Open Journal of Statistics, **2012, 2, 1-11**
http://dx.doi.org/10.4236/ojs.2012.21001 Published Online January 2012 (http://www.SciRP.org/journal/ojs)

An Approximate Hotelling T^2-Test for Heteroscedastic One-Way MANOVA

Jin-Ting Zhang
Department of Statistics and Applied Probability, National University of Singapore, Singapore
Email: stazjt@nus.edu.sg

Received December 9, 2011; revised December 28, 2011; accepted January 11, 2012

ABSTRACT

In this paper, we consider the general linear hypothesis testing (GLHT) problem in heteroscedastic one-way MANOVA. The well-known Wald-type test statistic is used. Its null distribution is approximated by a Hotelling T^2 distribution with one parameter estimated from the data, resulting in the so-called approximate Hotelling T^2 (AHT) test. The AHT test is shown to be invariant under affine transformation, different choices of the contrast matrix specifying the same hypothesis, and different labeling schemes of the mean vectors. The AHT test can be simply conducted using the usual F-distribution. Simulation studies and real data applications show that the AHT test substantially outperforms the test of [1] and is comparable to the parametric bootstrap (PB) test of [2] for the multivariate k-sample Behrens-Fisher problem which is a special case of the GLHT problem in heteroscedastic one-way MANOVA.

Keywords: Approximate Hotelling T^2 Test; Multivariate k-Sample Behrens-Fisher Problem; Wishart-Approximation; Wishart Mixture

1. Introduction

The problem of comparing the mean vectors of k multivariate populations based on k independent samples is referred to as multivariate analysis of variance (MANOVA). If the k covariance matrices are assumed to be equal, Wilks' likelihood ratio, Lawley-Hotelling trace, Bartlett-Nanda-Pillai's trace and Roy's largest root tests ([3], Ch. 8, Sec. 6) can be used. When $k = 2$, Hotelling's T^2 test is the uniformly most powerful affine invariant test. These tests, however, may become seriously biased when the assumption of equality of covariance matrices is violated. In real data analysis, such an assumption is often violated and is hard to check.

The problem for testing the difference between two normal mean vectors without assuming equality of covariance matrices is referred to as the multivariate Behrens-Fisher (BF) problem. This problem has been well addressed in the literature. Reference [4] essentially showed, via some intensive simulations, that when there is no information about the correctness of the assumption of the equality of the covariance matrices, it is better to directly proceed to make inference using some BF testing procedure which is robust against the violation of the assumption, e.g., using the modified Nel and van der Mere's (MNV) test proposed by [5]. Other such testing procedures include those proposed by [1,6-11], among others. Reference [12] compared seven tests and recommended the tests of [8,9]. However, Reference [5] noted that both [8,9]'s tests are not affine invariant. Further studies by [5,10,11] indicate that the MNV test is comparable to, or better than, other affine invariant tests.

When $k > 2$, and the covariance matrices are unknown and arbitrary, the problem of testing equality of the mean vectors is more complex, and only approximate solutions are available. Some of these solutions are obtained via generalizing the associated solutions to the univariate BF problem. For example, Reference [6]'s first and second-order tests are extensions of his series solutions to the univariate BF problem. Reference [1] generalized [13]'s univariate approximate degrees of freedom solution to heteroscedastic one-way MANOVA. Both tests are based on an affine-invariant test statistic but used different approaches to approximate its null distribution. Reference [14] proposed a generalized F-test. Reference [15] compared James's first and second-order tests, Johansen's test, and Bartlett-Nanda-Pillai's trace test and concluded that none of them is satisfactory for all sample sizes and parameter configurations. Overall, they recommended the James second-order test followed by the Johansen test. Reference [2] claimed, based on a preliminary study, that the James second-order test is computationally very involved, and is difficult to apply when $k = 3$ or more, and offered little improvement over the Johansen test. They then proposed a parametric bootstrap (PB) test to the multivariate k-sample BF problem or heteroscedastic one-way MANOVA, which is an extension of their test to the univariate k sample BF problem (see [16]). They compared their test, via some intensive simulations for various sample sizes and parameter configurations against the Johansen test and the generalized F-test of [14] and found that their PB test performed best, followed by the Johansen test while the generalized F-test performed worst.

In this paper, we consider the general linear hypothesis testing (GLHT) problem in heteroscedastic one-way MANOVA. The well-known Wald-type test statistic is used. Its null distribution is approximated by a Hotelling T^2 distribution with one parameter estimated from the data, resulting in the so-called approximate Hotelling T^2 (AHT) test. The AHT test can be regarded as a natural extension of [8]'s test and [5]'s MNV test from for the multivariate two-sample BF problem to for the GLHT problem in heteroscedastic one-way MANOVA. In view of the good performance of [8]'s test (see [12]), the MNV test (see [5]), and the AHT tests for one-way and two-way ANOVA (see [17,18]), we expect that the AHT test will also perform well for heteroscedastic one-way MANOVA. The AHT test is shown to be invariant under affine transformation, different choices of the contrast matrix used to specify the same hypothesis, and different labeling schemes of the mean vectors. It can be simply conducted using the usual F-distribution. Intensive simulations are conducted to compare the AHT test against the Johansen test and the PB test under various sample sizes and parameter configurations. The simulation results show that the AHT test indeed performs well and it outperforms the Johansen test substantially and is comparable to the PB test of [2].

The rest of the paper is organized as follows. In Section 2, the AHT test is developed. Simulation studies are presented in Section 3. An application to a real data set is given in Section 4. Technical proofs of the main results are outlined in Section 5.

2. Main Results

2.1. The Wald-Type Test Statistic

Given k independent normal samples

$$\boldsymbol{x}_{lj}, j=1,2,\cdots,n_l \sim N_p\left(\boldsymbol{\mu}_l,\Sigma_l\right), l=1,2,\cdots,k, \quad (1)$$

where and throughout, $N_p\left(\boldsymbol{\mu},\boldsymbol{V}\right)$ denotes a p-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$, we want to test whether the k mean vectors are equal:

$$H_0:\boldsymbol{\mu}_1=\boldsymbol{\mu}_2=\cdots=\boldsymbol{\mu}_k, \text{ versus } H_1:H_0 \text{ is not true}, \quad (2)$$

without assuming the equality of the covariance matrices $\Sigma_l, l=1,2,\cdots,k$. The above problem is usually referred to as the multivariate k-sample BF problem or the overall heteroscedastic one-way MANOVA test, which is a special case of the following GLHT problem in heteroscedastic one-way MANOVA:

$$H_0:\boldsymbol{C\mu}=\boldsymbol{c}, \text{ vs } H_1:\boldsymbol{C\mu}\neq\boldsymbol{c}, \quad (3)$$

where $\boldsymbol{\mu}=\left[\boldsymbol{\mu}_1^T,\boldsymbol{\mu}_2^T,\cdots,\boldsymbol{\mu}_k^T\right]^T$ is a long mean vector obtained via stacking all the population mean vectors of the k samples together into a single column vector, $\boldsymbol{C}:q\times(kp)$ is a known coefficient matrix with $\text{Rank}\left(\boldsymbol{C}\right)=q$, and $\boldsymbol{c}:q\times 1$ is a known constant vector. In fact, the GLHT problem (3) reduces to the multivariate k-sample BF problem (2) when we set $\boldsymbol{c}=\boldsymbol{0}$ and set $\boldsymbol{C}=\left[\boldsymbol{I}_{k-1},-1_{k-1}\right]\otimes\boldsymbol{I}_p$, a contrast matrix whose rows sum up to 0, where $\boldsymbol{I}_r$ and $\boldsymbol{1}_r$ denote the identity matrix of size $r\times r$ and a r-dimensional vector of ones, and $\otimes$ is the usual Kronecker product operator.

Remark 1 *The contrast matrix* $\boldsymbol{C}$ *for the null hypothesis in* (2) *is not unique. For example,* $\tilde{\boldsymbol{C}}=\left[-\boldsymbol{1}_{k-1},\boldsymbol{I}_{k-1}\right]\otimes\boldsymbol{I}_p$ *is also a contrast matrix for the null hypothesis in* (2). *However, it will be showed that the AHT test proposed in this paper will not depend on the choice of the contrast matrices specifying the same hypothesis.*

The GLHT problem (3) is very general. It includes not only the overall heteroscedastic one-way MANOVA test (2) but also various post hoc and contrast tests as special cases since any post hoc and contrast tests can be written in the form of (3). For example, when the overall heteroscedastic one-way MANOVA test is rejected, it is of interest to further test if $\boldsymbol{\mu}_1=2\boldsymbol{\mu}_2$ or if a contrast is zero, e.g., $\boldsymbol{\mu}_1-4\boldsymbol{\mu}_2+3\boldsymbol{\mu}_3=0$. In fact, these two testing problems can be written in the form of (3) with $\boldsymbol{c}=\boldsymbol{0}$ and $\boldsymbol{C}=\left(e_{1,k}-2e_{2,k}\right)^T\otimes\boldsymbol{I}_p$ and $\boldsymbol{C}=\left(e_{1,k}-4e_{2,k}+3e_{3,k}\right)^T\otimes\boldsymbol{I}_p$ respectively where and throughout $\boldsymbol{e}_{r,k}$ denotes a unit vector of length k with r-th entry being 1 and others 0.

Remark 2 *From the above various definitions of* C, *we have* $\boldsymbol{C}=\boldsymbol{C}_0\otimes\boldsymbol{I}_p$ *where* $\boldsymbol{C}_0$ *is a full rank matrix of size* $q_0\times k$ *so that we always have* $q=q_0p$. *If* $\boldsymbol{C}_0$ *is a contrast matrix, so is* $\boldsymbol{C}$.

To construct the test statistic for the GLHT problem (3), let $\hat{\boldsymbol{\mu}}_l=\bar{x}_l=n_l^{-1}\sum_{j=1}^{n_l}x_{lj}$ and $\hat{\Sigma}_l=\left(n_l-1\right)^{-1}\sum_{j=1}^{n_l}\left(x_{lj}-\hat{\boldsymbol{\mu}}_l\right)\left(x_{lj}-\hat{\boldsymbol{\mu}}_l\right)^T$ be the sample mean vector and sample covariance matrix of the l-th sample. Set $\hat{\boldsymbol{\mu}}=\left[\hat{\boldsymbol{\mu}}_1^T,\hat{\boldsymbol{\mu}}_2^T,\cdots,\hat{\boldsymbol{\mu}}_k^T\right]^T$ which is an unbiased estimator of $\boldsymbol{\mu}$. Then $\hat{\boldsymbol{\mu}}\sim N_{kp}\left(\boldsymbol{\mu},\Sigma\right)$ where $\Sigma=\text{diag}\left(\frac{\Sigma_1}{n_1},\frac{\Sigma_2}{n_2},\cdots,\frac{\Sigma_k}{n_k}\right)$. It follows that $\boldsymbol{C}\hat{\boldsymbol{\mu}}-\boldsymbol{c}\sim N_q\left(\boldsymbol{C\mu}-\boldsymbol{c},\boldsymbol{C}\Sigma\boldsymbol{C}^T\right)$. This suggests that a Wald-type test statistic can be constructed as

$$T=\left(\boldsymbol{C}\hat{\boldsymbol{\mu}}-\boldsymbol{c}\right)^T\left(\boldsymbol{C}\hat{\Sigma}\boldsymbol{C}^T\right)^{-1}\left(\boldsymbol{C}\hat{\boldsymbol{\mu}}-\boldsymbol{c}\right), \quad (4)$$

where $\hat{\Sigma} = \text{diag}\left(\frac{\hat{\Sigma}_1}{n_1}, \frac{\hat{\Sigma}_2}{n_2}, \cdots, \frac{\hat{\Sigma}_k}{n_k}\right)$. Notice that the distribution of T is very complicated and its closed-form distribution is generally not tractable in the context of heteroscedastic one-way MANOVA.

Remark 3 *When the covariance matrix homogeneity is valid and the sample covariance matrices $\hat{\Sigma}_l$ are replaced by their pooled sample covariance matrix $\sum_{l=1}^{k}(n_l-1)\hat{\Sigma}_l/(N-k)$ where $N=\sum_{l=1}^{k} n_l$ denotes the total sample size of the k samples, it is easy to show that $T/(N-k)$ follows the distribution of the well-known Lawley-Hotelling trace test statistic* ([3], *Ch.* 8, *Sec.* 6) *with q_0 and $N-k$ degrees of freedom where q_0 is defined in Remark* 2.

Remark 4 *When the covariance matrix homogeneity is actually valid, we still can apply the AHT test proposed in this paper, pretending that the k sample covariance matrices $\hat{\Sigma}_l$ were not the same. We will see that the AHT test is simpler than the Lawley-Hotelling trace test which in general does not have a closed-form formula for its null distribution; see* [3] (*Ch.* 8, *Sec.* 6). *The simulation results presented in Section* 3 *show that the AHT test works reasonably well for those covariance matrix homogeneity cases.*

To construct the AHT test based on T, following [5] and [10], we re-express T as

$$T = \boldsymbol{z}^T \boldsymbol{W}^{-1} \boldsymbol{z}, \tag{5}$$

where $\boldsymbol{z} = \left(\boldsymbol{C}\Sigma\boldsymbol{C}^T\right)^{-1/2}\left(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}\right)$ and $\boldsymbol{W} = \left(\boldsymbol{C}\Sigma\boldsymbol{C}^T\right)^{-1/2}\left(\boldsymbol{C}\hat{\Sigma}\boldsymbol{C}^T\right)\left(\boldsymbol{C}\Sigma\boldsymbol{C}^T\right)^{-1/2}$. Notice that the above re-expression theoretically helps the development of the AHT test but in practice we still use (4) to compute the value of T. We have $\boldsymbol{z} \sim N_q\left(\boldsymbol{\mu}_z, \boldsymbol{I}_q\right)$ where $\boldsymbol{\mu}_z = \left(\boldsymbol{C}\Sigma\boldsymbol{C}^T\right)^{-1/2}\left(\boldsymbol{C}\boldsymbol{\mu} - \boldsymbol{c}\right)$. Let $n_{\min} = \min_{l=1}^{k} n_l$ and $n_{\max} = \max_{l=1}^{k} n_l$. Let χ_q^2 denote a χ^2-distribution with q degrees of freedom.

Remark 5 *Assume that the sample sizes $n_1, n_2, \cdots, n_k$ tend to infinity proportionally. That is,*

$$n_l/n_{\min} \to r_l < \infty, l = 1, 2, \cdots, k, \text{as } n_{\min} \to \infty. \tag{6}$$

Then it is easy to show that as $n_{\min} \to \infty$, T asymptotically follows χ_q^2. However, we can show that the convergence rate of T to χ_q^2 is of order $n_{\min}^{-1/2}$ which is rather slow. Thus, the resulting χ^2-test is hardly useful for the heteroscedastic GLHT problem (3).

Remark 6 *When the assumption* (6) *is not satisfied, the ratio $n_{\max}/n_{\min}$ will tend to ∞ as $n_{\min} \to \infty$ so that the limit of $n_{\min}\Sigma$ is not a full rank matrix and hence the limit of $n_{\min}\boldsymbol{C}\Sigma\boldsymbol{C}^T$ is not invertible. In this case, the test statistic T is not well defined so that the AHT test proposed in this paper will not perform well.*

Let $W_r(m, \boldsymbol{V})$ denote a Wishart distribution of m degrees of freedom and with covariance matrix $\boldsymbol{V}: r \times r$. We first show that $\boldsymbol{W}$ is a Wishart mixture, *i.e.*, a linear combination of several independent Wishart random matrices. For this purpose, we decompose $\boldsymbol{C}$ into k blocks of size $q \times p$ so that $\boldsymbol{C} = \left[\boldsymbol{C}_1, \boldsymbol{C}_2, \cdots, \boldsymbol{C}_k\right]$ with $\boldsymbol{C}_1$ consisting of the first p columns of $\boldsymbol{C}$, $\boldsymbol{C}_2$ the second p-columns of $\boldsymbol{C}$, and so on.

Remark 7 *When $\boldsymbol{C} = \boldsymbol{C}_0 \otimes \boldsymbol{I}_p$ where $\boldsymbol{C}_0 = \left(\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_k\right): q_0 \times k$ with $\boldsymbol{c}_l$ being the l-th column of $\boldsymbol{C}_0$, we have $\boldsymbol{C}_l = \boldsymbol{c}_l \otimes \boldsymbol{I}_p, l = 1, 2, \cdots, k$.*

Set $\boldsymbol{H}_l = \left(\boldsymbol{C}\Sigma\boldsymbol{C}^T\right)^{-1/2}\boldsymbol{C}_l, l = 1, 2, \cdots, k$. Then $\boldsymbol{H} = \left(\boldsymbol{C}\Sigma\boldsymbol{C}^T\right)^{-1/2} = \left[\boldsymbol{H}_1, \boldsymbol{H}_2, \cdots, \boldsymbol{H}_k\right]$. Define the total variation of a random matrix $\boldsymbol{X} = \left(x_{ij}\right): m \times m$ as $V\left(\boldsymbol{X}\right) = \text{Etr}\left(\boldsymbol{X} - \text{E}\boldsymbol{X}\right)^2 = \sum_{i=1}^{m}\sum_{j=1}^{m}\text{Var}\left(x_{ij}\right)$, *i.e.*, the sum of the variances of all the entries of $\boldsymbol{X}$.

Theorem 1 *We have*

$$\boldsymbol{W} = \boldsymbol{H}\hat{\Sigma}\boldsymbol{H}^T = \sum_{l=1}^{k}\boldsymbol{W}_l, \tag{7}$$

where $\boldsymbol{W}_l = n_l^{-1}\boldsymbol{H}_l\hat{\Sigma}_l\boldsymbol{H}_l^T \sim W_q\left(n_l - 1, \frac{\boldsymbol{\Omega}_l}{n_l - 1}\right), l = 1, 2, \cdots, k$ are independent with $\Omega_l = n_l^{-1}\boldsymbol{H}_l\Sigma_l\boldsymbol{H}_l^T$. Furthermore,

$$\begin{aligned} E\left(\boldsymbol{W}\right) &= \sum_{l=1}^{k}\boldsymbol{\Omega}_l = \boldsymbol{I}_q, \\ V\left(\boldsymbol{W}\right) &= \sum_{l=1}^{k}\left(n_l - 1\right)^{-1}\left[tr\left(\boldsymbol{\Omega}_l^2\right) + tr^2\left(\boldsymbol{\Omega}_l\right)\right]. \end{aligned} \tag{8}$$

Theorem 1 is important for the AHT test. It says that $\boldsymbol{W}$ is a Wishart mixture and it gives the mean matrix and the total variation of $\boldsymbol{W}$.

2.2. The AHT Test

When $\boldsymbol{W} \sim W_q\left(d, \boldsymbol{I}_q/d\right)$ were valid with $d \geq q$, the random variable T given in (5) would follow $T_{q,d}^2$, a Hotelling T^2-distribution with parameters q and d. Theorem 1 shows that $\boldsymbol{W}$ is in general a Wishart mixture instead of a single Wishart random matrix. To overcome this difficulty, we may approximate the distribution of $\boldsymbol{W}$ by that of a single Wishart random matrix, say, $\boldsymbol{R} \sim W_q\left(d, \boldsymbol{\Omega}\right)$ where the unknown parameters d and $\boldsymbol{\Omega}$ are determined via matching the mean matrices and total variations of $\boldsymbol{W}$ and $\boldsymbol{R}$. That is, we solve the following two equations for d and $\boldsymbol{\Omega}$:

$$E\left(\boldsymbol{W}\right) = E\left(\boldsymbol{R}\right), V\left(\boldsymbol{W}\right) = V\left(\boldsymbol{R}\right). \tag{9}$$

The solution is given in Theorem 2 below together with the range of d.

Theorem 2 *The solution of* (9) *is given by*

$$\boldsymbol{\Omega} = \boldsymbol{I}_q/d, d = \frac{q\left(q+1\right)}{\sum_{l=1}^{k}\left(n_l - 1\right)^{-1}\left[tr\left(\boldsymbol{\Omega}_l^2\right) + tr^2\left(\boldsymbol{\Omega}_l\right)\right]}. \tag{10}$$

Moreover, d satisfies the following inequalities:

$$\frac{q+1}{p+1}\left(n_{\min}-1\right) \le d \le \frac{p\left(q+1\right)}{q\left(p+1\right)}\left(N-k\right). \quad (11)$$

Remark 8 *Theorem* 2 *indicates that* $d \ge q+1$ *provided* $n_{\min} \ge p+2$. *This guarantees that the distribution of the test statistic T defined in* (5) *can be approximated by* $T^2_{q,d}$. *That is why the test proposed here is called the AHT* (*approximate Hotelling* T^2) *test.*

Remark 9 *From* (10), *it is seen that when* $n_{\min}$ *becomes large*, *d generally becomes large*; *and when* $n_{\min} \to \infty$, *we have* $d \to \infty$ *so that* $T^2_{q,d}$ *weakly tends to* χ^2_q, *the limit distribution of T as pointed out in Remark* 5.

Remark 10 *The technique used to approximate a Wishart mixture* ***W*** *by a single Wishart random matrix* $\boldsymbol{R} \sim W_q\left(d,\boldsymbol{\Omega}\right)$ *may be referred to as the Wishart-approximation method. The original version of the Wishart-approximation method is due to* [8] *who determined the unknown parameters d and* $\boldsymbol{\Omega}$ *via matching the first two moments of* ***W*** *and* ***R***. *The article obtained a number different solutions to d*, *with the simplest one being the same as the one presented in Theorem* 2.

Remark 11 *The key idea of the Wishart-approximation method is very similar to that of the well-known* χ^2*-approximation method developed by* [19] *who approximated the distribution of a* χ^2*-mixture* (*see* [20]) *using that of a* χ^2*-random variable multiplied by a constant via matching the first two moments.*

Remark 12 *The first application of the Wishart-approximation method may be due to* [8] *who obtained an approximate test for the multivariate two-sample BF problem. The resulting test is not affine-invariant*, *as pointed out by* [5]. *The authors of* [5] *then modified Nel and van der Merwe*'*s test*, *resulting in the so-called MNV test. Recent applications of the Wishart-approximation method were given by* [17,18] *who studied tests of linear hypotheses in heteroscedastic one-way and two-way ANOVA. The AHT test proposed in this paper is a new application of the Wishart-approximation method.*

In real data application, the parameter d has to be estimated based on the data. A natural estimator of d is obtained via replacing $\boldsymbol{\Omega}_l, l=1,2,\cdots,k$ by their estimators:

$$\hat{\boldsymbol{\Omega}}_l = n_l^{-1}\left(\boldsymbol{C}\hat{\Sigma}\boldsymbol{C}^T\right)^{-1/2}\boldsymbol{C}_l\hat{\Sigma}_l\boldsymbol{C}_l^T\left(\boldsymbol{C}\hat{\Sigma}\boldsymbol{C}^T\right)^{-1/2}, l=1,2,\cdots,k, \quad (12)$$

so that

$$\hat{d} = \frac{q\left(q+1\right)}{\sum_{l=1}^{k}\left(n_l-1\right)^{-1}\left[tr\left(\hat{\boldsymbol{\Omega}}_l^2\right)+tr^2\left(\hat{\boldsymbol{\Omega}}_l\right)\right]}. \quad (13)$$

Notice that $\sum_{l=1}^{k}\hat{\boldsymbol{\Omega}}_l = \boldsymbol{I}_q$ so that the range of *d* given in (11) is also the range of $\hat{d}$.

Remark 13 *Under the assumption* (6), *it is standard to show that as* $n_{\min} \to \infty$, *we have* $\hat{d} \to d$. *In addition*, *we can show that* $E\left(T^2_{q,\hat{d}}\right) = E\left(T\right)\left[1+O\left(n_{\min}^{-2}\right)\right]$ *and* $Var\left(T^2_{q,\hat{d}}\right) = Var\left(T\right)\left[1+O\left(n_{\min}^{-1}\right)\right]$. *That is*, *the means of T and* $T^2_{q,\hat{d}}$ *are matched up to order* $n_{\min}^{-2}$ *while the variances of T and* $T^2_{q,\hat{d}}$ *are matched only up to order* $n_{\min}^{-1}$. *This is not bad since here we only use one tuning parameter* $\hat{d}$ *and the distribution of* $T^2_{q,\hat{d}}$ *is easy to use.*

In summary, the AHT test is based on approximating the distribution of the Wald-type test statistic *T* (4) by $T^2_{q,\hat{d}}$. It can be conducted using the usual *F*-distribution since

$$T^2_{q,\hat{d}} \overset{d}{=} \frac{q\hat{d}}{\hat{d}-q+1}F_{q,\hat{d}-q+1}, \quad (14)$$

where and throughout, the expression $X \overset{d}{=} Y$ means "*X* and *Y* have the same distribution". In other words, the critical value of the AHT test can be specified as $\frac{q\hat{d}}{\hat{d}-q+1}F_{q,\hat{d}-q+1}\left(1-\alpha\right)$ for the nominal significance level α. We reject the null hypothesis in (3) when this critical value is exceeded by the observed test statistic *T*. The AHT test can also be conducted via computing the P-value based on the approximate distribution specified in (14).

2.3. Minimum Sample Size Determination

Let $\left[a\right]$ denote the integer part of *a*. When $X \sim F_{q,v}$, it is easy to show that *X* has up to $\left(\left[v/2\right]-1\right)$ finite moments:

$$E\left(X^r\right) = \frac{v^r q\left(q+2\right)\cdots\left\{q+2\left(r-1\right)\right\}}{q^r\left(v-2\right)\left(v-4\right)\cdots\left(v-2r\right)}, \quad r=1,2,\cdots,\left[v/2\right]-1.$$

In general, *T* has some finite moments. If its approximate Hotelling T^2-distribution $T^2_{q,\hat{d}}$ is good, it should also have the same number of finite moments. To assure that $T^2_{q,\hat{d}}$ has up to *r* finite moments, by (14), the minimum sample size must satisfy

$$n_{\min} > \left(p+2\right)+\frac{2\left(p+1\right)\left(r-1\right)}{q+1}, \quad (15)$$

which is obtained via using the lower bound of *d* (and $\hat{d}$ as well) given in (11). The required minimum sample size may be defined as $\left[a\right]+1$ where *a* is the quantity given in the right-hand side of (15). It is seen that when *p* or *r* is large or when *q* is small, the required minimum sample size is also large. By Remark 2, we have

$q = q_0 p \geq p$. Thus, a sufficient condition to guarantee that the approximate Hotelling T^2-distribution $T^2_{q,\hat{d}}$ (14) has up to r finite moments is that $n_{\min} > p + 2r$.

Remark 14 *When* $n_{\min}$ *is too small, e.g.,* $n_{\min} < p + 2$, *the AHT test may not perform well since in this case, the first moment of* $T^2_{q,\hat{d}}$ *is not finite although the first moment of T is usually finite.*

2.4. Properties of the AHT Test

In practice, the observed response vectors in (1) are often re-centered or rescaled before any inference is conducted. It is desirable that the inference is invariant under the recentering or rescale transformation. They are two special cases of the following affine transformation of the observed response vectors $\boldsymbol{x}_{lj}$:

$$\tilde{\boldsymbol{x}}_{lj} = \boldsymbol{B}x_{lj} + b,\ j = 1,2,\cdots,n_l; l = 1,2,\cdots,k, \quad (16)$$

where $\boldsymbol{B}$ is any nonsingular matrix and $\boldsymbol{b}$ is any constant vector. The proposed AHT test is affine-invariant as stated in the theorem below.

Theorem 3 *The proposed AHT test is affine-invariant in the sense that both T and* $\hat{d}$ *are invariant under the affine-transformation* (16).

Remark 1 mentions that the contrast matrix $\boldsymbol{C}$ used to write (2) into the form of the GLHT problem (3) is not unique and the AHT test is invariant to various choices of the contrast matrix. This result follows from Theorem 4 below immediately if we notice a result from [21] (Ch. 5, Sec. 4), which states that for any two contrast matrices $\tilde{\boldsymbol{C}}$ and $\boldsymbol{C}$ defining the same hypothesis, there is a nonsingular matrix $\boldsymbol{P}$ such that $\tilde{\boldsymbol{C}} = \boldsymbol{PC}$.

Theorem 4 *The AHT test is invariant when the coefficient matrix* $\boldsymbol{C}$ *and the constant vector* $\boldsymbol{c}$ *in* (3) *are replaced with*

$$\tilde{\boldsymbol{C}} = \boldsymbol{PC} \text{ and } \tilde{\boldsymbol{c}} = \boldsymbol{Pc}, \quad (17)$$

respectively where $\boldsymbol{P}$ *is any nonsingular matrix.*

Finally, we have the following result.

Theorem 5 *The AHT test is invariant under different labeling schemes of the mean vectors* $\boldsymbol{\mu}_l, l = 1,2,\cdots,k$.

3. Simulation Studies

In this section, intensive simulations are conducted to compare the AHT test against the test of [1] and the PB test of [2]. All the three tests are affine-invariant. Reference [2] demonstrated that the PB test generally outperforms the test of [1] and the generalized F-test of [14] in terms of size controlling. The generalized F-test are generally very liberal and time consuming. Therefore, we shall not include it for comparison against the AHT test.

Following [2], for simplicity, we set $\Sigma_1 = I_p, \Sigma_2 = \mathrm{diag}(\lambda_1,\cdots,\lambda_p)$ and $\Sigma_l, l = 3,4,\cdots,k$ to be some positive definite matrices, where p, $\lambda_1,\cdots,\lambda_k$ and other tuning parameters are specified later. Let $\boldsymbol{n} = (n_1, n_2, \cdots, n_k)$ denote the vector consisting of the k sample sizes. For given $\boldsymbol{n}$ and $\Sigma_l, l = 1,2,\cdots,k$, we first generated k sample mean vectors $\hat{\boldsymbol{\mu}}_l, l = 1,\cdots,k$ and k sample covariance matrices $\hat{\Sigma}_l, l = 1,\cdots,k$ by

$$\hat{\boldsymbol{\mu}}_l \sim N_p(\mu_l, \Sigma_l/n_l),$$
$$\hat{\Sigma}_l \sim W_p(n_l - 1, \Sigma_l/(n_l - 1)), l = 1,2,\cdots,k$$

where the population mean vectors $\boldsymbol{\mu}_l = \boldsymbol{\mu}_1 + l\delta u, l = 2,\cdots,k$ with $\boldsymbol{\mu}_1$ being the first population mean vector, $\boldsymbol{u}$ a constant unit vector specifying the direction of the population mean differences, and δ a tuning parameter controlling the amount of the population mean differences. Without loss of generality, we specified $\boldsymbol{\mu}_1$ as **0** and $\boldsymbol{u}$ as $\boldsymbol{u}_0/\|\boldsymbol{u}_0\|$ where $\boldsymbol{u}_0 = (1,2,\cdots,p)^T$ for any p and $\|\boldsymbol{u}_0\|$ denotes the usual L^2-norm of $\boldsymbol{u}_0$. We then applied the Johansen, PB, and AHT tests to the generated sample mean vectors and the sample covariance matrices, and recorded their P-values. The empirical sizes and powers of the Johansen, PB, and AHT tests were computed based on 10000 runs and the number of inner loops for the PB test is 1000. In all the simulations conducted, the significance level was specified as 5% for simplicity.

The empirical sizes (associated with $\delta = 0$) and powers (associated with $\delta > 0$) of the Johansen, PB, and AHT tests for the multivariate k-sample BF problem (2), together with the associated tuning parameters, are presented in **Tables 1-3**, in the columns labeled with "Joh", "PB", and "AHT" respectively. As seen from the three tables, three sets of the tuning parameters for population covariance matrices are examined, with the first set specifying the homogeneous cases and seven sets of sample sizes are specified, with the first three sets specifying the balanced sample size cases. To measure the overall performance of a test in terms of maintaining the nominal size α, we define the average relative error as $\mathrm{ARE} = M^{-1}\sum_{j=1}^{M}|\hat{\alpha}_j - \alpha|/\alpha \times 100$ where $\hat{\alpha}_j$ denotes the j-th empirical size for $j = 1,2,\cdots,M$, $\alpha = 0.05$ and M is the number of empirical sizes under consideration. The smaller ARE value indicates the better overall performance of the associated test. Usually, when ARE ≤ 10, the test performs very well; when $10 < \mathrm{ARE} \leq 20$, the test performs reasonably well; and when $\mathrm{ARE} > 20$, the test does not perform well since its empirical sizes are either too liberal or too conservative. Notice that for a good test, the larger the sample sizes, the smaller the ARE values. Notice that for simplicity, in the specification of the covariance and sample size tuning parameters, we often use a_r to denote "a repeats r times", e.g., $(30)_2$ = (30, 30) and $(2_3, 4, 1_2)$ = (2, 2, 2, 4, 1, 1). **Tables 1-3** show the empirical sizes and powers of the Johansen, PB, and AHT tests for a bivariate case with $k = 2$, a

Table 1. Empirical sizes and powers of the Johansen, PB, and AHT tests for bivariate one-way MANOVA.

		$k = 2$			$\Sigma_1 = \boldsymbol{I}_2$			$\Sigma_2 = \mathrm{diag}(\lambda)$					
λ	n	$\delta = 0$			$\delta = 0.7$			$\delta = 1.4$			$\delta = 2.1$		
		Joh	PB	AHT	Joh	PB	AHT	Joh	PB	AHT	Joh	PB	AHT
	$n_{(1)}$	0.049	0.044	0.046	0.504	0.520	0.492	0.984	0.984	0.982	1.000	1.000	1.000
	$n_{(2)}$	0.050	0.050	0.050	0.724	0.716	0.721	0.999	0.999	0.999	1.000	1.000	1.000
	$n_{(3)}$	0.049	0.046	0.049	0.903	0.901	0.902	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda_{(1)}$	$n_{(4)}$	0.047	0.048	0.045	0.606	0.592	0.597	0.994	0.992	0.994	1.000	1.000	1.000
	$n_{(5)}$	0.046	0.044	0.046	0.972	0.965	0.972	1.000	1.000	1.000	1.000	1.000	1.000
	$n_{(6)}$	0.050	0.049	0.049	0.599	0.587	0.592	0.995	0.995	0.994	1.000	1.000	1.000
	$n_{(7)}$	0.047	0.045	0.047	0.967	0.974	0.967	1.000	1.000	1.000	1.000	1.000	1.000
	$n_{(1)}$	0.047	0.054	0.044	0.263	0.255	0.252	0.784	0.769	0.771	0.985	0.984	0.984
	$n_{(2)}$	0.047	0.049	0.046	0.386	0.381	0.382	0.939	0.943	0.938	0.999	0.999	0.999
	$n_{(3)}$	0.050	0.046	0.050	0.586	0.583	0.585	0.995	0.996	0.995	1.000	1.000	1.000
$\lambda_{(2)}$	$n_{(4)}$	0.048	0.044	0.047	0.335	0.321	0.329	0.901	0.895	0.898	0.999	0.999	0.998
	$n_{(5)}$	0.053	0.055	0.053	0.801	0.812	0.801	0.999	1.000	0.999	1.000	1.000	1.000
	$n_{(6)}$	0.052	0.055	0.049	0.285	0.279	0.276	0.827	0.811	0.816	0.990	0.991	0.988
	$n_{(7)}$	0.052	0.054	0.051	0.667	0.643	0.665	0.998	0.999	0.998	1.000	1.000	1.000
	$n_{(1)}$	0.050	0.050	0.047	0.201	0.190	0.192	0.639	0.654	0.625	0.948	0.941	0.943
	$n_{(2)}$	0.047	0.055	0.047	0.290	0.298	0.286	0.850	0.865	0.847	0.996	0.994	0.996
	$n_{(3)}$	0.047	0.050	0.047	0.454	0.455	0.454	0.970	0.972	0.970	0.999	1.000	0.999
$\lambda_{(3)}$	$n_{(4)}$	0.051	0.055	0.050	0.255	0.243	0.250	0.787	0.784	0.781	0.988	0.989	0.988
	$n_{(5)}$	0.044	0.050	0.044	0.668	0.666	0.668	0.999	1.000	0.999	1.000	1.000	1.000
	$n_{(6)}$	0.053	0.053	0.050	0.228	0.223	0.218	0.689	0.685	0.674	0.962	0.969	0.957
	$n_{(7)}$	0.054	0.059	0.053	0.523	0.526	0.522	0.989	0.989	0.989	1.000	1.000	1.000
		ARE	4.48	7.14	5.14								

$\lambda_{(1)} = (1_2)$, $\lambda_{(2)} = (1,5)$, $\lambda_{(3)} = (1,10)$, $n_{(1)} = (7_2)$, $n_{(2)} = (10_2)$, $n_{(3)} = (15_2)$, $n_{(4)} = (7,10)$, $n_{(5)} = (15,30)$, $n_{(6)} = (10,7)$ and $n_{(7)} = (30,15)$.

3-variate case with $k = 3$ and a 5-variate case with $k = 5$, respectively.

From **Table 1**, it is seen that for the two-sample BF problem, the Johansen, PB, and AHT tests performed very similarly with the Johansen test slightly outperforming the other two tests. However, from **Tables 2** and **3**, it is seen that with k increasing to 3 and 5, the Johansen test performed much worse than the PB and AHT tests. The later two tests were generally comparable for various sample sizes and parameter configurations. Since the PB test is much more computationally intensive, it is less attractive in real data analysis. The AHT test is then a nice alternative, especially when k is moderate or large.

4. Application to the Egyptian Skull Data

The Egyptian skull data set was recently analyzed by [2]. It can be downloaded freely at Statlib (http://lib.stat.cmu.edu/DASL/Stories/EgyptianSkullDevelopment.html). There are five samples of 30 skulls from the early pre-dynastic period (circa 4000 BC), the late pre-dynastic period (circa 3300 BC), the 12-th and 13-th dynasties (circa 1850 BC), the Ptolemaic period (circa 200 BC), and the Roman period (circa AD 150). Four measurements are available on each skull, namely, x_1 = maximum breadth, x_2 = borborygmatic height, x_3 = dentoalveolar length, and x_4 = nasal height (all in mm). To compare the AHT test with the test of [1] and the PB test of [2] in various cases, we applied these three tests to

Table 2. Empirical sizes and powers of the Johansen, PB, and AHT tests for trivariate one-way MANOVA.

		$k=3$			$\Sigma_1 = \boldsymbol{I}_3$			$\Sigma_2 = \mathrm{diag}(\lambda)$			$\Sigma_3 = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$		
(λ,ρ)	n	$\delta=0$			$\delta=0.4$			$\delta=0.8$			$\delta=1.2$		
		Joh	PB	AHT	Joh	PB	AHT	Joh	PB	AHT	Joh	PB	AHT
	$n_{(1)}$	0.069	0.040	0.037	0.272	0.188	0.178	0.796	0.721	0.683	0.989	0.975	0.972
	$n_{(2)}$	0.058	0.037	0.045	0.388	0.350	0.345	0.956	0.939	0.946	0.999	0.999	0.999
	$n_{(3)}$	0.049	0.041	0.046	0.603	0.596	0.590	0.998	0.997	0.998	1.000	1.000	1.000
$(\lambda_{(1)},\rho_{(1)})$	$n_{(4)}$	0.069	0.053	0.056	0.419	0.375	0.365	0.950	0.935	0.932	0.999	0.999	0.999
	$n_{(5)}$	0.063	0.064	0.059	0.644	0.637	0.627	0.998	0.998	0.998	1.000	1.000	1.000
	$n_{(6)}$	0.067	0.052	0.052	0.477	0.432	0.422	0.986	0.984	0.976	1.000	0.999	1.000
	$n_{(7)}$	0.062	0.052	0.057	0.775	0.751	0.762	1.000	1.000	1.000	1.000	1.000	1.000
	$n_{(1)}$	0.074	0.050	0.042	0.285	0.206	0.193	0.798	0.710	0.693	0.988	0.971	0.968
	$n_{(2)}$	0.058	0.039	0.045	0.397	0.338	0.352	0.956	0.948	0.944	0.999	1.000	0.999
	$n_{(3)}$	0.051	0.049	0.047	0.605	0.591	0.592	0.998	0.998	0.998	1.000	1.000	1.000
$(\lambda_{(2)},\rho_{(2)})$	$n_{(4)}$	0.072	0.059	0.057	0.414	0.360	0.363	0.943	0.922	0.922	0.999	0.999	0.999
	$n_{(5)}$	0.067	0.056	0.063	0.640	0.612	0.625	0.999	0.998	0.998	1.000	1.000	1.000
	$n_{(6)}$	0.070	0.050	0.054	0.516	0.440	0.454	0.995	0.990	0.991	1.000	1.000	1.000
	$n_{(7)}$	0.068	0.065	0.063	0.865	0.836	0.854	1.000	1.000	1.000	1.000	1.000	1.000
	$n_{(1)}$	0.075	0.044	0.043	0.284	0.205	0.192	0.800	0.715	0.684	0.987	0.967	0.965
	$n_{(2)}$	0.060	0.046	0.049	0.386	0.353	0.344	0.954	0.945	0.939	0.999	0.999	0.999
	$n_{(3)}$	0.050	0.053	0.046	0.605	0.599	0.589	0.997	0.996	0.997	1.000	1.000	1.000
$(\lambda_{(3)},\rho_{(3)})$	$n_{(4)}$	0.075	0.062	0.060	0.406	0.346	0.354	0.945	0.915	0.925	0.999	1.000	0.999
	$n_{(5)}$	0.060	0.052	0.056	0.623	0.589	0.607	0.998	0.998	0.997	1.000	1.000	1.000
	$n_{(6)}$	0.076	0.067	0.059	0.518	0.447	0.457	0.994	0.986	0.989	1.000	1.000	1.000
	$n_{(7)}$	0.061	0.058	0.056	0.866	0.851	0.856	1.000	1.000	1.000	1.000	1.000	1.000
	ARE	29.14	14.00	13.52									

$(\lambda_{(1)},\rho_{(1)})=(1_3,0)$, $(\lambda_{(2)},\rho_{(2)})=(1,5,0.1,0.05)$, $(\lambda_{(3)},\rho_{(3)})=(1,3,0.1,0.09)$, $n_{(1)}=(7_3)$, $n_{(2)}=(10_3)$, $n_{(3)}=(15_3)$, $n_{(4)}=(7,10,20)$, $n_{(5)}=(10,20,40)$, $n_{(6)}=(20,10,7)$ and $n_{(7)}=(40,20,10)$.

check the significance of the mean vector differences of the first k samples, using only the first $\boldsymbol{n}=(n_1,n_2,\cdots,n_k)$ observations for $\boldsymbol{n}=(10_k),(20_k)$ and (30_k) for k = 2, 3, 4 and 5. There are totally 12 cases under consideration. The number of bootstrap replications in the PB test is 10000 and hence the time spent by the PB test is about 10000 times of that spent by the other two tests. The P-values of the three tests for various cases are presented in **Table 4**.

From **Table 4**, it is seen that the P-values of the three tests are close to each other with the P-values of the Johansen test slightly smaller in almost all the cases. Reference [2] showed via intensive simulations that the PB test performed well for various parameter configurations. Therefore, we may use the P-values of the PB test as benchmark to compare the AHT test with the Johansen test. It is seen from **Table 4** that the P-values of the AHT tests are closer to the P-values of the PB test than those of the Johansen test. In this sense, the AHT test performed similar to the PB test and outperformed the Johansen test. This is in agreement with the conclusions drawn from the simulation results presented in the previous section.

Table 3. Empirical sizes and powers of the Johansen, PB, and AHT tests for 5-variate one-way MANOVA.

$k=5$	$\Sigma_1 = \boldsymbol{I}_5$	$\Sigma_2 = \text{diag}(\lambda)$			$\Sigma_3 = \text{diag}(\eta)$			$\Sigma_4 = \text{diag}(u)$			$\Sigma_5 = \text{diag}(v)$		
(λ,η,u,v)	n	$\delta=0$			$\delta=0.15$			$\delta=0.30$			$\delta=0.45$		
		Joh	PB	AHT	Joh	PB	AHT	Joh	PB	AHT	Joh	PB	AHT
	$n_{(1)}$	0.087	0.047	0.046	0.218	0.132	0.136	0.692	0.568	0.572	0.974	0.948	0.946
	$n_{(2)}$	0.057	0.054	0.049	0.299	0.259	0.270	0.930	0.914	0.919	0.999	0.999	0.999
	$n_{(3)}$	0.049	0.042	0.048	0.625	0.623	0.624	1.000	1.000	1.000	1.000	1.000	1.000
$(\lambda_{(1)},\eta_{(1)},u_{(1)},v_{(1)})$	$n_{(4)}$	0.061	0.052	0.055	0.358	0.320	0.339	0.966	0.960	0.962	1.000	1.000	1.000
	$n_{(5)}$	0.055	0.045	0.053	0.521	0.487	0.514	0.998	0.999	0.998	1.000	1.000	1.000
	$n_{(6)}$	0.066	0.057	0.060	0.411	0.374	0.391	0.988	0.980	0.986	1.000	1.000	1.000
	$n_{(7)}$	0.056	0.049	0.054	0.596	0.573	0.589	1.000	1.000	0.999	1.000	1.000	1.000
	$n_{(1)}$	0.096	0.047	0.055	0.128	0.068	0.072	0.230	0.137	0.143	0.458	0.300	0.325
	$n_{(2)}$	0.059	0.046	0.049	0.103	0.085	0.086	0.332	0.281	0.298	0.705	0.680	0.672
	$n_{(3)}$	0.053	0.054	0.052	0.168	0.156	0.166	0.674	0.684	0.672	0.984	0.981	0.984
$(\lambda_{(2)},\eta_{(2)},u_{(2)},v_{(2)})$	$n_{(4)}$	0.059	0.049	0.055	0.117	0.112	0.108	0.382	0.371	0.367	0.797	0.780	0.785
	$n_{(5)}$	0.054	0.049	0.053	0.135	0.128	0.131	0.537	0.519	0.532	0.937	0.935	0.936
	$n_{(6)}$	0.069	0.048	0.060	0.135	0.106	0.119	0.473	0.411	0.441	0.897	0.870	0.884
	$n_{(7)}$	0.053	0.042	0.049	0.159	0.142	0.154	0.657	0.642	0.646	0.983	0.981	0.981
	$n_{(1)}$	0.101	0.051	0.053	0.119	0.051	0.064	0.188	0.111	0.108	0.331	0.194	0.213
	$n_{(2)}$	0.066	0.053	0.056	0.093	0.072	0.077	0.221	0.178	0.191	0.503	0.469	0.464
	$n_{(3)}$	0.052	0.050	0.050	0.121	0.111	0.119	0.454	0.438	0.452	0.893	0.885	0.891
$(\lambda_{(3)},\eta_{(3)},u_{(3)},v_{(3)})$	$n_{(4)}$	0.056	0.044	0.052	0.101	0.086	0.094	0.273	0.256	0.260	0.646	0.608	0.631
	$n_{(5)}$	0.052	0.043	0.051	0.107	0.100	0.103	0.401	0.390	0.395	0.825	0.796	0.822
	$n_{(6)}$	0.064	0.048	0.055	0.107	0.085	0.094	0.283	0.236	0.256	0.624	0.581	0.591
	$n_{(7)}$	0.053	0.049	0.049	0.113	0.101	0.107	0.381	0.359	0.368	0.800	0.778	0.791
	ARE	25.71	6.95	7.05									

$\lambda_{(1)} = (1_5)$, $\eta_{(1)} = (1_5)$, $u_{(1)} = (1_5)$, $v_{(1)} = (1_5)$, $\lambda_{(2)} = (12_2, 1, 24, 1)$, $\eta_{(2)} = (1, 0.1, 2, 24, 21)$, $u_{(2)} = (1, 3, 9, 10)$, $v_{(2)} = (5, 15_2, 45, 50)$, $\lambda_{(3)} = (1, 3, 9_2, 5)$, $\eta_{(3)} = (5, 15, 45_3)$, $u_{(3)} = (1, 3_2, 9, 30)$, $v_{(3)} = (5, 15_2, 45, 100)$, $n_{(1)} = (15_5)$, $n_{(2)} = (25_5)$, $n_{(3)} = (50_5)$, $n_{(4)} = (20, 25, 35, 40, 50)$, $n_{(5)} = (30, 35, 40, 50, 70)$, $n_{(6)} = (50, 40, 35, 25, 20)$ and $n_{(7)} = (70, 50, 40, 35, 30)$.

Table 4. P-values of the Johansen, PB, and AHT tests for the Egyptian skull data example.

Null hypothesis	Joh	PB	AHT	Joh	PB	AHT	Joh	PB	AHT
$H_0 : \mu_1 = \mu_2$	0.6213	0.6412	0.6448	0.7156	0.7194	0.7227	0.8109	0.8182	0.8142
$H_0 : \mu_1 = \mu_2 = \mu_3$	0.5531	0.6107	0.6234	0.1948	0.2063	0.2071	0.0300	0.0326	0.0298
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$	0.0574	0.1050	0.1105	0.0202	0.0225	0.0227	0.0002	0.0002	0.0002
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$	0.0173	0.0502	0.0532	0.0021	0.0021	0.0025	0.0000	0.0000	0.0000

It is also seen that the first null hypothesis in **Table 4** is not significant, with the P-values of the three tests larger than 60% and increasing with increasing the sample sizes; the other three null hypotheses are significant, with the P-values of the three tests decreasing to less than 5% with increasing the sample sizes. These results suggest that the Egyptian skulls had little change in the early and late pre-dynastic periods but experienced a significant change over the later three periods.

5. Technical Proofs

Proof of Theorem 1 Notice first that if $\boldsymbol{Y} \sim W_p(n, \boldsymbol{V})$, then we have $E(\boldsymbol{Y}) = n\boldsymbol{V}$ and $V(\boldsymbol{Y}) = Etr(\boldsymbol{Y} - E\boldsymbol{Y})^2 = n\left[tr(\boldsymbol{V}^2) + tr^2(\boldsymbol{V})\right]$. In addition, it is well known that for $l = 1, 2, \cdots, k$, we have $\hat{\Sigma}_l \sim W_p\left(n_l - 1, \frac{\Sigma_l}{n_l - 1}\right)$. Thus

$$\boldsymbol{W}_l = n_l^{-1}\boldsymbol{H}_l\hat{\Sigma}_l\boldsymbol{H}_l^T \sim W_q\left(n_l - 1, \frac{\boldsymbol{\Omega}_l}{n_l - 1}\right)$$

where

$$\boldsymbol{\Omega}_l = E\boldsymbol{W}_l = n_l^{-1}\boldsymbol{H}_l\Sigma_l\boldsymbol{H}_l^T .$$

Therefore, we have $E(\boldsymbol{W}_l) = \boldsymbol{\Omega}_l$ and

$$V(\boldsymbol{W}_l) = (n_l - 1)^{-1}\left[tr(\boldsymbol{\Omega}_l^2) + tr^2(\boldsymbol{\Omega}_l)\right].$$

Since $\boldsymbol{W}_l$ are independent, we have

$$E(\boldsymbol{W}) = \sum_{l=1}^k \boldsymbol{\Omega}_l = \sum_{l=1}^k n_l^{-1}\boldsymbol{H}_l\Sigma_l\boldsymbol{H}_l^T = \boldsymbol{H}\Sigma\boldsymbol{H} = \boldsymbol{I}_q,$$

and

$$\begin{aligned} V(\boldsymbol{W}) &= \sum_{l=1}^k V(\boldsymbol{W}_l) \\ &= \sum_{l=1}^k (n_l - 1)^{-1}\left[tr(\boldsymbol{\Omega}_l^2) + tr^2(\boldsymbol{\Omega}_l)\right], \end{aligned}$$

as desired. The theorem is proved.

Proof of Theorem 2 By Theorem 1, $E(\boldsymbol{W}) = \boldsymbol{I}_q$ and

$$V(\boldsymbol{W}) = \sum_{l=1}^k (n_l - 1)^{-1}\left[tr(\boldsymbol{\Omega}_l^2) + tr^2(\boldsymbol{\Omega}_l)\right].$$

By the proof of Theorem 1, we have $E(\boldsymbol{R}) = d\boldsymbol{\Omega}$ and

$$V(\boldsymbol{R}) = d\left[tr(\boldsymbol{\Omega}^2) + tr^2(\boldsymbol{\Omega})\right].$$

Equating $E(\boldsymbol{W})$ and $E(\boldsymbol{R})$ leads to $\boldsymbol{\Omega} = \boldsymbol{I}_q/d$. It follows that $V(\boldsymbol{R}) = q(q+1)/d$. Equating $V(\boldsymbol{W})$ and $V(\boldsymbol{R})$ then leads to (10) as desired.

We first find the lower bound of *d*. This is equivalent to finding the upper bound of the denominator of *d*. For $l = 1, 2, \cdots, k$, set $\boldsymbol{B}_l = n_l^{-1/2}\boldsymbol{H}_l\Sigma_l^{1/2}$ which is a $q \times p$ full rank matrix. Then $\boldsymbol{\Omega}_l = \boldsymbol{B}_l\boldsymbol{B}_l^T$. It follows that $\boldsymbol{\Omega}_l$ are nonnegative, so are their eigenvalues. In addition, the matrix $\boldsymbol{\Omega}_l$ and the matrix $\boldsymbol{\Omega}_l = \boldsymbol{B}_l^T\boldsymbol{B}_l : p \times p$ have the same non-zero eigenvalues. Thus, $\boldsymbol{\Omega}_l$ has at most *p* nonzero eigenvalues. Denote the largest *p* eigenvalues of $\boldsymbol{\Omega}_l$ by $\lambda_{l,r}, r = 1, 2, \cdots, p$ which include all the nonzero eigenvalues of $\boldsymbol{\Omega}_l$. By Theorem 1, $\sum_{l=1}^k \boldsymbol{\Omega}_l = \boldsymbol{I}_q$. This leads to $\boldsymbol{I}_q - \boldsymbol{\Omega}_l = \sum_{r=1, r\neq l}^k \boldsymbol{\Omega}_r$, which implies that $\boldsymbol{I}_q - \boldsymbol{\Omega}_r$ is nonnegative. By singular value decomposition of $\boldsymbol{\Omega}_l$, it is easy to show that $\lambda_{l,r} \le 1, r = 1, 2, \cdots, p$. It follows that

$$tr(\boldsymbol{\Omega}_l) = \sum_{r=1}^p \lambda_{l,r} \le p ,$$

and

$$tr(\boldsymbol{\Omega}_l^2) = \sum_{r=1}^p \lambda_{l,r}^2 \le \sum_{r=1}^p \lambda_{l,r} = tr(\boldsymbol{\Omega}_l).$$

Therefore,

$$\begin{aligned} &\sum_{l=1}^k (n_l - 1)^{-1}\left[tr(\boldsymbol{\Omega}_l^2) + tr^2(\boldsymbol{\Omega}_l)\right] \\ &\le (n_{\min} - 1)^{-1}\sum_{l=1}^k\left[tr(\boldsymbol{\Omega}_l^2) + tr^2(\boldsymbol{\Omega}_l)\right] \\ &\le (n_{\min} - 1)^{-1}\sum_{l=1}^k\left[tr(\boldsymbol{\Omega}_l) + ptr(\boldsymbol{\Omega}_l)\right] \\ &\le (n_{\min} - 1)^{-1}(1 + p)\sum_{l=1}^k tr(\boldsymbol{\Omega}_l) \\ &= (n_{\min} - 1)^{-1} q(p+1). \end{aligned}$$

It follows that $d \ge \frac{q+1}{p+1}(n_{\min} - 1)$. The first inequality in (11) is proved.

We now find the upper bound for *d*. This is equivalent to finding the minimum value of the denominator of *d*. Using the eigenvalues of $\boldsymbol{\Omega}_l$ defined above, we have

$$tr(\boldsymbol{\Omega}_l^2) = \sum_{r=1}^p \lambda_{l,r}^2 \ge p\left[\sum_{r=1}^p \lambda_{l,r}/p\right]^2 = p^{-12}tr^2(\boldsymbol{\Omega}_l).$$

It follows that

$$\begin{aligned} &\sum_{l=1}^k (n_l - 1)^{-1}\left[tr(\boldsymbol{\Omega}_l^2) + tr^2(\boldsymbol{\Omega}_l)\right] \\ &\ge (1 + p^{-1})\sum_{l=1}^k (n_l - 1)^{-1} tr^2(\boldsymbol{\Omega}_l). \end{aligned}$$

For convenience, we now set $\delta_l = tr(\boldsymbol{\Omega}_l), l = 1, 2, \cdots, k$. Then by Theorem 1, we have $\sum_{l=1}^k \delta_l = tr\left(\sum_{l=1}^k \boldsymbol{\Omega}_l\right) = q.$ Set $g(\delta_1, \delta_2, \cdots, \delta_k) = \sum_{l=1}^k (n_l - 1)^{-1}\delta_l^2$ where $\delta_1, \cdots, \delta_{k-1}$ are linear independent but $\delta_k = q - \sum_{l=1}^{k-1}\delta_l$. Taking the partial derivatives of *g* with respect to $\delta_l, l = 1, 2, \cdots, k-1$ and setting them to 0 lead to the following normal equation system:

$$\frac{\partial g}{\partial \delta_l} = \frac{2\delta_l}{n_l - 1} - \frac{2\left(q - \sum_{r=1}^{k-1}\delta_r\right)}{n_k - 1} = 0,\ l = 1, 2, \cdots, k-1.$$

Solving the above equation system with respect to

$\delta_l, l=1,2,\cdots,k-1$, together with the fact $\delta_k = q-\sum_{l=1}^{k-1}\delta_l$, leads to

$$\delta_l = q\frac{n_l-1}{N-k}, l=1,2,\cdots,k, \qquad (18)$$

where $N=\sum_{l=1}^{k} n_l$ as defined before. Since for $r,l=1,2,\cdots,k-1$, we have

$$\frac{\partial^2 g}{\partial\delta_l\partial\delta_r} = \begin{cases} 2(n_l-1)^{-1}+2(n_k-1)^{-1}, & \text{if } r=l, \\ 2(n_k-1)^{-1}, & \text{if } r\neq l, \end{cases}$$

the associated Hessian matrix of g is positive definite. Thus, the function $g(\delta_1,\cdots,\delta_k)$ has minimum value $q^2/(N-k)$ when $\delta_l, l=1,2,\cdots,k$ take the values in (18). It follows that the upper bound of d is $\frac{p(q+1)}{q(p+1)}(N-k)$ as desired. The theorem is proved.

Proof of Theorem 3 Since $\boldsymbol{\mu}_l$ and Σ_l denote the mean vector and covariance matrix of x_{lj}, we let $\tilde{\boldsymbol{\mu}}_l$ and $\tilde{\Sigma}_l$ denote the mean vector and covariance matrix of the affine-transformed responses $\tilde{x}_{lj}$ given by (16). Then we have $\tilde{\boldsymbol{\mu}}_l = \boldsymbol{B}\mu_l+\boldsymbol{b}$ and $\tilde{\Sigma}_l = \boldsymbol{B}\Sigma_l\boldsymbol{B}^T$. It follows that $\boldsymbol{\mu}_l = \boldsymbol{B}^{-1}(\tilde{\boldsymbol{\mu}}_l-\boldsymbol{b}), l=1,2,\cdots,k$. As we defined the long mean vector μ and the big covariance matrix Σ in Section 2, we define $\tilde{\mu}$ and $\tilde{\Sigma}$ similarly. Then we have $\boldsymbol{\mu} = \tilde{\boldsymbol{B}}^{-1}(\tilde{\boldsymbol{\mu}}-\tilde{\boldsymbol{b}})$ and $\tilde{\Sigma} = \tilde{\boldsymbol{B}}\Sigma\tilde{\boldsymbol{B}}^T$ where $\tilde{\boldsymbol{B}} = \boldsymbol{I}_k\otimes\boldsymbol{B}$ and $\tilde{\boldsymbol{b}} = 1_k\otimes\boldsymbol{b}$. It follows that the GLHT problem (3) can be equivalently expressed as

$$\tilde{\boldsymbol{H}}_0 : \tilde{\boldsymbol{C}}\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{c}}, \text{ vs } \tilde{\boldsymbol{H}}_1 : \tilde{\boldsymbol{C}}\tilde{\boldsymbol{\mu}} \neq \tilde{\boldsymbol{c}},$$

where $\tilde{\boldsymbol{C}} = \boldsymbol{C}\tilde{\boldsymbol{B}}^{-1}$ and $\tilde{\boldsymbol{c}} = \boldsymbol{C}\tilde{\boldsymbol{B}}^{-1}\boldsymbol{b}+\boldsymbol{c}$.

Since $\hat{\boldsymbol{\mu}}_l$ and $\hat{\Sigma}_l$ denote the unbiased estimators of $\boldsymbol{\mu}_l$ and Σ_l for the original responses $x_{lj}, j=1,2,\cdots,n_l$, we define $\hat{\tilde{\boldsymbol{\mu}}}_l$ and $\hat{\tilde{\Sigma}}_l$ as the unbiased estimators of $\tilde{\boldsymbol{\mu}}_l$ and $\tilde{\Sigma}_l$ for the affine-transformed responses $\tilde{x}_{lj}, j=1,2,\cdots,n_l$. Then by the affine-transformation (16), it is easy to see that $\hat{\tilde{\boldsymbol{\mu}}}_l = \boldsymbol{B}\hat{\boldsymbol{\mu}}_l+\boldsymbol{b}$, and $\hat{\tilde{\Sigma}}_l = \boldsymbol{B}\hat{\Sigma}_l\boldsymbol{B}^T$ Therefore, $\hat{\tilde{\boldsymbol{\mu}}}_l = \tilde{\boldsymbol{B}}\hat{\boldsymbol{\mu}}_l+\tilde{\boldsymbol{b}}$ and $\hat{\tilde{\Sigma}}_l = \tilde{\boldsymbol{B}}\hat{\Sigma}_l\tilde{\boldsymbol{B}}^T$. Using the above, we have $\tilde{\boldsymbol{C}}\hat{\tilde{\boldsymbol{\mu}}}_l-\tilde{\boldsymbol{c}} = \boldsymbol{C}\hat{\boldsymbol{\mu}}_l-\boldsymbol{c}$ and $\tilde{\boldsymbol{C}}\hat{\tilde{\Sigma}}\tilde{\boldsymbol{C}}^T = \boldsymbol{C}\hat{\Sigma}\boldsymbol{C}^T$. The affine-invariance of T follows immediately.

To show that $\hat{d}$ is affine-invariant, by (13), it is sufficient to show that $tr(\Omega_l)$ and $tr(\tilde{\Omega}_l^2)$ are affine-invariant. Let $\boldsymbol{G}_l = n_l^{-1}\boldsymbol{C}_l\hat{\Sigma}_l\boldsymbol{C}_l^T, l=1,2,\cdots,k$ and $\boldsymbol{G}_l = \boldsymbol{C}\hat{\Sigma}\boldsymbol{C}^T$. Then we have $\boldsymbol{G} = \sum_{l=1}^{k}\boldsymbol{G}_l$ and $\hat{\Omega}_l = \boldsymbol{G}^{-1/2}\boldsymbol{G}_l\boldsymbol{G}^{-1/2}$. It follows that $tr(\hat{\Omega}_l) = tr(\boldsymbol{G}_l\boldsymbol{G}^{-1})$ and $tr(\hat{\Omega}_l^2) = tr([\boldsymbol{G}_l\boldsymbol{G}^{-1}]^2)$. Since $\boldsymbol{G}$ is affine-invariant, we only need to show that $\boldsymbol{G}_l, l=1,2,\cdots,k$ are affine-invariant. Since $\tilde{\boldsymbol{C}} = \boldsymbol{C}\tilde{\boldsymbol{B}}^{-1}$ implies $\tilde{\boldsymbol{C}}_l = \boldsymbol{C}_l\boldsymbol{B}^{-1}, l=1,2,\cdots,k$ and $\hat{\tilde{\Sigma}} = \tilde{\boldsymbol{B}}\hat{\Sigma}\tilde{\boldsymbol{B}}^T$ implies $\hat{\tilde{\Sigma}}_l = \tilde{\boldsymbol{B}}\hat{\Sigma}_l\tilde{\boldsymbol{B}}^T, l=1,2,\cdots,k$, the affine-invariance of $\boldsymbol{G}_l$ follows immediately. The theorem is then proved.

Proof of Theorem 4 First of all, under the transformation (17), we have $\tilde{\boldsymbol{C}}\hat{\boldsymbol{\mu}}-\tilde{\boldsymbol{c}} = \boldsymbol{P}(\boldsymbol{C}\hat{\boldsymbol{\mu}}-\boldsymbol{c})$ and $\tilde{\boldsymbol{C}}\hat{\Sigma}\tilde{\boldsymbol{C}}^T = \boldsymbol{P}(\boldsymbol{C}\hat{\Sigma}\boldsymbol{C}^T)\boldsymbol{P}^T$. The invariance of T under (17) follows immediately.

To show that $\hat{d}$ is invariant under the transformation (17), by (13), it is sufficient to show that $tr(\Omega_l)$ and $tr(\tilde{\Omega}_l^2)$ are invariant under (17). The transformation (17) implies that $\tilde{\boldsymbol{C}}_l = \boldsymbol{P}\boldsymbol{C}_l, l=1,2,\cdots,k$. Then we have $\tilde{\boldsymbol{G}}_l = \boldsymbol{P}\boldsymbol{G}_l\boldsymbol{P}^T, l=1,2,\cdots,k$ and $\tilde{\boldsymbol{G}} = \sum_{l=1}^{k}\boldsymbol{G}_l = \boldsymbol{P}\boldsymbol{G}\boldsymbol{P}^T$. It follows that $\tilde{\boldsymbol{G}}_l\tilde{\boldsymbol{G}}^{-1} = \boldsymbol{P}\boldsymbol{G}_l\boldsymbol{G}^{-1}\boldsymbol{P}^{-1}$ so that

$$\begin{aligned} tr(\hat{\Omega}_l) &= tr(\tilde{\boldsymbol{G}}_l\tilde{\boldsymbol{G}}^{-1}) = tr(\boldsymbol{P}\boldsymbol{G}_l\boldsymbol{G}^{-1}\boldsymbol{P}^{-1}) \\ &= tr(\boldsymbol{G}_l\boldsymbol{G}^{-1}) = tr(\Omega_l). \end{aligned}$$

Similarly, we can show that $tr(\tilde{\Omega}_l^2) = tr(\Omega_l^2)$. This proves that $\hat{d}$ is invariant under the transformation (17). The theorem is then proved.

Proof of Theorem 5 Let $l_1, l_2, \cdots, l_k$ be any permutation of $1,2,\cdots,k$. Then it is easy to see that

$$\begin{aligned} \sum_{l=1}^{k}\boldsymbol{C}_l\hat{\boldsymbol{\mu}}_l &= \sum_{u=1}^{k}\boldsymbol{C}_{l_u}\hat{\boldsymbol{\mu}}_{l_u}, \\ \boldsymbol{G}_l &= \sum_{l=1}^{k} n_l^{-1}\boldsymbol{C}_l\hat{\Sigma}_l^2\boldsymbol{C}_l^T = \sum_{u=1}^{k} n_{l_u}^{-1}\boldsymbol{C}_{l_u}\hat{\Sigma}_{l_u}^2\boldsymbol{C}_{l_u}^T, \end{aligned}$$

showing that $\boldsymbol{C}_l\hat{\boldsymbol{\mu}}_l = \sum_{l=1}^{k}\boldsymbol{C}_l\hat{\boldsymbol{\mu}}_l$, $\boldsymbol{G}_l, l=1,\cdots,k$, and $\boldsymbol{G} = \boldsymbol{C}\hat{\Sigma}\boldsymbol{C}^T = \sum_{l=1}^{k} G_l$ are invariant under different labeling schemes of the mean vectors and so is the Wald-type test statistic T.

To show that $\hat{d}$ is invariant under different labeling schemes of the group mean vectors, by (13), it is sufficient to show that the denominator of $\hat{d}$ has such a property. This is actually the case by noticing that the denominator of $\hat{d}$

$$\begin{aligned} &\sum_{l=1}^{k}(n_l-1)^{-1}\left[tr^2(\Omega_l)+tr(\Omega_l^2)\right] \\ &= \sum_{l=1}^{k}(n_l-1)^{-1}\left[tr^2(\boldsymbol{G}_l\boldsymbol{G}^{-1})+tr\left([\boldsymbol{G}_l\boldsymbol{G}^{-1}]^2\right)\right] \\ &= \sum_{u=1}^{k}(n_{l_u}-1)^{-1}\left[tr^2(\boldsymbol{G}_{l_u}\boldsymbol{G}^{-1})+tr\left([\boldsymbol{G}_{l_u}\boldsymbol{G}^{-1}]^2\right)\right]^2. \end{aligned}$$

This completes the proof of the theorem.

6. Acknowledgements

The work was supported by the National University of Singapore Academic Research Grant R-155-000-108-112. The author thanks the Editor for helpful comments and suggestions that help improve the presentation of the paper.

REFERENCES

[1] S. Johansen, "The Welch-James Approximation to the Distribution of the Residual Sum of Squares in a Weighted linear Regression," *Biometrika*, Vol. 67, No. 1,

1980, pp. 85-95. doi:10.1093/biomet/67.1.85

[2] K. Krishnamoorthy and F. Lu, "A Parametric Bootstrap Solution to the MANOVA under Heteroscedasticity," *Journal of Statistical Computation and Simulation*, Vol. 80, No. 8, 2010, pp. 873-887. doi:10.1080/00949650902822564

[3] T. W. Anderson, "An Introduction to Multivariate Statistical Analysis," Wiley, New York, 2003.

[4] K. Krishnamoorthy and Y. Xia, "On Selecting Tests for Equality of Two Normal Mean Vectors," *Multivariate Behavioral Research*, Vol. 41, No. 4, 2006, pp. 533-548. doi:10.1207/s15327906mbr4104_5

[5] K. Krishnamoorthy and J. Yu, "Modified Nel and van der Merwe Test for the Multivariate Behrens-Fisher Problem," *Statistics and Probability Letters*, Vol. 66, No. 2, 2004, pp. 161-169. doi:10.1016/j.spl.2003.10.012

[6] G. S. James, "Tests of Linear Hypotheses in Univariate and Multivariate Analysis When the Ratios of the Population Variances Are Unknown," *Biometrika*, Vol. 41, No. 1-2, 1954, pp. 19-43.

[7] Y. Yao, "An Approximate Degrees of Freedom Solution to the Multivariate Behrens-Fisher Problem," *Biometrika*, Vol. 52, 1965, pp. 139-147.

[8] D. G. Nel and C. A. van der Merwe, "A Solution to the Multivariate Behrens-Fisher Problem," *Communication Statistics*: *Theory and Methods*, Vol. 15, No. 12, 1986, pp. 3719-3735. doi:10.1080/03610928608829342

[9] S. Kim, "A Practical Solution to the Multivariate Behrens-Fisher Problem," *Biometrika*, Vol. 79, No. 1, 1992, pp. 171-176. doi:10.1093/biomet/79.1.171

[10] H. Yanagihara and K. H. Yuan, "Three Approximate Solutions to the Multivariate Behrens-Fisher Problem," *Communication Statistics*: *Simulation and Computation*, Vol. 34, No. 4, 2005, pp. 975-988. doi:10.1080/03610910500308396

[11] A. Belloni and G. Didier, "On the Behrens-Fisher Problem: A Globally Convergent Algorithm and a Finite-Sample Study of the Wald, LR and LM Tests," *Annals of Statistics*, Vol. 36, No. 5, 2008, pp. 2377-2408. doi:10.1214/07-AOS528

[12] 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," *Communication Statistics*: *Theory and Methods*, Vol. 26, 1997, pp. 1251-1273.

[13] B. L. Welch, "On the Comparison of Several Mean Values: An Alternative Approach," *Biometrika*, Vol. 38, 1951, pp. 330-336.

[14] J. Gamage, T. Mathew and S. Weerahandi, "Generalized p-Values and Generalized Confidence Regions for the Multivariate Behrens-Fisher Problem and MANOVA," *Journal of Multivariate Analysis*, Vol. 88, No. 1, 2004, pp. 177-189. doi:10.1016/S0047-259X(03)00065-4

[15] K. L. Tang and J. Algina, "Performing of Four Multivariate Tests under Variance-Covariance Heteroscedasticity," *Multivariate Behavioral Research*, Vol. 28, No. 4, 1993, pp. 391-405. doi:10.1207/s15327906mbr2804_1

[16] K. Krishnamoorthy, F. Lu and T. Mathew, "A Parametric Bootstrap Approach for ANOVA with Unequal Variances: Fixed and Random Models," *Computational Statistics and Data Analysis*, Vol. 51, No. 12, 2007, pp. 5731-5742. doi:10.1016/j.csda.2006.09.039

[17] J. T. Zhang, "Tests of Linear Hypotheses in the ANOVA under Heteroscedasticity," Manuscript, 2012.

[18] J. T. Zhang, "An Approximate Degrees of Freedom Test for Heteroscedastic Two-Way ANOVA," *Journal of Statistical Planning and Inference*, Vol. 142, 2012, pp. 336-346.

[19] F. E. Satterthwaite, "An Approximate Distribution of Estimate of Variance Components," *Biometrics Bulletin*, Vol. 2, No. 6, 1946, pp. 110-114. doi:10.2307/3002019

[20] J. T. Zhang, "Approximate and Asymptotic Distribution of χ^2-Type Mixtures with Application," *Journal of American Statistical Association*, Vol. 100, No. 469, 2005, pp. 273-285. doi:10.1198/016214504000000575

[21] A. M. Kshirsagar, "Multivariate Analysis," Marcel Decker, New York, 1972.

