Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

Inferences on a linear combination of K multivariate normal mean vectors

S. H. Lin [a] & R. S. Wang [b]

[a] National Taichung Institute of Technology , Taichung, Taiwan, Republic of China

[b] Institute of Statistics, National ChiaoTung University , Hsinchu, Taiwan, Republic of China

Published online: 16 Apr 2009.

To cite this article: S. H. Lin & R. S. Wang (2009) Inferences on a linear combination of K multivariate normal mean vectors, Journal of Applied Statistics, 36:4, 415-428, DOI: 10.1080/02664760802474231

To link to this article: http://dx.doi.org/10.1080/02664760802474231

Journal of Applied Statistics
Vol. 36, No. 4, April 2009, 415–428

Taylor & Francis
Taylor & Francis Group

Inferences on a linear combination of K multivariate normal mean vectors

S.H. Lin[a]* and R.S. Wang[b]

[a]*National Taichung Institute of Technology, Taichung, Taiwan, Republic of China;* [b]*Institute of Statistics, National ChiaoTung University, Hsinchu, Taiwan, Republic of China*

(*Received 7 December 2007; final version received 12 September 2008*)

In this paper, the hypothesis testing and confidence region construction for a linear combination of mean vectors for K independent multivariate normal populations are considered. A new generalized pivotal quantity and a new generalized test variable are derived based on the concepts of generalized p-values and generalized confidence regions. When only two populations are considered, our results are equivalent to those proposed by Gamage et al. [*Generalized p-values and confidence regions for the multivariate Behrens–Fisher problem and MANOVA*, J. Multivariate Aanal. 88 (2004), pp. 117–189] in the bivariate case, which is also known as the bivariate Behrens–Fisher problem. However, in some higher dimension cases, these two results are quite different. The generalized confidence region is illustrated with two numerical examples and the merits of the proposed method are numerically compared with those of the existing methods with respect to their expected areas, coverage probabilities under different scenarios.

Keywords: coverage probability; generalized confidence region; generalized pivotal quantity; generalized test variable; heteroscedasticity; type I error

1. Introduction

Suppose there exist K independent d-variate normal populations with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$, $i = 1, 2, \ldots, K$, where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are possibly unknown and unequal among groups. We want to make inferences on a linear combination of K mean vectors. This problem arises because sometimes there is a theoretical reason for believing some characteristics of these populations to be such that their mean vectors have some relationships or practitioners want to know some characteristics of a compound material. For example, in the Edgar Anderson's famous Iris data, there is a theoretical belief that the four gene structures of three species are such that the mean vectors of the three populations, (1) *Iris versicolor* (2) *Iris setosa* and (3) *Iris virginica*, are related to $3\boldsymbol{\mu}_1 = 2\boldsymbol{\mu}_2 + \boldsymbol{\mu}_3$ [1].

If the difference between the covariance matrices is small and the sample sizes are large, the Hotelling's T^2-test for testing the linear combination of mean vectors has good performance. However, if the covariance matrices are quite different and/or the sample sizes are small, the

*Corresponding author. Email: suelin@ntit.edu.tw

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760802474231
http://www.informaworld.com

nominal significance level may be distorted. Therefore, in this paper, we intend to develop a procedure to provide generalized inferences for the linear combination of the mean vectors, $\boldsymbol{\theta} = \mathbf{G}\boldsymbol{\mu}$, where $\mathbf{G}$ is a designed $d \times dK$ matrix, and $\boldsymbol{\mu}$ is the dK-variate mean vector with $\boldsymbol{\mu}' = (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_K')$. That is, we will provide a generalized confidence region for $\boldsymbol{\theta}$ and test the hypothesis

$$H_0 : \mathbf{G}\boldsymbol{\mu} = \boldsymbol{\theta}_0 \quad \text{vs.} \quad H_1 : \mathbf{G}\boldsymbol{\mu} \neq \boldsymbol{\theta}_0, \tag{1}$$

where $\boldsymbol{\theta}_0$ is a given vector. For example, in the Iris data, we can set $\mathbf{G} = (3\mathbf{I}_d, -2\mathbf{I}_d, -\mathbf{I}_d)$ and $\boldsymbol{\theta}_0 = \mathbf{0}$ to perform this hypothesis.

Suppose that $\mathbf{X}_{ij}$s are independent random vectors of sample size n_i. Define the ith sample mean vector and sample covariance matrix as

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

It can be shown that

$$\bar{\mathbf{X}}_i \sim N_d\left(\boldsymbol{\mu}_i, \frac{\boldsymbol{\Sigma}_i}{n_i}\right) \text{ and } \mathbf{A}_i = n_i\mathbf{S}_i \sim W_d(n_i - 1, \boldsymbol{\Sigma}_i), \quad i = 1, \ldots, K, \tag{3}$$

and both of them are independently distributed, where $N_d(\boldsymbol{\pi}, \boldsymbol{\Psi})$ denotes d-variate normal distribution with mean vector $\boldsymbol{\pi}$ and $W_d(r, \boldsymbol{\Psi})$ is the d-dimensional Wishart distribution with degrees of freedom r and scale matrix $\boldsymbol{\Psi}$. Furthermore, n_i is supposed to be greater than d, $n_i > d$, $i = 1, \ldots, K$, to ensure $\mathbf{S}_i^{-1}$ exist with probability one. Because the distributions of $\bar{\mathbf{X}}_i$ and $\mathbf{S}_i$ are affine invariant, we will test the problem (1) and construct a confidence region of $\boldsymbol{\theta}(=\mathbf{G}\boldsymbol{\mu})$ based on these judicious condensations of the data. Using the underlying distribution assumptions, our approach procedures are associated with an exact probability statement and a repeated sampling interpretation.

For $K = 2$, $\mathbf{G} = (\mathbf{I}_d, -\mathbf{I}_d)$ and $\boldsymbol{\theta}_0 = \mathbf{0}$, Equation (1) is reduced to the well-known multivariate Behrens–Fisher problem. For this topic, there are several exact as well as approximate tests considered in the literature for the past five decades. For example, Christensen and Rencher [3] compared seven solutions for their type I error rates and powers and suggested that Kim's [6] and Nel and Van der Merwe's [8] solutions had the highest powers among solutions whose Type I error rates were not inflated. Krishnamoorthy and Yu [7] modified the Nel and Van der Merwe's test and provided an approximate invariant solution for the problem. In addition to those approximate procedures, Bennett [2] provided an exact solution for the generalized Behrens–Fisher problem. However, the power obtained by Bennett's method was poor under unequal sample sizes because the method was not based on sufficient statistics. Johnson and Weerahandi [5] provided an exact Bayesian solution based on the Bayesian approach and Gamage et al. [4] provided the generalized p-values and generalized confidence region for the Behrens–Fisher problem.

In this paper, we would like to further consider K nonhomogeneous multivariate normal populations with equal and unequal sample sizes and unequal covariance matrices, and then provide an invariant generalized test variable (GTV) and construct a generalized confidence region for a linear combination of K multivariate normal mean vectors. In our proposed model, the multivariate Behrens–Fisher problem can be treated as a special case of our model.

Our inferential procedures are based on the generalized approach introduced by Tsui and Weerahandi [12] and then Weerahandi [13]; see the books by Weerahandi [14,15] for detailed discussions along with numerous examples. The concepts of generalized p-value and generalized confidence intervals have turned out to be extremely fruitful for obtaining tests and confidence intervals involving 'non-standard' parameters. Several articles have appeared in the literature

describing such applications. Therefore, we will use the idea to derive a new generalized pivot quantity that is simple to use for both hypothesis testing and confidence region estimation of $\mathbf{G}\boldsymbol{\mu}$.

The rest of the paper unfolds as follows. The theory of generalized p-values and generalized confidence interval will be briefly outlined in Section 2. Our procedures for hypothesis testing and the generalized confidence region of $\mathbf{G}\boldsymbol{\mu}$ construction are presented in Section 3. Several methods in the multivariate Behrens–Fisher problem are also briefly introduced in Section 3. Simulation studies are presented in Section 4 to compare the type I error rates, expected areas and the coverage probabilities in different combinations of sample sizes and covariance matrices for difference procedures, and then two sets of data will be illustrated for our procedures in Section 5. Finally, some conclusions are provided in Section 6.

2. Generalized *p*-value and generalized confidence interval

Let $\mathbf{X}$ be a random variable whose distribution depends on a vector of unknown parameters $\zeta = (\theta, \boldsymbol{\eta})$, where θ is the parameter of interest, and $\boldsymbol{\eta}$ is a vector of nuisance parameters. Suppose we are interested in testing

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0, \tag{4}$$

where θ_0 is a prespecified quantity.

The GTV of the form $T(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta})$ with $\mathbf{x}$ being the observed value of $\mathbf{X}$ is chosen to satisfy the following requirements:

(i) For fixed $\mathbf{x}$, the distribution of $T(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta})$ is free of the vector of nuisance parameters $\boldsymbol{\eta}$.
(ii) The value of $T(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta})$ at $\mathbf{X} = \mathbf{x}$ is free of any unknown parameters.
(iii) For fixed $\mathbf{x}$ and $\boldsymbol{\eta}$, $\Pr[T(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta}) \geq t]$ is either an increasing or a decreasing function of θ for any given t. (5)

Under the above conditions, if $T(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta})$ is stochastically increasing in θ, then a generalized extreme region is defined as $\mathrm{C} = \{\mathbf{X}: T(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta}) \geq T(\mathbf{x}; \mathbf{x}, \theta, \boldsymbol{\eta})\}$. The generalized p-values for testing the hypothesis in Equation (4) is defined as

$$p = \Pr(\mathrm{C}|\theta_0). \tag{6}$$

Under the same setup, a *generalized pivotal quantity* (GPQ), $T_1(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta})$, satisfies the following conditions:

(i) The distribution of $T_1(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta})$ is free of unknown parameters.
(ii) The observed value of $T_1(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta})$ is free of nuisance parameters $\boldsymbol{\eta}$. (7)

Let c_1 and c_2 be such that

$$\Pr[c_1 \leq T_1(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta}) \leq c_2] = 1 - \alpha, \tag{8}$$

then $\{\theta : c_1 \leq T_1(\mathbf{x}; \mathbf{x}, \theta, \boldsymbol{\eta}) \leq c_2\}$ is a $100(1-\alpha)\%$ generalized confidence interval for θ. Furthermore, if the value of $T_1(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta})$ at $\mathbf{X} = \mathbf{x}$ is θ, then $\{T_1(\mathbf{x}; \alpha/2), T_1(\mathbf{x}; 1-\alpha/2)\}$ is a $100(1-\alpha)\%$ confidence interval for θ, where $T_1(\mathbf{x}; \gamma)$ represents the rth quantile of $T_1(\mathbf{X}; \mathbf{x}, \theta, \boldsymbol{\eta})$.

3. Hypothesis testing and confidence region estimation for $\mathbf{G}\boldsymbol{\mu}$

Suppose we have K independent d-variate multivariate normal populations with mean vector $\boldsymbol{\mu}_i$ and unequal covariance matrices $\boldsymbol{\Sigma}_i$ for the ith sample. Let $\bar{\mathbf{X}}_i$ and $\mathbf{S}_i$ be the sample mean vector

and sample covariance matrix for the ith population, which are defined in Equation (2). We will consider the problem of estimating a linear combination of K multivariate normal mean vectors, $\mathbf{G}\boldsymbol{\mu}$, based on the minimal sufficient statistics $(\bar{\mathbf{X}}_1, \ldots, \bar{\mathbf{X}}_K, \mathbf{S}_1, \ldots, \mathbf{S}_K)$.

In this section, we first derive the generalized p-value and construct a generalized confidence region of $\mathbf{G}\boldsymbol{\mu}$ based on the generalized method in Section 3.1 and then review some commonly used methods in Section 3.2. For some special cases, especially the multivariate Behrens–Fisher problem, several methods are reviewed in Section 3.3.

3.1 *Solutions based on the generalized method*

It is noted that $\bar{\mathbf{X}}_i$ and $\mathbf{S}_i$ are mutually independent with $\bar{\mathbf{X}}_i \sim N_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i/n_i)$, $\mathbf{S}_i \sim W_d(n_i - 1, \boldsymbol{\Sigma}_i/n_i)$ and $\mathbf{A}_i = n_i\mathbf{S}_i \sim W_d(n_i - 1, \boldsymbol{\Sigma}_i)$, $i = 1, \ldots, K$. Let $\bar{\mathbf{X}}' = (\bar{\mathbf{X}}_1', \ldots, \bar{\mathbf{X}}_K')$ then the MLE (maximum likelihood estimator) of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \mathbf{G}\bar{\mathbf{X}} \sim N_d(\boldsymbol{\theta}, \mathbf{G}\boldsymbol{\Phi}\mathbf{G}'), \tag{9}$$

where the block diagonal matrix (Bdiag) $\boldsymbol{\Phi} = \text{Bdiag}(\boldsymbol{\Sigma}_1/n_1, \ldots, \boldsymbol{\Sigma}_K/n_K) \equiv \begin{pmatrix} n_1^{-1}\boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & n_K^{-1}\boldsymbol{\Sigma}_K \end{pmatrix}$.

If the covariance matrices $\boldsymbol{\Sigma}_i$'s are given, it is known that from Equation (9) we can get

$$\left(\mathbf{G}\boldsymbol{\Phi}\mathbf{G}'\right)^{-1/2}\mathbf{G}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \equiv \mathbf{Z}_d \sim N_d(\mathbf{0}, \mathbf{I}_d). \tag{10}$$

If the covariance matrix $\boldsymbol{\Sigma}_i$ for the ith population is unknown, we can define

$$\mathbf{R} = \left[\mathbf{s}^{-1/2}\boldsymbol{\Phi}\mathbf{s}^{-1/2}\right]^{-1/2}\left[\mathbf{s}^{-1/2}\mathbf{S}\mathbf{s}^{-1/2}\right]\left[\mathbf{s}^{-1/2}\boldsymbol{\Phi}\mathbf{s}^{-1/2}\right]^{-1/2}, \tag{11}$$

where $\mathbf{S} = \text{Bdiag}(\mathbf{S}_1, \ldots, \mathbf{S}_K)$ is the block diagonal matrix with the observed value $\mathbf{s} = \text{Bdiag}(\mathbf{s}_1, \ldots, \mathbf{s}_K)$ and $\mathbf{S}_i$s are the sample covariance matrices with the observed values $\mathbf{s}_i$s. $\boldsymbol{\Psi}^{1/2}$ means the positive definite square root of the positive definite matrix $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}^{-1/2} = (\boldsymbol{\Psi}^{1/2})^{-1}$. It should be noted that $\mathbf{R}$ also stands for a block diagonal matrix with $\mathbf{R} = \text{Bdiag}(\mathbf{R}_1, \ldots, \mathbf{R}_K)$, where

$$\mathbf{R}_i = \left[\mathbf{s}_i^{-1/2}\left(\frac{\boldsymbol{\Sigma}_i}{n_i}\right)\mathbf{s}_i^{-1/2}\right]^{-1/2}\left[\mathbf{s}_i^{-1/2}\mathbf{S}_i\mathbf{s}_i^{-1/2}\right]\left[\mathbf{s}_i^{-1/2}\left(\frac{\boldsymbol{\Sigma}_i}{n_i}\right)\mathbf{s}_i^{-1/2}\right]^{-1/2}. \tag{12}$$

Since $\mathbf{R}_i \sim W_d(n_i - 1, \mathbf{I}_d)$ is free of any unknown parameters, and for the fact that at $\mathbf{S} = \mathbf{s}$, the observed value $\mathbf{r}$ of $\mathbf{R}$ is $\left[\mathbf{s}^{-1/2}\boldsymbol{\Phi}\mathbf{s}^{-1/2}\right]^{-1}$, it is clear that $\mathbf{s}^{1/2}\mathbf{R}^{-1}\mathbf{s}^{1/2} = \boldsymbol{\Phi}$ at $\mathbf{S} = \mathbf{s}$. That means we can use the information of $\mathbf{s}$ and $\mathbf{R}$ to make an inference about the nuisance parameters, $\boldsymbol{\Phi}$. Furthermore, we can derive the generalized inferences for $\mathbf{G}\boldsymbol{\mu}$ based on $\bar{\mathbf{X}}$ and $\mathbf{R}$. Let $\bar{\mathbf{x}}$ and $\mathbf{r}$ be the corresponding observed values of $\bar{\mathbf{X}}$ and $\mathbf{R}$, respectively, the generalized pivot quantity can be expressed as

$$\begin{aligned}\mathbf{T}(\bar{\mathbf{X}}, \mathbf{R}; \bar{\mathbf{x}}, \mathbf{r}) &= \mathbf{G}\bar{\mathbf{x}} - \left(\mathbf{G}\mathbf{s}^{1/2}\mathbf{R}^{-1}\mathbf{s}^{1/2}\mathbf{G}'\right)^{1/2}\left(\mathbf{G}\boldsymbol{\Phi}\mathbf{G}'\right)^{-1/2}\mathbf{G}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \\ &= \mathbf{G}\bar{\mathbf{x}} - \left(\mathbf{G}\mathbf{s}^{1/2}\mathbf{R}^{-1}\mathbf{s}^{1/2}\mathbf{G}'\right)^{1/2}\mathbf{Z}_d.\end{aligned} \tag{13}$$

It is noted that the value of $\mathbf{T}$ in Equation (13) at $(\bar{\mathbf{X}}, \mathbf{S}) = (\bar{\mathbf{x}}, \mathbf{s})$ is $\mathbf{G}\boldsymbol{\mu}$, which is the parameter of interest. Furthermore, $\mathbf{T}$ is the function of the Wishart distributions $\mathbf{R}_i$s, the standard multivariate normal distribution $\mathbf{Z}_d$ and the observed values $(\bar{\mathbf{x}}, \mathbf{s})$. The distribution of $\mathbf{T}$ is independent of any unknown parameters, therefore, $\mathbf{T}$ in Equation (13) satisfies the two conditions in Equation (7) and is truly a GPQ, which can be used to construct the confidence region for $\mathbf{G}\boldsymbol{\mu}$.

3.1.1 *The generalized **p**-value*

For given$(\bar{\mathbf{x}}, \mathbf{s})$, the distribution in Equation (13) is independent of unknown parameters and hence the Monte Carlo method can be utilized to construct a confidence region of $\mathbf{G}\boldsymbol{\mu}$, and test the hypothesis

$$H_0 : \mathbf{G}\boldsymbol{\mu} = \boldsymbol{\theta}_0 \quad \text{vs.} \quad H_1 : \mathbf{G}\boldsymbol{\mu} \neq \boldsymbol{\theta}_0, \tag{14}$$

where $\boldsymbol{\theta}_0$ is a given vector. Suppose $\mathbf{m}_\mathbf{T}$ and $\mathbf{S}_\mathbf{T}$ are the mean and covariance matrix of $\mathbf{T}$, and $\tilde{\mathbf{T}} = \mathbf{S}_\mathbf{T}^{-1/2}(\mathbf{T} - \mathbf{m}_\mathbf{T})$ is the standardized expression of $\mathbf{T}$, then the generalized p-value for testing Equation (14) can be computed by

$$p = \Pr\{\|\tilde{\mathbf{T}}\| > \|\tilde{\boldsymbol{\theta}}_0\| | \bar{\mathbf{x}}, \mathbf{r}\}, \tag{15}$$

where $\tilde{\boldsymbol{\theta}}_0 = \mathbf{S}_\mathbf{T}^{-1/2}(\boldsymbol{\theta}_0 - \mathbf{m}_\mathbf{T})$, $\|\tilde{\mathbf{T}}\|$ and $\|\tilde{\boldsymbol{\theta}}_0\|$ are norms of $\tilde{\mathbf{T}}$ and $\tilde{\boldsymbol{\theta}}_0$, respectively, with $\|\tilde{\mathbf{T}}\| = \sqrt{\tilde{\mathbf{T}}'\tilde{\mathbf{T}}}$, and the null hypothesis (14) will be rejected whenever $p \leq \alpha$.

Furthermore, if we want to test the MANOVA problem of the form $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_K$ which can be expressed as $H_0 : \mathbf{G}^*\boldsymbol{\mu} = \mathbf{0}$. One convenient choice for $\mathbf{G}^*$ in this particular problem is

$$\mathbf{G}^* = \begin{bmatrix} \mathbf{I}_d & -\mathbf{I}_d & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{I}_d & \mathbf{0} & -\mathbf{I}_d & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & & & & & \\ \mathbf{I}_d & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & -\mathbf{I}_d \end{bmatrix} = \begin{bmatrix} \mathbf{G}^{(2)} \\ \mathbf{G}^{(3)} \\ \vdots \\ \mathbf{G}^{(K)} \end{bmatrix},$$

where

$$\mathbf{G}^{(i)} = (c_1^{(i)}\mathbf{I}_d, \dots, c_K^{(i)}\mathbf{I}_d),\ c_j^{(i)} = \begin{cases} 1 & j = 1 \\ -1 & j = i \\ 0 & \text{o.w.} \end{cases}$$

Similar to $\mathbf{T}$ in Equation (13), the generalized test variable can be expressed as $\mathbf{T}^* = \mathbf{G}^*\bar{\mathbf{x}} - \left(\mathbf{G}^*\mathbf{s}^{1/2}\mathbf{R}^{-1}\mathbf{s}^{1/2}\mathbf{G}^{*\prime}\right)^{1/2}\mathbf{Z}_{d(K-1)}$. And the p-value can also be computed in the similar way as Equation (15).

3.1.2 *The generalized confidence region*

If we are interested in constructing confidence interval of $\boldsymbol{\theta}$, since $\mathbf{T}$ in Equation (13) also fulfills two requirements of the generalized pivotal quantity and the observed value of $\mathbf{T}$ is $\boldsymbol{\theta}$, so it can be used to construct the confidence region of $\boldsymbol{\theta}$. Let $q_{\{\|\tilde{\mathbf{T}}\|;1-\alpha\}}$ be the $100(1-\alpha)$th percentile of $\|\tilde{\mathbf{T}}\|$, such that

$$\Pr\left\{\tilde{\mathbf{T}}'\tilde{\mathbf{T}} = (\mathbf{T} - \mathbf{m}_\mathbf{T})'\mathbf{S}_\mathbf{T}^{-1}(\mathbf{T} - \mathbf{m}_\mathbf{T}) \leq q^2_{\{\|\tilde{\mathbf{T}}\|;1-\alpha\}}\right\} = 1\text{-}\alpha, \tag{16}$$

Therefore, the $100(1-\alpha)$% confidence region of $\boldsymbol{\theta}$ can be solved through

$$\left\{\boldsymbol{\theta} : (\boldsymbol{\theta} - \mathbf{m}_\mathbf{T})'\mathbf{S}_\mathbf{T}^{-1}(\boldsymbol{\theta} - \mathbf{m}_\mathbf{T}) \leq q^2_{\{\|\tilde{\mathbf{T}}\|;1-\alpha\}}\right\}. \tag{17}$$

3.2 ***Solutions based on the classical methods***

In the classical procedure, the Hotelling's T^2 test and the Chi-square test are the commonly used methods. In Hotelling's T^2 test, we assume the population covariance matrices are the same, whereas in the classical Chi-square method, practitioners usually replace the population covariance matrices with the sample covariance matrices. We will briefly introduce these two methods to deal with our problem.

3.2.1 *The Hotelling's T^2 test*

In this method, we will assume that $\mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_K = \mathbf{\Sigma}$ and $\mathbf{G} = (c_1\mathbf{I}_d, \cdots, c_K\mathbf{I}_d)$, then the point estimator of $\boldsymbol{\theta} = \mathbf{G}\boldsymbol{\mu} = \sum_{i=1}^{K} c_i\boldsymbol{\mu}_i$ and the pool covariance matrix are

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{K} c_i\bar{\mathbf{X}}_i \quad \text{and} \quad \mathbf{S}_{\mathrm{H}} = \frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i\right)\left(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i\right)' = \frac{1}{N-K}\sum_{i=1}^{K} n_i\mathbf{S}_i, \quad (18)$$

respectively, where $N = \sum_{i=1}^{K} n_i$ and $\bar{\mathbf{X}}_i$ and $\mathbf{S}_i$ are defined in Equation (2), respectively. The criterion is

$$\begin{aligned} Q^2 &= \left(\sum_{i=1}^{K} c_i\bar{\mathbf{X}}_i - \boldsymbol{\theta}\right)' \left[\sum_{i=1}^{K} \frac{c_i^2\mathbf{S}_{\mathrm{H}}}{n_i}\right]^{-1} \left(\sum_{i=1}^{K} c_i\bar{\mathbf{X}}_i - \boldsymbol{\theta}\right) \\ &= (\hat{\boldsymbol{\mu}} - \boldsymbol{\theta})' \left(b\mathbf{S}_{\mathrm{H}}\right)^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{\theta}), \end{aligned}$$

where Q^2 has the Hotelling's T^2 distribution with $N - K$ degrees of freedom and $b = \sum_{i=1}^{K} c_i^2/n_i$. Thus

$$\frac{Q^2}{N-K} \times \frac{N-K-d+1}{d} \sim F_{d,N-K-d+1}, \quad (19)$$

so the p-value for testing $H_0 : \sum_{i=1}^{K} c_i\boldsymbol{\mu}_i = \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is a given vector, is

$$p = \Pr\left[F_{(d,N-K-d+1)} > \left(\sum_{i=1}^{K} c_i\bar{\mathbf{x}}_i - \boldsymbol{\theta}_0\right)' \mathbf{S}_{\mathrm{H}}^{-1} \left(\sum_{i=1}^{K} c_i\bar{\mathbf{x}}_i - \boldsymbol{\theta}_0\right) \cdot \frac{N-K-d+1}{bd(N-K)}\right], \quad (20)$$

and the $100(1-\alpha)\%$ confidence region of $\boldsymbol{\theta}$ can be solved through the inequality

$$\left\{\boldsymbol{\theta} : (\hat{\boldsymbol{\mu}} - \boldsymbol{\theta})'\mathbf{S}_{\mathrm{H}}^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\theta}) \le \frac{bd(N-K)}{N-K-d+1} F_{1-\alpha}(d, N-K-d+1)\right\}, \quad (21)$$

where $F_{1-\alpha}(d, N-K-d+1)$is the $100(1-\alpha)$th percentile of the $F_{d,N-K-d+1}$ distribution.

3.2.2 *The classical Chi-square test*

The classical Chi-square method is valid when the covariance matrices are known. The statistics $\mathrm{H}_d^2, \mathrm{H}_d^2 = (\hat{\boldsymbol{\mu}} - \boldsymbol{\theta})' \left[\sum_{i=1}^{K} c_i^2\mathbf{S}_i/(n_i - 1)\right]^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{\theta})$, is distributed approximately as a Chi-square distribution with degrees of freedom d when the sample sizes tend to infinity, where $\hat{\boldsymbol{\mu}} = \sum_{i=1}^{K} c_i\bar{\mathbf{X}}_i$ and $\boldsymbol{\theta} = \sum_{i=1}^{K} c_i\boldsymbol{\mu}_i$. The p-value for testing $H_0 : \sum_{i=1}^{K} c_i\boldsymbol{\mu}_i = \boldsymbol{\theta}_0$ is

$$p = \Pr\left[\chi_d^2 > \left(\sum_{i=1}^{K} c_i\bar{\mathbf{x}}_i - \boldsymbol{\theta}_0\right)' \left[\sum_{i=1}^{K} \frac{c_i^2\mathbf{s}_i}{(n_i-1)}\right]^{-1} \left(\sum_{i=1}^{K} c_i\bar{\mathbf{x}}_i - \boldsymbol{\theta}_0\right)\right], \quad (22)$$

and the approximate $100(1-\alpha)\%$ confidence region of $\boldsymbol{\theta}$ may be obtained by evaluating

$$\left\{\boldsymbol{\theta} : (\hat{\boldsymbol{\mu}} - \boldsymbol{\theta})' \left(\sum_{i=1}^{K} \frac{c_i^2\mathbf{s}_i}{(n_i-1)}\right)^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{\theta}) \le \chi_{1-\alpha}^2(d)\right\}, \quad (23)$$

where $\chi_{1-\alpha}^2(d)$ is the $100(1-\alpha)$th percentile of the χ^2 distribution with degrees of freedom d.

3.3 *The multivariate Behrens–Fisher problem*

If we are only interested in the multivariate Behrens–Fisher problem, that is, only two populations are related and $c_1 = 1$ and $c_2 = -1$, i.e., $\mathbf{G} = (\mathbf{I}_d, -\mathbf{I}_d)$, then Equation (13) for the generalized pivotal quantity becomes

$$\mathbf{T}_1(\bar{\mathbf{X}}, \mathbf{S}; \bar{\mathbf{x}}, \mathbf{s}) = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \left(\mathbf{s}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{s}_1^{1/2} + \mathbf{s}_2^{1/2}\mathbf{R}_2^{-1}\mathbf{s}_2^{1/2}\right)^{1/2} \mathbf{Z}_d. \tag{24}$$

The p-value for testing

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \tag{25}$$

is similar to Equation (15) by replacing $\tilde{\mathbf{T}}$ and $\tilde{\boldsymbol{\theta}}_0$ with $\tilde{\mathbf{T}}_1$ and $\tilde{\mathbf{0}}$, respectively.

Some other methods for dealing with the multivariate Behrens–Fisher problem are briefly reviewed in the following.

3.3.1 *Gamage, Mathew and Weerahandi*

The p-value for testing Equation (25) derived by Gamage et al. [4] is

$$p = \Pr\left\{ \mathrm{T}_{\mathrm{Gam}} \geq (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \left(\frac{\mathbf{s}_1}{\mathrm{n}_1 - 1} + \frac{\mathbf{s}_2}{\mathrm{n}_2 - 1} \right)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) | H_0 \right\}, \tag{26}$$

where $\mathrm{T}_{\mathrm{Gam}}$ is defined as

$$\mathrm{T}_{\mathrm{Gam}} = \mathbf{Z}'[\mathbf{v}_1^{1/2}\boldsymbol{\Psi}_1^{-1}\mathbf{v}_1^{1/2} + \mathbf{v}_2^{1/2}\boldsymbol{\Psi}_2^{-1}\mathbf{v}_2^{1/2}]\mathbf{Z}. \tag{27}$$

In Equation (27), $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I}_d)$, $\mathbf{V}_{\mathrm{i}} = \mathbf{V}_{\mathrm{i}}(\mathbf{S}_{\mathrm{i}}; \mathbf{s}_1, \mathbf{s}_2) = (\mathbf{s}_1/(\mathrm{n}_1 - 1) + \mathbf{s}_2/(\mathrm{n}_2 - 1))^{-1/2}\, \mathbf{S}_{\mathrm{i}}(\mathbf{s}_1/(\mathrm{n}_1 - 1) + \mathbf{s}_2/(\mathrm{n}_2 - 1))^{-1/2}$ with $\mathbf{v}_i = \mathbf{V}_{\mathrm{i}}(\mathbf{s}_{\mathrm{i}}; \mathbf{s}_1, \mathbf{s}_2)$ being the observed values of $\mathbf{V}_i$ and $\boldsymbol{\Psi}_{\mathrm{i}} \sim \mathrm{W}_{\mathrm{d}}(\mathrm{n}_{\mathrm{i}} - 1, \mathbf{I}_{\mathrm{d}})$, $i = 1, 2$.

Furthermore, they also defined $\mathrm{T}^*_{\mathrm{Gam}}/\mathrm{t}^*_{\mathrm{Gam}}$, to test the MANOVA problem of the form $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_K$, where $\mathrm{T}^*_{\mathrm{Gam}}(\Sigma_1, \ldots, \Sigma_K) = \sum_{\mathrm{i}=1}^{\mathrm{K}} n_i(\bar{\mathbf{X}}_{\mathrm{i}} - \hat{\boldsymbol{\mu}})'\Sigma_{\mathrm{i}}^{-1}(\bar{\mathbf{X}}_{\mathrm{i}} - \hat{\boldsymbol{\mu}})$, $\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{K} n_i \Sigma_i^{-1}\right)^{-1} \sum_{i=1}^{K} n_i \Sigma_i^{-1} \bar{\mathbf{X}}_{\mathrm{i}}$ and $\mathrm{t}^*_{\mathrm{Gam}}$ is the observed value of $\mathrm{T}^*_{\mathrm{Gam}}$. However, as the authors had mentioned in their paper, this new GTV $\mathrm{T}^*_{\mathrm{Gam}}/\mathrm{t}^*_{\mathrm{Gam}}$ was not invariant under nonsingular transformation [4].

3.3.2 *Krishnamoorthy and Yu*

Krishnamoorthy and Yu [7] modified the Nel and Van der Merwe's [8] test and provided an approximate invariant solution for the multivariate Behrens–Fisher problem. They obtained a nonsingular invariant statistic

$$\mathrm{T}_{\mathrm{Kri}} = \left[(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right]' \left[(n_1 - 1)^{-1}\mathbf{S}_1 + (n_2 - 1)^{-1}\mathbf{S}_2\right]^{-1} \left[(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right], \tag{28}$$

which is approximately distributed as $\omega d F_{d,\omega-d+1}/(\omega - d + 1)$ where $\omega = (d(d+1))/(n_1 - 1)^{-1}\left[\mathrm{tr}\Lambda_1^2 + (\mathrm{tr}\Lambda_1)^2\right] + (n_2 - 1)^{-1}\left[\mathrm{tr}\Lambda_2^2 + (\mathrm{tr}\Lambda_2)^2\right]$, $\Lambda_1 = \mathbf{S}_1/n_1 - 1(\mathbf{S}_1/n_1 - 1 + \mathbf{S}_2/n_2 - 1)^{-1}$, $\Lambda_2 = \mathbf{S}_2/n_2 - 1(\mathbf{S}_1/n_1 - 1 + \mathbf{S}_2/n_2 - 1)^{-1}$.

The p-value for testing Equation (25) is

$$p = \Pr\left\{ F_{d,\omega-d+1} \geq \frac{\omega - d + 1}{\omega d} \cdot (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \left(\frac{\mathbf{s}_1}{\mathrm{n}_1 - 1} + \frac{\mathbf{s}_2}{\mathrm{n}_2 - 1} \right)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) | H_0 \right\}. \tag{29}$$

4. Simulation studies

In this section, we first consider the multivariate Behrens–Fisher problem compared with five methods with their type I errors. Then, for the case of $K = 3(> 2)$, we present expected areas and coverage probabilities of three methods for various sample sizes and parameter configurations. According to the eigen decomposition theorem, for any positive definite matrix $\mathbf{\Sigma}$, there exists an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}'\mathbf{\Sigma}\mathbf{P}$ is diagonal. Without loss of generality, the normal random vectors are generated with zero mean vector and diagonal covariance matrices $\mathbf{\Sigma}_i$ with $\mathbf{\Sigma}_1 = \mathbf{I}_d$ in our simulation studies. The simulation was processed using MATLAB software.

4.1 *The multivariate Behrens–Fisher problem*

We apply five methods to calculate the type I error probabilities of multivariate Behrens–Fisher problem with various sample sizes (n_1, n_2), dimension numbers d and diagonal covariance matrices $\mathbf{\Sigma}_1 = \mathbf{I}_d$ and $\mathbf{\Sigma}_2 = a\mathbf{I}_d$, where a stands for the heterogeneity rate. We consider $d = 2$ and $d = 4$ with direct pairing and inverse pairing. The results are in Tables 1 and 2, respectively.

Table 1. Type I error with 10,000 iterations $\mathbf{\Sigma}_1 = \mathbf{I}_2$, $\mathbf{\Sigma}_2 = a\mathbf{I}_2$.

a	General	Hote	Chi	Gam	Kri
$n_1 = 10, n_2 = 15$					
9	0.045	0.032	0.082	0.045	0.053
25	0.046	0.034	0.098	0.046	0.056
100	0.048	0.030	0.097	0.048	0.048
400	0.051	0.034	0.103	0.050	0.053
$n_1 = 10, n_2 = 10$					
9	0.046	0.088	0.106	0.046	0.053
25	0.050	0.098	0.118	0.048	0.052
100	0.049	0.102	0.122	0.049	0.049
400	0.051	0.114	0.135	0.052	0.053
$n_1 = 15, n_2 = 10$					
9	0.048	0.116	0.112	0.048	0.055
25	0.050	0.192	0.124	0.049	0.051
100	0.048	0.208	0.129	0.048	0.050
400	0.050	0.219	0.132	0.051	0.051

Table 2. Type I error with 10,000 iterations $\mathbf{\Sigma}_1 = \mathbf{I}_4$, $\mathbf{\Sigma}_2 = a\mathbf{I}_4$.

a	General	Hote	Chi	Gam	Kri
$n_1 = 10, n_2 = 5$					
9	0.038	0.438	0.556	0	0.097
25	0.046	0.620	0.656	0	0.127
100	0.048	0.735	0.699	0.001	0.140
400	0.051	0.781	0.722	0.008	0.128
$n_1 = 10, n_2 = 10$					
9	0.033	0.136	0.227	0.032	0.057
25	0.042	0.170	0.258	0.042	0.056
100	0.048	0.190	0.281	0.048	0.048
400	0.053	0.193	0.284	0.053	0.049
$n_1 = 10, n_2 = 20$					
9	0.032	0.014	0.121	0.032	0.056
25	0.043	0.011	0.125	0.043	0.054
100	0.048	0.010	0.136	0.048	0.051
400	0.050	0.011	0.145	0.050	0.052

The smaller sample sizes associated with smaller variances is called direct pairing, whereas the smaller sample sizes associated with larger variances is called inverse pairing. Each combination is based on 10,000 replicates with $\alpha = 0.05$ and these comparisons presented correspond to:

(1) General: The generalized method proposed in this article.
(2) Hote: The classical Hotelling's method.
(3) Chi: Classical Chi-square test.
(4) Gam: Gamage, Mathew and Weerahandi.
(5) Kri: Krishnamoorthy and Yu.

The methods (1) and (4) are both based on 5000 runs in each simulation. From Table 1, it is interesting to find that the results based on our proposed method are very close to those proposed by Gamage et al. [4] except those obtained only by simulation and rounding off errors. Both of them have the type I error probabilities close to the nominal level. The method proposed by Krishnamoorthy and Yu [7] also has similar results except for a few combinations. However, in Table 2 with $d = 4$, excepting our proposed method, there are unanticipated results in the inverse pairing case $n_1 = 10, n_2 = 5$. The method proposed by Gamage et al. [4] tends to accept the null hypothesis (25) since the generalized p-values calculated by their test variable do not have a uniform distribution in this case while we use the standardized GTV to calculate the generalized p-values. And the type I error probabilities of the method proposed by Krishnamoorthy and Yu [7] range from 0.11 to 0.16. The type I error probabilities calculated based on the classical Hotelling's method are underestimated when direct pairing is considered and overestimated when two sample sizes are equal or inverse pairing is considered. Those obtained based on the classical Chi-square test are overestimated in all combinations and their performances grow worse as the degree of nonhomogeneity increases. This comes to a similar conclusion with a number of other problems solved based on generalized p-values, see Thursby [12], Zhou and Mathew [18] and many others. They found that when the covariance matrices are quite different, the nominal significance level obtained by the Hotelling's and the Chi-square methods may be distorted.

Although the method proposed by Krishnamoorthy and Yu [7] is a strong candidate for the multivariate Behrens–Fisher problem, it has some weaknesses for particular combinations of sample sizes, dimensions and parameter configurations. When inverse pairing is considered and the smallest sample size is close to the dimension, the type I error rates obtained by Krishnamoorthy and Yu [7] are especially higher than the nominal level. As shown in Table 2, the type I errors are about double to triple the nominal level 0.05 when variable number is 4 and sample sizes are (10, 5). Moreover, it can be used only in two populations. Furthermore, we also calculate the type I error probabilities for higher dimensions and large sample sizes. Since the results are similar to Tables 1 and 2, we show some of them in Table 3 with $d = 8, 10, 20$ and $n_1 = 30, 50, 100$. Thus, for overall comparisons from Tables 1 to 3, we conclude that our proposed method is a good alternative for the multivariate Behrens–Fisher problem especially when inverse pairing and heteroscedastity are considered.

4.2 *The expected areas and coverage probabilities*

In simulation studies, we used 1000 iterations to calculate the expected areas of the 95% confidence regions and the corresponding coverage probabilities of $c_1\boldsymbol{\mu}_1 + c_2\boldsymbol{\mu}_2 + c_3\boldsymbol{\mu}_3$ under different scenarios. First, we chose $(c_1, c_2, c_3) = (1, -1, 0)$, which is known as the multivariate Behrens–Fisher problem, and the results compared with five methods are in Table 4. Next, we choose $(c_1, c_2, c_3) = (0.5, 0.5, -1)$ and the results compared with three methods are in Table 5.

From Table 4, we find that the coverage probabilities obtained by the Hotelling's method are overestimated when large sample sizes are associated with large covariance matrices and vice

Table 3. Type I error with 10,000 iterations $\mathbf{\Sigma}_1 = \mathbf{I}_d$, $\mathbf{\Sigma}_2 = a\mathbf{I}_d$.

a	General	Hote	Chi	Gam	Kri
$d = 8;\ (n_1\ n_2) = (30\ 10)$					
9	0.039	0.715	0.718	0	0.096
25	0.045	0.875	0.787	0.001	0.100
100	0.049	0.939	0.816	0.014	0.088
400	0.050	0.961	0.830	0.038	0.072
$d = 10;\ (n_1\ n_2) = (50\ 15)$					
9	0.044	0.795	0.656	0.033	0.073
25	0.045	0.921	0.696	0.034	0.062
100	0.052	0.962	0.727	0.047	0.056
400	0.049	0.974	0.734	0.048	0.052
$d = 10;\ (n_1\ n_2) = (50\ 50)$					
9	0.046	0.090	0.142	0.045	0.054
25	0.050	0.102	0.165	0.050	0.050
100	0.051	0.111	0.176	0.051	0.051
400	0.050	0.114	0.175	0.050	0.047
$d = 20;\ (n_1\ n_2) = (100\ 25)$					
9	0.033	0.973	0.926	0	0.101
25	0.044	0.997	0.953	0.007	0.082
100	0.050	1.000	0.962	0.037	0.060
400	0.048	1.000	0.967	0.047	0.048
$d = 20;\ (n_1\ n_2) = (100\ 100)$					
9	0.045	0.092	0.173	0.044	0.051
25	0.048	0.113	0.203	0.048	0.050
100	0.053	0.128	0.224	0.053	0.053
400	0.050	0.131	0.226	0.048	0.049

Table 4. Expected areas of 95% confidence regions and coverage probabilities of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ under $\mathbf{\Sigma}_1 = \mathbf{I}_2$, and $\mathbf{\Sigma}_2 = (n_2/n_1)a\mathbf{I}_2$.

a	General	Hote	Chi	Gam	Kri
$n_1 = 10, n_2 = 20$					
15	36.758(0.963)	65.0438(0.993)	29.056(0.924)	36.765(0.963)	35.063(0.938)
25	59.280(0.958)	107.646(0.994)	47.089(0.922)	59.294(0.960)	57.072(0.957)
50	115.603(0.959)	214.151(0.996)	92.155(0.925)	115.664(0.961)	114.061(0.951)
100	228.298(0.959)	427.157(0.996)	182.275(0.926)	228.378(0.960)	224.332(0.953)
500	1129.820(0.959)	2131.21(0.997)	903.209(0.929)	1129.990(0.959)	1109.206(0.962)
$n_1 = 20, n_2 = 10$					
15	22.078(0.961)	9.413(0.820)	13.574(0.896)	22.089(0.961)	21.249(0.936)
25	36.033(0.961)	14.124(0.791)	21.928(0.892)	36.061(0.961)	34.617(0.950)
50	70.938(0.955)	25.845(0.769)	42.801(0.891)	70.982(0.956)	69.578(0.944)
100	140.755(0.957)	49.233(0.751)	84.539(0.892)	140.829(0.956)	139.740(0.938)
500	699.357(0.959)	236.168(0.741)	418.422(0.889)	699.641(0.959)	693.913(0.957)

versa. The coverage probabilities obtained by the Chi-square method are underestimated in all cases. On the other hand, the remaining three methods have good coverage probabilities and similar expected areas in all cases. In Table 5, although the Hotelling's method and the Chi-square test have smaller average areas of 95% confidence regions, their confidence regions are too small to ensure their coverage and probabilities are close to the nominal level of 0.95. On the contrary, these simulated results support that our method not only assures the level of the test in all cases, but also has good coverage probabilities comparing to those of the classical Hotelling's method and the classical Chi-square test.

Table 5. Expected areas of 95% confidence regions and coverage probabilities of $\boldsymbol{\mu}_1/2 + \boldsymbol{\mu}_2/2 - \boldsymbol{\mu}_3$ under $\boldsymbol{\Sigma}_1 = \mathbf{I}_2$, $\boldsymbol{\Sigma}_2 = 3\mathbf{I}_2$ and $\boldsymbol{\Sigma}_3 = a\mathbf{I}_2$.

a	General	Hote	Chi
$(n_1\ n_2\ n_3) = (10\ 8\ 5)$			
9	118.257(0.957)	18.319(0.716)	29.651(0.776)
25	297.648(0.954)	32.708(0.588)	71.844(0.738)
50	614.952(0.959)	56.745(0.536)	145.938(0.757)
100	1204.655(0.941)	101.422(0.486)	284.573(0.750)
500	5926.295(0.953)	452.362(0.463)	1388.306(0.756)
$(n_1\ n_2\ n_3) = (8\ 10\ 5)$			
9	110.010(0.959)	18.991(0.758)	27.637(0.790)
25	299.104(0.961)	34.243(0.602)	72.169(0.758)
50	616.133(0.954)	58.086(0.555)	145.488(0.755)
100	1208.969(0.953)	102.749(0.507)	284.487(0.770)
500	6063.718(0.952)	463.086(0.454)	1417.512(0.759)
$(n_1\ n_2\ n_3) = (5\ 10\ 8)$			
9	42.678(0.966)	20.266(0.884)	20.828(0.866)
25	106.023(0.954)	42.525(0.804)	52.563(0.846)
50	207.290(0.968)	77.877(0.807)	103.119(0.868)
100	401.150(0.941)	145.413(0.774)	200.280(0.841)
500	2061.426(0.945)	721.244(0.760)	1031.064(0.849)

5. Illustrative examples

5.1 *Example 1*

Zerbe [16] analysed the plasma inorganic phosphate flux data to study the association of hyperglycemia and relative hyperinsulinemia. The standard glucose tolerance tests were administered to 13 control (C) and 20 obese (O) patients on the Pediatric Clinical Research Ward of the University of Colorado Medical Center. Zerbe and Murphy [17] divided the 20 obese patients into two subgroups; the first 12 obese patients were nonhyperinsulinemic (NO) while the latter eight were hyperinsulinemic (HO). The sample means of plasma inorganic phosphate measurements determined from blood samples withdrawn 0, 0.5, 1, 1.5, 2, 3, 4 and 5 hours after a standard-dose oral glucose challenge are reported in Table 6. The researchers wanted to compare the mean curves separately over the first 3 and last 2 hours of the glucose tolerance test since the metabolic mechanism responsible for the liver changes.

We consider the multivariate Behrens–Fisher problem twice to see whether two mean vectors are equal. First, we want to test if the mean curves of the nonhyperinsulinemic obese group and the hyperinsulinemic obese group are the same. If we cannot reject this null hypothesis, we further discuss the equality of the mean curves of the control group and the obese group, and all results are in Table 7. We regard the ratio of determinants of sample covariance matrices as the crude index of the heteroscedasticity. From Table 7, $\mathbf{m_T}$ and $\mathbf{G}\bar{\mathbf{x}}$ are very close, and ratios donot display strong

Table 6. Sample means of plasma inorganic phosphate (mg/dl).

Group		Hours after glucose challenge							
		0	0.5	1	1.5	2	3	4	5
C		4.092	3.262	2.723	2.631	3.046	3.346	3.515	3.939
O		4.530	4.140	3.780	3.480	3.195	3.375	3.700	4.015
O	NO	4.358	4.033	3.567	3.292	3.100	3.333	3.708	4.000
O	HO	4.788	4.300	4.100	3.763	3.338	3.438	3.688	4.038

Table 7. Various comparisons of mean flux curves over selected time intervals following oral glucose challenge.

Groups	Interval	$\mathbf{m}'_{\mathbf{T}}$ and $(\mathbf{G}\bar{\mathbf{x}})'$	Ratios	*p*-value		
				General	Gam	Kri
(NO, HO)	0–3 hours	0.42, 0.26, 0.53, 0.47,0.23, 0.11 (0.43, 0.27, 0.53, 0.47, 0.24, 0.10)	(10, 1)	0.695	0.670	0.455
(NO, HO)	3–5 hours	0.10, −0.02, 0.04 (0.10, −0.02, 0.04)	(1, 1.9)	0.869	0.897	0.880
(C, O)	0–3 hours	0.44, 0.88, 1.06, 0.85, 0.15, 0.03 (0.44, 0.88, 1.06, 0.85, 0.15, −0.03)	(1, 1.6)	0.004	0.006	0.000
(C, O)	3–5 hours	0.028, 0.183, 0.078 (0.029, 0.185, 0.077)	(2.1, 1)	0.651	0.665	0.617
(C, O)	0–5 hours	0.4, 0.9, 1.1, 0.9, 0.1, 0.03, 0.19, 0.08 (0.4, 0.9, 1.1, 0.8, 0.1, 0.03, 0.18, 0.08)	(1, 2.0)	0.036	0.050	0.001
Other comparisons						
(C, NO)	0–3 hours	0.25, 0.76, 0.83, 0.65, 0.05, −0.02 (0.27, 0.77, 0.84, 0.66, 0.05, −0.01)	(1.2, 1)	0.021	0.023	0.007
(C, NO)	3–5 hours	−0.015, 0.190, 0.058 (−0.013, 0.193, 0.062)	(3.4, 1)	0.642	0.642	0.579
(C, HO)	0–3 hours	0.69, 1.04, 1.37, 1.13, 0.29, 0.09 (0.70, 1.04, 1.38, 1.13, 0.29, 0.09)	(12, 1)	0.007	0.014	0.001
(C, HO)	3-5 hours	0.095, 0.178, 0.108 (0.091, 0.172, 0.099)	(1.8, 1)	0.899	0.923	0.902

heteroscedasticity among groups. The *p*-values in Table 7 indicate that no significant evidence exists to reject the null hypothesis that the mean curve of the nonhyperinsulinemic obese group and that of the hyperinsulinemic obese group are equal. However, the mean curves of the control group and the obese group are the same in the 3–5 hours interval, but different in the 0–3 hours interval. Hence the metabolic mechanisms over the first 3 hours of the glucose tolerance test should be quite different between the control group and the obese group. We also run some tests with the similar conclusions as [17]. It should be noted that we used $\mathbf{G}^*$ to test the equality of the mean curves of three groups (C, NO, HO). In the 3–5 hours interval, the ratio of determinants is (3.37, 1, 1.92) and the *p*-value by our method is 0.905, which strongly supports the null hypothesis. In the 0–3 hours' interval, the ratio of determinants is (11.8, 9.88, 1) and the *p*-value is 0.035, which reject the null hypothesis.

5.2 *Example 2*

Sterczer et al. [10] studied the effect of tap water and three kinds of cholagogues, magnesium sulfate, clanobutin and cholecystokinin, on changes in the gallbladder volume (GBV) by two-dimensional ultrasonography in six healthy dogs. In this experiment, the dogs were treated with each test substance and GBV (ml) was measured immediately before the administration of each test substance and at 10-minute intervals for 120 minutes thereafter. They found that the changes in the GBV treated with magnesium sulfate were very similar to those treated with clanobutin. The GBV data was available in Reiczigel [9].

Studying the human medical literature about the effects exerted by tap water and clanobutin, a researcher experimented with cocktail therapy, mixing 70% tap water, 20% clanobutin and 10% cholecystokinin. The knowledge of $\mathbf{G}\boldsymbol{\mu}$ could help him to prevent the patients' uncomfortableness, or the threshold value $\boldsymbol{\theta}_0$. The ratio of canine GBV to human beings is about 3:1 (50:17.4), and the ratios of one minus the maximal reductions in canine GBV to human beings are 0.75 and 0.87, with respective to tap water and clanobutin. Hence he can set $\mathbf{G}\boldsymbol{\mu} = 3 * (0.7 * 0.75\boldsymbol{\mu}_1 + 0.2 * 0.87\boldsymbol{\mu}_2 + 0.1\boldsymbol{\mu}_3) = 1.575\boldsymbol{\mu}_1 + 0.522\boldsymbol{\mu}_2 + 0.3\boldsymbol{\mu}_3$. To ensure the inverse of the sample covariance

Table 8. Sample means of GBV and the 95% confidence region of $\mathbf{G}\boldsymbol{\mu}$.

	Minutes after treatment				
	20	40	60	80	100
Tap water	12.505	14.153	15.242	16.995	18.090
Clanobutin	12.082	13.248	13.890	14.480	15.232
Cholecystokinin	16.643	16.512	16.712	16.853	16.455
$\mathbf{m_T}$	30.654	33.817	35.821	39.016	40.977
	2014.1				
	2552.8	3285.4			
$\mathbf{S_T}$	2795.0	3588.4	3940.8		
	2901.5	3751.9	4110.0	4309.2	
	2984.2	3837.7	4212.5	4402.2	4520.6

The 95% confidence region of $\mathbf{G}\boldsymbol{\mu}$: $\left\{\mathbf{G}\boldsymbol{\mu} : (\mathbf{G}\boldsymbol{\mu} - \mathbf{m_T})'\mathbf{S_T}^{-1}(\mathbf{G}\boldsymbol{\mu} - \mathbf{m_T}) \leq 24.704\right\}$

matrix exists with probability one, the dimension of the measurements must be less than six. We take the first five measurements at 20-minute intervals for 100 minutes and the ratio of determinants is (1217.8, 1, 1.6). The 95% confidence region of $\mathbf{G}\boldsymbol{\mu}$ from Equation (17) and the summary data are in Table 8. The researcher can check to see if $\boldsymbol{\theta}_0$ is in the 95% confidence region with $q^2_{\{\|\tilde{\mathbf{T}}\|;95\%\}} = 24.704$.

In Example 1, we not only test the multivariate Behrens–Fisher problem twice but also test the MANOVA problem. We illustrate the process to find $\mathbf{G}$ and the procedure for constructing the 95% confidence region based on our proposed method in Example 2. It should be noted that in the Edgar Anderson's Iris data, the 95% confidence region of $3\boldsymbol{\mu}_1 - 2\boldsymbol{\mu}_2 - \boldsymbol{\mu}_3$ does not contain $\mathbf{0}$, which means that such a relationship among the three species does not exist.

6. Conclusions

In this paper, the generalized method provides an alternative way of dealing with a linear combination of mean vectors with different covariance matrices among multiple groups. We demonstrate the advantages of our proposed method when more measurements are taken and there are serious heteroscedasticities among groups. The traditional methods usually are restricted to the equal covariance matrices among group or known covariance matrices that are sometimes not available in practical applications. According to the numerical examples, our proposed method is recommended since the generalized p-values assure the level of the test in all simulated cases. Moreover, the coverage probabilities and the expected areas are satisfactory while the other methods become worse when the heteroscedasticities increased. Therefore, it is fair to say that our proposed method with fitting $\mathbf{G}$ is quite applicable for practical use, especially in compound material, mixed aqueous solution and cocktail therapy.

Acknowledgement

This research has been supported by grant NSC 95-2118-M-025-002 from the National Science Council of Taiwan.

References

[1] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley, Hoboken, NJ, 2003.
[2] B.M. Bennett, *Note on a solution of the generalized Nehrens–Fisher problem*, Ann. Inst. Statist. Math. 2 (1951), pp. 87–90.

[3] W.F. Christensen and A.C. Rencher, *A comparison of type I error rates and power levels for seven solutions to the multivariate Behrens–Fisher problem*, Commun. Statist. Simul. Comput. 26 (1997), pp. 1251–1273.
[4] J. Gamage, T. Mathew, and S. Weerahandi, *Generalized p-values and generalized confidence regions for the multivariate Behrens–Fisher problem and MANOVA*, J. Multivariate Anal. 88 (2004), pp. 177–189.
[5] A. Johnson and S. Weerahandi, *A Bayesian solution to the multivariate Behrens–Fisher problem*, J. Amer. Statist. Assoc. 83 (1988), pp. 145–149.
[6] S. Kim, *A practical solution to the multivariate Behrens–Fisher problem*, Biometrika 79 (1992), pp. 171–176.
[7] K. Krishnamoorthy and J. Yu, *Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem*, Stat. Probab. Lett. 66 (2004), pp. 161–169.
[8] D.G. Nel and C.A. Van der Merwe, *A solution to the multivariate Behrens–Fisher problem*, Communi. Stat. Theory Methods 15 (1986), pp. 3719–3735.
[9] J. Reiczigel, *Analysis of experimental data with repeated measurements*, Biometrics 55 (1999), pp. 1059–1063.
[10] A. Sterczer, K. Vörös, and F. Karsai, *Effect of cholagogues on the volume of the gallbladder of dogs*, Res. Vet. Sci. 60 (1996), pp. 44–47.
[11] K.W. Tsui and S. Weerahandi, *Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters*, J. Amer. Statist. Assoc. 84 (1989), pp. 602–607.
[12] J.G. Thursby, *A comparison of several exact and approximate tests for structural shift under heteroscedasticity*, J. Econometrics 53 (1992), pp. 363–386.
[13] S. Weerahandi, *Generalized confidence intervals*, J. Amer. Statist. Assoc. 88 (1993), pp. 899–905.
[14] S. Weerahandi, *Exact Statistical Methods for Data Analysis*, Springer-Verlag, New York, 1995.
[15] S. Weerahandi, *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models*, Wiley, New Jersey, 1995.
[16] G.O. Zerbe, *Randomization analysis of the completely randomized design extended to growth and response curves*, J. Amer. Statist. Assoc. 74 (1979), pp. 215–221.
[17] G.O. Zerbe and J.R. Murphy, *On multiple comparisons in the randomization analysis of growth and response curves*, Biometrics 42 (1986), pp. 795–804.
[18] L. Zhou and T. Mathew, *Some tests for variance components using generalized p-values*, Technometrics 36 (1994), pp. 394–402.