Assessing Performance of Tests for Equality of High-Dimensional Mean Vectors with Unequal Block Diagonal Covariances

Knavoot Jiamwattanapong[a]*, Nisanad Ingadapa[b], Jitkawee Krachangmek[c], Piyada Phrueksawatnon[d]

*knavoot.jia@rmutr.ac.th
[a,b,c]Faculty of Liberal Arts, Rajamangala Unversity of Technology Rattanakosin, 96 M.3 Phutthamonthon Sai 5, Salaya, Nakhon Pathom 73170, Thailand
[d]School of Science, University of Phayao, Mae Ka, Phayao 56000, Thailand

Abstract

In many high-dimensional applications, it is of interest to test whether the mean vectors of two populations are equal. However, the presence of unknown and unequal block diagonal covariance matrices can complicate the testing procedure, and choosing an appropriate test becomes challenging. The study aimed to investigate the performance of tests for equality of two high-dimensional mean vectors with unknown and unequal block diagonal covariance matrices. The study focused on three tests: T_1 proposed by Srivastava, Katayama, and Kano (2013), T_2 proposed by Hu, Bai, Wang, and Wang (2017), and T_3 proposed by Ahmad (2019). The study included both cases: equal and unequal sample sizes. The effect of block size in the covariance matrix on the performance of the tests was also studied. The data was gathered using two independent high-dimensional samples based on multivariate normality and unequal covariance matrices with block diagonal structure. The number of variables studied ranged from 50 - 500 and the sample size was 20 - 200. The results showed that, for equal sample sizes, both of the tests T_1 and T_2 performed well, and when the sample size exceeds 20, the test T_1 performed slightly higher than T_2. When two sample sizes were unequal, the test T_2 outperformed the tests T_1 and T_3. A study of the effect of block size discovered that larger block sizes resulted in poor test performance and the influence of block size diminishes as sample size increases.

Keywords: High-dimensional mean vectors; Unequal covariance matrices; Block diagonal covariances; Covariance matrix block size; high-dimensional data

1. Introduction

In recent years, rapid advancements in measurement technology, data storage, and processing have resulted in the proliferation of big datasets. These datasets play a crucial role in the efficient operation of numerous organizations in various domains and have important implications for research and the quality of human life (Laney, 2001). To draw reliable conclusions and reduce risks associated with decision-making, statistical methods for analyzing big data are vital. For example, predicting drug effectiveness in personalized medicine based on the DNA data of individual patients requires statistical analysis (Datta et al., 2018).

High-dimensional data, which are characterized by a large number of variables of interest and a small sample size, are increasingly being collected in various fields, including medicine, genetics, and economics (Fan & Lv,

2010). However, analyzing such data is complex and presents unique challenges. Existing multivariate data analysis methods, such as the comparison of mean vectors through Hotelling's T^2 test (Hotelling, 1931), are not applicable to high-dimensional data due to the curse of dimensionality and increased false positives (Pandit et al., 2020).

Equivalence testing of two high-dimensional population mean vectors typically relies on a multivariate distribution, and the structure of the common variance matrix can affect the performance of tests (Fan et al., 2020). Specifically, when the covariance matrix has a block diagonal structure, the block size can impact the performance of tests by introducing spurious correlations and inflating test statistics (Barnett & Onofrei, 2018). However, there is currently limited research on mean vector testing in this context, which can result in users choosing inappropriate methods and ultimately yielding unreliable conclusions (Shi et al., 2019).

To address these limitations, this research aims to study the performance of two tests for comparing mean vectors when the covariance matrices are unequal and possess block diagonal structures, as well as examining the impact of block size on the tests. The research will be conducted on two independent and randomly selected samples from a multivariate normally distributed population with an unequal and unknown covariance matrix and a block diagonal covariance structure.

2. Tests for Two High-Dimensional Mean Vectors with Unequal Covariance Matrices

2.1 Chen and Qin's Test

When comparing the mean vectors of two populations with unequal and unknown covariance matrices ($\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$), Chen and Qin (2010) proposed an important testing method. This method builds upon the work of Bai and Saranadasa (1996), who developed a method for the case where the covariance matrices of both populations are assumed to be equal in a high-dimensional setting. Chen and Qin extended this method to handle the case of unequal covariance matrices.

For testing the hypotheses $H: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $K: \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, the test proposed by Chen and Qin (2010), denoted by T_{CQ}, is defined as follows:

$$T_{CQ} = \frac{Q_n}{\hat{\sigma}_{CQ}} \ ,$$

where

$$Q_n = \left[(\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2)'(\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2) - \frac{tr(\mathbf{S}_1)}{n_1} - \frac{tr(\mathbf{S}_2)}{n_2} \right] / \sqrt{p} \ ,$$

$$\hat{\sigma}^2_{CQ} = \frac{1}{p}\left[\frac{2}{n_1(n_1-1)} tr(\boldsymbol{\Sigma}_1^2) + \frac{2}{n_2(n_2-1)} tr(\boldsymbol{\Sigma}_2^2) + \frac{4}{n_1 n_2} tr(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2) \right],$$

$$tr(\boldsymbol{\Sigma}_i^2) = \frac{1}{n_i(n_i-1)} \left\{ \sum_{j \neq k}^{n_i} (\mathbf{x}_{ij} - \tilde{\mathbf{x}}_{i(j,k)}) \mathbf{x}_{ij}^{'} (\mathbf{x}_{ij} - \tilde{\mathbf{x}}_{i(j,k)}) \mathbf{x}_{ik}^{'} \right\}, \ i = 1, 2,$$

$$tr(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2) = \frac{1}{n_1 n_2} tr\left[\sum_{j=1}^{n_1} \sum_{k=1}^{n_2} (\mathbf{x}_{1j} - \tilde{\mathbf{x}}_{1(j)}) \mathbf{x}'_{1j} (\mathbf{x}_{2k} - \tilde{\mathbf{x}}_{2(k)}) \mathbf{x}'_{2k} \right],$$

$$\tilde{\mathbf{x}}_{i(j,k)} = \frac{1}{n_i - 2}(n_i\tilde{\mathbf{x}}_i - \mathbf{x}_{ij} - \mathbf{x}_{ik}),\ i = 1,2;\ j,k = 1,2,...,n_i,$$

and $$\tilde{\mathbf{x}}_{2(k)} = \frac{1}{n_i}(n_i\tilde{\mathbf{x}}_i - \mathbf{x}_{ik}),\ i = 1,2;\ k = 1,2,...,n_i$$

The test using the T_{CQ} statistic will reject the null hypothesis at a significance level of α when the value of $T_{CQ} > Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)\%$ quantile of the standard normal distribution.

The assumptions of T_{CQ} are
(1) $n_1/(n_1 + n_1) \to k \in (0,1)$, where $n \to \infty$
(2) $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_i(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o(n^{-1}tr\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}), i = 1,2$

The test proposed by Chen and Qin (2010) is known to be invariant under orthogonal transformation. However, it is not invariant under scalar transformation, and the calculation to obtain the test statistic can be quite complex.

2.2 Modified Chen and Qin' s Test

Although the T_{CQ} test proposed by Chen and Qin (2010) is more widely applicable than the test of Bai and Saranadasa (1996) because it does not require the same covariance matrices for both populations, the estimator $\hat{\sigma}^2_{CQ}$ in (3) is still not very efficient. To improve the performance of the test, Srivastava et al. (2013) suggested using the UMVUE (Uniformly Minimum Variance Unbiased Estimator) under the normal distribution of $tr(\Sigma_i^2)\ /\ p$ instead of the estimator $\hat{\sigma}^2_{CQ}$ used in (3). The estimator proposed by Srivastava et al. (2013), denoted by $\hat{\sigma}^2_U$, is as follows:

$$\hat{\sigma}^2_U = \frac{2}{n_1^2}\hat{a}_{21} + \frac{2}{n_2^2}\hat{a}_{22} + \frac{4}{pn_1n_2}tr(\mathbf{S}_1\mathbf{S}_2),$$

where $$\hat{a}_{2i} = \frac{(n_1 - 1)^2}{pn_1(n_1 - 2)}\left[tr(\mathbf{S}_i^2) - \frac{1}{n_i - 1}\left(tr(\mathbf{S}_i)\right)^2\right],\ i = 1,2$$

The test proposed by Srivastava et al. (2013) modifies the estimator of the covariance matrix in the test statistic of Chen and Qin (2010) to improve its performance. The modified test statistic, denoted by T_{MCQ}, is defined as follows:

$$T_{MCQ} = \frac{Q_n}{\hat{\sigma}_{Q_n}},$$

where $$Q_n = (\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2)'(\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2) - \frac{tr(\mathbf{S}_1)}{n_1} - \frac{tr(\mathbf{S}_2)}{n_2},$$

$$\hat{\sigma}^2_{Q_n} = \frac{2}{n_1(n_1 - 1)}tr(\boldsymbol{\Sigma}_1^2) + \frac{2}{n_2(n_2 - 1)}tr(\boldsymbol{\Sigma}_2^2) + \frac{4}{n_1n_2}tr(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2),$$

$$tr(\boldsymbol{\Sigma}_i^2) = \frac{(n_i-1)^2}{(n_i+1)(n_i-2)}\left\{tr(\mathbf{S}_i^2) - \frac{1}{n_i-1}\left(tr(\mathbf{S}_i)\right)^2\right\},$$

$$\mathbf{S}_i = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)'(x_{ij}-\bar{x}_i) \ , i=1,2,$$

and $$tr(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) = tr(\mathbf{S}_1\mathbf{S}_2)$$

The test using the T_{MCQ} statistic will reject the null hypothesis at a significance level of α when the value of $T_{MCQ} > Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)\%$ quantile of the standard normal distribution.

While the T_{MCQ} test has fewer initial assumptions compared to the Bai and Saranadasa (1996) test, it is only invariant under orthogonal transformations and not invariant under scalar transformations, which can be a limitation. Additionally, the calculation to obtain the test statistic can be quite complex.

2.3 Srivastava, Katayama and Kano's Test

One important testing method proposed by Srivastava et al. (2013) is the T_1 test, which has the property of being invariant under scalar transformation. The T_1 test statistic is as follows:

$$T_1 = \frac{(\bar{x}_1-\bar{x}_2)'\hat{\mathbf{D}}^{-1}(\bar{x}_1-\bar{x}_2)-p}{\left[pVar(\hat{q}_n)(1+\frac{tr(\mathbf{R}^2)}{p^{3/2}})\right]^{1/2}},$$

where $$\hat{\mathbf{D}} = \frac{\hat{\mathbf{D}}_1}{n_1}+\frac{\hat{\mathbf{D}}_2}{n_2}, \ \hat{\mathbf{D}}_i = diag(s_{i11}, s_{i22}, ..., s_{ipp}) \ , \ i=1,2 \ ,$$

s_{ikk} are elements on the main diagonal of the pooled sample covariance matrix $\mathbf{S}_i$,

$$Var(\hat{q}_n) = \frac{2tr(\mathbf{R}^2)}{p} - \frac{2\left[tr(\hat{\mathbf{D}}^{-1}\mathbf{S}_1)\right]^2}{pn_1^2(n_1-1)} - \frac{2\left[tr(\hat{\mathbf{D}}^{-1}\mathbf{S}_2)\right]^2}{pn_2^2(n_2-1)},$$

and $$\mathbf{R} = \hat{\mathbf{D}}^{-1/2}\left(\frac{\mathbf{S}_1}{n_1}+\frac{\mathbf{S}_2}{n_2}\right)\hat{\mathbf{D}}^{-1/2}$$

The test using the T_1 statistic will reject the null hypothesis at a significance level of α when the value of $T_1 > Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)\%$ quantile of the standard normal distribution.

The assumptions of T_1 are

(1) $0 < c_1 < \min_{1\le k\le p}\sigma_{ikk} \le \max_{1\le k\le p}\sigma_{ikk} < c_2 < \infty$

(2) $\lim_{p\to\infty} tr(\mathrm{R}^4)/[\mathrm{tr}(\mathrm{R}^2)]^2 = 0$ $n \to \infty$

(3) $n_1/(n_1+n_2) \to k \in (0,1)$, where $n = n_1 + n_2 \to \infty$

(4) $n_{\min} = O(p^{\delta})$, $\delta > 1/2$, $n_{\min} = \min(n_1, n_2)$

The important advantage that both T_1 and T_{MCQ} tests have in common is that they do not rely on any assumptions about the distribution of the population, meaning that they can be applied to non-normally distributed data.

2.4 Hu et al.'s Test

Hu et al. (2017) developed a test that extends the work of Chen and Qin (2010) for testing mean vectors in a single population, two populations, and three or more populations. This method does not assume normality of the data, and for the case of two populations, it uses the same test statistic as proposed by Chen and Qin (2010). However, the method for calculating the mean and variance of the test statistic is simpler than that used by Chen and Qin.

The T_2 test statistic proposed by Hu et al. (2017) is as follows:

$$T_2 = \frac{T_n}{\hat{\sigma}_n},$$

where $$T_n = \sum_{i=1}^{2} \frac{1}{n_i(n_i-1)} \sum_{k \neq r}^{n_i} \sum^{n_i} \mathbf{x}'_{ik}\mathbf{x}_{ir} - \frac{2}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} \mathbf{x}'_{1k}\mathbf{x}_{2l}$$

and $$\hat{\sigma}_n^2 = \sum_{i=1}^{2} \frac{2}{n_i(n_i-1)} \left(\frac{(n_i-1)^2}{(n_i+1)(n_i-2)} \right) \left(tr(S_i^2) - \frac{[tr(S_i)]^2}{n_i - 1} \right) + \sum_{i<j}^{2} \frac{4tr(S_i S_j)}{n_i n_j}$$

The test using the T_2 statistic will reject the null hypothesis at a significance level of α when the value of $T_2 > Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)\%$ quantile of the standard normal distribution.

The assumptions of T_2 are

(1) $\mathbf{X}_{ij} = \mathbf{\Gamma}_i \mathbf{Z}_{ij} + \boldsymbol{\mu}_i$, $i = 1,...,k$, $j = 1,...,n_i$, where $\mathbf{\Gamma}_i$ is a $p \times m$ matrix, $m \geq p$ such that $\mathbf{\Gamma}_i \mathbf{\Gamma}_i' = \mathbf{\Sigma}_i$, and $\{\mathbf{Z}_{ij}\}_{j=1}^{n_i}$ is m-variate i.i.d. with $E(\mathbf{Z}_{ij}) = 0$ and $Var(\mathbf{Z}_{ij}) = \mathbf{I}_m$, $\mathbf{I}_m$ is an identity matrix.

(2) $\mathbf{Z}_{ij} = (z_{ij1},..., z_{ijm})'$, where $E(z_{ijl_1}^{\alpha_1} z_{ijl_2}^{\alpha_2} ... z_{ijl_q}^{\alpha_q}) = E(z_{ijl_1}^{\alpha_1}) E(z_{ijl_2}^{\alpha_2}) ... E(z_{ijl_q}^{\alpha_q})$ and

$E(z_{ijk}^4) < \infty$ for positive integer q, where $\sum_{l=1}^{q} \alpha_l \leq 8$ and $l_1 \neq l_2 \neq ... \neq l_q$

(3) $n_i / \sum_{i=1}^{k} n_i \to k_i \in (0,1),\ i = 1,...,k,$ where $n \to \infty$

(4) $tr(\boldsymbol{\Sigma}_l \boldsymbol{\Sigma}_d \boldsymbol{\Sigma}_l \boldsymbol{\Sigma}_h) = o[tr(\boldsymbol{\Sigma}_l \boldsymbol{\Sigma}_d) tr(\boldsymbol{\Sigma}_l \boldsymbol{\Sigma}_h)]$, $d, l, h \in \{1, 2, ..., k\}$

(5) $(\boldsymbol{\mu}_d - \boldsymbol{\mu}_l)' \boldsymbol{\Sigma}_d (\boldsymbol{\mu}_d - \boldsymbol{\mu}_h) = o\left[n^{-1} tr\left\{ (\sum_{i=1}^{k} \boldsymbol{\Sigma}_i)^2 \right\} \right],\ d, l, h \in \{1, 2, ..., k\}$

2.5 Ahmad's Test

Ahmad (2019) proposed an interesting test that does not rely on assumptions about the distribution of data. This method is based on the development of a test statistic that differs from previously mentioned methods. According to Ahmad (2019), the proposed test statistic relies on fewer initial assumptions than other tests. Additionally, Ahmad (2019) provided a method for testing the mean vector in cases where there are more than two populations.

The T_3 test statistic proposed by Ahmad (2019) is as follows:

$$T_3 = \frac{pQ_0 / Q^*}{\sqrt{\hat{V}}},$$

where $n = n_1 + n_2,\ Q_0 = U_0 / p$,

$$U_0 = \sum_{i=1}^{2} U_{n_i} - 2U_{n_1} U_{n_2},$$

$$U_{n_i} = \frac{1}{n_i (n_i - 1)} \sum_{k=1}^{n_i} \sum_{r=1, k \neq r}^{n_i} \mathbf{x}_{ik}' \mathbf{x}_{ir},\ i = 1, 2,$$

$$U_{n_1} U_{n_2} = \frac{1}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} \mathbf{x}_{1k}' \mathbf{x}_{2l},$$

$$Q^* = \frac{tr(S_1)}{n_1} + \frac{tr(S_2)}{n_2},$$

$$E_{2i} = \left[\frac{n_i - 1}{n_i (n_i - 2)(n_i - 3)} \right] \left[(n_i - 1)(n_i - 2) tr(S_i^2) + (tr(S_i))^2 - n_i Q_i \right],$$

$$Q_i = \sum_{k=1}^{n_i} (\tilde{x}_{ik}' \tilde{x}_{ik})^2 / (n_i - 1),\ \tilde{x}_{ik} = x_{ik} - \bar{x}_i,\ i = 1, 2,$$

$$E_{3i} = \left[\frac{n_i - 1}{n_i(n_i-2)(n_i-3)}\right]\left[2tr(S_i^2) + (n_i{}^2 - 3n_i + 1)(tr(S_i))^2 - n_i Q_i\right],$$

$$Q_i = \sum_{k=1}^{n_i} (\tilde{x}_{ik}' \tilde{x}_{ik})^2 / (n_i - 1), \quad \tilde{x}_{ik} = x_{ik} - \bar{x}_i,\ i = 1,2\ ,$$

$$\hat{V} = \frac{2a}{b}\ ,$$

$$a = \frac{n^2}{p^2}\left[\frac{E_{21}}{n_1^2} + \frac{E_{22}}{n_2^2} + \frac{2tr(S_1S_2)}{n_1 n_2}\right],$$

and $$b = \frac{E_{31}}{n_1^2} + \frac{E_{32}}{n_2^2} + \frac{2tr(S_1)tr(S_2)}{n_1 n_2}.$$

The test using the T_3 statistic will reject the null hypothesis at a significance level of α when the value of $T_3 > Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)\%$ quantile of the standard normal distribution.

The assumptions of T_3 are

(1) $E(X_{iks}^4) = \gamma_{is} \le \gamma < \infty, \forall s = 1,..., p,\ i = 1,2,\ \gamma \in \mathbb{R}^+$

(2) $\lim_{p\to\infty} \sum_{i=1}^{p} \nu_{is} = \sum_{s=1}^{\infty} \nu_{is} = \nu_{i0} \in \mathbb{R}^+,\ i = 1,2$, where $\nu_{is} = \lambda_{is} / p$ are eigen values of $\boldsymbol{\Sigma}_i / p, i = 1,2$

(3) $\lim_{n_i, p\to\infty} p / n_i = c_i = O(1), i = 1,2$

(4) $\lim_{n_i\to\infty} n / n_i = \rho_i = O(1), n = n_1 + n_2,\ i = 1,2$

(5) $\lim_{p\to\infty} \boldsymbol{\mu}_i' \boldsymbol{\Sigma}_k \boldsymbol{\mu}_j / p = \phi_{ijk} \le \phi = O(1),\ i, j, k = 1,2$

The comparison of mean vectors between two high-dimensional populations with unequal covariance matrices has been addressed by various methods. For instance, Sukcharoen and Chongcharoen (2019) proposed a test suitable for cases with block diagonal covariance matrices, while Thonghnunui et al. (2020) presented a method assuming knowledge of the covariance matrix for one population and an unknown covariance matrix for the other. These approaches are rooted in the work of Jiamwattanapong and Chongcharoen (2015, 2017).

Despite their relevance, these tests rely on specific assumptions about the data distribution and can pose implementation challenges. Consequently, they were not considered in the present study. Instead, this research focuses on evaluating the performance of three alternative tests proposed by Srivastava et al. (2013), Hu et al. (2017), and Ahmad (2019). These tests are designed to compare high-dimensional mean vectors of two populations with unequal and block diagonal covariance matrices, addressing a gap in previous literature.

3. Simulation Procedure

A simulation procedure was conducted to evaluate the performance of three tests for the equality of population mean vectors, namely the T_1 test proposed by Srivastava et al. (2013), the T_2 test proposed by Hu et al. (2017), and the T_3 test proposed by Ahmad (2019). The sample data were drawn from the normal population with mean vector and block diagonal covariance matrix. The covariance matrices of both populations were set to be unequal but they shared the same covariance structure.

3.1 Data Simulation

This study utilized R version 4.2.0 to simulate data under the main and alternative assumptions. The study is divided into two parts. Part 1 examines the performance of the test under multivariate normal distribution with a block diagonal structure for both equal and unequal sample sizes, when the two populations have unequal covariance matrices. Part 2 investigates the impact of the block size on the efficiency of the test for multivariate normal distribution with a common covariance matrix.

Part 1: Study and Comparison of Three Tests

The study relied on data simulation under multivariate normal distribution. The first random sample, denoted as $x_{11}, x_{12}, \dots, x_{1n_1}$, was generated from a p-dimensional normal distribution $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, while the second independent sample, denoted as $x_{21}, x_{22}, \dots, x_{2n_2}$, was generated from $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, where $p > n$; $n = n_1 + n_2 - 2$, with $n_1 = n_2$.

Under the null hypothesis, the mean vectors of both populations were set to be equal ($\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$), while the covariance matrices of the two populations had different values but shared the same block-diagonal structure as follows:

$$\boldsymbol{\Sigma}_1 = \mathbf{D}^{1/2}\mathbf{R}_1\mathbf{D}^{1/2},$$

where $\mathbf{D} = diag(d_1, \dots, d_p)$, $d_i = 2 + \frac{p-i+1}{p}, i = 1,2, \dots, p$, $\mathbf{R_1} = diag(\mathbf{R_{11}}, \mathbf{R_{12}}, \dots, \mathbf{R_{1m}})$, $\mathbf{R_{1k}} = (r_{ij})$, $k = 1,2, \dots, m$, $r_{ij} = \begin{cases} 1 & , i = j \\ (-1)^{i+j} c^{|i-j|} & , i \neq j, \end{cases}$ and $c = 0.2$.

The covariance of the second population, $\boldsymbol{\Sigma}_2$, was set to have the same block diagonal structure as $\boldsymbol{\Sigma}_1$, with the only difference being the value of c, which was changed from 0.2 to 0.9.

The number of blocks in $\mathbf{R}_i, i = 1,2$ was m and the common block size was q, which meant that the size of each block (except for the last block) was q. The first m-1 blocks had the same size, which was q, while the size of the last block was $p - (m-1)q$.

Under the alternative hypothesis, we set $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\boldsymbol{\mu}_2 = (\delta_1, \delta_2, \dots, \delta_p)'$, where $\delta_{2k-1} = 0$ and $\delta_{2k} \sim U(-0.5, 0.5)$ for $k = 1,2, \dots p/2$. Each condition was iterated 5,000 times with a nominal significance level of 0.05. The study considered both cases where the sample sizes were equal and unequal as follows:

(1) In the case of equal sample sizes, the number of variables (p) and sample sizes (n_i) were set as follows: for p = 50, $n_1 = n_2$ = 20; for p = 100, $n_1 = n_2 \in \{20, 40\}$; for $p \in \{200, 300\}$, $n_1 = n_2 \in \{20, 40, 60\}$; for p = 400, $n_1 = n_2 \in \{20, 40, 60, 80\}$; for p = 500, $n_1 = n_2 \in \{20, 40, 60, 80, 100\}$ and the size of the common blocks (except the last block) in the covariance matrix was set to q, where $q = n_i/4$.

(2) In the case of unequal sample sizes, set $n_2 = 2n_1$ and the number of variables (p) and sample sizes (n_i) were set as follows: for p = 100, n_1 = 20; for $p \in \{200, 300\}$, $n_1 \in \{20, 40, 60\}$; for p = 400, $n_1 \in$ {20, 40,

60, 80}; for p = 500, n_1 ∈{20, 40, 60, 80, 100} and the size of the common block (except the last block) in the covariance matrix was set to q, where q is the maximum integer such that $q \leq (n_i - 1)/5$.

Part 2: To investigate the impact of the block size in the covariance matrix on the performance of the tests

Under the null and alternative hypotheses as in Part 1, the mean vectors were set the same values as in Part 1 and examined both cases where the sample sizes were equal and unequal as follows:

(1) In the case where the sample sizes were equal, we set the number of variables (p), the sample size (n_i), and the common block size (q) as follows: for p = 200, n_1 = 40, we set q ∈{5, 10}; for p = 400, we set n_1∈{40, 60} and q ∈{5, 10}. The size of each block (except for the last block) in the covariance matrix was equal to q, where $q = n_i/4$.

(2) In the case of unequal sample sizes, we set $n_2 = 2n_1$ and determined the number of variables (p), sample sizes (n_i), and block size (q) as follows. For p = 200 and n_1 = 40, we varied q among {3, 5, 7}. For p = 400 and n_1 = 40, we also varied q among {3, 5, 7}, whereas for p = 400 and n_1 = 60, we extended the set of q values to include {3, 5, 7, 11}. We divided the variables into blocks of common size q, where $q \leq (n_i - 1)/5$ with the exception of the last block, whose size was determined by the remaining variables.

3.2 Performance of the Test

In this study, the performance of the tests was measured using the attained significance level (ASL) and empirical power. The performance of the tests was evaluated by examining whether the ASL was within an acceptable range around the nominal level (i.e., 0.05) and by comparing the empirical power of different tests. Tests with ASL within an acceptable range were compared, and the test with the higher empirical power was considered more effective.

The ASL and empirical power were defined as follows:

(1) ASL is defined as

$$\text{ASL} = (number\ of\ times\ t_H > c)/m\ ,$$

where t_H is the test statistic obtained from simulating the data under the null hypothesis,
c is the critical value of the test,
and m is the total number of iterations in the simulation.

(2) Empirical power is defined as

$$\text{Empirical power} = (number\ of\ times\ t_K > c)/m\ ,$$

where t_K is the test statistic obtained from simulating the data under the alternative hypothesis,
c is the critical value of the test,
and m is the total number of iterations in the simulation.

When considering the ASL, Cochran's criterion (Cochran, 1954) was used at a nominal significance level. In this study, when the ASL value fell within the range of 0.040 to 0.060, the test was considered to have good performance in terms of controlling the type 1 error rate. In comparison, a test with higher empirical power is deemed better if it fell within the acceptable range of ASL.

4. Results

The study results are divided into 2 parts as follows. Part 1 is a study of the effectiveness of testing for equality of population mean vectors between 2 sets using 3 methods: test statistics T_1, T_2 and T_3. The results are obtained at a significance level of 0.05, when the sample sizes are equal as shown in Table 1, and when the sample sizes are unequal as shown in Table 2. Part 2 is a study of the impact of block size in the covariance matrix on the testing methods of population mean vectors using 3 methods: test statistics T_1, T_2 and T_3. The

results are shown in Tables 3 and 4.

Table 1. ASL and Empirical Power of the Tests with Equal Sample Sizes and Nominal Level 0.05

p	n_i $(n_1 = n_2)$	q	ASL			Empirical power		
			T_1	T_2	T_3	T_1	T_2	T_3
50	20	5	0.0620	0.0558	0.0242	0.1734	0.1662	0.0900
100	20	5	0.0800	0.0630	0.0142	0.2602	0.2210	0.0662
	40	10	0.0634	0.0682	0.0398	0.3980	0.4116	0.2880
200	20	5	0.0864	0.0624	0.0032	0.3988	0.3290	0.0392
	40	10	0.0628	0.0620	0.0202	0.6496	0.6426	0.3860
	60	15	0.0544	0.0568	0.0284	0.8838	0.8876	0.7734
300	20	5	0.0944	0.0590	0.0010	0.4996	0.4104	0.0198
	40	10	0.0682	0.0682	0.0116	0.8106	0.7938	0.4502
	60	15	0.0530	0.0578	0.0194	0.9704	0.9712	0.8898
400	20	5	0.0872	0.0532	0.0000	0.6016	0.4982	0.0168
	40	10	0.0632	0.0578	0.0044	0.8994	0.8878	0.5052
	60	15	0.0600	0.0608	0.0128	0.9940	0.9932	0.9422
	80	20	0.0538	0.0608	0.0196	1.0000	1.0000	0.9988
500	20	5	0.0894	0.0466	0.0000	0.6898	0.5788	0.0074
	40	10	0.0698	0.0644	0.0030	0.9530	0.9406	0.5466
	60	15	0.0652	0.0680	0.0122	0.9999	0.9988	0.9724
	80	20	0.0516	0.0556	0.0134	1.0000	1.0000	0.9996
	100	25	0.0530	0.0590	0.0216	1.0000	1.0000	1.0000

Table 1 presents the results of three tests (T_1, T_2, and T_3) with equal sample sizes ($n_1 = n_2$) and a nominal level of 0.05. The results show that as p increases, the ASL values generally decrease, indicating that it becomes more difficult to detect differences between the mean vectors of the two populations as the number of variables increases. Based on the ASL value, the T_2 test is closer to the nominal level of 0.05 compared to the other tests. Additionally, when considering the empirical power of T_2, it was found to increase towards 1 as the sample size increased. On the other hand, the T_1 test had a different ASL value from 0.05 when the sample size was 20, but its performance improved as the sample size increased beyond 20. The empirical power values generally increase as the sample size increases. The empirical power of T_2 was slightly higher than T_1 when the number of variables and sample size were the same.

As for the T_3 test, its performance was found to be unacceptable in all studied scenarios. Overall, the results suggest that T_1 and T_2 are more reliable tests for detecting differences between the mean vectors than T_3, particularly when the sample size is relatively small and the number of variables is large.

Table 2. ASL and Empirical Power of the Tests with Unequal Sample Sizes and Nominal Level 0.05

p	n_1	n_2	q	ASL			Empirical power		
				T_1	T_2	T_3	T_1	T_2	T_3
100	20	40	3	0.0052	0.0660	0.0056	0.1050	0.3878	0.1030
200	20	40	3	0.0024	0.0572	0.0010	0.1332	0.5850	0.0646
	40	80	7	0.0032	0.0534	0.0072	0.6126	0.9294	0.7310
	60	120	11	0.0034	0.0596	0.0158	0.9250	0.9974	0.980
300	20	40	3	0.0010	0.0554	0.0002	0.1532	0.7178	0.0418
	40	80	7	0.0018	0.0570	0.0036	0.7888	0.9868	0.8240
	60	120	11	0.0018	0.0550	0.0082	0.9902	0.9996	0.9976
400	20	40	3	0.0004	0.0582	0.0000	0.1550	0.8154	0.0246
	40	80	7	0.0016	0.0514	0.0012	0.9042	0.9964	0.8910
	60	120	11	0.0016	0.0530	0.0060	0.9970	1.0000	0.9992
	80	160	15	0.0020	0.0548	0.0096	1.0000	1.0000	1.0000
500	20	40	3	0.0000	0.0566	0.0000	0.1474	0.8758	0.0180
	40	80	7	0.0012	0.0524	0.0010	0.9494	0.9996	0.9234
	60	120	11	0.0008	0.0608	0.0026	0.9998	1.0000	1.0000
	80	160	15	0.0018	0.0558	0.0068	1.0000	1.0000	1.0000
	100	200	19	0.0026	0.0576	0.0122	1.0000	1.0000	1.0000

Table 2 presents the results of three tests with unequal sample sizes and a nominal level of 0.05. It was found that only the T_2 test has an ASL value close to the nominal level of 0.05 compared to other tests. When considering the empirical power of the T_2 test, it was found to increase close to 1 as the sample size increases. On the other hand, both T_1 and T_3 tests had performance that did not meet acceptable criteria when the sample sizes were not the same in all studied situations.

Table 3. ASL and Empirical Power of the Tests with Varying Block Sizes for Equal Sample Sizes and Nominal Level 0.05

p	n_i $(n_1 = n_2)$	q	ASL			Empirical power		
			T_1	T_2	T_3	T_1	T_2	T_3
200	40	5	0.0638	0.0620	0.0122	0.7422	0.7264	0.4280
		10	0.0628	0.0636	0.0174	0.6472	0.6464	0.3814
400	40	5	0.0676	0.0560	0.0014	0.9540	0.9442	0.5660
		10	0.0660	0.0612	0.0044	0.8982	0.8824	0.5020
	60	5	0.0646	0.0588	0.0074	0.9986	0.9988	0.9758
		10	0.0610	0.0602	0.0110	0.9944	0.9932	0.9510
		15	0.0576	0.0614	0.0136	0.9928	0.9910	0.9460

Table 3 displays the ASL and empirical power of three tests with varying block sizes, equal sample sizes, and a nominal level of 0.05. The results indicate that both T_1 and T_2 exhibit acceptable performance. When keeping the number of variables (p) and sample size (n_i) constant, increasing the block size enhances the ASL but diminishes the power of the test. This effect is evident when the sample size is 40. However, when the sample size is 60, the block size has a smaller impact on the power of the test. Hence, the optimal block size choice depends on the sample size. If the sample size is restricted, the block size should be chosen to be smaller.

Overall, the findings suggest that selecting an appropriate block size is crucial for the performance of the tests. In general, larger block sizes tend to result in better performance, as demonstrated by the higher empirical power values for tests with larger q values. However, there may be a trade-off between performance and computational efficiency, as larger block sizes may require more computational resources to compute the test statistics.

Table 4. ASL and Empirical Power of the Tests with Varying Block Sizes for Unequal Sample Sizes and Nominal Level 0.05

p	n_1	n_2	q	ASL			Empirical power		
				T_1	T_2	T_3	T_1	T_2	T_3
200	40	80	3	0.0034	0.0518	0.0034	0.7592	0.9584	0.7690
			5	0.0036	0.0620	0.0078	0.6702	0.9404	0.7482
			7	0.0036	0.0588	0.0078	0.6052	0.9300	0.7212
400	40	80	3	0.0014	0.0564	0.0004	0.9586	0.9980	0.9042
			5	0.0018	0.0546	0.0008	0.9234	0.9980	0.8928
			7	0.0016	0.0578	0.0010	0.9006	0.9968	0.8874
	60	120	3	0.0036	0.0556	0.0028	1.0000	1.0000	0.9998
			5	0.0016	0.0538	0.0034	0.9996	1.0000	0.9996
			7	0.0024	0.0568	0.0050	0.9992	1.0000	0.9996
			11	0.0020	0.0586	0.0046	0.9988	1.0000	0.9998

Table 4 shows the results of three different tests (T_1, T_2, and T_3) conducted under unequal sample sizes, varying block sizes (q) for two different number of variables p (200 and 400), and a nominal level of 0.05. The results show that only the T_2 test has acceptable performance when the sample sizes are unequal. In addition, the impact of block size on the power of the test is dependent on the sample size, where increasing the block size leads to a decrease in power. This relationship is more evident when the sample size is 40, as opposed to when the sample size is 60 where the effect of block size is relatively smaller.

In summary, the results suggest that the choice of block size can impact the power of the tests, and a larger block size may lead to a more powerful test. However, this relationship is dependent on the sample size, and the effect of block size may be smaller for larger sample sizes.

5. Conclusion and Discussion

5.1 Conclusion

1) For two sets of data with equal sample sizes, the T_2 test proposed by Hu et al. (2017) is more efficient than other tests, and the empirical power of T_2 increases to near 1 as the sample size increases. The second best test is T_1 (proposed by Srivastava et al. (2013), which is more efficient when the sample size is larger than 20. In the same situation (same number of variables and sample size), the power of the T_1 test is slightly higher than that of the T_2 test. As for the T_3 test, its performance is not acceptable in all situations studied.

For two sets of data with unequal sample sizes, the T_2 test proposed by Hu et al. (2017) is more efficient than tests, and the empirical power of T_2 increases to near 1 as the sample size increases. As for the T_1 and T_3 test statistics, their performance is not acceptable in all situations studied when the sample sizes are different.

2) For two sets of data with equal sample size and a fixed number of variables, increasing the block size adjustment does not greatly affect the performance of the T_1 and T_2 tests. However, increasing the block size leads to an increase in the ASL values, and the power decreases. In larger sample sizes, such as when the sample size is of 60 and the number of variables is 400, the impact of block size on testing performance is relatively small. Therefore, if the sample size is limited, the size of the block should be chosen accordingly.

For two sets of data with unequal sample sizes and a fixed number of variables, only the T_2 test showed acceptable performance when the block size was adjusted. Increasing the block size led to a decrease in the power of the test, and in larger sample sizes (e.g., when the first sample size is 60, the second sample size is 120, and the number of variables is 400), the size of the block had a small impact on testing performance.

3) To test the equality of high-dimensional mean vectors with a multivariate normal distribution, the choice of test depends on the sample size and whether the sample sizes are equal or unequal. If the sample sizes are equal and greater than 20, the test from Srivastava et al. (2013) should be used, while for sample sizes less than or equal to 20, the test from Hu et al. (2017) is recommended. If the sample sizes are unequal, the test from Hu et al. (2017) should be used. When the covariance matrix has a block diagonal structure, the largest possible block size should be chosen to preserve the data from the sample covariance matrix and obtain the most common variance of the samples.

5.2 Discussion

1) Based on the results showing unsatisfactory performance of the T_1 test proposed by Srivastava et al. (2013) in some situations, it is possible that the test performance may depend on having a sufficiently large sample size and meeting the assumption $n_1/(n_1+n_2) \to k \in (0,1)$, where $n = n_1 + n_2 \to \infty$ and $n_{min} = O(p^{\delta}), \delta > \frac{1}{2}$, $n_{min} = \min(n_1, n_2)$. It is assumed that T_1 test will exhibit higher efficacy with a large sample size.

2) The study found that the T_3 test proposed by Ahmad (2019) did not achieve the acceptable criteria, possibly due to the total sample size being not greater than the individual sample size. Hence, it may not be compliant with the assumption $\lim_{n_i \to \infty} (n/n_i) = \rho_i = O(1), n = n_1 + n_2, i = 1,2$.

3) An increase in block size can lead to a decrease in power of the test, and therefore, the selection of block size should be dependent on the sample size. As recommended by Krishnamoorthy and Yu (2004), the size of the block should not exceed $n_i/4, i = 1, 2$ for the case of equal sample sizes and should not exceed $\min((n_1 - 1)/5, (n_2 - 1)/5)$ for the case of unequal sample sizes.

Acknowledgements

We are grateful for the support provided by Rajamangala University of Technology Rattanakosin (RMUTR) through the research funds that enabled this study.

References

Ahmad, M. R. (2019). A unified approach to testing mean vectors with large dimensions. AStA Advances in Statistical Analysis, 103(4), 593-618.

Bai, Z., & Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. Statistica Sinica, 311-329.

Barnett, D. E., & Onofrei, D. (2018). Block diagonal covariance matrices and multivariate equivalence testing. Journal of Multivariate Analysis, 167, 152-167.

Chen, S. X., & Qin, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics, 38(2), 808-835.

Cochran, W. G. (1954). Some methods for strengthening the common tests. Biometrics, 10(4), 417-451.

Datta, S., Deb, B., & Mukhopadhyay, A. (2018). A review of machine learning in personalized medicine: Research efforts and future directions. Proceedings of the IEEE, 106(8), 1284-1299.

Fan, J., & Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. Statistical Science, 25(3), 371-391.

Fan, J., Li, R., & Wang, Y. (2020). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. Proceedings of the National Academy of Sciences, 117(16), 8918-8925.

Hotelling, H. (1931). The economics of exhaustible resources. Journal of Political Economy, 39(2), 137-175.

Hu, J., Bai, Z., Wang, C., & Wang, W. (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. Annals of the Institute of Statistical Mathematics, 69(2), 365-387.

Jiamwattanapong, K., & Chongcharoen, S. (2015). A new test for the mean vector in high-dimensional data. Songklanakarin Journal of Science & Technology, 37(4).

Jiamwattanapong, K., & Chongcharoen, S. (2017). A two-sample test for mean vectors in high-dimensional data. Applied Science and Innovative Research, 1(2), 118-130.

Krishnamoorthy, K., & Yu, J. (2004). Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem. Statistics & Probability Letters, 66(2), 161-169.

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. META Group Research Note, 6(70), 1-3.

Pandit, D., Zhang, J., & Li, X. (2020). Testing for high-dimensional mean vectors with block-wise dependence. Biometrics, 76(4), 1226-1236.

Shi, K., Yu, Z., & Xie, M. (2019). A review on statistical inference methods for high-dimensional data. Mathematical Biosciences and Engineering, 16(6), 7482-7501.

Srivastava, M. S., Katayama, S., & Kano, Y. (2013). A two sample test in high dimensional data. Journal of Multivariate Analysis, 114, 349-358.

Sukcharoen, P., & Chongcharoen, S. (2019). A Test on the Multivariate Behrens–Fisher Problem in High–Dimensional Data by Block Covariance Estimation. Journal of Mathematics and Statistics, 15(1), 44-54.

Thonghnunui, N., Chongcharoen, S., & Jiamwattanapong, K. (2020). Two-sample multivariate tests for high-dimensional data when one covariance matrix is unknown. Communications in Statistics-Simulation and Computation, 1-21.