Thailand Statistician
April 2022; 20(2): 412-419
http://statassoc.or.th
Contributed paper

One-Sided Multivariate Test for Two Population Means with Common Unknown Covariance Matrices of High-Dimensional Data

Samruam Chongcharoen [a], Pawat Paksaranuwat* [a] and Kunjira Kingphai [b]
[a] Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand.
[b] Department of Mathematics and Statistics, Faculty of Science and Technology,
Chiang Mai Rajabhat University, Chiang Mai, Thailand.
*Corresponding author; e-mail: pawat.pak@cmu.ac.th

Received: 7 December 2020
Revised: 5 May 2021
Accepted: 30 May 2021

Abstract

In this paper, a novel multivariate test for analyzing multivariate datasets with fewer observations than the dimension is developed. More specifically, we consider the problem of the one-sided hypothesis testing of mean vectors from two multivariate normal populations when the covariance matrices are commonly unknown. As we knew that on high-dimensional data, the sample covariance matrix is singular, a new test is proposed based on the classical Hotelling's T^2 test and the idea of keeping more information from the sample covariance matrices as much as possible. The simulation results showed that the proposed test gave the attained significance level of the proposed test close to setting nominal significance level satisfactorily and also gave high the attained power of the test. Furthermore, the performance of the proposed test is also shown by an empirical analysis of the DNA microarray data set.

Keywords: Block diagonal matrix structure, high-dimensional data, hypothesis testing, two-sample mean vectors.

1. Introduction

High-dimensional data that have dimensions a lot larger than the sample size data are increasingly encountered in statistical application in many areas but mostly in biology and finance. When high-dimensional data is obtained, the classical multivariate statistic cannot be computed because its' sample covariance matrix becomes singular (Eaton and Perlman 1973). In one sample case, Chongcharoen (2012) studied the one-sided multivariate tests for high-dimensional data for one sample with unknown covariance matrices by combining the high-dimensional tests of Dempter (1958, 1960) test, and also he applied one sample version of (Bai and Saranadasa 1996, Srivastata and Du 2008) to Follmann's test (Follmann 1996). From the high-dimensional tests proposed by Dempter (1958), Bai and Saranadasa (1996), and Srivastata and Du (2008), their tests involved only elements on diagonal elements of sample covariance matrices which made their testing statistics lose some information from the sample covariance matrices**.** Motivated by this kind of data and the previous literature, we proposed one-sided multivariate tests of two independent samples with unknown but equal covariance matrices for high-dimensional data based on the idea of keeping information from those two sample covariance matrices as much as possible by modifying with Follmann's test.

2. Materials and Methods

Throughout this paper, we suppose that $X_{i1}, X_{i2}, ..., X_{in_i}$; $i = 1,2$ are independent random samples from a p-dimensional multivariate normal distribution with unknown mean vector μ_i and unknown common positive definite covariance matrix $\mathbf{\Sigma}_i$, or $X_{ij} \sim N_p(\boldsymbol{\mu}_i, \mathbf{\Sigma}_i)$, $i = 1,2$ and $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$. We consider to compare two multivariate means of two independent multivariate normal populations with an alternative one-sided test in high-dimensional data. That is, we want to test the hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $H_1 : \boldsymbol{\mu}_1 > \boldsymbol{\mu}_2$ when both $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are unknown and $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ when the data in hand has the number of dimensions larger than the number of observations.

Originally, the unrestricted test statistic for testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ when two independent random samples $X_{i1}, X_{i2}, ..., X_{in_i}$, $i = 1,2$ were drawn from two independent p-dimensional multivariate normal distribution with unknown mean vector $\boldsymbol{\mu}_i$ and unknown common positive definite covariance matrix $\mathbf{\Sigma}_i$, or $X_{ij} \sim N_p(\boldsymbol{\mu}_i, \mathbf{\Sigma}_i)$, $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, was tested by the Hotelling's T^2 test statistic when $p \leq n_1 + n_2 - 2$ defined as

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}\left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\right)' \mathbf{S}_p^{-1}\left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\right), \tag{1}$$

where the sample mean vectors $\left(\bar{\mathbf{X}}_i\right)$ and the sample covariance matrix $\left(\mathbf{S}_i\right)$ are defined respectively by

$$\bar{\mathbf{X}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{X}_{ij} \text{ and } \mathbf{S}_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(X_{ij} - \bar{\mathbf{X}}_i\right)\left(X_{ij} - \bar{\mathbf{X}}_i\right)', \ i = 1,2. \tag{2}$$

It is well known that S_i are positive definite matrices with probability one. The pooled sample covariance matrix $(\mathbf{S}_p)$ is used as the estimates of the common population covariance given by

$$\mathbf{S}_p = \frac{1}{v}\sum_{i=1}^{2}(n_i - 1)\mathbf{S}_i,$$

where $v = n_1 + n_2 - 2$ with the dimension $p \times p$ with $p \leq v$. The Hotelling's T^2 test is a uniformly most powerful test, and the statistic T^2 can be converted to a central F-distribution as $(v - p + 1)T^2 / vp \sim F_{p, v-p+1}$, (Davis 2002, Anderson 2003). However, it can be easily seen from (1) that the Hotelling's T^2 statistic requires the pooled sample covariance matrix $\mathbf{S}_p$ invertible. As mentioned before, for the high-dimensional data $\mathbf{S}_p$ is not invertible, so $\mathbf{S}_p^{-1}$ does not exist. Thus, the Hotelling's statistic T^2 cannot be applied.

Jiamwattanapong and Chongcharoen (2017) proposed a test statistic for testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ for high- dimensional data as

$$T_q = \frac{T_n - \sum_{j=1}^{m} \frac{v q_j}{v - q_j - 1}}{\sqrt{\sum_{j=1}^{m} \frac{2v^2(v-1)q_j}{(v-q_j+1)^2(v-q_j-3)}}},$$

where $T_n = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)'\mathbf{D}_q^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$; $\bar{\mathbf{X}}_i,\ i = 1, 2$ defined in (2) and $\mathbf{D}_q$ is a block diagonal matrix as $\mathbf{D}_q = \text{diag}(\mathbf{S}_{11}, \mathbf{S}_{22,\ldots,} \mathbf{S}_{mm})$, where $\mathbf{S}_{jj},\ j = 1, 2, \ldots, m,\ m \le p,$ are submatrices obtained from the pooled sample covariance matrix $\mathbf{S}_p$ giving

$$\mathbf{D}_q = \begin{pmatrix} \mathbf{S}_{11} & 0 & \cdots & 0 \\ 0 & \mathbf{S}_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{S}_{mm} \end{pmatrix}, \tag{3}$$

which $\mathbf{S}_{jj}$ are the sample covariance submatrices of dimension $q_j \times q_j$ with $q_j < n_1 + n_2 - 2$ and $\sum_{j=1}^{m} q_j = p$. Since $q_j < n_1 + n_2 - 2$ then $\mathbf{S}_{jj}, j = 1, 2, \ldots, m,\ m \le p$ are all invertible (Eaton and Perlman, 1973). As a result, $\mathbf{D}_q$ is also invertible, and the inverse of $\mathbf{D}_q$ can be obtained as: $\mathbf{D}_q^{-1} = \text{diag}(\mathbf{S}_{11}^{-1}, \mathbf{S}_{22}^{-1}, \ldots, \mathbf{S}_{mm}^{-1})$. Then T_n can be computed, consequently, T_q statistic can be obtained. Under $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, Jiamwattanapong and Chongcharoen (2017) showed that T_q asymptotically distributed as $N(0,1)$. This method based on the idea of keeping more or as much as possible of the information from the sample covariance matrix that differs from those proposed by Dempster (1958), Bai and Saranadasa (1996) and Srivastava and Du (2008) that do not need the inverse of the sample covariance matrix but used only information on the diagonal element of the sample covariance matrix. Jiamwattanapong and Chongcharoen (2017) also showed that T_q test is invariant under a group of scalar transformation $\mathbf{X} \to \mathbf{DX}$, where $\mathbf{D} = \text{diag}(c_1, c_2, \ldots, c_p)$ and $c_i \ne 0$ for all $i = 1, 2, \ldots, p$. They recommended from their simulation results that the appropriate block sizes q_j should be equal to $v - 6$ where $v = n_1 + n_2 - 2$.

Motivated by this kind of data, the previous literature and Follmann's test (Follmann 1996), so we combined the unrestricted alternative test for high dimensional multivariate tests T_q with Follmann's idea to propose a test statistic for testing one-sided multivariate hypothesis, $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $H_1 : \boldsymbol{\mu}_1 > \boldsymbol{\mu}_2$ for high-dimensional data as

$$T_q = \frac{T_n - \sum_{j=1}^{m} \frac{v q_j}{v - q_j - 1}}{\sqrt{\sum_{j=1}^{m} \frac{2v^2(v-1)q_j}{(v-q_j+1)^2(v-q_j-3)}}} \text{ and } \sum_{i=1}^{p}\left(\bar{\mathbf{X}}_{1i} - \bar{\mathbf{X}}_{2i}\right),$$

and $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ will be rejected at a significant level α if

$$T_q \ge z_{1-2\alpha} \text{ and } \sum_{i=1}^{p}\left(\bar{\mathbf{X}}_{1i} - \bar{\mathbf{X}}_{2i}\right) > 0, \tag{4}$$

where $z_{1-2\alpha}$ is the $(1-2\alpha)^{\text{th}}$ quantile of standard normal distribution**.** By Theorem 2.1 of Follmann (1996), one has $1'\left(\underline{\mu}_1-\underline{\mu}_2\right)=0,$ where $\underline{\mu}_i;\ i=1,2$ is p-dimensional population mean vectors, and the significance level is approximated by

$$P\left(T_q \ge z_{1-2\alpha} \cap 1'\left(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2\right)>0\right)=P\left(T_q \ge z_{1-2\alpha}\right)P\left(1'\left(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2\right)>0\right)=\left(2\alpha\right)\left(\frac{1}{2}\right)=\alpha.$$

The performance of this proposed test will be studied by simulation techniques.

3. Simulation Results

In this section, the performance of the proposed test was evaluated using a simulation study. Firstly, for each setting, we generate the data $X_{ij} \sim N_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i),\ i=1,2$ with $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}$ then we compute the pooled sample covariance matrix $\mathbf{S}_p$. We set up the block diagonal matrix $\mathbf{D}_q$ with m submatrices $\mathbf{S}_{jj},\ j=1,2,\ldots,m,\ m\le p,$ obtained from the pooled sample covariance matrix $\mathbf{S}_p$ as in (3). The first $m-1$ submatrices of $\mathbf{S}_{jj}$ has dimension $q_j \times q_j$ when $q_1=q_2=,\ldots,=q_{m-1}=\nu-6$ and the last submatrix $\mathbf{S}_{mm}$ with dimension $q_m \times q_m$ which q_m is the positive integer that $\sum_{j=1}^{m} q_j = p.$ This set up are guaranteed that all submatrices $\mathbf{S}_{jj},\ j=1,2,\ldots,m$ are invertible which makes $\mathbf{D}_q^{-1}$ exist as the recommendations of Jiamwattanapong and Chongcharoen (2017). Generally, there are many forms of the population covariance matrix structures, so we grouped those forms of them into three possible types of them including in our study as:

Type I: $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}$ that $\boldsymbol{\Sigma}$ are of diagonal matrix form as $\operatorname{diag}\left(\boldsymbol{\Sigma}_{11},\boldsymbol{\Sigma}_{22},\ldots,\boldsymbol{\Sigma}_{mm}\right)$ where $\Sigma_{jj},$ $j=1,2,\ldots,m,\ m\le p,$ is a $q_j\times q_j$ matrix; $\Sigma_{jj}=c_jI+\left(1-c_j\right)J,\ c_j\sim U\left(0,1\right);$ J is $q_j\times q_j$ matrix of 1's ; I is $q_j\times 1$ vector of 1's.

Type II: $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}$ that $\boldsymbol{\Sigma}$ are of diagonal matrix form as $\operatorname{diag}\left(\boldsymbol{\Sigma}_{11},\boldsymbol{\Sigma}_{22},\ldots,\boldsymbol{\Sigma}_{\tilde{m}\tilde{m}}\right)$ where Σ_{jj} is a $\tilde{q}_j\times\tilde{q}_j$ matrix with $j=1,2,\ldots,\tilde{m},$ where $\tilde{m}$ is random positive number from $\{2,3,\ldots,p\}$ and $\tilde{q}_j$ are random positive numbers such that $\sum_{j=1}^{\tilde{m}}\tilde{q}_j=p;\ \Sigma_{jj}=c_jI+\left(1-c_j\right)J,\ c_j\sim U\left(0,1\right);\ J$ is $\tilde{q}_j\times\tilde{q}_j$ matrix of 1's ; I is $\tilde{q}_j\times 1$ vector of 1's**.**

Type III: $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}$ that $\boldsymbol{\Sigma}$ are not of diagonal matrix form as $\boldsymbol{\Sigma}_i=\mathbf{D}_\sigma^{1/2}\Re\mathbf{D}_\sigma^{1/2}$ where $\mathbf{D}_\sigma^{1/2}=\operatorname{diag}\left(\sigma_1,\sigma_2,\ldots,\sigma_p\right),$ $\sigma_j=2+\left(-1\right)^{j-1}\left(p-j+1\right)/p$ and $\Re=\left[r_{ij}\right]_{p\times p},$ $r_{ij}=\left(-1\right)^{i+j}\left(c^{|i-j|}\right),$ where $c\sim U\left(0,1\right).$

For mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, we also considered in 3 possible types in our study as (1) $\boldsymbol{\mu}_1=\boldsymbol{\mu}_2=\underline{\mathbf{0}},$ (2) $\boldsymbol{\mu}_1=\boldsymbol{\mu}_2=\underline{u},$ and (3) $\boldsymbol{\mu}_2=u$ and $\boldsymbol{\mu}_1=\boldsymbol{\mu}_2+\underline{v}$ with $\underline{u}=\left[\mathbf{u_1}\quad\mathbf{u_2}\quad\ldots\quad\mathbf{u_p}\right]',$ $\mathbf{u}_i\sim \mathrm{U}\left(-1,1\right)$ and $\underline{v}=\left[\mathbf{v_1}\quad\mathbf{v_2}\quad\ldots\quad\mathbf{v_p}\right]',\mathbf{v}_i\sim\mathrm{U}\left(0.1,0.5\right).$ The estimated size of type I error and the estimated power of the proposed test statistics for $H_0:\boldsymbol{\mu}_1=\boldsymbol{\mu}_2$ against $H_1:\boldsymbol{\mu}_1>\boldsymbol{\mu}_2$ were evaluated by computing as $\frac{\text{number of rejection}}{h}$, where h is a number of iterations of each data sets simulated under the null hypothesis or the alternative hypothesis**.** All of this simulation studies, we set the nominal significance level as $\alpha = 0.05$. The simulated data of each combination of designed means and designed covariance matrices were generated on the dimension as $p \in \{50, 100, 200, 400\}$ with both equal and unequal sample size, (n_1, n_2) then the test statistic from the proposed test is computed and the number of rejections is counted. Repeating this procedure $h = 10,000$ times, the attained significance level and the attained power of the proposed test are obtained as in Tables 1 and 3 follows:

Table 1 The attained significance level and the attained power when common covariance matrices $\boldsymbol{\Sigma}$ as setting Type I

p	(n_1, n_2)	q_j	Type I error		Power of the test
			$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \underline{\mathbf{0}}$	$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \underline{u}$	$\boldsymbol{\mu}_2 = \underline{u}$ and $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 + \underline{v}$
50	(20, 20)	5	0.0497	0.0410	0.9866
100	(20, 20)	5	0.0487	0.0421	0.9885
	(35, 40)	5	0.0500	0.0469	0.9706
	(60, 60)	10	0.0475	0.0485	0.9777
200	(20, 20)	5	0.0502	0.0524	0.9891
	(35, 40)	5	0.0497	0.0533	0.9720
	(60, 60)	10	0.0513	0.0526	0.9945
	(80, 100)	10	0.0493	0.0521	0.9858
	(120, 120)	20	0.0476	0.0485	0.9960
400	(20, 20)	5	0.0485	0.0495	0.9931
	(35, 40)	5	0.0493	0.0503	0.9838
	(60, 60)	10	0.0482	0.0480	1.0000
	(80, 100)	10	0.0504	0.0477	1.0000
	(120, 120)	20	0.0479	0.0489	1.0000
	(210, 250)	20	0.0497	0.0505	1.0000
	(350, 350)	40	0.0502	0.0471	1.0000

From Tables 1 and 2, when unknown common covariance matrices $\boldsymbol{\Sigma}$ are block diagonal matrices, all attained significance levels are close to the nominal significance level setting and all attained powers are close to 1 in every situation considered. These shown that the proposed test gave a reasonable results and excellent power.

From Table 3, when unknown common covariance matrices $\boldsymbol{\Sigma}$ are not block diagonal matrices as real-life situation, the proposed test still gave all attained significance level close to the nominal significance level setting and all attained power close to 1 in every situation considered.

As all results from simulation studies, we may recommend using the proposed test for one-sided alternative multivariate test in high-dimensional data when two independent high-dimensional random samples are obtained from a p-dimensional multivariate normal distribution with unknown mean vector and common unknown positive definite covariance matrix.

Table 2 The attained significance level and the attained power when common covariance matrices $\mathbf{\Sigma}$ as setting Type II

p	(n_1, n_2)	q_j	Type I error		Power of the test
			$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \underline{\mathbf{0}}$	$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \underline{u}$	$\boldsymbol{\mu}_2 = \underline{u}$ and $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 + \underline{v}$
50	(20, 20)	5	0.0523	0.0451	0.9744
100	(20, 20)	5	0.0522	0.0426	0.9907
	(35, 40)	5	0.0525	0.0534	0.9849
	(60, 60)	10	0.0497	0.0482	0.9687
200	(20, 20)	5	0.0516	0.0489	0.9912
	(35, 40)	5	0.0483	0.0546	0.9940
	(60, 60)	10	0.0538	0.0573	0.9970
	(80, 100)	10	0.0503	0.0593	0.9989
	(120, 120)	20	0.0511	0.0509	1.0000
400	(20, 20)	5	0.0507	0.0561	0.9939
	(35, 40)	5	0.0495	0.0494	1.0000
	(60, 60)	10	0.0529	0.0523	1.0000
	(80, 100)	10	0.0540	0.0507	1.0000
	(120, 120)	20	0.0508	0.0510	1.0000
	(210, 250)	20	0.0525	0.0494	1.0000
	(350, 350)	40	0.0516	0.0575	1.0000

Table 3 The attained significance level and the attained power when common covariance matrices $\mathbf{\Sigma}$ as setting Type III

p	(n_1, n_2)	q_j	Type I error		Power of the test
			$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \underline{\mathbf{0}}$	$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \underline{u}$	$\boldsymbol{\mu}_2 = \underline{u}$ and $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 + \underline{v}$
50	(20, 20)	5	0.0472	0.0444	0.9092
100	(20, 20)	5	0.0474	0.0447	0.9705
	(35, 40)	5	0.0513	0.0504	0.9918
	(60, 60)	10	0.0495	0.0492	0.9974
200	(20, 20)	5	0.0497	0.0512	0.9796
	(35, 40)	5	0.0488	0.0522	0.9843
	(60, 60)	10	0.0515	0.0523	0.9866
	(80, 100)	10	0.0534	0.0520	0.9833
	(120, 120)	20	0.0486	0.0483	0.9967
400	(20, 20)	5	0.0529	0.0529	0.9906
	(35, 40)	5	0.0495	0.0487	1.0000
	(60, 60)	10	0.0490	0.0498	1.0000
	(80, 100)	10	0.0479	0.0466	1.0000
	(120, 120)	20	0.0476	0.0483	1.0000
	(210, 250)	20	0.0473	0.0464	1.0000
	(350, 350)	40	0.0555	0.0520	1.0000

4. A Real Data Example

An example to demonstrate the proposed test is an analysis of DNA microarray data from an oncology study. The data, published by Notterman et al. (2001) were retrieved on April 20, 2015 from the Princeton University Gene Expression Project website (http://genomics-pubs.Princeton.Edu/oncology). A selection of 100 genes (p) was used to test the mean vectors of two independent groups, tumor tissue, and normal tissue. The first and the second group are the sample size of 15 and 20, respectively, provided that $v = n_1 + n_2 - 2 = 33.$ Before calculating the proposed test statistic, the equality of covariance matrices is tested, using the method presented by Chaipitak and Chongcharoen (2013) leading to the conclusion of equal covariance matrices. The proposed test statistic as in (4) is computed and shown in Table 4. With a large value of testing statistic T_q and $\sum_{i=1}^{p}\left(\bar{X}_{1i} - \bar{X}_{2i}\right) > 0,$ this leads to the rejection of the null hypothesis and concludes that the tumor tissue means of 100 gene expression levels are significantly greater than those from normal tissue means at the significance level of 0.05.

Table 4 Statistic values for testing tumor tissue mean of 100 genes expression

	T_q	$\sum_{i=1}^{p}\left(\bar{\mathbf{X}}_{1i} - \bar{\mathbf{X}}_{2i}\right)$
Test Statistic	183.4251	506.3389

5. Conclusions

In this study, we proposed a test statistic for testing the hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $H_1 : \boldsymbol{\mu}_1 > \boldsymbol{\mu}_2$ from two multivariate normal distributed data with the common unknown covariance matrices in high-dimensional data by applying unrestricted test statistic proposed by Jiamwattanapong and Chongcharoen (2017) for high dimensional data with Follmann's test (Follman 1996). The simulation results indicate that our proposed test gave good performance when both common unknown covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are block diagonal matrix. Moreover, in general situations, when both common unknown covariance matrices are not the block diagonal matrix, our proposed test also performs very well and gives excellent power as well. So we recommend this proposed test statistic for a one-sided alternative multivariate hypothesis test where the data collected from two independent multivariate normal distributions with common unknown covariance matrices structure in high dimensional data.

References

Anderson TW. An introduction to multivariate statistical analysis. New Jersey: John Wiley & Sons; 2003.

Bai Z, Saranadasa H. Effect of high dimension: by an example of a two sample problem. Stat Sin. 1996; 311-329.

Chongchharoen S. One-sided multivariate tests for high dimensional data. J Math Stat. 2012; 8(2): 274-282.

Chaipitak S., Chongchharoen S. A test for testing the equality of two covariance matrices for high-dimensional data. J Appl Sci. 2013; 13(2): 270-277.

Davis CS. Statistical methods for the analysis of repeated measurements. New York: Springer; 2002.

Dempster AP. A high dimensional two sample significance test. Ann Math Stat. 1958; 995-1010.

Dempster AP. A significance test for the separation of two highly multivariate small samples. Biometrics. 1960; 16(1): 41-50.

Eaton ML, Perlman MD. The non-singularity of generalized sample covariance matrices. Ann Stat. 1973; 1(4): 710-717.

Follmann D. A simple multivariate test for one-sided alternatives. J Am Stat Assoc. 1996; 91(434): 854-861.

Jiamwattanapong K, Chongchharoen S. A new test for the mean vector in high-dimensional data. Songklanakarin J Sci Technol 2017; 37(4): 477-484.

Notterman, DA, Alon U, Sierk AJ, Levine AJ. Transcriptional gene expression profiles of colorectal adenoma adenocarcinoma and normal tissue examined by oligonucleotide arrays. Cancer Research. 2001; 61: 3124-3130.

Srivatava MS, DU M. A test for the mean vector with fewer observations than the dimension. J Multivar Anal. 2008; 99(3): 386-402.