48	32.48	32.47	.950	.950	.936
800	5	31.95	31.95	31.95	.950	.950	.943
25	10	48.81	51.37	39.05	.963	.847	.661
30	10	44.62	46.14	38.43	.960	.884	.727
35	10	42.30	43.20	37.91	.956	.904	.767
40	10	40.72	41.31	37.46	.955	.916	.797
45	10	39.58	39.98	37.06	.953	.924	.818
50	10	38.77	39.00	36.72	.952	.929	.832
100	10	35.20	35.14	34.71	.949	.945	.895
200	10	33.41	33.36	33.29	.949	.949	.923
400	10	32.43	32.43	32.42	.950	.950	.937
800	10	31.97	31.94	31.93	.950	.950	.943
			$\alpha = 0.01$				
25	5	68.70	68.33	54.82	.990	.958	.760
30	5	60.40	60.18	52.59	.990	.972	.828
35	5	55.87	55.65	50.87	.990	.979	.868
40	5	53.09	52.74	49.50	.989	.982	.893
45	5	50.88	50.69	48.39	.990	.984	.910
50	5	49.28	49.17	47.46	.990	.986	.923
100	5	43.21	43.14	42.90	.990	.989	.966
200	5	40.43	40.37	40.34	.990	.990	.980
400	5	38.96	38.99	38.98	.990	.990	.986
800	5	38.36	38.28	38.28	.990	.990	.988
25	10	62.84	66.03	47.90	.993	.944	.820
30	10	55.76	58.17	47.06	.993	.965	.870
35	10	52.34	53.87	46.34	.992	.973	.899
40	10	50.17	51.16	45.72	.992	.978	.917
45	10	48.57	49.28	45.18	.991	.981	.930
50	10	47.40	47.90	44.71	.991	.983	.938
100	10	42.55	42.58	41.98	.990	.989	.970
200	10	40.24	40.17	40.07	.990	.990	.981
400	10	38.95	38.92	38.91	.990	.990	.986
800	10	38.36	38.26	38.26	.990	.990	.988

Note. $\widetilde{\text{CP}}_{\text{YS}\cdot\text{L}2} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L}2}(\alpha))$, $\widetilde{\text{CP}}_{\text{YS}\cdot F2} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot F2}(\alpha))$, $\widetilde{\text{CP}}_{\chi^2} = \widetilde{\text{CP}}(\chi^2_{p,\alpha})$, $\chi^2_{20,0.05} = 31.41$, $\chi^2_{20,0.01} = 37.57$.

Table 5: Simulated and approximate values and coverage probabilities for pairwise comparisons when $m = 6$ and $(p_1, p_2, p_3, p_4, p_5) = (15, 12, 9, 6, 3)$

Sample Size		Upper Percentile				Coverage Probability		
$n_1^{(\ell)}$	$n_2^{(\ell)} = n_3^{(\ell)} = \cdots = n_5^{(\ell)}$	$\widetilde{t}^2_{\text{simu}\cdot\text{p}}$	$\widetilde{t}^2_{\text{simu}\cdot\text{Bon}}$	$t^2_{\text{YS}\cdot\text{L6}}$	$t^2_{\text{YS}\cdot F6}$	$\widetilde{\text{CP}}_{\text{YS}\cdot\text{L6}}$	$\widetilde{\text{CP}}_{\text{YS}\cdot F6}$	$\widetilde{\text{CP}}_{\chi^2}$
			$\alpha = 0.05$					
25	5	39.86	40.62	40.48	39.78	.957	.949	.828
30	5	38.82	39.40	39.40	38.98	.957	.952	.855
35	5	38.04	38.56	38.65	38.37	.957	.954	.873
40	5	37.52	38.04	38.09	37.89	.957	.954	.886
45	5	37.08	37.65	37.65	37.51	.957	.955	.895
50	5	36.76	37.34	37.31	37.20	.957	.955	.902
100	5	35.24	35.77	35.73	35.71	.956	.956	.931
200	5	34.42	34.87	34.91	34.91	.957	.957	.945
400	5	34.04	34.55	34.49	34.49	.956	.956	.950
800	5	33.78	34.21	34.28	34.28	.957	.957	.954
25	10	38.92	39.50	39.77	38.37	.959	.943	.852
30	10	38.09	38.73	38.82	37.89	.958	.947	.872
35	10	37.48	38.02	38.16	37.51	.958	.950	.886
40	10	37.03	37.62	37.67	37.20	.958	.952	.896
45	10	36.72	37.18	37.30	36.94	.957	.953	.903
50	10	36.44	36.97	37.00	36.72	.957	.954	.909
100	10	35.10	35.65	35.62	35.57	.957	.956	.934
200	10	34.38	34.83	34.88	34.87	.957	.956	.945
400	10	33.98	34.49	34.48	34.48	.957	.957	.951
800	10	33.80	34.30	34.28	34.28	.957	.957	.954
			$\alpha = 0.01$					
25	5	46.58	46.98	46.88	45.99	.991	.988	.938
30	5	45.25	45.41	45.52	44.98	.991	.989	.951
35	5	44.31	44.29	44.57	44.22	.991	.990	.959
40	5	43.62	43.92	43.87	43.63	.991	.990	.965
45	5	43.06	43.36	43.33	43.15	.991	.990	.969
50	5	42.65	43.18	42.90	42.77	.991	.990	.971
100	5	40.73	41.01	40.94	40.92	.991	.991	.982
200	5	39.71	40.08	39.93	39.93	.991	.991	.987
400	5	39.23	39.55	39.41	39.41	.991	.991	.989
800	5	38.89	39.27	39.15	39.15	.991	.991	.990
25	10	45.36	45.85	45.99	44.22	.992	.987	.950
30	10	44.30	44.85	44.79	43.63	.991	.988	.959
35	10	43.54	43.61	43.96	43.15	.991	.989	.965
40	10	42.99	43.30	43.36	42.77	.991	.989	.969
45	10	42.65	42.75	42.89	42.44	.991	.989	.972
50	10	42.21	42.42	42.51	42.17	.991	.990	.974
100	10	40.61	40.76	40.80	40.74	.991	.990	.983
200	10	39.66	39.64	39.89	39.88	.991	.991	.987
400	10	39.14	39.44	39.40	39.40	.991	.991	.989
800	10	38.92	39.01	39.15	39.15	.991	.991	.990

Note. $\widetilde{\text{CP}}_{\text{YS}\cdot\text{L6}} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L6}}(\alpha_\text{p}))$, $\widetilde{\text{CP}}_{\text{YS}\cdot F6} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot F6}(\alpha_\text{p}))$, $\widetilde{\text{CP}}_{\chi^2} = \widetilde{\text{CP}}(\chi^2_{p,\alpha_\text{p}})$, $\alpha_\text{p} = \alpha/15$, $\chi^2_{15,0.05/15} = 34.07$, $\chi^2_{15,0.01/15} = 38.89$.

Table 6: Simulated and approximate values and coverage probabilities for pairwise comparisons when $m = 10$ and $(p_1, p_2, p_3, p_4, p_5) = (15, 12, 9, 6, 3)$

Sample Size		Upper Percentile				Coverage Probability		
$n_1^{(\ell)}$	$n_2^{(\ell)} = n_3^{(\ell)} = \cdots = n_5^{(\ell)}$	$\widetilde{t}^2_{\text{simu}\cdot\text{p}}$	$\widetilde{t}^2_{\text{simu}\cdot\text{Bon}}$	$t^2_{\text{YS}\cdot\text{L10}}$	$t^2_{\text{YS}\cdot F10}$	$\widetilde{\text{CP}}_{\text{YS}\cdot\text{L10}}$	$\widetilde{\text{CP}}_{\text{YS}\cdot F10}$	$\widetilde{\text{CP}}_{\chi^2}$
				$\alpha = 0.05$				
25	5	41.42	41.91	42.14	41.66	.959	.953	.865
30	5	40.74	41.44	41.38	41.08	.958	.954	.883
35	5	40.21	40.74	40.84	40.64	.958	.955	.896
40	5	39.78	40.36	40.43	40.29	.958	.956	.905
45	5	39.52	40.26	40.11	40.01	.958	.956	.911
50	5	39.26	39.94	39.85	39.78	.958	.957	.917
100	5	38.14	38.82	38.67	38.66	.957	.957	.938
200	5	37.49	37.94	38.04	38.04	.958	.958	.948
400	5	37.21	37.59	37.72	37.72	.957	.957	.953
800	5	37.06	37.66	37.55	37.55	.957	.957	.955
25	10	40.72	41.30	41.61	40.64	.960	.949	.884
30	10	40.16	40.75	40.94	40.29	.960	.952	.897
35	10	39.73	40.30	40.47	40.01	.959	.954	.907
40	10	39.42	39.97	40.12	39.78	.959	.955	.913
45	10	39.20	39.81	39.84	39.58	.958	.955	.918
50	10	38.98	39.63	39.62	39.42	.958	.956	.922
100	10	38.05	38.43	38.58	38.55	.957	.957	.939
200	10	37.48	38.05	38.02	38.01	.957	.957	.949
400	10	37.18	37.71	37.71	37.71	.957	.957	.953
800	10	37.06	37.63	37.55	37.55	.957	.957	.955
				$\alpha = 0.01$				
25	5	47.37	47.72	47.75	47.18	.991	.989	.957
30	5	46.47	46.93	46.83	46.47	.991	.990	.965
35	5	45.87	46.13	46.18	45.94	.991	.990	.970
40	5	45.36	45.40	45.69	45.52	.991	.990	.973
45	5	45.05	45.41	45.30	45.18	.991	.990	.975
50	5	44.75	45.20	44.99	44.90	.991	.990	.977
100	5	43.34	43.74	43.57	43.55	.991	.991	.985
200	5	42.60	42.50	42.82	42.82	.991	.991	.988
400	5	42.18	42.55	42.43	42.43	.991	.991	.989
800	5	42.00	42.48	42.23	42.23	.991	.991	.990
25	10	46.47	46.82	47.12	45.94	.992	.988	.965
30	10	45.77	46.13	46.31	45.52	.992	.989	.970
35	10	45.30	45.62	45.74	45.18	.991	.990	.974
40	10	44.91	45.17	45.31	44.90	.991	.990	.976
45	10	44.61	44.81	44.98	44.67	.991	.990	.978
50	10	44.43	44.87	44.71	44.47	.991	.990	.979
100	10	43.24	43.29	43.46	43.42	.991	.991	.985
200	10	42.59	43.27	42.79	42.78	.991	.991	.988
400	10	42.17	42.29	42.42	42.42	.991	.991	.990
800	10	41.96	42.11	42.23	42.23	.991	.991	.990

Note. $\widetilde{\text{CP}}_{\text{YS}\cdot\text{L10}} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L10}}(\alpha_\text{p}))$, $\widetilde{\text{CP}}_{\text{YS}\cdot F10} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot F10}(\alpha_\text{p}))$, $\widetilde{\text{CP}}_{\chi^2} = \widetilde{\text{CP}}(\chi^2_{p,\alpha_\text{p}})$, $\alpha_\text{p} = \alpha/45$, $\chi^2_{15,0.05/45} = 37.39$, $\chi^2_{15,0.01/45} = 42.03$.

Table 7: Simulated and approximate values and coverage probabilities for comparisons with a control when $m = 6$ and $(p_1, p_2, p_3, p_4, p_5) = (15, 12, 9, 6, 3)$

Sample Size		Upper Percentile				Coverage Probability		
$n_1^{(\ell)}$	$n_2^{(\ell)} = n_3^{(\ell)} = \cdots = n_5^{(\ell)}$	$\widetilde{t}^2_{\text{simu}\cdot\text{c}}$	$\widetilde{t}^2_{\text{simu}\cdot\text{Bon}}$	$t^2_{\text{YS}\cdot\text{L6}}$	$t^2_{\text{YS}\cdot F6}$	$\widetilde{\text{CP}}_{\text{YS}\cdot\text{L6}}$	$\widetilde{\text{CP}}_{\text{YS}\cdot F6}$	$\widetilde{\text{CP}}_{\chi^2}$
			$\alpha = 0.05$					
25	5	35.54	36.06	35.96	35.38	.955	.948	.857
30	5	34.59	35.15	35.06	34.71	.955	.951	.878
35	5	33.96	34.47	34.43	34.20	.955	.953	.891
40	5	33.54	33.99	33.96	33.80	.955	.953	.900
45	5	33.19	33.66	33.60	33.48	.955	.954	.906
50	5	32.89	33.35	33.31	33.22	.955	.954	.912
100	5	31.60	32.03	31.98	31.97	.955	.955	.935
200	5	30.86	31.29	31.29	31.29	.955	.955	.946
400	5	30.54	30.89	30.94	30.94	.955	.955	.951
800	5	30.34	30.74	30.76	30.76	.956	.956	.953
25	10	34.72	35.20	35.36	34.20	.957	.944	.875
30	10	34.04	34.55	34.57	33.80	.956	.947	.890
35	10	33.52	33.97	34.02	33.48	.956	.949	.900
40	10	33.11	33.62	33.61	33.22	.956	.951	.908
45	10	32.82	33.30	33.30	33.00	.956	.952	.913
50	10	32.62	33.06	33.05	32.82	.955	.953	.917
100	10	31.46	31.85	31.89	31.85	.955	.955	.937
200	10	30.85	31.29	31.26	31.26	.955	.955	.946
400	10	30.53	30.94	30.93	30.93	.955	.955	.951
800	10	30.37	30.75	30.76	30.76	.955	.955	.953
			$\alpha = 0.01$					
25	5	42.34	42.64	42.53	41.78	.990	.989	.951
30	5	41.16	41.50	41.37	40.91	.991	.989	.961
35	5	40.31	40.55	40.55	40.25	.991	.990	.967
40	5	39.74	39.96	39.95	39.74	.991	.990	.970
45	5	39.35	39.44	39.48	39.33	.990	.990	.973
50	5	38.91	39.17	39.11	39.00	.991	.990	.976
100	5	37.31	37.42	37.41	37.40	.990	.990	.984
200	5	36.43	36.62	36.54	36.53	.990	.990	.987
400	5	35.93	36.04	36.09	36.09	.991	.991	.989
800	5	35.63	35.84	35.86	35.86	.991	.991	.990
25	10	41.37	41.54	41.76	40.25	.991	.987	.960
30	10	40.45	40.70	40.74	39.74	.991	.988	.966
35	10	39.73	39.82	40.03	39.33	.991	.989	.971
40	10	39.17	39.34	39.50	39.00	.991	.989	.974
45	10	38.84	39.08	39.10	38.72	.991	.990	.976
50	10	38.60	38.79	38.77	38.48	.991	.990	.977
100	10	37.13	37.30	37.29	37.24	.991	.990	.984
200	10	36.34	36.46	36.50	36.49	.990	.990	.987
400	10	35.95	36.13	36.08	36.08	.990	.990	.989
800	10	35.65	35.82	35.86	35.86	.991	.991	.990

Note. $\widetilde{\text{CP}}_{\text{YS}\cdot\text{L6}} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot\text{L6}}(\alpha_\text{c}))$, $\widetilde{\text{CP}}_{\text{YS}\cdot F6} = \widetilde{\text{CP}}(t^2_{\text{YS}\cdot F6}(\alpha_\text{c}))$, $\widetilde{\text{CP}}_{\chi^2} = \widetilde{\text{CP}}(\chi^2_{p,\alpha_\text{c}})$, $\alpha_\text{c} = \alpha/5$, $\chi^2_{15,0.05/5} = 30.58$, $\chi^2_{15,0.01/5} = 35.63$.

Table 8: Simulated and approximate values and coverage probabilities for comparisons with a control when $m = 10$ and $(p_1, p_2, p_3, p_4, p_5) = (15, 12, 9, 6, 3)$

Sample Size		Upper Percentile				Coverage Probability		
$n_1^{(\ell)}$	$n_2^{(\ell)} = n_3^{(\ell)} = \cdots = n_5^{(\ell)}$	$\widetilde{t}^2_{\mathrm{simu}\cdot\mathrm{c}}$	$\widetilde{t}^2_{\mathrm{simu}\cdot\mathrm{Bon}}$	$t^2_{\mathrm{YS}\cdot\mathrm{L10}}$	$t^2_{\mathrm{YS}\cdot F10}$	$\widetilde{\mathrm{CP}}_{\mathrm{YS}\cdot\mathrm{L10}}$	$\widetilde{\mathrm{CP}}_{\mathrm{YS}\cdot F10}$	$\widetilde{\mathrm{CP}}_{\chi^2}$
			$\alpha = 0.05$					
25	5	35.64	36.24	36.29	35.91	.957	.953	.894
30	5	35.07	35.71	35.68	35.44	.957	.954	.906
35	5	34.65	35.28	35.25	35.09	.957	.955	.914
40	5	34.33	34.88	34.92	34.81	.957	.956	.920
45	5	34.07	34.66	34.66	34.58	.957	.956	.925
50	5	33.92	34.54	34.46	34.40	.957	.956	.927
100	5	32.96	33.52	33.50	33.49	.957	.957	.943
200	5	32.43	32.98	33.00	33.00	.958	.958	.951
400	5	32.19	32.72	32.74	32.74	.957	.957	.954
800	5	32.10	32.65	32.60	32.60	.957	.957	.955
25	10	35.06	35.63	35.86	35.09	.959	.950	.906
30	10	34.59	35.23	35.33	34.81	.959	.953	.915
35	10	34.32	34.91	34.95	34.58	.958	.953	.920
40	10	34.05	34.67	34.67	34.40	.957	.954	.925
45	10	33.84	34.43	34.45	34.24	.957	.955	.928
50	10	33.67	34.25	34.27	34.11	.957	.956	.931
100	10	32.87	33.40	33.43	33.41	.957	.957	.944
200	10	32.44	32.99	32.98	32.97	.957	.957	.950
400	10	32.23	32.76	32.73	32.73	.957	.957	.953
800	10	32.07	32.64	32.60	32.60	.957	.957	.955
			$\alpha = 0.01$					
25	5	41.78	41.96	42.14	41.66	.991	.990	.968
30	5	41.19	41.60	41.38	41.08	.991	.990	.972
35	5	40.57	40.91	40.84	40.64	.991	.990	.976
40	5	40.12	40.33	40.43	40.29	.991	.990	.978
45	5	39.85	39.97	40.11	40.01	.991	.990	.980
50	5	39.60	39.88	39.85	39.78	.991	.990	.981
100	5	38.45	38.67	38.67	38.66	.991	.991	.986
200	5	37.78	38.02	38.04	38.04	.991	.991	.989
400	5	37.45	37.82	37.72	37.72	.991	.991	.990
800	5	37.34	37.44	37.55	37.55	.991	.991	.990
25	10	41.07	41.21	41.61	40.64	.991	.989	.973
30	10	40.48	40.76	40.94	40.29	.991	.989	.976
35	10	40.12	40.32	40.47	40.01	.991	.990	.978
40	10	39.82	39.98	40.12	39.78	.991	.990	.980
45	10	39.52	39.76	39.84	39.58	.991	.990	.981
50	10	39.32	39.43	39.62	39.42	.991	.990	.982
100	10	38.36	38.64	38.58	38.55	.991	.991	.987
200	10	37.79	38.04	38.02	38.01	.991	.991	.989
400	10	37.51	37.69	37.71	37.71	.991	.991	.990
800	10	37.30	37.62	37.55	37.55	.991	.991	.990

Note. $\widetilde{\mathrm{CP}}_{\mathrm{YS}\cdot\mathrm{L10}} = \widetilde{\mathrm{CP}}(t^2_{\mathrm{YS}\cdot\mathrm{L10}}(\alpha_\mathrm{c}))$, $\widetilde{\mathrm{CP}}_{\mathrm{YS}\cdot F10} = \widetilde{\mathrm{CP}}(t^2_{\mathrm{YS}\cdot F10}(\alpha_\mathrm{c}))$, $\widetilde{\mathrm{CP}}_{\chi^2} = \widetilde{\mathrm{CP}}(\chi^2_{p,\alpha_\mathrm{c}})$, $\alpha_\mathrm{c} = \alpha/9$, $\chi^2_{15,0.05/9} = 32.47$, $\chi^2_{15,0.01/9} = 37.39$.

Acknowledgments

The second author's research was in part supported by Grant-in-Aid for Scientific Research (C) (26330050).

References

[1] Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, **52**, 200–203.

[2] Bhargava, R. (1962). Multivariate tests of hypotheses with incomplete data. *Technical report No.3, Applied Mathematics and Statistics Laboratories, Stanford University*, Stanford, California.

[3] Chang, W. -Y. and Richards, D. St. P. (2009). Finite-sample inference with monotone incomplete multivariate normal data, I. *Journal of Multivariate Analysis*, **100**, 1883–1899.

[4] Jinadasa, K. G. and Tracy, D. S. (1992). Maximum likelihood estimation for multivariate normal distribution with monotone sample. *Communications in Statistics – Theory and Methods*, **21**, 41–50.

[5] Kanda, T. and Fujikoshi, Y. (1998). Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data. *American Journal of Mathematical and Management Sciences*, **18**, 161–190.

[6] Krishnamoorthy, K. and Pannala, M. K. (1999). Confidence estimation of a normal mean vector with incomplete data. *The Canadian Journal of Statistics*, **27**, 395–407.

[7] Rao, C. R. (1956). Analysis of dispersion with incomplete observations on one of the characters. *Journal of the Royal Statistical Society: Series B*, **18**, 259–264.

[8] Seko, N. (2012). Tests for mean vectors with two-step monotone missing data for the k-sample problem. *SUT Journal of Mathematics*, **48**, 213–229.

[9] Seko, N., Kawasaki, T. and Seo, T. (2011). Testing equality of two mean vectors with two-step monotone missing data. *American Journal of Mathematical and Management Sciences*, **31**, 117–135.

[10] Seko, N., Yamazaki, A. and Seo, T. (2012). Tests for mean vector with two-step monotone missing data. *SUT Journal of Mathematics*, **48**, 13–36.

[11] Seo, T., Mano, S. and Fujikoshi, Y. (1994). A generalized Tukey conjecture for multiple comparisons among mean vectors. *Journal of the American Statistical Association*, **89**, 676–679.

[12] Seo, T. and Siotani, M. (1992). The multivariate Studentized range and its upper percentiles. *Journal of the Japan Statistical Society*, **22**, 123–137.

[13] Yagi, A. and Seo, T. (2014). A test for mean vector and simultaneous confidence intervals with three-step monotone missing data. *American Journal of Mathematical and Management Sciences*, **33**, 161–175.

[14] Yagi, A. and Seo, T. (2015a). Tests for mean vectors with two-step and three-step monotone samples. *Josai Mathematical Monographs*, **8**, 49–71.

[15] Yagi, A. and Seo, T. (2015b). Tests for equality of mean vectors and simultaneous confidence intervals with two-step or three-step monotone missing data patterns. to appear in *American Journal of Mathematical and Management Sciences*.

[16] Yu, J., Krishnamoorthy, K. and Pannala, K. M. (2006). Two-sample inference for normal mean vectors based on monotone missing data. *Journal of Multivariate Analysis*, **97**, 2162–2176.