Bias Correction for T^2 Type Statistic with Two-step Monotone Missing Data

Tamae Kawasaki[a], Takashi Seo[b]

[a] *Department of Mathematical Information Science, Graduate School of Science, Tokyo University of Science, Tokyo, Japan;* [b] *Department of Mathematical Information Science Faculty of Science, Tokyo University of Science, Tokyo, Japan*

In this paper, we consider the problem of testing for mean vector when the data have two-step monotone missing pattern. We derive the stochastic expansion of Hotelling's T^2 type statistic for the case where the sample size is large. Asymptotic expectation and variance are obtained using the stochastic expansion. Further, we propose the Bartlett corrected statistics for two-step monotone missing data. Finally, we present a numerical comparison based on Monte Carlo simulation for some selected parameters.

Keywords: Asymptotic expansion; Bias correction; Monte Carlo simulation; Stochastic expansion; Two-step monotone missing data.

AMS Subject Classification : Primary : 62H20, 62H10; Secondary : 62E17

1 Introduction

In almost all statistical analyses, missing data is a constantly occurring problem. Many statistical methods have been developed to analyze data with missing values. In particular, Anderson [1] developed an approach to derive the maximum likelihood estimators (MLEs) of the mean and covariance matrix by solving the likelihood equation for monotone missing data. Kanda and Fujikoshi [2] showed the properties of the MLEs based on k-step monotone missing samples for $k = 2, 3$ and general k. Moreover, closed forms are obtained for the MLEs in the case where a k-step monotone pattern using matrix derivatives is considered, as proposed by Jinadasa and Tracy [3]. Srivastava [4] derived the MLEs for missing data from a growth curve and several other models, and LR tests were developed. By using the idea of Srivastava [4], Srivastava and Carter [5] and Shutoh et al. [6] obtained the MLEs for the mean vector and covariance matrix using the Newton-Raphson method with general missing data. Yu, Krishnamoorthy, and Pannala [7] obtained Hotelling's T^2 type statistic and derived the F-approximations to the distribution of Hotelling's T^2 type statistic when the data have three-step monotone missing data. For a two-step monotone missing pattern, Seko, Kawasaki, and Seo [8] derived the T^2 type statistic of testing for two normal mean vectors and its approximate upper percentile. We note, among many other papers, Bhargava [9], Krishnamoorthy and Pannala [10], Seo and Srivastava [11], and Wu, Genton, and Stefanski [12]. They noted that it is difficult to derive the exact distribution of Hotelling's T^2 type statistic. In order to obtain the bias corrected transformation statistic using Bartlett correction, we derive the stochastic expansion of Hotelling's T^2 type statistic.

In this paper, we consider two-step monotone missing data drawn from a multivariate normal population that is of the form

$$\begin{pmatrix} x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1p_1}^{(1)} & x_{1p_1+1}^{(1)} & \cdots & x_{1p}^{(1)} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ x_{n_11}^{(1)} & x_{n_12}^{(1)} & \cdots & x_{n_1p_1}^{(1)} & x_{n_1p_1+1}^{(1)} & \cdots & x_{n_1p}^{(1)} \\ x_{11}^{(2)} & x_{12}^{(2)} & \cdots & x_{1p_1}^{(2)} & * & \cdots & * \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ x_{n_21}^{(2)} & x_{n_22}^{(2)} & \cdots & x_{n_2p_1}^{(2)} & * & \cdots & * \end{pmatrix},$$

where $n = n_1 + n_2$ and $p = p_1 + p_2$. "$*$" indicates missing data. That is, we have complete data for n_1 mutually independent observations with p dimensions and incomplete data for n_2 mutually independent observations with p_1 dimensions. Any analysis of the data is then dependent on assumptions about the missing data mechanism. However, because the causes for missing data are not our concern, we ignore the process that causes missing data by assuming that the data are missing at random (MAR). Rubin [13] pointed out that in the case of MAR and the observed data that are observed at random (OAR) or, more simply, the missing data that are missing completely at random (MCAR), the missing data mechanism can be ignored for likelihood based inferences. For the formal definition and exposition of MAR, OAR, and MCAR, we refer to Little and Rubin [14].

Let $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_{n_1}^{(1)}$ be distributed as the multivariate normal $N_p(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{x}_1^{(2)}, \ldots, \boldsymbol{x}_{n_2}^{(2)}$ be distributed as the multivariate normal $N_{p_1}(\boldsymbol{\mu}_1, \Sigma_{11})$, where each $\boldsymbol{x}_j^{(1)} = (x_{j1}^{(1)}, \ldots, x_{jp}^{(1)})', j = 1, \ldots, n_1$ is $p \times 1$ and each $\boldsymbol{x}_j^{(2)} = (x_{j1}^{(2)}, \ldots, x_{jp_1}^{(2)})', j = 1, \ldots, n_2$ is $p_1 \times 1$, and

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

We partition $\boldsymbol{x}_j^{(1)}$ into a $p_1 \times 1$ random vector and a $p_2 \times 1$ random vector as $\boldsymbol{x}_j^{(1)} = (\boldsymbol{x}_{1j}^{(1)'}, \boldsymbol{x}_{2j}^{(1)'})'$, where

$\boldsymbol{x}_{ij}^{(1)} : p_i \times 1, i = 1, 2, j = 1, \ldots, n_1$.

We define the sample means as

$$\overline{\boldsymbol{x}}_1^{(1)} = \frac{1}{n_1}\sum_{j=1}^{n_1}\boldsymbol{x}_{1j}^{(1)}, \quad \overline{\boldsymbol{x}}_2^{(1)} = \frac{1}{n_1}\sum_{j=1}^{n_1}\boldsymbol{x}_{2j}^{(1)}, \quad \overline{\boldsymbol{x}}^{(2)} = \frac{1}{n_2}\sum_{j=1}^{n_2}\boldsymbol{x}_{j}^{(2)},$$

where $\overline{\boldsymbol{x}}^{(1)} = (\overline{\boldsymbol{x}}_1^{(1)'}, \overline{\boldsymbol{x}}_2^{(1)'})'$, and the sample covariance matrices are given as

$$S^{(1)} = \frac{1}{N_1}\sum_{j=1}^{n_1}(\boldsymbol{x}_j^{(1)} - \overline{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_j^{(1)} - \overline{\boldsymbol{x}}^{(1)})' = \begin{pmatrix} S_{11}^{(1)} & S_{12}^{(1)} \\ S_{21}^{(1)} & S_{22}^{(1)} \end{pmatrix},$$

$$S^{(2)} = \frac{1}{N_2}\sum_{j=1}^{n_2}(\boldsymbol{x}_j^{(2)} - \overline{\boldsymbol{x}}^{(2)})(\boldsymbol{x}_j^{(2)} - \overline{\boldsymbol{x}}^{(2)})',$$

where $N_i = n_i - 1$.

Let the MLEs of $\boldsymbol{\mu}$ and Σ be denoted by $\widehat{\boldsymbol{\mu}}$ and $\widehat{\Sigma}$, which are partitioned in the same manner as $\boldsymbol{\mu}$ and Σ. We assume that the observation vectors are distributed as $N_p(\boldsymbol{\mu}, \Sigma)$ and $n_1 > p$, which is a necessary and sufficient condition for the existence and uniqueness of the MLEs of $\boldsymbol{\mu}$ and Σ. Anderson and Olkin [15] derived the MLEs of $(\boldsymbol{\mu}, \Sigma)$ (see Kanda and Fujikoshi [2], Chang and Richards [16], Seko, Yamazaki, and Seo [17]) as

$$\widehat{\boldsymbol{\mu}} = \begin{pmatrix} \widehat{\boldsymbol{\mu}}_1 \\ \widehat{\boldsymbol{\mu}}_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}(n_1\overline{\boldsymbol{x}}_1^{(1)} + n_2\overline{\boldsymbol{x}}^{(2)}) \\ \overline{\boldsymbol{x}}_2^{(1)} - \widehat{\Sigma}_{21}\widehat{\Sigma}_{11}^{-1}(\overline{\boldsymbol{x}}_1^{(1)} - \widehat{\boldsymbol{\mu}}_1) \end{pmatrix}, \ \widehat{\Sigma} = \begin{pmatrix} \frac{1}{n}(W_{11}^{(1)} + W^{(2)}) & \widehat{\Sigma}_{11}(W_{11}^{(1)})^{-1}W_{12}^{(1)}W_{12}^{(1)} \\ W_{21}^{(1)}(W_{11}^{(1)})^{-1}\widehat{\Sigma}_{11} & \frac{1}{n_1}W_{22\cdot 1} + \widehat{\Sigma}_{21}\widehat{\Sigma}_{11}^{-1}\widehat{\Sigma}_{12} \end{pmatrix},$$

where

$$W^{(1)} = N_1 S^{(1)} = \begin{pmatrix} W_{11}^{(1)} & W_{12}^{(1)} \\ W_{21}^{(1)} & W_{22}^{(1)} \end{pmatrix}, \quad W_{22\cdot 1}^{(1)} = W_{22}^{(1)} - W_{21}^{(1)}(W_{11}^{(1)})^{-1}W_{12}^{(1)},$$

$$W^{(2)} = N_2 S^{(2)} + \frac{n_1 n_2}{n}(\overline{\boldsymbol{x}}_1^{(1)} - \overline{\boldsymbol{x}}^{(2)})(\overline{\boldsymbol{x}}_1^{(1)} - \overline{\boldsymbol{x}}^{(2)})'.$$

Then, the mean vector and covariance matrix of $\widehat{\boldsymbol{\mu}}$ are given by

$$\mathrm{E}[\widehat{\boldsymbol{\mu}}] = \boldsymbol{\mu}, \quad \mathrm{Cov}[\widehat{\boldsymbol{\mu}}] = \begin{pmatrix} \frac{1}{n}\Sigma_{11} & \frac{1}{n}\Sigma_{12} \\ \frac{1}{n}\Sigma_{21} & \mathrm{Cov}[\widehat{\boldsymbol{\mu}}_2] \end{pmatrix},$$

respectively, when

$$\mathrm{Cov}[\widehat{\boldsymbol{\mu}}_2] = \frac{1}{n_1}\left(\Sigma_{22} - \frac{n_2}{n}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right) + \frac{n_2 p_1}{n n_1(n_1 - p_1 - 2)}\Sigma_{22\cdot 1}, \ (n_1 > p_1 + 2).$$

In this paper, we consider the test for mean vector based on two-step monotone missing data. In order to derive asymptotic expectation and variance, the stochastic expansion of Hotelling's T^2 type statistic is calculated. Further, we propose Bartlett correction statistics for two-step monotone missing data. Using Monte Carlo simulations, we investigate the first and second moments, variance, and the mean square error (MSE) for the T^2 type statistic. Moreover, we show the type I error rate when the null hypothesis is rejected using χ_p^2 under the simulated transformation statistics. Note that under H_0, the T^2 type statistic is asymptotically distributed as χ^2 with the degree of freedom p when $n_1, n \to \infty$ with $n_1/n \to \delta \in (0, 1]$ (see Chang and Richards [16]).

The remainder of this paper is organized as follows. In Section 2, the stochastic expansion of the T^2 type statistic is derived, and we calculate its expectation. Section 3 provides the stochastic expansion of the T^4 type statistic, and we present asymptotic variance. In Section 4, we propose the Bartlett corrected statistic and transformation statistic for two-step monotone missing data. Finally, we perform the Monte Carlo simulation and discuss the results in Section 5.

2 Asymptotic Expectation of T^2 Type Statistic

The problem of testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0$ is a specified vector, has been studied extensively for two-step monotone missing data. Seko, Yamazaki, and Seo [17] obtained the approximate upper percentiles of Hotelling's T^2 type statistic and the likelihood ratio test statistic for two-step monotone missing data in one sample problem.

For two-step monotone missing data, it is easy to construct a test statistic based on Hotelling's T^2 statistic structure:

$$T^2 = (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)' \left\{ \widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}) \right\}^{-1} (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0),$$

where

$$\widehat{\mathrm{Cov}}[\widehat{\boldsymbol{\mu}}] = \begin{pmatrix} \frac{1}{n}\widehat{\Sigma}_{11} & \frac{1}{n}\widehat{\Sigma}_{12} \\ \frac{1}{n}\widehat{\Sigma}_{21} & \widehat{\mathrm{Cov}}[\widehat{\boldsymbol{\mu}}_2] \end{pmatrix}.$$

Kanda and Fujikoshi [2] considered two- or three-step monotone missing data and obtained the asymptotic expansion of the distributions of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\Sigma}$ when n_1 and n_2 are large, and the MLE's of $\boldsymbol{\mu}$ and the usual transformed matrix of Σ are given in explicit forms for a general k. In this paper, we consider the asymptotic expansion of the first and second order moments of T^2 when n_1 and n_2 are large. In our derivations, we consider the stochastic expansions of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\Sigma}$ in terms of

$$\boldsymbol{z}^{(1)} = \begin{pmatrix} \boldsymbol{z}_1^{(1)} \\ \boldsymbol{z}_2^{(1)} \end{pmatrix} = \sqrt{N_1}\begin{pmatrix} \overline{\boldsymbol{x}}_1^{(1)} - \boldsymbol{\mu}_1 \\ \overline{\boldsymbol{x}}_2^{(1)} - \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{z}^{(2)} = \sqrt{N_2}(\overline{\boldsymbol{x}}^{(2)} - \boldsymbol{\mu}_1),$$
$$V^{(1)} = \sqrt{N_1}(S^{(1)} - \Sigma) = \begin{pmatrix} V_{11}^{(1)} & V_{12}^{(1)} \\ V_{21}^{(1)} & V_{22}^{(1)} \end{pmatrix}, \quad V^{(2)} = \sqrt{N_2}(S^{(2)} - \Sigma_{11}).$$

T^2 can be written as

$$T^2 = \left(\sqrt{N}\widehat{\boldsymbol{\mu}}\right)' \left\{N\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})\right\}^{-1}\left(\sqrt{N}\widehat{\boldsymbol{\mu}}\right)$$

and $\sqrt{N}\,(\widehat{\boldsymbol{\mu}})$ and $N\widehat{\mathrm{Cov}}\,(\widehat{\boldsymbol{\mu}})$ can be expanded as

$$\sqrt{N}\widehat{\boldsymbol{\mu}} = \begin{pmatrix} \sqrt{\gamma_1}\boldsymbol{z}_1^{(1)} + \sqrt{\gamma_2}\boldsymbol{z}^{(2)} \\ \dfrac{1}{\sqrt{\gamma_1}}\boldsymbol{z}_2^{(1)} \end{pmatrix} + \frac{1}{\sqrt{N}}\begin{pmatrix} \boldsymbol{0} \\ \dfrac{\sqrt{\gamma_2}}{\sqrt{\gamma_1}}V_{21}^{(1)}\boldsymbol{z}^{(2)} - \dfrac{\gamma_2}{\gamma_1}V_{21}^{(1)}\boldsymbol{z}_1^{(1)} \end{pmatrix}$$
$$+\frac{1}{N}\begin{pmatrix} \left(\dfrac{1}{\sqrt{\gamma_1}} - 2\sqrt{\gamma_1}\right)\boldsymbol{z}_1^{(1)} + \left(\dfrac{1}{\sqrt{\gamma_2}} - 2\sqrt{\gamma_2}\right)\boldsymbol{z}^{(2)} \\ \dfrac{\gamma_2}{\gamma_1\sqrt{\gamma_1}}V_{21}^{(1)}V_{11}^{(1)}\boldsymbol{z}_1^{(1)} - \dfrac{\sqrt{\gamma_2}}{\gamma_1}V_{21}^{(1)}V_{11}^{(1)}\boldsymbol{z}^{(2)} \end{pmatrix} + \mathrm{O}_p(N^{-\frac{3}{2}}),$$

$$N\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}_2) = \frac{1}{\gamma_1}I_{p_2} + \frac{1}{\sqrt{N}}\cdot\frac{1}{\gamma_1\sqrt{\gamma_1}}V_{22}^{(1)} + \frac{1}{N}\left(\frac{\gamma_2 p_1 - 2}{\gamma_1^2}I_{p_2} - \frac{\gamma_2}{\gamma_1^2}V_{21}^{(1)}V_{12}^{(1)}\right) + \mathrm{O}_p(N^{-\frac{3}{2}}),$$

respectively, where $\gamma_i = N_i/N(\ i = 1, 2)$. Therefore, putting

$$\Delta = \begin{pmatrix} I_{p_1} & O_{p_{12}} \\ O_{p_{21}} & \gamma_1 I_{p_2} \end{pmatrix}, O_{p_{12}} : p_1 \times p_1, \text{and } O_{p_{21}} : p_2 \times p_1,$$

we have

$$T^2 = \left(\sqrt{N}\Delta^{\frac{1}{2}}\widehat{\boldsymbol{\mu}}\right)' \left\{N\Delta^{\frac{1}{2}'}\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})\Delta^{\frac{1}{2}}\right\}^{-1}\left(\sqrt{N}\Delta^{\frac{1}{2}}\widehat{\boldsymbol{\mu}}\right),$$

and

$$\Delta^{\frac{1}{2}}(\sqrt{N}\widehat{\boldsymbol{\mu}}) = \boldsymbol{u}^{[0]} + \frac{1}{\sqrt{N}}\boldsymbol{u}^{[1]} + \frac{1}{N}\boldsymbol{u}^{[2]} + \mathrm{O}_p(N^{-\frac{3}{2}}),$$
$$\Delta^{\frac{1}{2}'}\left\{N\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}_2)\right\}\Delta^{\frac{1}{2}} = I_p + \frac{1}{\sqrt{N}}M^{[1]} + \frac{1}{N}M^{[2]} + \mathrm{O}_p(N^{-\frac{3}{2}}),$$

where

$$\boldsymbol{u}^{[0]} = \begin{pmatrix} \sqrt{\gamma_1}\boldsymbol{z}_1^{(1)} + \sqrt{\gamma_2}\boldsymbol{z}^{(2)} \\ \boldsymbol{z}_2^{(1)} \end{pmatrix}, \quad \boldsymbol{u}^{[1]} = \begin{pmatrix} \boldsymbol{0} \\ \sqrt{\gamma_2}V_{21}^{(1)}\boldsymbol{z}^{(2)} - \dfrac{\gamma_2}{\sqrt{\gamma_1}}V_{21}^{(1)}\boldsymbol{z}_1^{(1)} \end{pmatrix},$$

$$\boldsymbol{u}^{[2]} = \begin{pmatrix} \left(\dfrac{1}{\sqrt{\gamma_1}} - 2\sqrt{\gamma_1}\right)\boldsymbol{z}_1^{(1)} + \left(\dfrac{1}{\sqrt{\gamma_2}} - 2\sqrt{\gamma_2}\right)\boldsymbol{z}^{(2)} \\ \dfrac{\gamma_2}{\gamma_1}V_{21}^{(1)}V_{11}^{(1)}\boldsymbol{z}_1^{(1)} - \sqrt{\dfrac{\gamma_2}{\gamma_1}}V_{21}^{(1)}V_{11}^{(1)}\boldsymbol{z}^{(2)} \end{pmatrix},$$

$$M^{[1]} = \begin{pmatrix} \sqrt{\gamma_1}V_{11}^{(1)} + \sqrt{\gamma_2}V^{(2)} & V_{12}^{(1)} \\ V_{21}^{(1)} & \dfrac{1}{\sqrt{\gamma_1}}V_{22}^{(1)} \end{pmatrix} \quad M^{[2]} = \begin{pmatrix} M_{11}^{[2]} & M_{21}^{[2]} \\ M_{12}^{[2]} & M_{22}^{[2]} \end{pmatrix},$$

$$M_{11}^{[2]} = \gamma_1\gamma_2\left(\frac{1}{\sqrt{\gamma_1}}\boldsymbol{z}_1^{(1)} - \frac{1}{\sqrt{\gamma_2}}\boldsymbol{z}^{(2)}\right)\left(\frac{1}{\sqrt{\gamma_1}}\boldsymbol{z}_1^{(1)} - \frac{1}{\sqrt{\gamma_2}}\boldsymbol{z}^{(2)}\right)' - 4I_{p_1},$$

$$M_{12}^{[2]} = \sqrt{\gamma_2}V^{(2)}V_{12}^{(1)} - \frac{\gamma_2}{\sqrt{\gamma_1}}V_{11}^{(1)}V_{12}^{(1)},$$

$$M_{22}^{[2]} = \frac{\gamma_2 p_1 - 2}{\gamma_1}I_{p_2} - \frac{\gamma_2}{\gamma_1}V_{21}^{(1)}V_{12}^{(1)}.$$

Further, since

$$\left[\Delta^{\frac{1}{2}'}\left\{N\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}_2)\right\}\Delta^{\frac{1}{2}}\right]^{-1} = I_p - \frac{1}{\sqrt{N}}M^{[1]} + \frac{1}{N}M^{[3]} + \mathrm{O}_p(N^{-\frac{3}{2}}),$$

where

$$M^{[3]} = \begin{pmatrix} M_{11}^{[3]} & M_{12}^{[3]} \\ M_{21}^{[3]} & M_{22}^{[3]} \end{pmatrix},$$

$$M_{11}^{[3]} = (\sqrt{\gamma_1}V_{11}^{(1)} + \sqrt{\gamma_2}V^{(2)})^2 + V_{12}^{(1)}V_{21}^{(1)} - \gamma_1\gamma_2\left(\frac{1}{\sqrt{\gamma_1}}\boldsymbol{z}_1^{(1)} - \frac{1}{\sqrt{\gamma_2}}\boldsymbol{z}^{(2)}\right)\left(\frac{1}{\sqrt{\gamma_1}}\boldsymbol{z}_1^{(1)} - \frac{1}{\sqrt{\gamma_2}}\boldsymbol{z}^{(2)}\right)' + 4I_{p_1},$$

$$M_{12}^{[3]} = M_{12}^{[3]'} = \frac{1}{\sqrt{\gamma_1}}V_{21}^{(1)}V_{11}^{(1)} + \frac{1}{\sqrt{\gamma_1}}V_{22}^{(1)}V_{21}^{(1)},$$

$$M_{22}^{[3]} = \frac{1}{\gamma_1}\left(V_{21}^{(1)}V_{12}^{(1)} + (V_{22}^{(1)})^2\right) - \frac{\gamma_2 p_1 - 2}{\gamma_1}I_{p_2},$$

we can obtain a stochastic expansion of T^2 given by

$$T^2 = Q_0 + \boldsymbol{z}_2^{(1)'}\boldsymbol{z}_2^{(1)} - \frac{1}{\sqrt{N}}Q_1 + \frac{1}{N}Q_2 + \mathrm{O}_p(N^{-\frac{3}{2}}),$$

where

$$
\begin{aligned}
Q_0 &= \mathrm{tr}\Lambda, \\
Q_1 &= \frac{1}{\sqrt{\gamma_1}}(\boldsymbol{z}_1^{(1)\prime}V_{12}^{(1)}\boldsymbol{z}_2^{(1)} + \boldsymbol{z}_2^{(1)\prime}V_{21}^{(1)}\boldsymbol{z}_1^{(1)} + \boldsymbol{z}_2^{(1)\prime}V_{22}^{(1)}\boldsymbol{z}_2^{(1)}) + \mathrm{tr}\left\{(\sqrt{\gamma_1}V_{11}^{(1)} + \sqrt{\gamma_2}V^{(2)})\Lambda\right\}, \\
Q_2 &= 2(\boldsymbol{z}_1^{(1)\prime}\boldsymbol{z}_1^{(1)} + \boldsymbol{z}^{(2)\prime}\boldsymbol{z}^{(2)}) + \frac{1}{\sqrt{\gamma_1\gamma_2}}(\boldsymbol{z}_1^{(1)\prime}\boldsymbol{z}^{(2)} + \boldsymbol{z}^{(2)\prime}\boldsymbol{z}_1^{(1)}) - \frac{\gamma_2 p_1 - 2}{\gamma_1}\boldsymbol{z}_2^{(1)\prime}\boldsymbol{z}_2^{(1)} \\
&\quad + \mathrm{tr}\left\{(\sqrt{\gamma_1}V_{11}^{(1)} + \sqrt{\gamma_2}V^{(2)})^2\Lambda\right\} - \gamma_1\gamma_2\mathrm{tr}\left\{\left(\frac{1}{\sqrt{\gamma_1}}\boldsymbol{z}_1^{(1)} - \frac{1}{\sqrt{\gamma_2}}\boldsymbol{z}^{(2)}\right)\left(\frac{1}{\sqrt{\gamma_1}}\boldsymbol{z}_1^{(1)} - \frac{1}{\sqrt{\gamma_2}}\boldsymbol{z}^{(2)}\right)'\Lambda\right\} \\
&\quad + \frac{1}{\gamma_1}\left(\boldsymbol{z}_1^{(1)\prime}V_{11}^{(1)}V_{12}^{(1)}\boldsymbol{z}_2^{(1)} + \boldsymbol{z}_1^{(1)\prime}V_{12}^{(1)}V_{21}^{(1)}\boldsymbol{z}_1^{(1)} + \boldsymbol{z}_1^{(1)\prime}V_{12}^{(1)}V_{22}^{(1)}\boldsymbol{z}_2^{(1)} + \boldsymbol{z}_2^{(1)\prime}V_{21}^{(1)}V_{11}^{(1)}\boldsymbol{z}_1^{(1)}\right. \\
&\quad \left. + \boldsymbol{z}_2^{(1)\prime}V_{22}^{(1)}V_{21}^{(1)}\boldsymbol{z}_1^{(1)} + \boldsymbol{z}_2^{(1)\prime}V_{21}^{(1)}V_{12}^{(1)}\boldsymbol{z}_2^{(1)} + \boldsymbol{z}_2^{(1)\prime}(V_{22}^{(1)})^2\boldsymbol{z}_2^{(1)}\right), \\
\Lambda &= (\sqrt{\gamma_1}\boldsymbol{z}_1^{(1)} + \sqrt{\gamma_2}\boldsymbol{z}^{(2)})(\sqrt{\gamma_1}\boldsymbol{z}_1^{(1)} + \sqrt{\gamma_2}\boldsymbol{z}^{(2)})'.
\end{aligned}
$$

By calculating the expectation of T^2, we have the following theorem.

Theorem 2.1. *Suppose that the data have two-step monotone missing pattern. If $\gamma_1(= N_1/N)$ is a positive constant as $N_i's$ tend to infinity, then the expectation of T^2 can be expanded as*

$$
\mathrm{E}[T^2] = \xi_1 + \mathrm{O}(N^{-2}),
$$

where

$$
\xi_1 = p + \frac{1}{N}c_1, \quad c_1 = p_1(p_1+4) + \frac{p_2(p + \gamma_1 p_1 + 3)}{\gamma_1}.
$$

3 Asymptotic Variance of T^2 Type Statistic

Similar to the asymptotic expansion of the expectation of T^2 in Section 2, we have the following result. We can expand T^4 as

$$
T^4 = (Q_0 + \boldsymbol{z}_2^{(1)\prime}\boldsymbol{z}_2^{(1)})^2 - \frac{2}{\sqrt{N}}R_1 + \frac{1}{N}R_2 + \mathrm{O}_p(N^{\frac{3}{2}}),
$$

where

$$
\begin{aligned}
R_1 &= (\mathrm{tr}\Lambda + \boldsymbol{z}_2^{(1)'}\boldsymbol{z}_2^{(1)})Q_1, \\
R_2 &= 2(\mathrm{tr}\Lambda + \boldsymbol{z}_2^{(1)'}\boldsymbol{z}_2^{(1)}) \times \Bigg\{ 2\boldsymbol{z}_1^{(1)'}\boldsymbol{z}_1^{(1)} + 2\boldsymbol{z}^{(2)'}\boldsymbol{z}^{(2)} + \frac{1}{\sqrt{\gamma_1\gamma_2}}(\boldsymbol{z}_1^{(1)'}\boldsymbol{z}^{(2)} + \boldsymbol{z}^{(2)'}\boldsymbol{z}_1^{(1)}) - \frac{\gamma_2 p_1 - 2}{\gamma_1}\boldsymbol{z}_2^{(1)'}\boldsymbol{z}_2^{(1)} \\
&\quad + \mathrm{tr}(\sqrt{\gamma_1}V_{11}^{(1)} + \sqrt{\gamma_2}V^{(2)})^2\Lambda - \gamma_1\gamma_2 \mathrm{tr}\left(\frac{1}{\sqrt{\gamma_1}}\boldsymbol{z}_1^{(1)} - \frac{1}{\sqrt{\gamma_2}}\boldsymbol{z}^{(2)}\right)\left(\frac{1}{\sqrt{\gamma_1}}\boldsymbol{z}_1^{(1)} - \frac{1}{\sqrt{\gamma_2}}\boldsymbol{z}^{(2)}\right)'\Lambda \\
&\quad + \frac{1}{\gamma_1}\Big(\boldsymbol{z}_1^{(1)'}V_{11}^{(1)}V_{12}^{(1)}\boldsymbol{z}_2^{(1)} + \boldsymbol{z}_1^{(1)'}V_{12}^{(1)}V_{21}^{(1)}\boldsymbol{z}_1^{(1)} + \boldsymbol{z}_1^{(1)'}V_{12}^{(1)}V_{22}^{(1)}\boldsymbol{z}_2^{(1)} + \boldsymbol{z}_2^{(1)'}V_{21}^{(1)}V_{11}^{(1)}\boldsymbol{z}_1^{(1)} \\
&\quad + \boldsymbol{z}_2^{(1)'}V_{22}^{(1)}V_{21}^{(1)}\boldsymbol{z}_1^{(1)} + \boldsymbol{z}_2^{(1)'}V_{21}^{(1)}V_{12}^{(1)}\boldsymbol{z}_2^{(1)} + \boldsymbol{z}_2^{(1)'}(V_{11}^{(1)})^2\boldsymbol{z}_2^{(1)}\Big)\Bigg\} + Q_1^2.
\end{aligned}
$$

By calculating the expectation of T^4, we have

$$\mathrm{E}[T^4] = \xi_2 + \mathrm{O}(N^{-2}),$$

where

$$
\begin{aligned}
\xi_2 &= p(p+2) + \frac{1}{N}c_2, \\
c_2 &= 2p_1\{(p+3)(p_1+2)+4\} + \frac{2p_2}{\gamma_1}\left\{(p+2)(p+4) + 2p_1(\gamma_1-\gamma_2) + p_1(p-\gamma_2 p+1)\right\}.
\end{aligned}
$$

Therefore, using the asymptotic expansions of $\mathrm{E}[T^2]$ and $\mathrm{E}[T^4]$, we obtain the following result.

Theorem 3.1. *Suppose that the data have two-step monotone missing pattern. Then, an asymptotic variance of T^2 is given by*

$$\mathrm{Var}[T^2] = \nu + \mathrm{o}(N^{-2}),$$

where

$$\nu = 2p + \frac{1}{N}(c_2 - 2pc_1) - \frac{1}{N^2}c_1^2.$$

4 Transformation Statistic

In this section, we discuss the transformation statistics following the Bartlett correction. Bartlett [18] proposed the well-known Bartlett correction, and the modified Bartlett correction was developed by Fujikoshi [19]. They are configured to vanish in the terms of order N^{-1} from the expectation of the statistic. The Bartlett corrected statistic for the T^2 type statistic is given by

$$T_{BC}^2 = \left(1 - \frac{1}{Np}c_1\right)T^2, \; Np > c_1.$$

For the statistic T_{BC}^2, we have

$$\mathrm{E}[T_{BC}^2] = p + \mathrm{O}(N^{-2}),$$
$$\mathrm{Var}[T_{BC}^2] = 2p + \frac{1}{N}\{c_2 - 2(p+2)c_1\} + \mathrm{O}(N^{-2}).$$

By using the idea of Fujikoshi (2000), consider the following transformation:

$$T_T^2 = \frac{Np(p+2)}{c_1}\log\left(1 + \frac{c_1}{Np(p+2)}T^2\right),$$

for the statistic T_T^2, we have

$$\mathrm{E}[T_T^2] = p + \mathrm{O}(N^{-2}),$$
$$\mathrm{Var}[T_T^2] = 2p + \frac{1}{N}\{c_2 - (p+4)c_1\} + \mathrm{O}(N^{-2}).$$

5 Numerical Studies

This section investigates the first and second moments, variance, MSE, and type I errors of T^2, T_{BC}^2, and T_T^2 evaluated via a Monte Carlo Simulation. Meanwhile, we calculate the asymptotic results of ξ_1, ξ_2, and ν stated in Section 2 and 3, and η. Note that

$$\mathrm{MSE} = \eta + \mathrm{o}(N^{-2}),$$

where

$$\eta = \mathrm{Var}[T^2] + (p - E[T^2])^2.$$

Computations are carried out for $p = 4, 8, 20$, where the equal missing pattern has $p_1 = p_2$, and the various conditions of p, n, n_1, n_2, and $\alpha = 0.05, 0.01$. We generate two-step missing data form $N_p(\mathbf{0}, I_p)$ and as numerical studies, we carry out 1,000,000 replications. The simulation results are shown in Tables 1-3.

Table 1. First and second moments of T^2 type statistic.

						First Moment				Second Moment			
p	p_1	p_2	n	n_1	n_2	$\mathrm{E}[T^2]$	$\mathrm{E}[T^2_{BC}]$	$\mathrm{E}[T^2_T]$	ξ_1	$\mathrm{E}[T^4]$	$\mathrm{E}[T^4_{BC}]$	$\mathrm{E}[T^4_T]$	ξ_2
4	2	2	20	10	10	8.21	3.19	5.12	6.44	176.81	26.74	39.55	56.89
$\mathrm{E}[\chi^2_4] = 4$			40	20	20	5.25	3.73	4.44	5.16	46.52	23.48	29.16	39.58
$\mathrm{E}[\chi^4_4] = 24$			80	40	40	4.52	3.88	4.19	4.56	31.97	23.59	26.10	31.59
			100	50	50	4.41	3.91	4.15	4.45	30.02	23.66	25.63	30.04
			160	80	80	4.25	3.95	4.10	4.28	27.56	23.86	25.04	27.75
			200	100	100	4.19	3.96	4.08	4.22	26.75	23.86	24.79	26.99
			300	150	150	4.12	3.97	4.04	4.15	25.70	23.84	24.45	25.99
			400	200	200	4.09	3.98	4.03	4.11	25.25	23.87	24.32	25.49
			30	20	10	5.39	3.62	4.45	5.31	49.38	22.36	29.13	41.68
			60	40	20	4.58	3.85	4.20	4.63	32.99	23.35	26.22	32.58
			120	80	40	4.27	3.94	4.10	4.31	27.95	23.75	25.08	28.23
			240	160	80	4.13	3.97	4.05	4.16	25.90	23.93	24.56	26.10
			480	320	160	4.06	3.98	4.02	4.08	24.88	23.93	24.24	25.05
			30	10	20	7.78	3.64	5.12	6.13	168.22	36.88	40.40	52.44
			60	20	40	5.11	3.82	4.41	5.01	43.58	24.31	29.00	37.55
			120	40	80	4.47	3.91	4.18	4.49	31.06	23.86	26.03	30.62
			240	80	160	4.21	3.96	4.08	4.24	27.08	23.87	24.90	27.27
			480	160	320	4.10	3.98	4.04	4.12	25.44	23.92	24.42	25.63
8	4	4	20	10	10	416.06	23.11	18.34	15.56	2.42‡	7.47†	474.41	242.67
$\mathrm{E}[\chi^2_8] = 8$			40	20	20	13.70	7.57	10.05	11.58	282.90	86.40	126.98	157.05
$\mathrm{E}[\chi^4_4] = 80$			80	40	40	9.95	7.78	8.79	9.74	129.99	79.50	96.18	117.54
			100	50	50	9.47	7.83	8.61	9.39	116.22	79.40	92.20	109.88
			160	80	80	8.84	7.89	8.36	8.86	99.76	79.44	86.96	98.53
			200	100	100	8.66	7.92	8.28	8.69	95.28	79.62	85.51	94.79
			300	150	150	8.43	7.95	8.19	8.46	89.70	79.76	83.59	89.83
			400	200	200	8.32	7.96	8.14	8.34	87.09	79.81	82.64	87.36
			30	20	10	14.40	7.15	10.12	12.03	315.25	77.63	127.90	167.13
			60	40	20	10.18	7.69	8.85	9.96	136.95	78.18	97.34	122.28
			120	80	40	8.94	7.87	8.39	8.96	102.37	79.19	87.68	100.83
			240	160	80	8.44	7.94	8.19	8.48	90.05	79.60	83.61	90.34
			480	320	160	8.22	7.98	8.10	8.24	84.90	79.92	81.87	85.15
			30	10	20	385.37	67.29	18.89	14.60	2.28‡	6.96††	523.95	221.40
			60	20	40	13.08	7.94	9.98	11.14	255.53	94.17	126.37	147.33
			120	40	80	9.74	7.87	8.74	9.53	123.88	80.90	95.23	112.88
			240	80	160	8.76	7.93	8.33	8.76	97.45	79.84	86.42	96.25
			480	160	320	8.35	7.96	8.15	8.38	87.82	79.74	82.87	88.08
20	10	10	80	40	40	37.35	20.59	26.90	28.97	1692.59	514.40	804.10	849.23
$\mathrm{E}[\chi^2_{20}] = 20$			100	50	50	31.30	20.12	24.77	27.14	1142.22	472.04	677.70	765.71
$\mathrm{E}[\chi^4_{20}] = 440$			160	80	80	25.54	19.88	22.47	24.43	735.71	445.86	555.22	642.03
			200	100	100	24.14	19.87	21.87	23.54	652.59	442.27	525.58	601.21
			300	150	150	22.54	19.89	21.17	22.45	564.74	439.87	492.50	552.22
			400	200	200	21.83	19.91	20.84	21.76	528.02	439.24	477.26	520.20
			120	80	40	26.20	19.72	22.67	24.95	776.86	440.18	565.09	666.05
			240	160	80	22.61	19.84	21.17	22.45	568.86	437.78	492.50	552.22
			480	320	160	21.22	19.93	20.56	21.22	498.02	438.97	464.77	495.91
			120	40	80	35.45	21.39	26.63	27.93	1518.27	552.85	791.36	800.56
			240	80	160	24.94	20.05	22.30	23.92	699.54	452.21	546.84	618.20
			480	160	320	22.11	19.95	21.00	21.95	542.11	441.62	484.22	528.59

Note : The symbol "†", "††", and "‡" means that value $\times 10^6$, value $\times 10^7$, and value $\times 10^9$, respectively.

Table 2. Variance and MSE of T^2 Type Stastic.

						Variance				MSE			
p	p_1	p_2	n	n_1	n_2	Var$[T^2]$	Var$[T^2_{BC}]$	Var$[T^2_T]$	ν	Var$[T^2]$	Var$[T^2_{BC}]$	Var$[T^2_T]$	η
4	2	2	20	10	10	109.35	16.54	13.35	15.36	127.11	17.19	14.60	21.33
Var$[\chi^2_4] = 8$			40	20	20	18.92	9.55	9.47	12.98	20.49	9.62	9.67	14.32
			80	40	40	11.56	8.53	8.54	10.76	11.82	8.54	8.58	11.08
			100	50	50	10.61	8.36	8.38	10.25	10.78	8.37	8.41	10.45
			160	80	80	9.54	8.26	8.26	9.44	9.60	8.26	8.27	9.52
			200	100	100	9.17	8.18	8.18	9.16	9.20	8.18	8.19	9.21
			300	150	150	8.72	8.09	8.09	8.78	8.73	8.09	8.09	8.81
			400	200	200	8.52	8.06	8.06	8.59	8.53	8.06	8.06	8.60
			30	20	10	20.37	9.22	9.35	13.51	22.29	9.36	9.55	15.22
			60	40	20	12.03	8.52	8.55	11.10	12.37	8.54	8.59	11.50
			120	80	40	9.72	8.26	8.27	9.63	9.79	8.26	8.28	9.72
			240	160	80	8.83	8.16	8.16	8.83	8.85	8.16	8.16	8.86
			480	320	160	8.37	8.05	8.06	8.42	8.38	8.05	8.06	8.43
			30	10	20	107.77	23.63	14.14	14.90	122.02	23.76	15.40	19.43
			60	20	40	17.47	9.74	9.53	12.42	18.70	9.78	9.70	13.44
			120	40	80	11.12	8.54	8.54	10.42	11.34	8.55	8.57	10.66
			240	80	160	9.32	8.22	8.22	9.26	9.37	8.22	8.23	9.31
			480	160	320	8.60	8.08	8.09	8.64	8.61	8.08	8.09	8.65
8	4	4	20	10	10	2.42‡	7.47†	138.20	242.67	2.42‡	7.47†	245.03	57.78
Var$[\chi^2_8] = 16$			40	20	20	95.19	29.07	26.08	22.98	127.69	29.25	30.26	35.79
			80	40	40	31.04	18.98	18.90	22.60	34.83	19.03	19.52	25.64
			100	50	50	26.53	18.12	18.11	21.75	28.69	18.15	18.48	23.67
			160	80	80	21.53	17.15	17.15	20.02	22.24	17.16	17.28	20.76
			200	100	100	20.21	16.89	16.89	19.33	20.65	16.90	16.97	19.80
			300	150	150	18.61	16.55	16.55	18.32	18.80	16.55	16.59	18.52
			400	200	200	17.92	16.42	16.42	17.77	18.02	16.42	16.44	17.89
			30	20	10	107.93	26.58	25.46	22.41	148.87	27.31	29.96	38.65
			60	40	20	33.36	19.04	19.05	23.16	38.10	19.14	19.77	26.99
			120	80	40	22.37	17.31	17.31	20.48	23.26	17.32	17.46	21.41
			240	160	80	18.78	16.60	16.60	18.46	18.97	16.60	16.64	18.69
			480	320	160	17.33	16.31	16.32	17.28	17.38	16.31	16.32	17.34
			30	10	20	2.28‡	6.96††	166.99	8.14	2.28‡	6.96††	285.65	51.75
			60	20	40	84.33	31.08	26.68	23.15	110.18	31.08	30.62	33.03
			120	40	80	29.06	18.98	18.82	21.96	32.08	18.99	19.37	24.32
			240	80	160	20.74	16.99	16.99	19.54	21.31	17.00	17.11	20.11
			480	160	320	18.09	16.43	16.43	17.90	18.22	16.43	16.45	18.04
20	10	10	80	40	40	297.85	90.52	80.35	9.72	598.74	90.87	128.00	90.26
Var$[\chi^2_{20}] = 40$			100	50	50	162.63	67.21	64.12	28.98	290.29	67.23	86.88	80.00
			160	80	80	83.42	50.56	50.23	45.18	114.11	50.57	56.34	64.81
			200	100	100	69.85	47.34	47.22	47.30	86.99	47.35	50.72	59.80
			300	150	150	56.79	44.23	44.21	47.64	63.23	44.24	45.57	53.15
			400	200	200	51.61	42.93	42.93	46.76	54.94	42.94	43.64	49.85
			120	80	40	90.33	51.18	50.98	43.79	128.79	51.26	58.13	68.24
			240	160	80	57.52	44.27	44.26	47.99	64.35	44.29	45.64	54.02
			480	320	160	47.52	41.89	41.89	45.49	49.02	41.89	42.21	46.98
			120	40	80	261.77	95.32	82.17	126.37	500.38	97.25	126.13	83.30
			240	80	160	77.51	50.11	49.65	46.04	101.92	50.11	54.93	61.41
			480	160	320	53.33	43.44	43.42	46.85	57.77	43.44	44.41	50.64

Note : The symbol "†", "††", and "‡" means that value $\times 10^6$, value $\times 10^7$, and value $\times 10^9$, respectively.

Table 3. Type I error of T^2 Type Statistic

α							0.05			0.01	
p	p_1	p_2	n	n_1	n_2	T^2	T^2_{BC}	T^2_T	T^2	T^2_{BC}	T^2_T
4	2	2	20	10	10	0.264	0.048	0.115	0.159	0.022	0.034
			40	20	20	0.131	0.051	0.070	0.052	0.016	0.016
			80	40	40	0.084	0.051	0.058	0.025	0.012	0.012
			100	50	50	0.076	0.050	0.056	0.021	0.012	0.011
			160	80	80	0.066	0.051	0.054	0.016	0.011	0.011
			200	100	100	0.062	0.050	0.053	0.015	0.011	0.011
			300	150	150	0.058	0.050	0.052	0.013	0.010	0.010
			400	200	200	0.056	0.050	0.051	0.012	0.010	0.010
			30	20	10	0.140	0.048	0.070	0.058	0.015	0.015
			60	40	20	0.088	0.050	0.058	0.027	0.012	0.012
			120	80	40	0.067	0.050	0.054	0.017	0.011	0.011
			240	160	80	0.058	0.050	0.052	0.013	0.011	0.011
			480	320	160	0.054	0.050	0.051	0.011	0.010	0.010
			30	10	20	0.243	0.061	0.113	0.140	0.029	0.038
			60	20	40	0.122	0.054	0.070	0.047	0.016	0.016
			120	40	80	0.080	0.051	0.058	0.023	0.012	0.012
			240	80	160	0.064	0.050	0.053	0.015	0.011	0.011
			480	160	320	0.056	0.050	0.051	0.012	0.010	0.010
8	4	4	20	10	10	0.773	0.089	0.497	0.688	0.070	0.338
			40	20	20	0.306	0.071	0.135	0.177	0.031	0.044
			80	40	40	0.142	0.057	0.077	0.054	0.016	0.018
			100	50	50	0.118	0.056	0.070	0.041	0.015	0.016
			160	80	80	0.088	0.053	0.061	0.025	0.013	0.013
			200	100	100	0.079	0.052	0.058	0.021	0.012	0.012
			300	150	150	0.069	0.052	0.056	0.017	0.011	0.011
			400	200	200	0.063	0.051	0.054	0.015	0.011	0.011
			30	20	10	0.335	0.061	0.137	0.201	0.026	0.043
			60	40	20	0.154	0.056	0.079	0.062	0.017	0.019
			120	80	40	0.093	0.053	0.062	0.028	0.013	0.013
			240	160	80	0.069	0.052	0.056	0.017	0.012	0.012
			480	320	160	0.059	0.051	0.053	0.013	0.011	0.011
			30	10	20	0.742	0.225	0.499	0.653	0.180	0.350
			60	20	40	0.281	0.080	0.133	0.156	0.035	0.045
			120	40	80	0.131	0.058	0.075	0.048	0.017	0.018
			240	80	160	0.083	0.053	0.059	0.023	0.012	0.012
			480	160	320	0.065	0.051	0.054	0.015	0.011	0.011
20	10	10	80	40	40	0.572	0.115	0.272	0.406	0.055	0.119
			100	50	50	0.424	0.090	0.188	0.256	0.036	0.068
			160	80	80	0.229	0.067	0.109	0.100	0.020	0.030
			200	100	100	0.178	0.062	0.092	0.068	0.017	0.023
			300	150	150	0.122	0.057	0.074	0.039	0.014	0.017
			400	200	200	0.099	0.055	0.067	0.028	0.013	0.015
			120	80	40	0.253	0.066	0.116	0.118	0.020	0.032
			240	160	80	0.125	0.056	0.074	0.040	0.014	0.017
			480	320	160	0.082	0.053	0.061	0.021	0.012	0.013
			120	40	80	0.526	0.130	0.261	0.359	0.063	0.115
			240	80	160	0.206	0.069	0.105	0.086	0.020	0.028
			480	160	320	0.108	0.056	0.070	0.032	0.013	0.016

From these numerical studies, it can be observed that our asymptotic results are considerably closer to the simulated values of T^2 than any transformation statistic stated in Tables 1 and 2. Meanwhile, T_T^2 values are stable and closer to our asymptotic results than T_{BC}^2 when the sample sizes are not large. Table 3 presents the type I error rate under the simulated T^2 type statistic when the null hypothesis is rejected using χ_p^2. From Table 3, we note that T_{BC}^2 is a very good value in any case, and T_T^2 is a good value when n_1 and n_2 are large. In particular, if n_1 is small and n_2 is considerably smaller than n_1, type I errors are less than the true significance level when $p = 4$. It can be observed from Table 3 that the null distribution of the transformation statistics are closer to chi-squared distribution.

6 Concluding Remarks

In this paper, we considered the problem of testing for mean vector when the data have two-step monotone missing pattern. We obtained an stochastic expansion of Hotelling's T^2 type statistic with the use of MLEs developed by Kanda and Fujikoshi [2], and we showed asymptotic expectation and variance. Moreover, we proposed a new transformation statistic T_T^2 and summarized these results with the Bartlett corrected statistic T_{BC}^2. We presented the first and second moments, variance, and MSE of T^2, T_{BC}^2, and T_T^2 through the Monte Carlo simulation and the asymptotic results. Additionally, we showed type I errors when the null hypothesis was rejected using χ_p^2 under the simulated T^2, T_{BC}^2, and T_T^2. In conclusion, it may be seen from the tables that the asymptotic results of first and second moments, variance, and MSE are considerably closer to the simulated values than any transformation statistics. T_{BC}^2 is a very good value in any case, T_T^2 is a stable value, and the type I error rates show that the null distribution of the transformation statistics are closer to chi-square distribution.

Acknowledgments

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 where some observations are missing. *J. Amer. Statist. Assoc.*, 52, 200–203.

[2] Kanda, T. and Fujikoshi, Y. (1998). Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data. *Amer. J. Math. Management Sci.*, 18, 161–190.

[3] Jinadasa, K. G. and Tracy, D.S. (1992). Maximum likelihood estimation for multivariate normal distribution with monotone sample. *Comm. Statist. Theory Methods*, 21(1), 41–50.

[4] Srivastava, M. S. (1985). Multivariate data with missing observations. *Comm. Statist. Theory Methods*, 14, 775–792.

[5] Srivastava, M. S. and Carter, E. M. (1986). The maximum likelihood method for non-response in sample survey. *Survey Methodology*, 12, 61–72.

[6] Shutoh, N., Kusumi, M., Morinaga, W., Yamada, S. and Seo, T. (2010). Testing equality of mean vectors in two sample problems with missing data. *Comm. Statist. Simulation Comput.*, 39, 487–500.

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

[8] Seko, N., Kawasaki, T. and Seo, T. (2011). Testing equality of two mean vectors with two-step monotone missing data. *Amer. J. Math. Management Sci.*, 31, 117–135.

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

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

[11] Seo, T. and Srivastava, M. S. (2000). Testing equality of means and simultaneous confidence intervals in repeated measures with missing data. *Biometrical J.*, 42, 981–993.

[12] Wu, Y., Genton, M.G. and Stefanski, A. (2006). A multivariate two-sample mean test for small sample size and missing data. *Biometrics*, 62, 877–885.

[13] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581–592.

[14] Little, R. J. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley.

[15] Anderson, T. W. and Olkin, I. (1985). Maximum-likelihood estimation of the parameters of a multivariate normal distribution. *Linear Algebra Appl.*, 70, 147–171.

[16] Chang, W.-Y. and Richard, D. St. P. (2009). Finite-sample inference with monotone incomplete multivariate normal data I. *J. Multivariate Anal.*, 100, 1883–1899.

[17] Seko, N., Yamazaki, A. and Seo, T. (2012). Tests for mean vector with two-step monotone missing data. *SUT J. Math.*. 48, No.1, 13–38.

[18] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. London Ser. A*, 160, 268–282.

[19] Fujikoshi, Y. (2000). Transformations with improved chi-squared approximations. *J. Multivariate Anal.*, 72, 249–263.