Testing equality of mean vectors based on three-step monotone missing 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 paper, we consider the problem of testing equality of two mean vectors when the data have three-step monotone pattern missing observations. We propose an approximate upper percentile of the Hotelling's T^2 type statistic where each of the data set has a three-step monotone missing data and the population covariance matrices are equal. This result is an extension of one sample problem given in Yagi and Seo (2014). Further, we consider multivariate multiple comparisons procedure of pairwise differences of mean vectors with three-step monotone missing data. Approximate simultaneous confidence intervals for pairwise multiple comparisons among mean vectors are obtained by using Bonferroni's approximate upper percentiles of $T^2_{\max}$ type statistic. Finally, the accuracy of the approximation is investigated by Monte Carlo simulation.

Key Words and Phrases: T^2 type statistic; Maximum likelihood estimator; Three-step monotone missing data; Simultaneous confidence intervals; Two-sample problem

1. Introduction

Consider the problem of testing for mean vectors with a three-step monotone missing data. Suppose that we have a complete data set for n_1 observations with p_1 dimensions and two incomplete data sets which have n_2 observations with p_2 dimensions and n_3 observations with p_3 dimensions. Such a data set is called a three-step monotone missing data pattern. The case when the missing observations are of the monotone type has been considered by many authors (see Rao (1956), Anderson (1957) and Bhargava (1962) among others). The closed form expressions for the maximum likelihood estimators (MLEs) of the mean vector and the covariance matrix in the case of the 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 the k-step monotone missing data. Yagi and Seo (2014) gave a simplified Hotelling's T^2 type statistic for one-sample problem and its approximate upper percentiles when the data have a three-step monotone missing data. This paper extends one-sample problem in Yagi and Seo (2014) to two-sample or m-sample problem in the case of the three-step monotone missing data. That is, using their idea, we develop an approximate upper percentile of T^2 statistic for two-sample problem. At first, in this paper, we consider the problem of testing $H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$ vs. $H_1 : \boldsymbol{\mu}^{(1)} \neq \boldsymbol{\mu}^{(2)}$ when the data have monotone pattern missing observations. In the case of a two-step monotone missing data, Seko, Kawasaki and Seo (2011) derived a Hotelling's T^2 type statistic and the likelihood ratio test statistic and their approximate upper percentiles. Also, Yu, Krishnamoorthy and Pannala (2006) discussed the Hotelling's T^2 type statistic and its approximate distribution by another approach. Recently, Seko (2012) discussed the tests for mean vectors with two-step monotone missing data for m-sample problem. Under two-sample or m-sample problem, in this paper, we propose a simplified Hotelling's T^2 type statistic and its approximate upper percentile in the case of a three-step monotone missing data, similar to that in the case of a two-step monotone missing data.

The remainder of this paper is organized as follows. In Section 2, we present the some notations as preliminaries and the MLEs of the mean vector and the covariance matrix for m-sample problem that includes two-sample problem. In Section 3, we give the Hotelling's T^2 type statistic for testing the equality of two mean vectors and its approximate upper percentiles. In Section 4, we discuss the simultaneous confidence intervals for all pairwise differences of mean vectors. In order to obtain the simultaneous confidence intervals, we derive an approximate upper percentile of $T^2_{\max}$ type statistic by Bonferroni's approximation method. In Section 5, we also give some simulation results.

2. Three-step monotone missing data and MLE

As preliminaries, we present some notations on vector and matrix needed to express the three-step monotone missing data for two-sample or a general m-sample problem. Let $\boldsymbol{x}$ be distributed as $N_p(\boldsymbol{\mu}, \Sigma)$, and let $\boldsymbol{x}_i = (\boldsymbol{x})_i$ be the vector of the first p_i elements of $\boldsymbol{x}$. Then, $\boldsymbol{x}_i (= (x_1, x_2, \ldots, x_{p_i})')$ is distributed as $N_{p_i}(\boldsymbol{\mu}_i, \Sigma_i)$, $i = 1, 2, 3$, where $\boldsymbol{\mu}_i = (\boldsymbol{\mu})_i = (\mu_1, \ldots, \mu_{p_i})'$ and Σ_i is the $p_i \times p_i$ principal submatrix of $\Sigma (= \Sigma_1)$. Further, let $(\Sigma_i)_j$ be the principal submatrix of Σ_i of order $p_j \times p_j$, $1 \leq i < j \leq 3$. As the notations for the covariance matrix, we define

$$\Sigma_{i+1} = (\Sigma_1)_{i+1},\ \Sigma_1 = \Sigma = \begin{pmatrix} \Sigma_{i+1} & \Sigma_{i+1,2} \\ \Sigma'_{i+1,2} & \Sigma_{i+1,3} \end{pmatrix}$$

and

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

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

Let $\boldsymbol{x}_{i1}^{(\ell)}, \ldots, \boldsymbol{x}_{in_i^{(\ell)}}^{(\ell)}$ be distributed as $N_{p_i}(\boldsymbol{\mu}_i^{(\ell)}, \Sigma_i)$ for $i = 1, 2, 3$ and $\ell = 1, 2, \ldots, m$, where $\boldsymbol{\mu}_i^{(\ell)} = (\mu_1^{(\ell)}, \mu_2^{(\ell)}, \ldots, \mu_{p_i}^{(\ell)})'$ and Σ_i is $p_i \times p_i$ covariance matrix, where $p = p_1 > p_2 > p_3 > 0$ and $n_1^{(\ell)} > p$. Let

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

$$E_i^{(\ell)} = \sum_{j=1}^{n_i^{(\ell)}} (\boldsymbol{x}_{ij}^{(\ell)} - \overline{\boldsymbol{x}}_i^{(\ell)})(\boldsymbol{x}_{ij}^{(\ell)} - \overline{\boldsymbol{x}}_i^{(\ell)})',\ i = 1, 2, 3.$$

Then we have the following theorem.

Theorem 1. *Let $\boldsymbol{x}_{ij}^{(\ell)}$, $i = 1, 2, 3$, $j = 1, 2, \ldots, n_i^{(\ell)}$, $\ell = 1, 2, \ldots, m$ be the j-th random vector from ℓ-th population distributed as $N_{p_i}(\boldsymbol{\mu}_i^{(\ell)}, \Sigma_i)$. Then the MLEs of $\boldsymbol{\mu}_i^{(\ell)}$, $\ell = 1, 2, \ldots, m$ and Σ are given by*

$$\widehat{\boldsymbol{\mu}}^{(\ell)} = \overline{\boldsymbol{x}}_1^{(\ell)} + \widehat{T}_2^{[pl]} \boldsymbol{d}_2^{(\ell)} + \widehat{T}_2^{[pl]} \widehat{T}_3^{[pl]} \boldsymbol{d}_3^{(\ell)},$$

$$\widehat{\Sigma}^{[pl]} = \frac{1}{M_2}\sum_{\ell=1}^{m} E_1^{(\ell)}$$
$$+ \frac{1}{M_3}\left[\sum_{\ell=1}^{m} G_2^{(\ell)}\left\{E_2^{(\ell)} + \frac{N_2^{(\ell)}N_3^{(\ell)}}{n_2^{(\ell)}}\boldsymbol{d}_2^{(\ell)}\boldsymbol{d}_2^{(\ell)\prime} - \frac{n_2^{(\ell)}}{N_2^{(\ell)}}L_{11}^{(\ell)}\right\}G_2^{(\ell)\prime}\right]$$
$$+ \frac{1}{M_4}\left[\sum_{\ell=1}^{m} G_2^{(\ell)}G_3^{(\ell)}\left\{E_3^{(\ell)} + \frac{N_3^{(\ell)}N_4^{(\ell)}}{n_3^{(\ell)}}\boldsymbol{d}_3^{(\ell)}\boldsymbol{d}_3^{(\ell)\prime} - \frac{n_3^{(\ell)}}{N_3^{(\ell)}}L_{21}^{(\ell)}\right\}G_3^{(\ell)\prime}G_2^{(\ell)\prime}\right],$$

where

$$\boldsymbol{d}_2^{(\ell)} = \frac{n_2^{(\ell)}}{N_3^{(\ell)}}\left\{\overline{\boldsymbol{x}}_2^{(\ell)} - (\overline{\boldsymbol{x}}_1^{(\ell)})_2\right\},\ \boldsymbol{d}_3^{(\ell)} = \frac{n_3^{(\ell)}}{N_4^{(\ell)}}\left[\overline{\boldsymbol{x}}_3^{(\ell)} - \frac{1}{N_3^{(\ell)}}\left\{n_1^{(\ell)}(\overline{\boldsymbol{x}}_1^{(\ell)})_3 + n_2^{(\ell)}(\overline{\boldsymbol{x}}_2^{(\ell)})_3\right\}\right],$$

$$N_{i+1}^{(\ell)} = \sum_{j=1}^{i} n_j^{(\ell)},\ M_r = \sum_{\ell=1}^{m} N_r^{(\ell)},\ r = 2, 3, 4,$$

$$\widehat{T}_2^{[pl]} = \begin{pmatrix} I_{p_2} \\ (\widehat{\Sigma}_{(1,2)}^{[pl]})'(\widehat{\Sigma}_2^{[pl]})^{-1} \end{pmatrix},\ \widehat{T}_3^{[pl]} = \begin{pmatrix} I_{p_3} \\ (\widehat{\Sigma}_{(2,2)}^{[pl]})'(\widehat{\Sigma}_3^{[pl]})^{-1} \end{pmatrix},$$

$$G_2^{(\ell)} = \begin{pmatrix} I_{p_2} \\ L_{12}^{(\ell)\prime}(L_{11}^{(\ell)})^{-1} \end{pmatrix},\ G_3^{(\ell)} = \begin{pmatrix} I_{p_3} \\ L_{22}^{(\ell)\prime}(L_{21}^{(\ell)})^{-1} \end{pmatrix},$$

$$L_1^{(\ell)} = E_1^{(\ell)},\ L_2^{(\ell)} = L_{11}^{(\ell)} + E_2^{(\ell)} + \frac{N_2^{(\ell)}N_3^{(\ell)}}{n_2^{(\ell)}}\boldsymbol{d}_2^{(\ell)}\boldsymbol{d}_2^{(\ell)\prime},$$

$$L_i^{(\ell)} = \begin{pmatrix} L_{i1}^{(\ell)} & L_{i2}^{(\ell)} \\ L_{i2}^{(\ell)\prime} & L_{i3}^{(\ell)} \end{pmatrix},\ i = 1, 2.$$

We note that the result of Theorem 1 is obtained in a straight forward method using one-sample problem case of Yagi and Seo (2014). In the next section, in order to obtain a Hotelling's T^2 type statistic for testing equality of two mean vectors, we give the covariance matrix of $\widetilde{\boldsymbol{\mu}}^{(1)} - \widetilde{\boldsymbol{\mu}}^{(2)}$, where $\widetilde{\boldsymbol{\mu}}^{(\ell)} = \overline{\boldsymbol{x}}_1^{(\ell)} + T_2\boldsymbol{d}_2^{(\ell)} + T_2T_3\boldsymbol{d}_3^{(\ell)}$. We note that the Hotelling's T^2 type statistic with $\widehat{\mathrm{Cov}}(\widetilde{\boldsymbol{\mu}}^{(1)} - \widetilde{\boldsymbol{\mu}}^{(2)})$ instead of $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})$ is adopted since $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})$ is complicated and $\widehat{\mathrm{Cov}}(\widetilde{\boldsymbol{\mu}}^{(1)} - \widetilde{\boldsymbol{\mu}}^{(2)})$ is asymptotically equivalent to $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})$. For the related idea for the case of one-sample problem, see Yagi and Seo (2014).

3. Hotelling's T^2 type statistic

In this section, we consider the hypothesis test, $H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$ vs. $H_1 : \boldsymbol{\mu}^{(1)} \neq \boldsymbol{\mu}^{(2)}$ when the data have a three-step monotone missing data pattern. In order to test the hypothesis H_0, under the assumption of common population covariance matrix, we adopt the Hotelling's T^2 type statistic given by

$$T^2_{\mathrm{YS}} = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})' \, \widetilde{\Gamma}^{-1}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}),$$

where $\widehat{\boldsymbol{\mu}}^{(\ell)} = \overline{\boldsymbol{x}}_1^{(\ell)} + \widehat{T}_2^{[pl]}\boldsymbol{d}_2^{(\ell)} + \widehat{T}_2^{[pl]}\widehat{T}_3^{[pl]}\boldsymbol{d}_3^{(\ell)},\ \ell = 1, 2$, and $\widetilde{\Gamma}$ is an estimator of $\Gamma = \mathrm{Cov}(\widetilde{\boldsymbol{\mu}}^{(1)} - \widetilde{\boldsymbol{\mu}}^{(2)})$, $\widetilde{\boldsymbol{\mu}}^{(\ell)} = \overline{\boldsymbol{x}}_1^{(\ell)} + T_2\boldsymbol{d}_2^{(\ell)} + T_2T_3\boldsymbol{d}_3^{(\ell)},\ \ell = 1, 2$. Then, using the result for the one-sample problem in Yagi and Seo (2014), we have

$$\widetilde{\Gamma} = \left(\sum_{\ell=1}^{2} \frac{1}{N_2^{(\ell)}}\right) \widehat{\Sigma}_1^{[pl]} - \left(\sum_{\ell=1}^{2} \frac{n_2^{(\ell)}}{N_2^{(\ell)}N_3^{(\ell)}}\right) \widehat{U}^{[pl]} - \left(\sum_{\ell=1}^{2} \frac{\{(n_1^{(\ell)})^2 - n_1^{(\ell)}n_2^{(\ell)} + 2(n_2^{(\ell)})^2\}n_3^{(\ell)}}{N_2^{(\ell)}(N_3^{(\ell)})^2N_4^{(\ell)}}\right) \widehat{V}^{[pl]},$$

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

$$\widehat{U}^{[pl]} = \begin{pmatrix} \widehat{\Sigma}_2^{[pl]} \\ (\widehat{\Sigma}_{22}^{[pl]})' \end{pmatrix} (\widehat{T}_2^{[pl]})', \ \widehat{V}^{[pl]} = \begin{pmatrix} \widehat{\Sigma}_3^{[pl]} \\ (\widehat{\Sigma}_{32}^{[pl]})' \end{pmatrix} (\widehat{T}_3^{[pl]})'(\widehat{T}_2^{[pl]})'.$$

We note that, under H_0, the T^2_{YS} is asymptotically distributed as a χ^2 distribution with p degrees of freedom when $n_1^{(\ell)}, N_4^{(\ell)} \to \infty$ with $n_1^{(\ell)}/N_4^{(\ell)} \to \delta^{(\ell)} \in (0, 1],\ \ell = 1, 2$. However, it is noted that χ^2 approximation is not a good approximate upper percentile of the T^2_{YS} when the sample size is not large. In order to give a good approximation even for small sample, we have the following theorem.

Theorem 2. *Suppose that two data sets have the same three-step monotone missing data pattern. Then, the two approximate upper 100α percentile of the T^2_{YS} statistic is given by*

$$t^2_{\mathrm{YS}\cdot\mathrm{MLE}}(\alpha) = T^2_{\mathrm{MLE}\cdot n,\alpha} - \frac{(n_2^{(1)} + n_2^{(2)})p_2 + (n_3^{(1)} + n_3^{(2)})p_3}{(n_2^{(1)} + n_2^{(2)} + n_3^{(1)} + n_3^{(2)})p_1} \left(T^2_{\mathrm{MLE}\cdot n,\alpha} - T^2_{\mathrm{MLE}\cdot N,\alpha}\right),$$
$$t^2_{\mathrm{YS}\cdot\mathrm{UNB}}(\alpha) = T^2_{\mathrm{UNB}\cdot n,\alpha} - \frac{(n_2^{(1)} + n_2^{(2)})p_2 + (n_3^{(1)} + n_3^{(2)})p_3}{(n_2^{(1)} + n_2^{(2)} + n_3^{(1)} + n_3^{(2)})p_1} \left(T^2_{\mathrm{UNB}\cdot n,\alpha} - T^2_{\mathrm{UNB}\cdot N,\alpha}\right),$$

where

$$T^2_{\mathrm{MLE}\cdot n,\alpha} = \frac{np_1}{n - p_1 - 1}F_{p_1, n-p_1-1,\alpha},\ T^2_{\mathrm{MLE}\cdot N,\alpha} = \frac{Np_1}{N - p_1 - 1}F_{p_1, N-p_1-1,\alpha},$$
$$T^2_{\mathrm{UNB}\cdot n,\alpha} = \frac{(n-2)p_1}{n - p_1 - 1}F_{p_1, n-p_1-1,\alpha},\ T^2_{\mathrm{UNB}\cdot N,\alpha} = \frac{(N-2)p_1}{N - p_1 - 1}F_{p_1, N-p_1-1,\alpha},$$

$$n = n_1^{(1)} + n_1^{(2)}, \ N = N_4^{(1)} + N_4^{(2)},$$

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

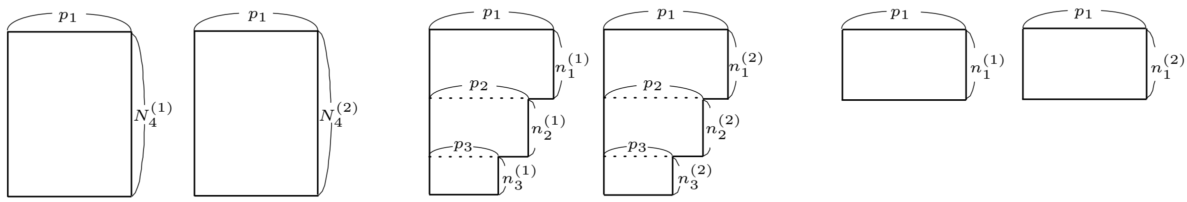

Figure 1: Approximation for the upper percentiles of $T^2_{\rm YS}$

As in Figure 1, $T^2_{{\rm MLE}\cdot N,\alpha}$ (or $T^2_{{\rm UNB}\cdot N,\alpha}$) and $T^2_{{\rm MLE}\cdot n,\alpha}$ (or $T^2_{{\rm UNB}\cdot n,\alpha}$) are calculated from complete data sets of left hand side $((N_4^{(1)} \times p_1) + (N_4^{(2)} \times p_1))$ and complete data sets of right hand side $((n_1^{(1)} \times p_1) + (n_1^{(2)} \times p_1))$, respectively. We note from Figure 1 that the upper percentiles of $T^2_{\rm YS}$ may be approximately between $T^2_{{\rm MLE}\cdot N,\alpha}$ (or $T^2_{{\rm UNB}\cdot N,\alpha}$) and $T^2_{{\rm MLE}\cdot n,\alpha}$ (or $T^2_{{\rm UNB}\cdot n,\alpha}$). Then, we propose $t^2_{{\rm YS}\cdot{\rm MLE}}(\alpha)$ and $t^2_{{\rm YS}\cdot{\rm UNB}}(\alpha)$ as approximate upper percentiles of $T^2_{\rm YS}$ using the linear interpolations for the coordinates $((n_1^{(1)} + n_1^{(2)})p_1, T^2_{{\rm MLE}\cdot n,\alpha})$ and $((N_4^{(1)} + N_4^{(2)})p_1, T^2_{{\rm MLE}\cdot N,\alpha})$ and for the coordinates $((n_1^{(1)} + n_1^{(2)})p_1, T^2_{{\rm UNB}\cdot n,\alpha})$ and $((N_4^{(1)} + N_4^{(2)})p_1, T^2_{{\rm UNB}\cdot N,\alpha})$, respectively. This approach is essentially based on the ones given in Yagi and Seo (2014) and Seko, Kawasaki and Seo (2011). In addition, using the idea in Seko, Kawasaki and Seo (2011), we can propose another approximations, which have a slightly different coefficient, are given by

$$\begin{aligned} t^2_{{\rm S}\cdot{\rm MLE}}(\alpha) =& T^2_{{\rm MLE}\cdot n,\alpha} \\ & - \left\{ 1 - \frac{(n_2^{(1)} + n_2^{(2)})(p_1 - p_2) + (n_3^{(1)} + n_3^{(2)})(p_1 - p_3)}{Np_1} \right\} \left(T^2_{{\rm MLE}\cdot n,\alpha} - T^2_{{\rm MLE}\cdot N,\alpha} \right), \\ t^2_{{\rm S}\cdot{\rm UNB}}(\alpha) =& T^2_{{\rm UNB}\cdot n,\alpha} \\ & - \left\{ 1 - \frac{(n_2^{(1)} + n_2^{(2)})(p_1 - p_2) + (n_3^{(1)} + n_3^{(2)})(p_1 - p_3)}{Np_1} \right\} \left(T^2_{{\rm UNB}\cdot n,\alpha} - T^2_{{\rm UNB}\cdot N,\alpha} \right). \end{aligned}$$

Further, as another approach by adjusting the degrees of freedom of F distribution, we can propose two approximate upper percentiles of $T^2_{\rm YS}$ given by

$$\begin{aligned} t^2_{F\cdot{\rm MLE}}(\alpha) &= \frac{n^* p_1}{n^* - p_1 - 1} F_{p_1, n^* - p_1 - 1, \alpha}, \\ t^2_{F\cdot{\rm UNB}}(\alpha) &= \frac{(n^* - 2) p_1}{n^* - p_1 - 1} F_{p_1, n^* - p_1 - 1, \alpha}, \end{aligned}$$

where

$$n^* = \frac{1}{p_1} \sum_{\ell=1}^{2} \left\{ n_1^{(\ell)} p_1 + n_2^{(\ell)} p_2 + n_3^{(\ell)} p_3 \right\}.$$

Figure 2 shows that $n^* = (n_*^{(1)} + n_*^{(2)})$ is the solution to the equation

$$\sum_{\ell=1}^{2} \left\{ n_1^{(\ell)} p_1 + n_2^{(\ell)} p_2 + n_3^{(\ell)} p_3 \right\} = (n_*^{(1)} + n_*^{(2)}) p_1.$$

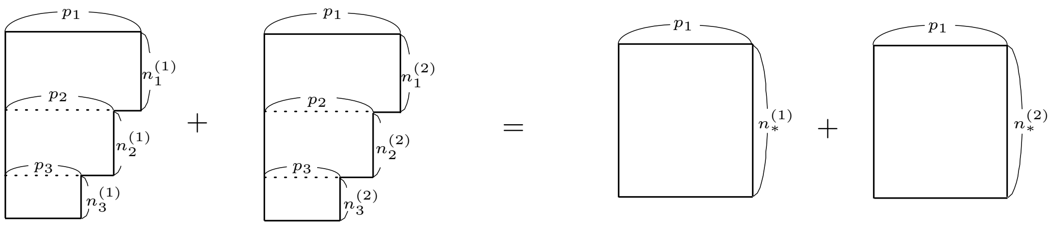

Figure 2: Approximation by adjusting the degrees of freedom of F distribution

4. Simultaneous confidence intervals for pairwise comparisons among mean vectors

We consider the simultaneous confidence intervals for all pairwise differences of mean vectors when each of the data set has three-step monotone missing observations.

Let $\boldsymbol{x}_{i1}^{(\ell)}, \ldots, \boldsymbol{x}_{in_i^{(\ell)}}^{(\ell)}$ be distributed as $N_{p_i}(\boldsymbol{\mu}_i^{(\ell)}, \Sigma_i)$ for $i = 1, 2, 3$ and $\ell = 1, 2, \ldots, m$. Then, for any nonnull vector $\boldsymbol{c} = (c_1, \ldots, c_p)'$, the simultaneous confidence intervals for all the linear combination $\boldsymbol{c}'(\boldsymbol{\mu}^{(a)} - \boldsymbol{\mu}^{(b)})$ of $\boldsymbol{\mu}^{(a)} - \boldsymbol{\mu}^{(b)}$, $1 \le a < b \le m$ are given by

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

where the values of $t_{\max \cdot p}^2(\alpha)$ is the upper percentiles of $T_{\max \cdot p}^2$ statistic

$$T_{\max \cdot p}^2 = \max_{1 \le a < b \le m} (\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)})' (\widetilde{\Gamma}^{[pl]})^{-1} (\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)}).$$

We note that $\widetilde{\Gamma}^{[pl]}$ is an estimator of $\widetilde{\Gamma} = \mathrm{Cov}(\widetilde{\boldsymbol{\mu}}^{(a)} - \widetilde{\boldsymbol{\mu}}^{(b)})$ and is calculated using the pooled estimator $\widehat{\Sigma}^{[pl]}$ for m-sample problem in Theorem 1.

However, it is not easy to get the upper percentiles of $T^2_{\max\cdot p}$ statistic. Therefore, we give Bonferroni's approximation which is one of the solution to this problem. Since the distributions of all T^2_{ab} are identical for $n_i^{(\ell)} = n_i,\ i = 1, 2, 3,\ \ell = 1, 2, \ldots, m$. Under the assumption that the number of observations $n_i^{(\ell)}$ is equal for m populations, the upper percentile of T^2_{YS} with Bonferroni's approximation can be the solution to the equation

$$\Pr\{T^2_{ab} > t^2_{\mathrm{Bon}}(\alpha^*)\} = \alpha^*,$$

where

$$T^2_{ab} = (\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)})'(\widetilde{\Gamma}^{[pl]})^{-1}(\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)}),\ \alpha^* = \frac{2\alpha}{m(m-1)}.$$

Using the approximations in Section 3 to $t^2_{\mathrm{Bon}}(\alpha^*)$, we can obtain the approximate simultaneous confidence intervals for all pairwise difference of mean vectors.

5. Simulation studies

To investigate the accuracy of some approximations, we compute the upper percentiles of the T^2_{YS} statistic and $T^2_{\max\cdot p}$ statistic using the Monte Carlo simulation. For each parameter, the simulation was carried out for 1,000,000 trials based on three-step monotone missing data sets.

Simulation results related to the upper percentiles of T^2_{YS} and their approximations are summarized in Tables 1 $\sim$ 3. Computations are made for the following three cases:

$$\begin{aligned}
\text{Case I} &: (p_1, p_2, p_3) = (12, 8, 4),\ n_1^{(1)} = n_1^{(2)} = 13, 15, 20, 25,\\
&\quad (n_2^{(1)}, n_3^{(1)}) = (n_2^{(2)}, n_3^{(2)}) = (5, 5), (10, 10),\ \alpha = 0.05, 0.01,\\
\text{Case II} &: (p_1, p_2, p_3) = (12, 8, 4),\ n_1^{(1)} = n_1^{(2)} = 30, 50, 100, 200, 300,\\
&\quad (n_2^{(1)}, n_3^{(1)}) = (n_2^{(2)}, n_3^{(2)}) = (10, 10), (20, 20),\ \alpha = 0.05, 0.01,\\
\text{Case III} &: (p_1, p_2, p_3) = (12, 8, 4),\ (n_1^{(1)}, n_1^{(2)}) = (13, 37), (15, 35), (20, 30), (25, 25),\\
&\quad (n_2^{(1)}, n_3^{(1)}) = (n_2^{(2)}, n_3^{(2)}) = (5, 5), (10, 10),\ \alpha = 0.05, 0.01,
\end{aligned}$$

Tables 1 $\sim$ 3 represent the results of Cases I $\sim$ III, respectively. Tables 1 $\sim$ 3 list the simulated upper percentiles of T^2_{YS} statistic ($t^2_{\mathrm{simu}}(\alpha)$), the approximate upper percentiles of T^2_{YS} ($t^2_{\mathrm{YS\cdot MLE}}(\alpha)$, $t^2_{\mathrm{YS\cdot UNB}}(\alpha)$, $t^2_{F\cdot\mathrm{MLE}}(\alpha)$, $t^2_{F\cdot\mathrm{UNB}}(\alpha)$, $t^2_{\mathrm{S\cdot MLE}}(\alpha)$, $t^2_{\mathrm{S\cdot UNB}}(\alpha)$), and the upper percentiles of χ^2 distribution with p degrees of freedom ($\chi^2_{p,\alpha}$). It may be noted from Tables 1 and 2 that the simulated values are not close to the upper percentiles of χ^2 distribution even when the sample size $n_1^{(\ell)}$ is moderately large. However, it is seen that the proposed approximations are good even for cases where $n_1^{(\ell)}$ is not large. In particular, it is noted that the values of $t^2_{\mathrm{YS\cdot MLE}}(\alpha)$ is considerably good for all cases. That is, the simulated coverage probabilities for $t^2_{\mathrm{YS\cdot MLE}}(\alpha)$ are considerably close to the nominal level $1-\alpha$. Further, Table 3 gives the simulated result and its approximations for the unbalanced cases where $n_1^{(1)} \neq n_1^{(2)}$ such that $n_1^{(1)} + n_1^{(2)} = 50$. It may be noted that the proposed approximations are good and $t^2_{\mathrm{YS\cdot MLE}}(\alpha)$ is also considerably close to the simulated value $t^2_{\mathrm{simu}}(\alpha)$. Therefore, it can be concluded that the approximation $t^2_{\mathrm{YS\cdot MLE}}(\alpha)$ is very good even for small samples and unbalanced cases.

Next, in order to investigate the accuracy of the approximation to the upper percentile of $T^2_{\max\cdot p}$ statistic and to compare the some approximate values with the simulated value, the Monte Carlo simulation was made for two cases that $m = 3, 6$. For parameters, we set them as follows:

Case IV $(m = 3)$: $(p_1, p_2, p_3) = (12, 8, 4)$, $\alpha = 0.05, 0.01$,

(i) $n_1^{(\ell)} = 13, 15, 20, 25$, $(n_2^{(\ell)}, n_3^{(\ell)}) = (5, 5), (10, 10)$,

(ii) $n_1^{(\ell)} = 30, 50, 100, 200$, $(n_2^{(\ell)}, n_3^{(\ell)}) = (10, 10), (20, 20)$,

Case V $(m = 3)$: $(p_1, p_2, p_3) = (20, 12, 6)$,

$n_1^{(\ell)} = 30, 50, 100, 200$, $(n_2^{(\ell)}, n_3^{(\ell)}) = (10, 10), (20, 20)$, $\alpha = 0.05, 0.01$,

Case VI $(m = 6)$: $(p_1, p_2, p_3) = (20, 12, 6)$,

$n_1^{(\ell)} = 30, 60, 100$, $(n_2^{(\ell)}, n_3^{(\ell)}) = (10, 10)$, $\alpha = 0.05, 0.01$,

Case VII $(m = 6)$: $(p_1, p_2, p_3) = (30, 20, 10)$,

$n_1^{(\ell)} = 40, 60, 100$, $(n_2^{(\ell)}, n_3^{(\ell)}) = (10, 10)$, $\alpha = 0.05, 0.01$, Case VIII $(m = 6) : (p_1, p_2, p_3) = (50, 30, 15),$

$$n_1^{(\ell)} = 60, 100, \ (n_2^{(\ell)}, n_3^{(\ell)}) = (10, 10), \ \alpha = 0.05, 0.01.$$

Tables 4-a, 4-b, 5 $\sim$ 8 list the simulated upper 100α percentiles of $T^2_{\max\cdot p}$ statistic $(t^2_{\text{max}\cdot\text{simu}}(\alpha))$, the simulated upper $100\alpha^*$ percentiles of T^2_{ab} statistic $(t^2_{\text{Bon}\cdot\text{simu}}(\alpha^*))$, that is the Bonferroni's approximation by Monte Carlo simulation, the approximate upper $100\alpha^*$ percentiles of T^2_{ab} statistic $(t^2_{\text{YS}\cdot\text{MLE}}(\alpha^*)$, $t^2_{\text{YS}\cdot\text{UNB}}(\alpha^*)$, $t^2_{F\cdot\text{MLE}}(\alpha^*)$ and $t^2_{F\cdot\text{UNB}}(\alpha^*))$ and the upper $100\alpha^*$ percentiles of χ^2 distribution with p degrees of freedom (χ^2_{p,α^*}). It may be noted from Tables 4-a, 4-b and 5 that the simulated values $t^2_{\text{Bon}\cdot\text{simu}}(\alpha^*)$ are larger than the simulated values $t^2_{\text{max}\cdot\text{simu}}(\alpha)$. Because it holds theoretically that the Bonferroni's approximation is always an overestimate for $T^2_{\max}$ type statistic. It may be seen from Tables that the simulated value $t^2_{\text{max}\cdot\text{simu}}(\alpha)$ is closer to the value of $t^2_{\text{YS}\cdot\text{UNB}}(\alpha^*)$ than that of $t^2_{\text{YS}\cdot\text{MLE}}(\alpha^*)$ for all cases. Tables 6 $\sim$ 8 give the simulated results and their approximations for the case where the dimensions are large. In these cases, it may be noted that the approximations are too conservative. The simulation studies show that the $t^2_{\text{YS}\cdot\text{UNB}}(\alpha^*)$ is close to $t^2_{\text{max}\cdot\text{simu}}(\alpha)$ and a good approximation in most cases.

Table 1: Simulated and approximate values and coverage probabilities for Case I

Sample size			Upper percentile (Coverage probability)							
$n_1^{(\ell)}$	$n_2^{(\ell)}$	$n_3^{(\ell)}$	$t^2_{\mathrm{simu}}(\alpha)$	$t^2_{\mathrm{YS \cdot MLE}}(\alpha)$	$t^2_{\mathrm{YS \cdot UNB}}(\alpha)$	$t^2_{F \cdot \mathrm{MLE}}(\alpha)$	$t^2_{F \cdot \mathrm{UNB}}(\alpha)$	$t^2_{\mathrm{S \cdot MLE}}(\alpha)$	$t^2_{\mathrm{S \cdot UNB}}(\alpha)$	$\chi^2_{p,\alpha}$
$\alpha = 0.05$										
13	5	5	50.71	48.47	45.32	41.39	39.09	40.55	38.33	21.03
				(0.94)	(0.93)	(0.90)	(0.88)	(0.90)	(0.88)	(0.53)
15	5	5	43.28	41.62	39.29	37.91	36.01	36.35	34.63	21.03
				(0.94)	(0.93)	(0.92)	(0.90)	(0.91)	(0.89)	(0.60)
20	5	5	34.72	34.00	32.55	32.83	31.52	31.40	30.25	21.03
				(0.95)	(0.93)	(0.94)	(0.92)	(0.92)	(0.91)	(0.71)
25	5	5	30.93	30.61	29.55	30.10	29.09	29.02	28.14	21.03
				(0.95)	(0.94)	(0.94)	(0.93)	(0.93)	(0.92)	(0.77)
13	10	10	46.24	45.74	42.90	34.45	32.95	39.14	37.07	21.03
				(0.95)	(0.93)	(0.87)	(0.85)	(0.91)	(0.90)	(0.59)
15	10	10	40.21	39.40	37.32	32.83	31.52	34.68	33.14	21.03
				(0.95)	(0.93)	(0.89)	(0.88)	(0.91)	(0.90)	(0.65)
20	10	10	32.91	32.56	31.28	30.10	29.09	29.89	28.91	21.03
				(0.95)	(0.94)	(0.93)	(0.91)	(0.92)	(0.91)	(0.74)
25	10	10	29.66	29.60	28.65	28.39	27.58	27.81	27.06	21.03
				(0.95)	(0.94)	(0.94)	(0.93)	(0.93)	(0.92)	(0.80)
$\alpha = 0.01$										
13	5	5	76.54	70.75	66.09	57.74	54.53	57.02	53.85	26.22
				(0.99)	(0.98)	(0.97)	(0.96)	(0.97)	(0.96)	(0.69)
15	5	5	62.29	58.54	55.22	52.01	49.41	49.76	47.37	26.22
				(0.99)	(0.98)	(0.98)	(0.97)	(0.97)	(0.97)	(0.75)
20	5	5	47.29	45.83	43.87	43.90	42.15	41.71	40.18	26.22
				(0.99)	(0.98)	(0.98)	(0.98)	(0.98)	(0.98)	(0.85)
25	5	5	41.09	40.47	39.06	39.65	38.33	38.02	36.86	26.22
				(0.99)	(0.99)	(0.99)	(0.98)	(0.98)	(0.98)	(0.90)
13	10	10	69.73	66.50	62.27	46.45	44.43	55.25	52.24	26.22
				(0.99)	(0.98)	(0.95)	(0.94)	(0.97)	(0.97)	(0.74)
15	10	10	57.73	55.10	52.14	43.90	42.15	47.36	45.21	26.22
				(0.99)	(0.98)	(0.97)	(0.96)	(0.97)	(0.97)	(0.79)
20	10	10	44.71	43.64	41.90	39.65	38.33	39.46	38.15	26.22
				(0.99)	(0.99)	(0.98)	(0.98)	(0.98)	(0.98)	(0.87)
25	10	10	39.27	38.95	37.70	37.04	35.98	36.21	35.23	26.22
				(0.99)	(0.99)	(0.99)	(0.98)	(0.98)	(0.98)	(0.91)

Table 2: Simulated and approximate values and coverage probabilities for Case II

Sample size			Upper percentile (Coverage probability)							
$n_1^{(\ell)}$	$n_2^{(\ell)}$	$n_3^{(\ell)}$	$t^2_{\text{simu}}(\alpha)$	$t^2_{\text{YS}\cdot\text{MLE}}(\alpha)$	$t^2_{\text{YS}\cdot\text{UNB}}(\alpha)$	$t^2_{F\cdot\text{MLE}}(\alpha)$	$t^2_{F\cdot\text{UNB}}(\alpha)$	$t^2_{\text{S}\cdot\text{MLE}}(\alpha)$	$t^2_{\text{S}\cdot\text{UNB}}(\alpha)$	$\chi^2_{p,\alpha}$
			$\alpha = 0.05$							
30	10	10	27.85	27.91	27.15	27.22	26.54	26.60	25.99	21.03
				(0.95)	(0.94)	(0.94)	(0.94)	(0.94)	(0.93)	(0.83)
50	10	10	24.86	24.96	24.53	24.81	24.40	24.42	24.05	21.03
				(0.95)	(0.95)	(0.95)	(0.94)	(0.94)	(0.94)	(0.89)
100	10	10	22.90	22.96	22.75	22.95	22.74	22.81	22.61	21.03
				(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.92)
200	10	10	21.99	22.00	21.89	21.99	21.89	21.95	21.85	21.03
				(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.94)
300	10	10	21.64	21.67	21.60	21.67	21.60	21.65	21.58	21.03
				(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.94)
30	20	20	26.89	27.15	26.47	25.73	25.21	25.88	25.35	21.03
				(0.95)	(0.95)	(0.94)	(0.93)	(0.94)	(0.93)	(0.85)
50	20	20	24.14	24.57	24.18	24.20	23.85	23.93	23.61	21.03
				(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.94)	(0.90)
100	20	20	22.57	22.83	22.64	22.77	22.58	22.60	22.43	21.03
				(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.93)
200	20	20	21.87	21.96	21.85	21.95	21.85	21.88	21.79	21.03
				(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.94)
300	20	20	21.60	21.65	21.59	21.65	21.58	21.62	21.55	21.03
				(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.95)	(0.94)
			$\alpha = 0.01$							
30	10	10	36.45	36.35	35.36	35.28	34.39	34.37	33.57	26.22
				(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.98)	(0.93)
50	10	10	31.76	31.92	31.37	31.70	31.17	31.11	30.64	26.22
				(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.96)
100	10	10	28.94	29.00	28.73	28.97	28.71	28.77	28.52	26.22
				(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.98)
200	10	10	27.61	27.60	27.47	27.60	27.47	27.54	27.41	26.22
				(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.98)
300	10	10	27.13	27.14	27.05	27.14	27.05	27.11	27.03	26.22
				(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)
30	20	20	35.00	35.22	34.34	33.05	32.39	33.32	32.63	26.22
				(0.99)	(0.99)	(0.99)	(0.98)	(0.99)	(0.98)	(0.94)
50	20	20	30.90	31.35	30.86	30.79	30.35	30.41	30.01	26.22
				(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.97)
100	20	20	28.47	28.81	28.56	28.72	28.48	28.47	28.26	26.22
				(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.98)
200	20	20	27.44	27.54	27.42	27.53	27.41	27.44	27.32	26.22
				(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)
300	20	20	27.05	27.11	27.03	27.11	27.02	27.06	26.98	26.22
				(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.99)

Table 3: Simulated and approximate values and coverage probabilities for Case Ⅲ

Sample size						Upper percentile (Coverage probability)							
$n_1^{(1)}$	$n_2^{(1)}$	$n_3^{(1)}$	$n_1^{(2)}$	$n_2^{(2)}$	$n_3^{(2)}$	$t^2_{\mathrm{simu}}(\alpha)$	$t^2_{\mathrm{YS\cdot MLE}}(\alpha)$	$t^2_{\mathrm{YS\cdot UNB}}(\alpha)$	$t^2_{F\cdot\mathrm{MLE}}(\alpha)$	$t^2_{F\cdot\mathrm{UNB}}(\alpha)$	$t^2_{\mathrm{S\cdot MLE}}(\alpha)$	$t^2_{\mathrm{S\cdot UNB}}(\alpha)$	$\chi^2_{p,\alpha}$
						$\alpha = 0.05$							
13	5	5	37	5	5	30.14	30.61	29.55	30.10	29.09	29.02	28.14	21.03
							(0.95)	(0.94)	(0.95)	(0.94)	(0.94)	(0.93)	(0.79)
15	5	5	35	5	5	30.47	30.61	29.55	30.10	29.09	29.02	28.14	21.03
							(0.95)	(0.94)	(0.95)	(0.94)	(0.94)	(0.93)	(0.78)
20	5	5	30	5	5	30.88	30.61	29.55	30.10	29.09	29.02	28.14	21.03
							(0.95)	(0.94)	(0.94)	(0.93)	(0.93)	(0.92)	(0.77)
25	5	5	25	5	5	30.93	30.61	29.55	30.10	29.09	29.02	28.14	21.03
							(0.95)	(0.94)	(0.94)	(0.93)	(0.93)	(0.92)	(0.77)
13	10	10	37	10	10	29.00	29.60	28.65	28.39	27.58	27.81	27.06	21.03
							(0.95)	(0.95)	(0.94)	(0.94)	(0.94)	(0.93)	(0.81)
15	10	10	35	10	10	29.31	29.60	28.65	28.39	27.58	27.81	27.06	21.03
							(0.95)	(0.94)	(0.94)	(0.93)	(0.94)	(0.93)	(0.81)
20	10	10	30	10	10	29.57	29.60	28.65	28.39	27.58	27.81	27.06	21.03
							(0.95)	(0.94)	(0.94)	(0.93)	(0.93)	(0.92)	(0.80)
25	10	10	25	10	10	29.66	29.60	28.65	28.39	27.58	27.81	27.06	21.03
							(0.95)	(0.94)	(0.94)	(0.93)	(0.93)	(0.92)	(0.80)
						$\alpha = 0.01$							
13	5	5	37	5	5	40.06	40.47	39.06	39.65	38.33	38.02	36.86	26.22
							(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.98)	(0.91)
15	5	5	35	5	5	40.48	40.47	39.06	39.65	38.33	38.02	36.86	26.22
							(0.99)	(0.99)	(0.99)	(0.99)	(0.99)	(0.98)	(0.90)
20	5	5	30	5	5	40.94	40.47	39.06	39.65	38.33	38.02	36.86	26.22
							(0.99)	(0.99)	(0.99)	(0.98)	(0.98)	(0.98)	(0.90)
25	5	5	25	5	5	41.09	40.47	39.06	39.65	38.33	38.02	36.86	26.22
							(0.99)	(0.99)	(0.99)	(0.98)	(0.98)	(0.98)	(0.90)
13	10	10	37	10	10	38.39	38.95	37.70	37.04	35.98	36.21	35.23	26.22
							(0.99)	(0.99)	(0.99)	(0.98)	(0.99)	(0.98)	(0.92)
15	10	10	35	10	10	38.80	38.95	37.70	37.04	35.98	36.21	35.23	26.22
							(0.99)	(0.99)	(0.99)	(0.98)	(0.98)	(0.98)	(0.92)
20	10	10	30	10	10	39.17	38.95	37.70	37.04	35.98	36.21	35.23	26.22
							(0.99)	(0.99)	(0.99)	(0.98)	(0.98)	(0.98)	(0.91)
25	10	10	25	10	10	39.27	38.95	37.70	37.04	35.98	36.21	35.23	26.22
							(0.99)	(0.99)	(0.99)	(0.98)	(0.98)	(0.98)	(0.91)

Table 4-a: Simulated and approximate values and coverage probabilities for Case IV ($\alpha = 0.05$)

Sample size			Upper percentile (Coverage probability)						
$n_1^{(\ell)}$	$n_2^{(\ell)}$	$n_3^{(\ell)}$	$t^2_{\max\cdot\text{simu}}(\alpha)$	$t^2_{\text{Bon}\cdot\text{simu}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{MLE}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{UNB}}(\alpha^*)$	$t^2_{F\cdot\text{MLE}}(\alpha^*)$	$t^2_{F\cdot\text{UNB}}(\alpha^*)$	χ^2_{p,α^*}
					$\alpha = 0.05$				
13	5	5	40.35	41.29	43.11	40.38	52.40	49.49	24.63
					(0.96)	(0.95)	(0.99)	(0.98)	(0.66)
15	5	5	37.87	38.71	39.53	37.35	47.45	45.07	24.63
					(0.96)	(0.95)	(0.99)	(0.98)	(0.69)
20	5	5	33.95	34.65	34.81	33.34	40.37	38.76	24.63
					(0.96)	(0.94)	(0.98)	(0.98)	(0.77)
25	5	5	31.73	32.26	32.43	31.31	36.63	35.41	24.63
					(0.96)	(0.95)	(0.98)	(0.98)	(0.81)
13	10	10	36.98	37.83	41.11	38.68	42.61	40.75	24.63
					(0.97)	(0.96)	(0.98)	(0.97)	(0.73)
15	10	10	35.22	35.96	37.88	35.94	40.37	38.76	24.63
					(0.97)	(0.96)	(0.98)	(0.97)	(0.75)
20	10	10	32.28	32.86	33.70	32.39	36.63	35.41	24.63
					(0.96)	(0.95)	(0.98)	(0.97)	(0.81)
25	10	10	30.52	31.05	31.63	30.62	34.32	33.34	24.63
					(0.96)	(0.95)	(0.98)	(0.97)	(0.84)
30	10	10	29.42	29.85	30.37	29.55	32.75	31.93	24.63
					(0.96)	(0.95)	(0.98)	(0.97)	(0.86)
50	10	10	27.39	27.77	28.02	27.54	29.56	29.07	24.63
					(0.96)	(0.95)	(0.97)	(0.97)	(0.90)
100	10	10	25.92	26.21	26.34	26.10	27.12	26.87	24.63
					(0.96)	(0.95)	(0.96)	(0.96)	(0.93)
200	10	10	25.14	25.45	25.49	25.37	25.88	25.76	24.63
					(0.95)	(0.95)	(0.96)	(0.96)	(0.94)
30	20	20	28.51	28.92	29.73	29.00	30.77	30.15	24.63
					(0.96)	(0.96)	(0.97)	(0.97)	(0.88)
50	20	20	26.71	27.05	27.69	27.26	28.75	28.34	24.63
					(0.96)	(0.96)	(0.97)	(0.97)	(0.92)
100	20	20	25.60	25.90	26.22	26.00	26.89	26.67	24.63
					(0.96)	(0.96)	(0.97)	(0.96)	(0.93)
200	20	20	25.07	25.33	25.46	25.34	25.82	25.70	24.63
					(0.96)	(0.95)	(0.96)	(0.96)	(0.94)

Note: $\alpha^* = 2\alpha/[m(m-1)]$.

Table 4-b: Simulated and approximate values and coverage probabilities for Case IV ($\alpha = 0.01$)

Sample size			Upper percentile (Coverage probability)						
$n_1^{(\ell)}$	$n_2^{(\ell)}$	$n_3^{(\ell)}$	$t^2_{\text{max}\cdot\text{simu}}(\alpha)$	$t^2_{\text{Bon}\cdot\text{simu}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{MLE}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{UNB}}(\alpha^*)$	$t^2_{F\cdot\text{MLE}}(\alpha^*)$	$t^2_{F\cdot\text{UNB}}(\alpha^*)$	χ^2_{p,α^*}
					$\alpha = 0.01$				
13	5	5	53.57	54.11	55.72	52.17	69.83	65.95	29.49
					(0.99)	(0.99)	(1.00)	(1.00)	(0.81)
15	5	5	49.17	49.52	50.37	47.58	62.22	59.11	29.49
					(0.99)	(0.99)	(1.00)	(1.00)	(0.84)
20	5	5	43.12	43.46	43.53	41.68	51.64	49.57	29.49
					(0.99)	(0.99)	(1.00)	(1.00)	(0.89)
25	5	5	39.75	40.17	40.15	38.76	46.19	44.65	29.49
					(0.99)	(0.99)	(1.00)	(1.00)	(0.92)
13	10	10	48.81	49.35	52.92	49.75	54.94	52.55	29.49
					(0.99)	(0.99)	(1.00)	(0.99)	(0.86)
15	10	10	45.73	46.07	48.06	45.58	51.64	49.57	29.49
					(0.99)	(0.99)	(1.00)	(0.99)	(0.88)
20	10	10	40.76	41.16	41.99	40.35	46.19	44.65	29.49
					(0.99)	(0.99)	(1.00)	(1.00)	(0.92)
25	10	10	38.16	38.40	39.05	37.81	42.88	41.66	29.49
					(0.99)	(0.99)	(1.00)	(1.00)	(0.94)
30	10	10	36.55	36.83	37.29	36.28	40.66	39.65	29.49
					(0.99)	(0.99)	(1.00)	(1.00)	(0.95)
50	10	10	33.58	33.80	34.06	33.48	36.20	35.60	29.49
					(0.99)	(0.99)	(1.00)	(0.99)	(0.97)
100	10	10	31.53	31.62	31.78	31.49	32.84	32.55	29.49
					(0.99)	(0.99)	(0.99)	(0.99)	(0.98)
200	10	10	30.51	30.61	30.64	30.49	31.17	31.02	29.49
					(0.99)	(0.99)	(0.99)	(0.99)	(0.99)
30	20	20	35.26	35.42	36.43	35.53	37.88	37.12	29.49
					(0.99)	(0.99)	(0.99)	(0.99)	(0.96)
50	20	20	32.72	32.87	33.61	33.09	35.08	34.58	29.49
					(0.99)	(0.99)	(0.99)	(0.99)	(0.98)
100	20	20	31.09	31.18	31.62	31.35	32.54	32.27	29.49
					(0.99)	(0.99)	(0.99)	(0.99)	(0.98)
200	20	20	30.37	30.54	30.59	30.45	31.09	30.94	29.49
					(0.99)	(0.99)	(0.99)	(0.99)	(0.99)

Note: $\alpha^* = 2\alpha/[m(m-1)]$.

Table 5: Simulated and approximate values and coverage probabilities for Case V

Sample size			Upper percentile (Coverage probability)						
$n_1^{(\ell)}$	$n_2^{(\ell)}$	$n_3^{(\ell)}$	$t^2_{\max\cdot\text{simu}}(\alpha)$	$t^2_{\text{Bon}\cdot\text{simu}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{MLE}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{UNB}}(\alpha^*)$	$t^2_{F\cdot\text{MLE}}(\alpha^*)$	$t^2_{F\cdot\text{UNB}}(\alpha^*)$	χ^2_{p,α^*}
					$\alpha = 0.05$				
30	10	10	48.62	49.32	49.62	48.23	56.71	55.25	35.70
					(0.96)	(0.95)	(0.99)	(0.98)	(0.71)
50	10	10	42.80	43.29	43.43	42.66	47.47	46.67	35.70
					(0.96)	(0.95)	(0.98)	(0.98)	(0.83)
100	10	10	39.00	39.39	39.43	39.07	41.30	40.92	35.70
					(0.95)	(0.95)	(0.97)	(0.97)	(0.90)
200	10	10	37.21	37.50	37.55	37.37	38.43	38.25	35.70
					(0.95)	(0.95)	(0.96)	(0.96)	(0.93)
30	20	20	46.65	47.19	48.27	47.02	51.23	50.16	35.70
					(0.96)	(0.95)	(0.98)	(0.97)	(0.75)
50	20	20	41.62	42.06	42.75	42.05	45.52	44.86	35.70
					(0.96)	(0.95)	(0.98)	(0.97)	(0.85)
100	20	20	38.49	38.89	39.20	38.86	40.82	40.47	35.70
					(0.96)	(0.95)	(0.97)	(0.97)	(0.91)
200	20	20	37.01	37.33	37.48	37.31	38.31	38.14	35.70
					(0.96)	(0.95)	(0.96)	(0.96)	(0.93)
					$\alpha = 0.01$				
30	10	10	58.85	59.25	59.46	57.79	69.02	67.25	41.37
					(0.99)	(0.99)	(1.00)	(1.00)	(0.86)
50	10	10	50.86	51.07	51.29	50.39	56.63	55.67	41.37
					(0.99)	(0.99)	(1.00)	(1.00)	(0.93)
100	10	10	45.87	46.01	46.13	45.70	48.55	48.11	41.37
					(0.99)	(0.99)	(1.00)	(0.99)	(0.97)
200	10	10	43.62	43.69	43.72	43.51	44.86	44.64	41.37
					(0.99)	(0.99)	(0.99)	(0.99)	(0.98)
30	20	20	56.41	56.51	57.70	56.20	61.63	60.35	41.37
					(0.99)	(0.99)	(1.00)	(0.99)	(0.89)
50	20	20	49.42	49.69	50.41	49.59	54.06	53.27	41.37
					(0.99)	(0.99)	(1.00)	(1.00)	(0.95)
100	20	20	45.32	45.54	45.83	45.43	47.93	47.52	41.37
					(0.99)	(0.99)	(0.99)	(0.99)	(0.97)
200	20	20	43.38	43.48	43.63	43.43	44.70	44.50	41.37
					(0.99)	(0.99)	(0.99)	(0.99)	(0.98)

Note: $\alpha^* = 2\alpha/[m(m-1)]$.

Table 6: Simulated and approximate values and coverage probabilities for Case VI

Sample size			Upper percentile (Coverage probability)						
$n_1^{(\ell)}$	$n_2^{(\ell)}$	$n_3^{(\ell)}$	$t^2_{\max\cdot\text{simu}}(\alpha)$	$t^2_{\text{Bon}\cdot\text{simu}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{MLE}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{UNB}}(\alpha^*)$	$t^2_{F\cdot\text{MLE}}(\alpha^*)$	$t^2_{F\cdot\text{UNB}}(\alpha^*)$	χ^2_{p,α^*}
					$\alpha = 0.05$				
30	10	10	46.96	47.53	49.65	48.28	69.02	67.25	41.37
					(0.97)	(0.96)	(1.00)	(1.00)	(0.85)
60	10	10	44.44	44.99	45.47	44.79	53.83	53.05	41.37
					(0.96)	(0.95)	(1.00)	(0.99)	(0.90)
100	10	10	43.15	43.73	43.85	43.44	48.55	48.11	41.37
					(0.96)	(0.95)	(0.99)	(0.99)	(0.92)
					$\alpha = 0.01$				
30	10	10	54.15	54.32	56.69	55.12	81.37	79.29	46.60
					(0.99)	(0.99)	(1.00)	(1.00)	(0.95)
60	10	10	50.79	50.90	51.56	50.80	61.99	61.10	46.60
					(0.99)	(0.99)	(1.00)	(1.00)	(0.97)
100	10	10	49.21	49.35	49.59	49.13	55.41	54.90	46.60
					(0.99)	(0.99)	(1.00)	(1.00)	(0.98)

Note: $\alpha^* = 2\alpha/[m(m-1)]$.

Table 7: Simulated and approximate values and coverage probabilities for Case VII

Sample size			Upper percentile (Coverage probability)						
$n_1^{(\ell)}$	$n_2^{(\ell)}$	$n_3^{(\ell)}$	$t^2_{\max\cdot\text{simu}}(\alpha)$	$t^2_{\text{Bon}\cdot\text{simu}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{MLE}}(\alpha^*)$	$t^2_{\text{YS}\cdot\text{UNB}}(\alpha^*)$	$t^2_{F\cdot\text{MLE}}(\alpha^*)$	$t^2_{F\cdot\text{UNB}}(\alpha^*)$	χ^2_{p,α^*}
					$\alpha = 0.05$				
40	10	10	63.01	63.78	66.40	65.01	96.53	94.60	55.24
					(0.97)	(0.97)	(1.00)	(1.00)	(0.81)
60	10	10	61.04	61.74	62.60	61.68	80.04	78.90	55.24
					(0.96)	(0.96)	(1.00)	(1.00)	(0.85)
100	10	10	58.78	59.38	59.65	59.11	69.00	68.37	55.24
					(0.96)	(0.95)	(1.00)	(0.99)	(0.90)
					$\alpha = 0.01$				
40	10	10	71.13	71.42	74.37	72.81	111.42	109.19	61.15
					(0.99)	(0.99)	(1.00)	(1.00)	(0.93)
60	10	10	68.55	68.77	69.84	68.82	91.00	89.70	61.15
					(0.99)	(0.99)	(1.00)	(1.00)	(0.95)
100	10	10	65.79	66.19	66.35	65.74	77.57	76.87	61.15
					(0.99)	(0.99)	(1.00)	(1.00)	(0.97)

Note: $\alpha^* = 2\alpha/[m(m-1)]$.

Table 8: Simulated and approximate values and coverage probabilities for Case VIII

Sample size			Upper percentile (Coverage probability)						
$n_1^{(\ell)}$	$n_2^{(\ell)}$	$n_3^{(\ell)}$	$t^2_{\mathrm{max\cdot simu}}(\alpha)$	$t^2_{\mathrm{Bon\cdot simu}}(\alpha^*)$	$t^2_{\mathrm{YS\cdot MLE}}(\alpha^*)$	$t^2_{\mathrm{YS\cdot UNB}}(\alpha^*)$	$t^2_{F\cdot\mathrm{MLE}}(\alpha^*)$	$t^2_{F\cdot\mathrm{UNB}}(\alpha^*)$	χ^2_{p,α^*}
				$\alpha = 0.05$					
60	10	10	95.34	96.22	98.68 (0.97)	97.22 (0.96)	153.99 (1.00)	151.76 (1.00)	81.36 (0.69)
100	10	10	90.19	91.08	91.41 (0.96)	90.56 (0.95)	116.84 (1.00)	115.76 (1.00)	81.36 (0.81)
				$\alpha = 0.01$					
60	10	10	105.22	105.50	108.25 (0.99)	106.64 (0.99)	173.75 (1.00)	171.24 (1.00)	88.37 (0.86)
100	10	10	99.00	99.62	99.86 (0.99)	98.94 (0.99)	129.50 (1.00)	128.31 (1.00)	88.37 (0.93)

Note: $\alpha^* = 2\alpha/[m(m-1)]$.

Acknowledgements

The authors wish to thank Ms. Miki Hosoya for her help with the Monte Carlo simulation in numerical examinations. The second author's research was in part supported by Grant-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] Anderson, T. W. and Olkin, I. (1985). Maximum-likelihood estimation of the parameters of a multivariate normal distribution. *Linear Algebra and Its Applications*, **70**, 147–171.

[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. *Communication 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. *Canadian Journal of Statistics*, **27**, 395–407.

[7] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data, 2nd ed.* Wiley, New York.

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

[9] 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.

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

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

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

[13] 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.