Stand-in Procedure to Multivariate Behrens-Fisher Problem

G.M. Oyeyemi* and P.O. Adebayo

Department of Statistics, University of Ilorin, Ilorin, Nigeria.

E-mail: gmoyeyemi@gmail.com*
bayooni3@gmail.com

Telephone:+2348065804084
+2347019964334

ABSTRACT

This work considers the problem of comparing two multivariate normal mean vectors under the heteroscedasticity of dispersion matrices. We develop a new procedure using approximate degree of freedom method by Satterthwaite [23] and broaden it to Multivariate Behrens-Fisher. The New procedure is compared with existing ones via R package simulation and Data used by James [8] and Yao [31]. We ascertain that, new procedure is better in terms of power of the test and type I error rate than all existing procedure mull over when the sample sizes are not equal, but the propose procedure perform the same with the selected procedure when sample sizes are equal.

(Keywords: multivariate Behrens-Fisher problem, Type 1 error rate, power of the test, heteroscedasticity)

INTRODUCTION

The statistic use to test the hypothesis that two mean vectors are equal:

$(H_o: \mu_1 = \mu_2)$ is Hotelling's T^2.

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{x}_1 - \bar{x}_2)' S_{pl}^{-1}(\bar{x}_1 - \bar{x}_2) \quad \textbf{(1)}$$

Where,

$$S_{pl} = \frac{1}{n_1 + n_2 - 2}[(n_1 - 1)S_1 + (n_2 - 1)S_2] \quad \textbf{(2)}$$

and $\bar{x}_i$ and S_i are the sample mean vector and sample variance –covariance matrix of the *i*th sample.

Hotelling's T^2, has three basic assumptions that are fundamental to the statistical theory: independent, multivariate normality and equality of variance-covariance matrices. A statistical test procedure is said to be robust or insensitive if departures from these assumptions do not greatly affect the significance level or power of the test. To use Hotelling's T^2 one must assume that the two samples are independent and that their variance-covariance matrices are equal $(\Sigma_1 = \Sigma_2 = \Sigma)$. When variance –covariance matrices are not homogeneous, the test statistic will not be distributed as a T^2. This predicament is known as the multivariate Behrens-Fisher problem.

The Behrens-Fisher Problem is the problem of interval estimation and hypothesis testing concerning the differences between the means of two normally distributed populations when the variances of the two populations are not assumed to be equal. While Multivariate Behrens-Fisher problem deal with testing the equality of two normal mean vector under heteroscedasticity of dispersion matrices.

The problem of comparing independent sample means arising from two populations with unequal variances has been studied for many years and there is a sizable literature. Historically, this problem has come to be known as the Behrens-fisher problem. The comparison of the means of two populations on the basis of two independent samples is one of the oldest problems in statistics. Indeed, it has been a testing ground for many methods of inference as well as for a variety of analytic approaches to practical problems.

We have quite number of scholars that worked on this multivariate Behrens-fisher problem; James [8], Jonhanson [10], Yao [31], Nel and van der Merwe [18], Gamage et al [6], Krishnamoorthy and Yu [13], Yanagihara and Yuan [30], Kim [12], Weerahandi [25], Kawasaki and Seo [11], and Park and Sinha [20]. All the methods used in approximating Behrens-Fisher problems are basically classified into four categories: approximate degree of freedom test, series expansion–based test, simulation–based test, and transformation based–test.

This paper aims at developing an alternative procedure to multivariate Behrens–Fisher problem by extending Satterthwaite's procedure [23], from univariate to multivariate Behrens–Fisher problem and compare the propose procedure with the existing ones in terms of power and type I error rate using real life data and R package to simulate under different conditions which are: (i) when the sample size equal and not ($n_1 \neq n_2$ and $n_1 = n_2$), (ii) when dealing with various sample sizes (small, medium and large).

METHODOLOGY

Consider two p –variate normal populations $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ where μ_1 and μ_2 are unknown $p \times 1$ vectors and Σ_1 and Σ_2 are unknown $p \times p$ positive definite matrices. Let$X_{\alpha 1} \sim N(\mu_1, \Sigma_1), \alpha = 1, 2, \ . \ . \ . \ , n_1,$ and $X_{\alpha 2} \sim N(\mu_2, \Sigma_2), \alpha = 1, 2, \ . \ . \ . \ , n_2,$ denote random samples from these two populations, respectively. We are interested in the testing problem:

$$H_o : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2$$

For $i = 1, 2,$ let

$$\bar{X}_i = \frac{1}{n_i} \sum_{\alpha=1}^{n_i} X_{\alpha i}$$

$$A_i = \sum_{n=1}^{n_i} (X_{\alpha i} - \bar{X}_i)(X_{\alpha i} - \bar{X}_i)'$$

$$S_i = A_i/(n_i - 1), \ i = 1, 2.$$

Then $\bar{X}_1$, $\bar{X}_2$, A_1 and A_2, which are sufficient for the mean vectors and dispersion matrices, are independent random variables having the distributions:

$$\bar{X}_i \sim N\left(\mu_i, \frac{\Sigma_i}{n_i}\right), and \ A_i \sim W_p(n_i - 1, \Sigma_i) \ , i = 1, 2 \quad \textbf{(7)}$$

Where $W_p(r, \Sigma)$denotes the p - dimensional wishart distribution with $df = r$ and scale matrix Σ.

$$\tilde{S}_i = S_i/n_i, \ \ i = 1, 2, \tilde{S} = \tilde{S}_1 + \tilde{S}_2, \ T^2 = (\bar{X}_1 - \bar{X}_2)' \tilde{S}^{-1} (\bar{X}_1 - \bar{X}_2) \quad \textbf{(8)}$$

The following are the review of the existing procedures or solutions to Multivariate Behrens-Fisher problem:

1. James [8], Oyeyemi [19] expressed the critical value for T_A^2 as a series of terms in descending order of magnitude. The 1st order approximation of the critical value is given by $r(A + rB)$ where r is the $1 - \alpha$ percentile point of the central chi – square distribution with p degrees of freedom,

$$A = 1 + \frac{1}{2p} \sum_{i=1}^{2} \frac{1}{n_i - 1} \left[tr(\tilde{S}^{-1} \tilde{S}_i)\right]^2$$

$$B = \frac{1}{2p(p+2)} \sum_{i=1}^{2} \frac{1}{n_i - 1} \left\{tr \left[2(\tilde{S}^{-1} \tilde{S}_i)^2\right] + \left[tr(\tilde{S}^{-1} \tilde{S}_i)\right]^2\right\}$$

2. Yao's [31], Ajit .C.T and Brent .R.L [1] invariant test. This is a multivariate extension of the Welch 'approximate degree of freedom' solution provided by Turkey and his test statistic based on a transformation of T_A^2. And is based on $T^2 \sim (vp/(v-p+1))F_{p,v-p+1}$ with the *d.f.v* given by:

$$\frac{1}{v} = \frac{1}{(T_U^2)^2}\sum_{i=1}^{2}\frac{1}{n_i-1}\left[(\bar{x}_1-\bar{x}_2)'\tilde{S}^{-1}\tilde{S}_i\tilde{S}^{-1}(\bar{x}_1-\bar{x}_2)\right]^2$$

$$T_{Yao} = \frac{(v-p+1)T_U^2}{pv}$$

3. Johansen's [10], invariant test, Park .J and Sinha. B.[20],. We use $T^2 \sim qF_{pv}$ where:

$$q = p + 2D - 6D/[p(p-1) + 2],\ v = p(p+2)/3D$$

$$D = \frac{1}{2}\sum_{i=1}^{2}\left\{tr\left[(I - (\tilde{S}_1^{-1} + \tilde{S}_2^{-1})^{-1}\tilde{S}_i^{-1})^2\right] + tr\left[\left(I - (\tilde{S}_1^{-1} + \tilde{S}_2^{-1})^{-1}\tilde{S}_i^{-1}\right)\right]^2\right\}\Big/ n_i$$

And his proposed test statistic

$$T_{Johan} = \frac{T_U^2}{q}$$

4. Krishnamoorthy and Yu 's [13], Lin and Wang [15] modified Nel/ Van der Merwe invariant solution. We use the idea as before, namely:

$T^2 \sim (v_{kY}p/(v_{kY}-p+1))F_{p,v_{kY}-p+1}$ with the *d.f.v* defined by:

$$v_{KY} = (p+p^2)/C(\tilde{S}_1,\tilde{S}_2)$$

$$(\tilde{S}_1,\tilde{S}_2) = \frac{1}{n_1}\left\{tr\left[(\tilde{S}_1\tilde{S}^{-1})^2\right] + \left[tr(\tilde{S}_1\tilde{S}^{-1})\right]^2\right\} + \frac{1}{n_2}\left\{tr\left[(\tilde{S}_2\tilde{S}^{-1})^2\right] + \left[tr(\tilde{S}_2\tilde{S}^{-1})\right]^2\right\}$$

$$T_{krish} = \frac{(v_{kY}-p+1)T_U^2}{pv_{KY}}$$

4. Yanagihara and Yuan [30], Kawasaki and Takashi [26]:

We use $T_{Yana} = \frac{n-2-\hat{\theta}_1}{(n-2)p}T \sim F_{p,\hat{v}}$ where:

$$\hat{v} = \frac{(n-2-\hat{\theta}_1)^2}{(n-2)\hat{\theta}_2 - \hat{\theta}_1}$$

$\hat{\theta}_1 = \frac{p\hat{\psi}_1 + (p-2)\hat{\psi}_2}{p(p+2)}$ and $\hat{\theta}_2 = \frac{\hat{\psi}_1 + 2\hat{\psi}_2}{p(p+2)}$

$$\hat{\psi}_1 = \frac{n_2^2(n-2)}{n^2(n_1-1)}\{tr(S_1\bar{S}^{-1})\}^2 + \frac{n_1^2(n-2)}{n^2(n_2-1)}\{tr(S_2\bar{S}^{-1})\}^2$$

$$\hat{\psi}_2 = \frac{n_2^2(n-2)}{n^2(n_1-1)}tr(S_1\bar{S}^{-1}S_1\bar{S}^{-1}) + \frac{n_1^2(n-2)}{n^2(n_2-1)}tr(S_2\bar{S}^{-1}S_2\bar{S}^{-1})$$

$$\bar{S} = \frac{n_2}{n}S_1 + \frac{n_1}{n}S_2$$

$\hat{v}$ is identical to the degrees of freedom in welch's [29] approximation. Therefore, T_{Yana} is a direct extension of Welch's procedure.

PROPOSED METHOD

The entire aforementioned scholars worked on the degree of freedom by using various methods to get approximate degree of freedom to the test statistic, which we are proposing to do the same by extending Satterthwaite's procedure (two moment solution to the Behrens-Fisher problem) in univariate to multivariate Behrens-Fisher problem. In 1946 Satterthwaite proposed a method to estimate the distribution of a linear combination of independent chi – square random variables with a chi – square distribution. . Let $L = \sum a_i U_i$ where a_i are known constants, while U_i are independent random variables such that:

$U_i = \frac{(n_i-1)s_i^2}{\sigma_i^2} \sim \chi^2_{(n_i-1)}$ and $a_i = \frac{c_i^2\sigma_i^2}{n_i(n_i-1)}$, for $i = 1, 2$.

Since linear combination of random variable does not, in general, possess a chi– square distribution. Satterthwaite [23] suggested the use of a chi–square distribution, Say $\chi^2_{(f)}$ as an approximation to the distribution of $\frac{f.L}{E[L]}$. This notion is compactly written as:

$$\frac{f.L}{E[L]} \dot{\sim} \chi^2_{(f)} \qquad \textbf{(9)}$$

Where " $\dot{\sim}$ " is taken to mean " is approximately distributed as." From an intuitive standpoint, the distribution of $\frac{f.L}{E[L]}$ should have characteristics similar to some member of the chi–square family of densities. But recall that if a chi–square distribution has degrees of freedom $(n_i - 1)$, then its mean is $(n_i - 1)$ and variance is $2(n_i - 1)$.

Symbolically, this requires that, the first moment of the statistic is:

$$E\left[\frac{f.L}{E[L]}\right] = f \qquad \textbf{(10)}$$

The means that we should use a chi–square with f degrees of freedom. Let consider the second moment. The variance of the statistic is:

$$Var\left[\frac{f.L}{E[L]}\right] = 2f \qquad \textbf{(11)}$$

Then we find first two central moment of L. Consider the following linear combination of these random variables:

$$L = a_1U_1 + a_2U_2 \qquad \textbf{(12)}$$

Where,

$$U_i = \frac{(n_i-1)s_i^2}{\sigma_i^2}$$

and

$$a_i = \frac{c_i^2\sigma_i^2}{n_i(n_i-1)} \qquad \textbf{(13)}$$

$$E[L] = E[a_1U_1 + a_2U_2]$$ (14)

Put Equation (13) into Equation (14):

$$E[L] = E\left[\frac{c_1^2\sigma_1^2}{n_1(n_1-1)} \cdot \frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{c_2^2\sigma_2^2}{n_2(n_2-1)} \cdot \frac{(n_2-1)S_2^2}{\sigma_2^2}\right]$$

Note that $E(S_i^2) = \sigma_i^2$ and

$E\left[\frac{(n_i-1)S_i^2}{\sigma_i^2}\right] = n_i - 1.$ then we have

$$E[L] = \frac{c_1^2\sigma_1^2}{n_1} + \frac{c_2^2\sigma_2^2}{n_2} \quad \textbf{(15)}$$

$$Var[L] = Var[a_1U_1 + a_2U_2] \quad \textbf{(16)}$$

Put Equation (13) into Equation (16):

$$Var[L] = Var\left[\frac{c_1^2\sigma_1^2}{n_1(n_1-1)} \cdot \frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{c_2^2\sigma_2^2}{n_2(n_2-1)} \cdot \frac{(n_2-1)S_2^2}{\sigma_2^2}\right]$$

Recall that $Var(S_i^2) = \frac{2\sigma_i^2}{n_i-1}$

and

$$Var\left[\frac{(n_i-1)S_i^2}{\sigma_i^2}\right] = 2(n_i - 1)$$

$$Var[L] = 2\left[\frac{c_1^4\sigma_1^4}{n_1^2(n_1-1)} + \frac{c_2^4\sigma_2^4}{n_2^2(n_2-1)}\right] \quad \textbf{(17)}$$

Substitute Equation (15) and (17) into Equation (11):

$$2f = \frac{f^2}{E[L]^2} \cdot Var[L]$$

When $c_1 = c_2 = 1$, and σ_1 and σ_2 are replaced by their respective best estimators, S_1 and S_2.

$$f = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}$$

We shall consider the test statistic $y'S^{-1}y$ and use Univariate Satterthwaite approximation of degrees of freedom method to suggest multivariate generalization based on the T^2 – distribution. We have:

$S = S_1 + S_2$ and $y = \bar{X}_1 - \bar{X}_2$

$$y \sim N(0, \Sigma)$$

If S were a Wishart matrix $(n_i - 1)S \sim wishart(n_i - 1, \Sigma)$

then for an arbitrary constant vector b we should have:

$$b'y \sim N(0, b'\Sigma b)$$

$$(n_i - 1)(b'Sb) \sim (b'\Sigma b)\chi^2_{(n_i-1)}$$

That is $m_i = \frac{(n_i-1)b'S_ib}{b'\Sigma_i b} \sim \chi^2_{(n_i-1)}$

and

$$r_i = \frac{d'_i d_i b'\Sigma_i b}{n_i(n_i-1)} \quad \textbf{(18)}$$

Equation (18) is the multivariate version of Equation (13). A linear combination of *p* (random) variables

$$h = r_1m_1 + r_2m_2$$

$$E[h] = E[r_1m_1 + r_2m_2] \quad \textbf{(19)}$$

Substitute Equation (18) into Equation (19):

$$E[h] = E\left[\frac{d'_1d_1b'\Sigma_1b}{n_1(n_1-1)} \cdot \frac{(n_1-1)b'S_1b}{b'\Sigma_1b} + \frac{d'_2d_2b'\Sigma_2b}{n_2(n_2-1)} \cdot \frac{(n_2-1)b'S_2b}{b'\Sigma_2b}\right]$$

Note that $E\left(\frac{(n_i-1)b'S_ib}{b'\Sigma_ib}\right) = (n_i - 1)$

$$E[h] = \frac{d'_1 d_1 b' \Sigma_1 b}{n_1(n_1-1)}.(n_1-1) + \frac{d'_2 d_2 b' \Sigma_2 b}{n_2(n_2-1)}.(n_2-1)$$

$$E[h] = \frac{d'_1 d_1 b' \Sigma_1 b}{n_1} + \frac{d'_2 d_2 b' \Sigma_2 b}{n_2} \quad \textbf{(21)}$$

$$Var[h] = var[r_1 m_1 + r_2 m_2] \quad \textbf{(22)}$$

Substitute Equation (18) into Equation (22)

$$Var[h] = Var\left[\frac{d'_1 d_1 b' \Sigma_1 b}{n_1(n_1-1)}.\frac{(n_1-1)b' S_1 b}{b' \Sigma_1 b} + \frac{d'_2 d_2 b' \Sigma_2 b}{n_2(n_2-1)}.\frac{(n_2-1)b' S_2 b}{b' \Sigma_2 b}\right]$$

Recall that $Var\left(\frac{(n_i-1)b' S_i b}{b' \Sigma_i b}\right) = 2(n_i - 1)$

$$Var[h] = \frac{(d'_1 d_1)^2 (b' \Sigma_1 b)^2}{n_1^2 (n_1-1)^2}.2(n_1-1) + \frac{(d'_2 d_2)^2 (b' \Sigma_2 b)^2}{n_2^2 (n_2-1)^2}.2(n_2-1)$$

$$Var[h] = 2\left[\frac{(d'_1 d_1)^2 (b' \Sigma_1 b)^2}{n_1^2 (n_1-1)} + \frac{(d'_2 d_2)^2 (b' \Sigma_2 b)^2}{n_2^2 (n_2-1)}\right] \quad \textbf{(23)}$$

Substitute equation (21) and (23) into equation (11)

$$2f = \frac{f^2.2\left[\frac{(d'_1 d_1)^2 (b' \Sigma_1 b)^2}{n_1^2 (n_1-1)} + \frac{(d'_2 d_2)^2 (b' \Sigma_2 b)^2}{n_2^2 (n_2-1)}\right]}{\left[\frac{d'_1 d_1 b' \Sigma_1 b}{n_1} + \frac{d'_2 d_2 b' \Sigma_2 b}{n_2}\right]^2}$$

$$f = \frac{\left[\frac{d'_1 d_1 b' \Sigma_1 b}{n_1} + \frac{d'_2 d_2 b' \Sigma_2 b}{n_2}\right]^2}{\left[\frac{(d'_1 d_1)^2 (b' \Sigma_1 b)^2}{n_1^2 (n_1-1)} + \frac{(d'_2 d_2)^2 (b' \Sigma_2 b)^2}{n_2^2 (n_2-1)}\right]} \quad (24)$$

Yao [31] showed that $w_b = \frac{(b'y)^2}{(b'Sb)} \sim t^2_{(n-1)}$

And also it was shown (Bush & Olkin,[3]) that

$sup(w_b) = w_{b^*} = \frac{(b^{*\prime} y)^2}{(b^{*\prime} S b^*)} = y' S^{-1} y$,

Where the maximizing $b^* = S^{-1} y$ and $d'_i d_i$ is Orthogonal matrix Then equation (24) become

$$f = \frac{\left(\frac{yS^{-1}S_1 S^{-1} y}{n_1} + \frac{yS^{-1}S_2 S^{-1} y}{n_2}\right)^2}{\frac{(yS^{-1}S_1 S^{-1} y)^2}{n_1^2 (n_1-1)} + \frac{(yS^{-1}S_2 S^{-1} y)^2}{n_2^2 (n_2-1)}}$$

$$f = \frac{\left(\Sigma \frac{1}{n_i}(yS^{-1}S_i S^{-1} y)\right)^2}{\Sigma \frac{1}{n_i^2 (n_i-1)} (yS^{-1}S_i S^{-1} y)^2} \quad \textbf{(25)}$$

When $y = \bar{X}_1 - \bar{X}_2$ Equation (25) becomes:

$$f = \frac{\left(\Sigma \frac{1}{n_i}((\bar{X}_1 - \bar{X}_2)S^{-1}S_i S^{-1}(\bar{X}_1 - \bar{X}_2))\right)^2}{\Sigma \frac{1}{n_i^2 (n_i-1)} ((\bar{X}_1 - \bar{X}_2)S^{-1}S_i S^{-1}(\bar{X}_1 - \bar{X}_2))^2}$$

Therefore $T \sim \frac{fp}{f-p+1} F_{p,f-p+1}$ approximately.

<u>Testing Equality of Variance – Covariance</u>

For a test of equality covariance matrices, we used the statistic:

$$M = (N-g)log|S| - \sum_{i=1}^{g} v_i log|S_i|$$

S is the the pooled-within estimate of the variance –covariance matrix and g denotes the number of groups:

$$S = \frac{1}{N-g}\sum_{i=1}^{g} v_i S_i$$

Where

$$N = \sum_{i=1}^{g} N_i \quad , v_i = N_i - 1$$

is fromed. Anderson [32] and Kullback [33,34] use this statistic to test equality of variance-covariance and however, multiplying M by 1 – C, where:

$$C = \frac{2p^2 + 3p - 1}{6(p+1)(g-1)}\left(\sum_{i=1}^{g}\frac{1}{v_i} - \frac{1}{N-g}\right)$$

$$\chi^2_B = (1-C)M$$

More rapidly approximates a chi-square distribution with degrees of freedom

$$v = \frac{p(p-1)(g-1)}{2}$$

H_o is rejected at the significance level α if,

$$\chi_B^2 > \chi_{\alpha(v)}^2$$

SIMULATION

A simulation using R statistics was conducted in order to estimate the power of the test and Type I error rate for the previously discussed procedure of multivariate Behrens-Fisher problem (James, Yao, Johanson, Krishnamoorthy and Yanagihara).

The simulations are carried out when the null hypothesis is true, and the distribution is Multivariate normal. Sample size small (15 and 15, 10 and 20), medium (100 and 100, 100 and 120) and large (1000 and 1000, 600 and 1000) and dimensionality (p) used were p = 2 and 3. For each of the above combinations, an $n_i \times p$ data matrix X_i $(i = 1 \; and \; 2)$ were replicated 1,000.

The comparison criteria; type I error rate and power of the test were therefore obtained and the results were presented in the Table 1.

Table 1: Test Results.

			Power					
			Jam	Johan	Yao	Krish	Yarah	Propose
P = 2	Equal sample	15,15	0.2201	0.5637	0.1464	0.1475	0.1642	0.1464
		100,100	0.3943	0.6208	0.2355	0.2356	0.2397	0.2355
		1000,1000	0.9726	0.9977	0.8480	0.8480	0.8480	0.8480
	Unequal sample	10,20	0.2174	0.5663	0.1446	0.1458	0.1629	0.1401*
		100,200	0.4885	0.7116	0.2967	0.2967	0.3002	0.2955*
		600,1000	0.9488	0.9932	0.7854	0.7854	0.7858	0.7853*
			Type I error rate					
			Jam	Johan	Yao	Krish	Yarah	Propose
P = 2	Equal sample	15,15	0.081	0.687	0.077	0.079	0.099	0.077
		100,100	0.233	0.794	0.233	0.233	0.247	0.233
		1000,1000	0.996	1.000	0.996	0.996	0.996	0.996
	Unequal sample	10,20	0.079	0.690	0.077	0.077	0.116	0.064*
		100,200	0.380	0.866	0.380	0.380	0.389	0.379*
		600,1000	0.983	1.000	0.983	0.983	0.983	0.983
			Power					
			Jam	Johan	Yao	Krish	Yarah	Propose
P = 3	Equal sample	15,15	0.2434	0.7883	0.1188	0.1203	0.1390	0.1188
		100,100	0.3929	0.7258	0.1662	0.1663	0.1701	0.1662
		1000,1000	0.9460	0.9975	0.6037	0.6037	0.6038	0.6037
	Unequal sample	15,20	0.2561	0.7712	0.1225	0.1236	0.1412	0.1214*
		100,200	0.4476	0.7763	0.1889	0.1889	0.1932	0.1874*
		600,1000	0.8856	0.9880	0.5008	0.5008	0.5010	0.5004*
			Type I error rate					
			Jam	Johan	Yao	Krish	Yarah	Propose
P = 3	Equal sample	15,15	0.062	0.944	0.056	0.058	0.107	0.056
		100,100	0.180	0.914	0.179	0.180	0.190	0.179
		1000,1000	0.974	1.000	0.974	0.974	0.974	0.974
	Unequal sample	15,20	0.084	0.928	0.072	0.079	0.122	0.069*
		100,200	0.247	0.943	0.245	0.246	0.258	0.244*
		600,1000	0.913	0.998	0.913	0.913	0.913	0.913

Table 1 showed that the power of the test statistics and type I error rate for Yao and the new proposed method are equal when the sample size are equal and random variable p = 2 and 3, for the three scenario; small, median, and large. But when sample size are not equal, the new procedure performed better in terms of power of the test and type I error rate in the three scenario, for random variable p = 2 and 3. The asterisk show where the new procedure method performed better than all other ones, but when sample size are very high for both condition (Equal and not equal; 1000 and 1000, 600 and 1000). The new procedure, Yao, Krishnamoorthy and Yanagihara behave the same in both power of the test and type I error rate.

<u>illustrated Example</u>

We will use the numerical example given by James [8] and Yao [31] to compare the six procedure namely: James, Yoa, Johanson, Krishnamoorthy, Yanagaharia, and the propose procedure. The sample means and their covariances are:

$$\bar{x}_1 = \begin{pmatrix} 9.82 \\ 15.06 \end{pmatrix} , \qquad \bar{x}_2 = \begin{pmatrix} 13.05 \\ 22.57 \end{pmatrix}$$

$$S_1 = \begin{pmatrix} 7.500 & -1.019 \\ -1.019 & 1.112 \end{pmatrix}, \; S_2 = \begin{pmatrix} 7.436 & 2.918 \\ 2.918 & 4.891 \end{pmatrix}$$

$$n_1 = 15, \qquad n_2 = 10$$

The difference between the means is:

$$x = \bar{x}_1 - \bar{x}_2 = \begin{pmatrix} -3.23 \\ -7.51 \end{pmatrix}$$

and the sample estimate for the covariance matrix is:

$$S = S_1 + S_2 = \begin{pmatrix} 14.936 & 1.899 \\ 1.899 & 6.003 \end{pmatrix},$$

While,

$$S^{-1} = \begin{pmatrix} 0.06976 & -0.02207 \\ -0.02207 & 0.17357 \end{pmatrix}$$

$$T^2 = x's^{-1}x$$

$$T^2 = 9.4462$$

Table 2: Solution Performance.

α = 0.05						
	Jam	**Johan**	**Yao**	**Krish**	**Yarah**	**Propose**
Critical value	7.2309	7.3773	8.1968	7.4605	6.4100	8.8773
Pvalue	0.0219	0.0271	0.0373	0.0302	0.0147	0.0459
Power	0.2870	0.5983	0.6329	0.6104	0.5470	0.9147
α = 0.025						
	Jam	**Johan**	**Yao**	**Krish**	**Yarah**	**Propose**
Critical value	9.1751	9.4647	10.6948	9.5595	7.9867	11.7667
Pvalue	0.0235	0.0261	0.0361	0.0292	0.0140	0.0446
Power	0.4167	0.7143	0.7476	0.7256	0.6594	0.9585
α = 0.01						
	Jam	**Johan**	**Yao**	**Krish**	**Yarah**	**Propose**
Critical value	11.9017	12.4835	14.4335	12.5874	10.1276	16.2272
Pvalue	0.0263	0.0255	0.0355	0.0286	0.0136	0.0438
Power	0.5872	0.8273	0.8548	0.8363	0.7757	0.9842

Table 2 shows that the propose solution performed better than all other solutions because it has highest critical value and power of the test followed by Yoa. And all the procedure accept alternative hypothesis at significant level α = 0.05. James and Yanagihara accept alternative hypothesis at significant level α = 0.025 while propose procedure, Krishnamoorthy, Yao and Johanson reject alternative hypothesis. But at α = 0.01 all the procedure reject alternative hypothesis. The power of the test for the propose procedure is better than all other procedure at the three significant level (0.05, 0.025 and 0.01).

REFERENCES

1. Ajit, C.T. and R.L. Bret. 2002. "Accurate Critical Constants for the One–Sided Approximate Likelihood Ratio Test of a Normal Mean Vector when the Covariance Matrix is Estimated". *Biometrics*. 58(3):650–656.

2. Bennett, B.M. 1951. "Note on a Solution of the Generalized Behrens-Fisher Problem". *Annals of the Institute of Statistical Mathematics*. 2:87–90.

3. Bush, K.A. and I. Olkin. 1959. "Extrema of Quadratic Forms with Applications to Statistics". *Biometrika.* 46:483–6.

4. Fco. J.G. and C. del Castillo. 2010. "The Multivariate Behrens–Fisher Distribution". *Journal of Multivariate Analysis*. 101(9):2091–2102.

5. Fisher, R.A. 1935. "The Fiducial Argument in Statistical Inference". *Annals of Eugenics*. 6:391 – 398.

6. Gamage, J., T. Mathew, and S. Weerahandi. 2004. "Generalized P–Values and Generalized Confidence Regions for Multivariate Behrens-Fisher Problem and MANOVA". *Journal of Multivariate Analysis*. 88:177–189.

7. Hotelling, H. 1931. "A Generalized T- Test and Measure of Multivariate Dispersion". *Proceedings of the Second Berkeley on Mathematics and Statistics*. 23 - 41.

8. Reed, J.F., III. 2003. "Solutions to the Behrens–Fisher Problem". *Computer Methods and Programs in Biomedicine* . 70(3):259–263.

9. James, G.S. 1954. "Tests of Linear Hypothesis in Univariate and Multivariate Analysis when the Ratios of the Population Variance are Unknown". *Biometrika*, 41:19–43.

10. Johansen, S. 1980. "The Welch – James Approximation to the Distribution of the Residual Sum of Squares in a Weighted Linear Regression". *Biometrika*. 67:85 – 95.

11. Kawasaki, T. and T. Seo. 2012. "A Two Sample Test for Mean Vectors with Unequal Covariance Matrices". *American Journal of Mathematical and Management Sciences*. 31(1-2):117-135.

12. Kim, S. 1992. "A Practical Solution to the Multivariate Behrens – Fisher Problem". *Biometrika*. 79:171 – 176.

13. Krishnamoorthy, K. and J. Yu. 2004. "Modified Nel and Van der Merwe test for the Multivariate Behrens–Fisher Problem". *Statistica and Probability Letters.* 66:161 – 169.

14. Krishnamoorthy, K. and J. Yu. 2012. "Multivariate Behrens-Fisher Problem with Mussing Data". *J. Multivariate Anal.*105:141 -150.

15. Lin, S.H. and R.S. Wang. 2009. "Inferences on a Linear Combination of K Multivariate Normal Mean Vectors". *Journal of Applied Statistics*. 36(4):415 – 428.

16. Linnik, Y.V. 1966. "Latest Investigations on Behrens-Fisher Problem". *Sankhya*. 28:15 – 24.

17. Lisa, M., H.J. Keselman, and M.H. Aynslie. 2005. "Robust Tests for the Multivariate Behrens–Fisher Problem". *Computer Methods and Programs in Biomedicine.* 77(2):129–139.

18. Nel, D.G. and C.A. Van der Merwe. 1986. "A Solution to the Multivariate Behrens–Fisher Problem," *Communications in Statistics: Theory and Methods.* 15:3719 – 3735.

19. Oyeyemi, G.M. and P.O. Adebayo. 2016. "Comparison of the Invariant Solution to the Multivariate Behrens-Fisher Problem". *Journal of the Nigerian Association of Mathematical Physics*. 34:199 – 208. Nigeria Association of Mathematical Physics.

20. de Siqueira, P.R. and F.F. Danie. 2010. "A Bayesian Solution to the Multivariate Behrens–Fisher Problem". *Computational Statistics & Data Analysis*.54:1622–1633.

21. Park, J. and B. Sinha. 2009. "Some Aspects of Multivariate Behrens-Fisher Problem". *Calcutta Statistical Association Bulletin*. journals.sagepub.com

22. Rencher, A.C. 2002. *Methods of Multivariate Analysis Second Edition*. John Wiley: New York, NY.

23. Satterthwaite, F.A. 1946. “An Approximate Distribution of Sstimates and Variance Components”. *Biometircs*. 2:110-114.

24. Seko, N., T. Kawasaki, and T. Seo, 2011). “Testing Equality of Two Mean Vectors with Two-Step Monotone Missing Data”. *Amer. J. Math. Management Sci*.31:117–135.

25. Scheffe’, H. 1943. “On Solutions of the Behrens-Fisher Problem, Based on the t – Distribution”. *The Annals of Mathematical Statistics*. 14:35 – 44.

26. Kawasaki, T. and T. Seo. 2015. “A Two Sample Test for Mean Vectors with Unequal Covariance Matrices”. *Communications in Statistics - Simulation and Computation.* 44(7):1850-1866.

27. Weerahandi, S. 1993. “General Confidence Intervals”. *J. Amer. Statist. Assoc*. 88: 899 – 905.

28. Weerahandi, S. 1995. *Exact Statistical Methods for Data Analysis*. Springer-Verlag: New York: NY.

29. Welch, B.L. 1947. “The Generalization of ‘Student’s’ Problem when Several Different Population Variances are Involved”. *Biometrika*. 34:28 – 35.

30. Yanagihara, H. and K. Yuan. 2005. “Three Approximate Solutions to the Multivariate Behrens-Fisher Problem”. *Comm. Statist., Simulation Comput*. 34:975 – 988.

31. Yao, Y. 1965. “An Approximation Degrees of Freedom Solution to the Multivariate Behrens – Fisher Problem”. *Biometrika.* 52:139–147.

32. Anderson, T. W. 2003. *An introduction to Multivariate Statistical Analysis. Third edition*. Wiley–Interscience: New York, NY.

33. Kullback, S. 1959, *Information Theory and Statistics*. John Wiley and Sons: New York, NY.

34. Kullback, S. 1967. “On Testing Correlation Matrices”. *Applied Statistics*. 16:80 – 85.

SUGGESTED CITATION

Oyeyemi, G.M. and P.O. Adebayo. 2018. “Stand-in Procedure to Multivariate Behrens-Fisher Problem”. *Pacific Journal of Science and Technology*. 19(1):141-150.

[Pacific Journal of Science and Technology](#)