Communications in Statistics - Simulation and Computation 33 (2004) 431-446

COMPARISON OF FIVE TESTS FOR THE COMMON MEAN OF SEVERAL MULTIVARIATE NORMAL POPULATIONS

K. Krishnamoorthy and Yong Lu
Department of Mathematics
University of Louisiana at Lafayette
Lafayette, LA 70504-1010
krishna@louisiana.edu

Key Words: affine invariant; Graybill-Deal estimator; Hotelling T^2 statistic; size; power

ABSTRACT

In this article, we evaluate the merits of some tests for testing about the common mean vector of several multivariate normal populations with different covariance matrices. Five tests, including the Fisher's exact test, are compared with respect to their powers. Powers are estimated using Monte Carlo simulation. Based on the comparison studies, we found that the Fisher's test and one of the test due to Zhou and Mathew (1994, *Journal of Multivariate Analysis*, 51, 265–276) are in general more powerful than other three tests. The tests are illustrated using simulated data.

1. INTRODUCTION

The problem of making inference about the common mean of several populations arises in situations where independent samples are collected from different studies or using different instruments. If independent samples are collected from different multivariate normal populations with a common mean but possibly with different covariance matrices, then the problem of interest is to combine the summary statistics of the samples to estimate or to test the common mean vector. Practical examples for the univariate case are given in Meier (1) and Eberhardt, Reeve and Spiegelman (2). For the multivariate case, Zhou and Mathew (3) pointed out such common mean problems arise in inter-block analysis of block designs. To describe the present problem, let us suppose that we have k p-variate normal populations with common mean vector μ and unknown covariance matrices $\Sigma_1, ..., \Sigma_k$. Let us denote the ith population by $N_p(\mu, \Sigma_i)$, $i = 1, ..., k$. Let $X_{i1}, ..., X_{in_i}$ be a sample from $N_p(\mu, \Sigma_i)$, $i = 1, ..., k$. The sample mean vector $\bar{X}_i$ and sample covariance matrix S_i are defined as

$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij} \quad \text{and} \quad S_i = \frac{1}{n_i - 1}\sum_{i=1}^{n}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'.$$

We note that $\bar{X}_i$'s and S_i's are mutually independent with

$$\bar{X}_i \sim N_p\left(\mu_i, \frac{1}{n_i}\Sigma_i\right) \quad \text{and} \quad (n_i - 1)S_i = A_i \sim W_p(n_i - 1, \Sigma_i),\ i = 1, ..., k. \tag{1.1}$$

Furthermore, the sample means and covariance matrices are minimal sufficient statistics, and hence the inferential methods are usually developed using these minimal sufficient statistics.

For the univariate case, the problem of estimating μ has received considerable attention. A popular result is due to Graybill and Deal (4) who first showed, for $k = 2$, that the combined estimator

$$\hat{\mu} = \left(\sum_{i=1}^{k} n_i S_i^{-1}\right)^{-1} \sum_{j=1}^{k} n_j S_j^{-1} \bar{X}_j \tag{1.2}$$

has a smaller variance than either sample mean provided both sample sizes are at least 11. Since then many authors considered the problems of hypothesis testing and interval estimation of μ. For some recent results, see Cohen and Sackrowitz (5), Zhou and Mathew (6), Krishnamoorthy and Lu (7), and the references therein. The problem of point estimation in the multivariate case has been addressed in a few articles. On the contrary to the univariate result, it has been shown by Chiou and Cohen (8) that the estimator in (1.2) dominates neither $\bar{X}_1$ nor $\bar{X}_2$ with respect to the covariance criterion. Other decision theoretic results about point estimation are given in Loh (9), Kubokawa (10) and George (11).

Regarding hypothesis testing about μ, Mathew and Zhou (3) proposed several combined tests including the Fisher's test which are all obtained by combining the p-values of the individual tests based on the Hotelling T^2 statistics. Although these tests are simple to use, it is difficult to invert them to find confidence regions for the common mean vector μ. Jordan and Krishnamoorthy (12) proposed a combined procedure that produces confidence region centered at the Graybill-Deal estimator in (1.2). These authors also provided a multiple comparison procedure for the components of μ.

The main purpose of this article is to compare the powers of the different combined tests and identify the better ones for practical use. In the following section, we describe five combined tests that will be considered for comparison. In Section 3, the powers of the tests are estimated using Monte Carlo simulation consisting of 100,000 runs. The powers are estimated for the cases of $p = 2$, $k = 2$ (Tables 1(a-c)), $p = 2, k = 7$ (Table 2a and 2b), and for the case of $p = 3, k = 2$ (Table 3). We first observe from the table values that none of the tests dominate the others uniformly. Based on simplicity and efficiency consideration, our study recommends the Fisher's exact test followed by one of the Zhou and Mathew's (3) test for practical applications. All the five tests are illustrated using simulated data in Section 4.

2. THE COMBINED TESTS

We shall now describe the combined tests for μ. Since the testing problem is location invariant, without loss of generality, we consider the hypotheses

$$H_0 : \mu = 0 \text{ vs. } H_a : \mu \neq 0. \tag{2.1}$$

The Hotelling T^2 statistic based on the ith sample is given by

$$T_i^2 = n_i \bar{X}_i' S_i^{-1} \bar{X}_i \quad \text{and} \quad T_i^2 \sim \frac{p(n_i - 1)}{n_i - p} F_{p, n_i - p}, \tag{2.2}$$

where $F_{m,n}$ denotes the F random variable with the numerator degrees of freedom (df) m and the denominator df $= n$.

For an observed value T_{i0}^2 of T_i^2, let P_i denote the p-value of the Hotelling T_i^2 test, $i = 1, ..., k$. That is,

$$P_i = P\left(F_{p, n_i - p} > \frac{n_i - p}{p(n_i - 1)} T_{i0}^2)\right), \ i = 1, ..., k. \tag{2.3}$$

Fisher's Test

The combined test based on the Fisher's method of combining independent tests reject the null hypotheses whenever

$$-2\sum_{i=1}^{k}\ln(P_i) > \chi^2_{2k}(1-\alpha),$$

where $\chi^2_m(a)$ denotes the $100a$th percentile of a chi-square distribution with df $= m$. In the sequel, we refer to this test as the Fisher's test.

Zhou and Mathew's (1994) Tests

Zhou and Mathew (3) proposed several combined tests for testing hypotheses in (2.1). These combined tests are obtained by combining the P_i's with weights inversely proportional to a scalar valued function of an estimate of Σ_i, and directly proportional to the sample size n_i, $i = 1, ..., k$. To describe these tests, let

$$W = -\sum_{i=1}^{k}\gamma_i \ln(P_i),\ \hat{\Sigma}_i = (A_i + n_i\bar{X}_i\bar{X}_i')/n_i \text{ and } Q_i = n_i\hat{\Sigma}_i^{-1},\ i = 1, ..., k, \tag{2.4}$$

where the weights γ_i's are positive real valued functions of Q_i chosen so that $\sum_{i=1}^{k}\gamma_i = 1$, and $A_i = (n_i - 1)S_i$, $i = 1, ..., k$. The combined test due to Zhou and Mathew (1994) rejects the null hypothesis when

$$\sum_{i=1}^{k}\frac{\gamma_i^{k-1}e^{-w/\gamma_i}}{\prod_{j=1;j\neq i}^{k}(\gamma_i-\gamma_j)} \leq \alpha(1+\eta), \tag{2.5}$$

where

$$\eta = \frac{\sum_{i<j}\eta_{ij}}{k(k-1)/2} \quad \text{and} \quad \eta_{ij} = \frac{\bar{X}_i'\bar{X}_j}{\|\bar{X}_i\|\|\bar{X}_j\|}.$$

It should be noted that the test based on (2.5) with $\eta = 0$ is also exact. Since our preliminary simulation studies and the comparison studies in Krishnamoorthy and Lu (7) for the univariate case showed that the test with $\eta = 0$ is, in general, less powerful than the one with η given above. For this reason, we do not consider the test with $\eta = 0$ for comparison. In the sequel, we consider the tests based on (2.5) with the following weights given in Zhou and Mathew (3):

(i) *Weights based on determinants:*

$$\gamma_i = \frac{|Q_i|}{\sum_{j=1}^{k}|Q_j|},\ i = 1, ..., k.$$

(ii) *Weights based on traces:*

$$\gamma_i = \frac{\mathrm{tr}(Q_i(\sum_{j=1}^k Q_j)^{-1})}{\sum_{i=1}^k \mathrm{tr}(Q_i(\sum_{j=1}^k Q_j)^{-1})} = \frac{\mathrm{tr}(Q_i(\sum_{j=1}^k Q_j)^{-1})}{p},\ i = 1, ..., k.$$

(iii) *Weights based on eigenvalues:*

$$\gamma_i = \frac{\lambda(Q_i(\sum_{j=1}^k Q_j)^{-1})}{\sum_{i=1}^k \lambda(Q_i(\sum_{j=1}^k Q_j)^{-1})},\ i = 1, ..., k,$$

where $\lambda(A)$ denotes the largest eigenvalue of A.

Jordan and Krishnamoorthy's (1996) Test

The combined test due to Jordan and Krishnamoorthy (12) is based on the test statistic which is a weighted average of the T_i^2's, and is given by

$$\sum_{i=1}^k c_i T_i^2,\ \text{where } c_i = \frac{\{\mathrm{Var}(T_i^2)\}^{-1}}{\sum_{j=1}^k \{\mathrm{Var}(T_j^2)\}^{-1}},\ i = 1, ..., k,$$

and

$$\mathrm{Var}(T_i^2) = \frac{2p(n_i-1)^2(n_i-2)}{(n_i-p-2)^2(n_i-p-4)},\ i = 1, ..., k.$$

These authors provided exact critical values of the combined test for the cases of $k = 2$ and $p = 2, 3$ and 4. For $k \geq 3$, they provided a moment approximation method to compute the critical values. Specifically, they showed that $\sum_{i=1} c_i T_i^2 \sim dF_{kp,\nu}$ approximately, where

$$\nu = \frac{4M_2kp - 2M_1^2(kp+2)}{M_2kp - M_1^2(kp+2)},\quad d = M_1\frac{\nu-2}{\nu},\ M_1 = p\sum_{i=1}^k \frac{c_i(n_i-1)}{n_i-p-2}$$

and

$$M_2 = p(p+2)\sum_{i=1}^k \frac{c_i(n_i-1)^2}{(n_i-p-2)(n_i-p-4)} + 2p^2\sum_{i>j}\frac{c_ic_j(n_i-1)(n_j-1)}{(n_i-p-2)(n_j-p-2)}.$$

3. POWER STUDIES AND DISCUSSION

It should be noted that the powers of the Fisher's test and Jordan and Krishnamoorthy's test can be estimated by generating non-central F variates. To explain this, let $F_{m,n}(c)$ denotes the non-central F random variable with the numerator df $= m$, denominator df $= n$, and the noncentrality parameter c. Note that under $H_a : \mu \neq 0$, the T_i^2 in (2.2) distributed as the non-central $F_{p,n_i-p}(\delta_i)$ random variable, where $\delta_i = n_i\mu'\Sigma_i^{-1}\mu$, $i = 1, ..., k$. Therefore, the following algorithm can be used to estimate the powers of the Fisher's test.

Algorithm 1

For given $n_1, ..., n_k$, p, $\delta_1, ..., \delta_k$:

do i = 1, m

do j = 1, k

generate $V_j \sim F_{p,n_j-p}(\delta_j)$

set $P_j = P(F_{p,n_j-p} > V_j)$

end do

set $W = -2\sum_{j=1}^{k} \ln(P_j)$

if $(W > \chi^2_{2k}(1-\alpha))$ power = power + $1/m$

end do

To generate noncentral $F_{m,n}(\delta)$ variates, we note that $F_{m,n}(\delta) \sim (n\chi^2_m(\delta))/(m\chi^2_n)$, where $\chi^2_a(c)$ denotes the noncentral chi-square random variable with degrees of freedom = m, and noncentrality parameter c, and the random variables in the numerator and in the denominator are independent. The noncentral chi-square random variates can be generated using the fact that $\chi^2_m(\sum_{i=1}^{m} a_i^2) = \sum_{i=1}^{m} Z_i^2$, where Z_i's are independent normal random variables with mean a_i and variance 1.

The power of the test due to Jordan and Krishnamoorthy (12) also depends on the parameter space via $(\delta_1, ..., \delta_k)$, and it can be expressed as

$$P\left(\sum_{i=1}^{k} c_i \frac{p(n_i-1)}{n_i-p} F_{p,n_i-p}(\delta_i) > a\right),$$

where a is the $100(1-\alpha)$th percentile of $\sum_{i=1}^{k} c_i \frac{p(n_i-1)}{n_i-p} F_{p,n_i-p}$.

Thus, we see that the powers of the Fisher's test and Jordan and Krishnamoorthy's test can be estimated without generating Wishart variates and computing their inverses. But to estimate the powers of Zhou and Mathew's tests, it is necessary to generate Wishart variates and compute their inverses. We used subroutine RNNOA for generating normal random numbers and the subroutine due to Smith and Hockings (13) for generating Wishart matrices. To understand the validity of the simulation, we also reported the sizes of the tests. The powers are reported for the tests:

(a) Fisher's test

(b) Zhou and Mathew's test with determinant weights

(c) Zhou and Mathew's test with trace weights

(d) Zhou and Mathew's test with eigenvalue weights

(e) Jordan and Krishnamoorthy's test

For $k = 2$, we note that there exists a nonsingular matrix M such that $MM' = \Sigma_2$ and $M\text{diag}(\lambda_1, ..., \lambda_p)M' = \Sigma_1$, where $\lambda_1, ..., \lambda_p$ are the eigenvalues of $\Sigma_1^{-1}\Sigma_2$. Since all the tests are affine invariant, we can take without loss of generality that $\Sigma_2 = I_p$ and $\Sigma_1 = \text{diag}(\lambda_1, ..., \lambda_p)$ for power comparisons. Furthermore, the powers of all the tests depend on the covariance matrices of the sample mean vectors (that is, Σ_i/n_i's), and hence without loss of generality we can take $n_1 = ... = n_k$ for power comparisons.

The estimated powers are presented in Table 1 for the case of $k = 2$ and $p = 2$, in Tables 2(a,b) for the case of $k = 7$ and $p = 2$, and in Table 3 for the case of $k = 2$ and $p = 3$. Comparisons among Zhou and Mathew's tests (b), (c) and (d) show that the test (c), the one with trace weights, is in general more powerful than (b) and (d). The test by Jordan and Krishnamoorthy is in general inferior to the Fisher's test, and is slightly better than the test (c) when $k = 7$. Therefore, here afterwards, we compare only the Fisher test and test (c).

When the sample sizes are equal, and the mean is proportional to the vector of ones, the test (c) is more powerful than the Fisher's test when Σ_1 is not far away from Σ_2; if Σ_1 and Σ_2 are drastically different, then the Fisher's test is better than the test (c). See the second block of columns in Table 1. If the components of μ are different, then the Fisher's test seems to be the best among all the tests for all Σ_1 and Σ_2. See the first block of columns in Table 1. Note that for $k \geq 3$, the Σ_i's are not simultaneously diagnolizable; the powers need to be computed for arbitrary positive definite matrices Σ_i's. However, for simplicity we consider only correlation matrices (Table 2a) and diagonal matrices (Table 2b) when $k = 7$ and $p = 2$. These two table values clearly indicate that the Fisher's test is in general more powerful than other tests. The power comparisons for $p = 3$ in Table 3 also indicate the Fisher's test is superior to other tests.

Table 1. *Simulated Powers of the Tests*

$k = 2,\ p = 2;\ n_1 = n_2 = 11;\ \alpha = 0.05;\ \Sigma_1 = \mathrm{diag}(\lambda_1, \lambda_2);\ \Sigma_2 = I_2$

		μ										
$(\sqrt{\lambda_1}, \sqrt{\lambda_2})$	Tests	$\binom{0}{0}$	$\binom{.1}{.2}$	$\binom{.3}{.1}$	$\binom{.3}{.2}$	$\binom{.8}{.2}$	$\binom{.9}{.1}$	$\binom{.3}{.3}$	$\binom{.4}{.4}$	$\binom{.5}{.5}$	$\binom{.6}{.6}$	$\binom{.8}{.8}$
(1.0,1.0)	(a)		.09	.14	.18	.77	.85	.24	.42	.61	.79	.97
	(b)		.10	.13	.18	.70	.79	.25	.42	.61	.78	.96
	(c)		.11	.14	.19	.75	.83	.26	.45	.65	.82	.98
	(d)		.11	.14	.18	.73	.81	.26	.44	.64	.81	.97
	(e)	.05	.09	.14	.17	.74	.82	.23	.39	.59	.76	.96
(1.0,2.0)	(a)		.08	.14	.16	.76	.85	.20	.33	.51	.69	.93
	(b)		.09	.12	.15	.64	.73	.20	.34	.51	.68	.91
	(c)		.09	.13	.16	.72	.81	.21	.36	.54	.72	.94
	(d)		.09	.13	.16	.69	.78	.20	.34	.52	.70	.93
	(e)	.05	.08	.13	.16	.73	.83	.19	.32	.49	.66	.91
(1.0,4.0)	(a)		.08	.14	.16	.75	.85	.19	.32	.48	.66	.91
	(b)		.08	.12	.14	.57	.65	.18	.29	.44	.60	.86
	(c)		.08	.13	.15	.68	.78	.20	.32	.49	.65	.90
	(d)		.08	.13	.15	.67	.76	.18	.30	.47	.63	.89
	(e)	.05	.08	.14	.15	.73	.82	.18	.30	.46	.63	.89
(1.0,7.0)	(a)		.08	.14	.16	.75	.85	.18	.31	.48	.65	.90
	(b)		.08	.11	.13	.54	.62	.17	.28	.41	.55	.81
	(c)		.08	.13	.15	.67	.76	.18	.30	.46	.62	.87
	(d)		.08	.12	.14	.67	.76	.17	.29	.44	.60	.86
	(e)	.05	.08	.13	.15	.73	.82	.18	.30	.46	.63	.89
(1.0,9.0)	(a)		.07	.14	.16	.75	.85	.18	.31	.47	.65	.90
	(b)		.08	.11	.13	.53	.61	.17	.27	.40	.54	.79
	(c)		.08	.12	.15	.66	.76	.18	.30	.45	.61	.86
	(d)		.07	.12	.14	.66	.76	.17	.28	.43	.59	.85
	(e)	.05	.07	.13	.15	.72	.82	.18	.30	.45	.62	.88

Table 2a. *Simulated Powers of the Tests*

$$k = 7,\ p = 2;\ \alpha = 0.05;\ n_i = 15, \Sigma_i = \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix},\ i = 1, ..., 7$$

		μ						μ			
$(\rho_1, ..., \rho_7)$	Tests	$\binom{0}{0}$	$\binom{.1}{.2}$	$\binom{.3}{.1}$	$\binom{.3}{.2}$	$\binom{.4}{.2}$	$\binom{.5}{.3}$	$\binom{.2}{.2}$	$\binom{.3}{.3}$	$\binom{.4}{.4}$	$\binom{.5}{.5}$
(.1,.1,.1,.1,.1,.1,.1)	(a)	.05	.19	.39	.50	.74	.95	.30	.67	.94	1
	(b)	.05	.15	.29	.38	.60	.87	.22	.53	.84	.97
	(c)	.05	.17	.35	.45	.69	.93	.26	.62	.91	.99
	(d)	.05	.16	.34	.44	.68	.92	.26	.60	.90	.99
	(e)	.05	.17	.36	.46	.70	.94	.27	.63	.91	.99
(.1,.2,.3,.4,.5,.6,.7)	(a)	.05	.17	.40	.42	.69	.91	.23	.54	.86	.98
	(b)	.05	.13	.30	.30	.53	.79	.17	.39	.69	.90
	(c)	.05	.15	.35	.36	.63	.87	.20	.48	.80	.96
	(d)	.05	.14	.33	.33	.58	.84	.19	.43	.75	.94
	(e)	.05	.16	.36	.38	.65	.89	.21	.50	.82	.97
(.1,.1,.2,.2,.9,.9,.9)	(a)	.05	.23	.64	.48	.85	.97	.23	.54	.85	.98
	(b)	.05	.18	.55	.29	.69	.83	.13	.27	.49	.73
	(c)	.05	.21	.62	.40	.80	.94	.18	.42	.73	.93
	(d)	.05	.18	.57	.30	.70	.84	.13	.27	.50	.74
	(e)	.05	.21	.60	.44	.81	.95	.21	.50	.82	.97
(.1,.1,.5,.5,.7,.9,.9)	(a)	.05	.21	.57	.44	.80	.95	.21	.51	.82	.97
	(b)	.05	.16	.49	.29	.64	.80	.13	.27	.50	.74
	(c)	.05	.19	.55	.37	.75	.91	.17	.40	.70	.92
	(d)	.05	.17	.51	.29	.66	.82	.13	.28	.51	.75
	(e)	.05	.19	.53	.41	.76	.93	.20	.47	.79	.96
(-.9,-.2,-.1,0,.1,.2,.9)	(a)	.05	.43	.80	.90	.99	1	.66	.97	1	1
	(b)	.05	.41	.75	.74	.91	.96	.53	.78	.89	.94
	(c)	.05	.48	.84	.90	.99	1	.68	.96	1	1
	(d)	.05	.37	.71	.61	.83	.89	.43	.59	.71	.84
	(e)	.05	.42	.79	.90	.99	1	.66	.98	1	1

Table 2b. *Simulated Powers of the Tests*

$k=7,\ p=2;\ \alpha=0.05;\ n_i=15, \Sigma_i=\mathrm{diag}(1,\sigma_i^2),\ i=1,...,7$

		μ						μ			
$(\sigma_1^2,...,\sigma_7^2)$	Tests	$\binom{0}{0}$	$\binom{.1}{.2}$	$\binom{.3}{.1}$	$\binom{.3}{.2}$	$\binom{.4}{.2}$	$\binom{.5}{.3}$	$\binom{.2}{.2}$	$\binom{.3}{.3}$	$\binom{.4}{.4}$	$\binom{.5}{.5}$
(1,1,1,1,1,1,1)	(a)	.05	.20	.41	.54	.78	.97	.32	.72	.96	1
	(b)	.05	.16	.31	.42	.64	.91	.24	.58	.88	.99
	(c)	.05	.18	.37	.49	.73	.95	.29	.67	.94	1
	(d)	.05	.17	.35	.47	.71	.95	.28	.65	.93	1
	(e)	.05	.19	.38	.50	.74	.96	.30	.68	.94	1
(1,2,2,3,2,4,2)	(a)	.05	.13	.39	.46	.73	.94	.24	.57	.87	.98
	(b)	.05	.11	.28	.33	.56	.83	.18	.43	.74	.93
	(c)	.05	.12	.34	.41	.67	.91	.22	.52	.83	.97
	(d)	.05	.12	.32	.38	.63	.89	.20	.49	.81	.97
	(e)	.05	.12	.36	.42	.69	.92	.22	.53	.84	.98
(1,2,3,4,5,6,7)	(a)	.05	.11	.38	.44	.71	.93	.22	.51	.83	.97
	(b)	.05	.10	.25	.30	.49	.76	.17	.38	.66	.88
	(c)	.05	.11	.32	.37	.62	.88	.20	.46	.77	.95
	(d)	.05	.10	.31	.36	.61	.87	.18	.44	.75	.94
	(e)	.05	.11	.35	.40	.66	.91	.20	.48	.79	.96
(1,1,1,1,9,9,9)	(a)	.05	.14	.39	.47	.74	.95	.26	.60	.90	.99
	(b)	.05	.13	.25	.33	.51	.80	.20	.45	.76	.94
	(c)	.05	.14	.33	.42	.66	.91	.24	.55	.86	.98
	(d)	.05	.13	.32	.40	.64	.91	.22	.53	.84	.98
	(e)	.05	.13	.36	.44	.70	.93	.24	.56	.87	.98
(1,1,3,3,9,9,9)	(a)	.05	.12	.39	.44	.71	.93	.23	.53	.85	.98
	(b)	.05	.11	.24	.29	.47	.74	.17	.39	.67	.89
	(c)	.05	.12	.31	.38	.62	.88	.21	.48	.80	.96
	(d)	.05	.11	.31	.37	.61	.88	.19	.46	.78	.95
	(e)	.05	.11	.35	.41	.67	.91	.21	.49	.81	.97

Table 3. *Simulated Powers of the Tests*

$k=2,\ p=3;\ n_1=n_2=11;\ \alpha=0.05;\ \Sigma_1=\text{diag}(\lambda_1,\lambda_2,\lambda_3);\ \Sigma_2=I_3$

$(\sqrt{\lambda}_1,\sqrt{\lambda}_2,\sqrt{\lambda}_3)$	Tests	μ = $\begin{pmatrix}0\\0\\0\end{pmatrix}$	$\begin{pmatrix}.3\\.2\\.1\end{pmatrix}$	$\begin{pmatrix}.5\\.3\\.1\end{pmatrix}$	$\begin{pmatrix}.7\\.4\\.2\end{pmatrix}$	$\begin{pmatrix}.8\\.6\\.2\end{pmatrix}$	$\begin{pmatrix}.9\\.5\\.3\end{pmatrix}$
(1.0,1.0,1.0)	(a)		.15	.34	.64	.84	.88
	(b)		.13	.29	.56	.76	.82
	(c)		.14	.33	.63	.83	.87
	(d)		.14	.31	.60	.80	.85
	(e)	.05	.14	.32	.61	.81	.85
(1.0,2.0,3.0)	(a)		.13	.30	.58	.77	.83
	(b)		.12	.23	.44	.62	.67
	(c)		.13	.28	.54	.73	.79
	(d)		.12	.27	.51	.71	.76
	(e)	.05	.13	.28	.54	.73	.80
(1.0,3.0,3.0)	(a)		.13	.30	.57	.75	.82
	(b)		.12	.23	.40	.60	.66
	(c)		.12	.27	.52	.71	.78
	(d)		.12	.26	.50	.69	.75
	(e)	.05	.12	.28	.54	.72	.79
(1.0,2.0,4.0)	(a)		.13	.30	.58	.76	.83
	(b)		.11	.23	.43	.61	.65
	(c)		.12	.28	.53	.72	.78
	(d)		.12	.27	.51	.70	.76
	(e)	.05	.13	.28	.55	.73	.80
(1.0,5.0,7.0)	(a)		.13	.29	.56	.74	.81
	(b)		.12	.23	.42	.59	.64
	(c)		.12	.26	.49	.68	.74
	(d)		.12	.26	.49	.68	.73
	(e)	.05	.12	.27	.53	.71	.79

Table 3 continued.

$(\sqrt{\lambda_1},\sqrt{\lambda_2},\sqrt{\lambda_3})$	Tests	μ $\begin{pmatrix}0\\0\\0\end{pmatrix}$	$\begin{pmatrix}.1\\.1\\.1\end{pmatrix}$	$\begin{pmatrix}.2\\.2\\.2\end{pmatrix}$	$\begin{pmatrix}.3\\.3\\.3\end{pmatrix}$	$\begin{pmatrix}.4\\.4\\.4\end{pmatrix}$	$\begin{pmatrix}.5\\.5\\.5\end{pmatrix}$
(1.0,1.0,1.0)	(a)		.07	.13	.26	.46	.68
	(b)		.07	.13	.25	.44	.64
	(c)		.07	.14	.28	.49	.71
	(d)		.07	.14	.27	.47	.68
	(e)	.05	.07	.13	.25	.44	.65
(1.0,2.0,3.0)	(a)		.06	.11	.19	.33	.51
	(b)		.06	.11	.20	.32	.49
	(c)		.06	.11	.21	.36	.54
	(d)		.06	.11	.20	.34	.51
	(e)	.05	.06	.10	.19	.31	.49
(1.0,3.0,3.0)	(a)		.06	.10	.19	.32	.49
	(b)		.06	.11	.19	.32	.48
	(c)		.06	.11	.21	.35	.53
	(d)		.06	.11	.19	.33	.50
	(e)	.05	.06	.10	.18	.30	.47
(1.0,2.0,4.0)	(a)		.06	.10	.19	.33	.50
	(b)		.06	.11	.19	.32	.47
	(c)		.06	.11	.21	.35	.53
	(d)		.06	.11	.19	.33	.50
	(e)	.05	.06	.10	.18	.31	.48
(1.0,5.0,7.0)	(a)		.06	.10	.18	.31	.47
	(b)		.06	.10	.19	.30	.45
	(c)		.06	.11	.20	.33	.49
	(d)		.06	.10	.18	.31	.47
	(e)	.05	.06	.10	.17	.30	.45

Complete comparison studies for larger k are not feasible because the powers of the tests (b), (c) and (d) need to be computed over a set of positive definite matrices. However, our limited comparison studies clearly showed that no test dominates the Fisher test uniformly; the Fisher test indeed performs better than other tests for the parameter configurations considered. Thus, if hypothesis test about the common mean is of primary interest, then on the basis of simplicity and efficiency, we recommend the Fisher's test followed by Zhou and Mathew's (3) test with trace weights for practical applications. Jordan and Krishnamoorthy's approach is useful for confidence estimation of μ and for making simultaneous inference about the components of μ.

4. AN ILLUSTRATIVE EXAMPLE

We shall now illustrate the tests considered in Section 2 using simulated data. The samples given in Table 4 are generated from four bivariate normal populations with common mean (0.6, 0.2) and covariance matrices

$$\Sigma_1 = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 4 \end{pmatrix}, \ \Sigma_2 = \begin{pmatrix} 1 & 2 \\ 2 & 9 \end{pmatrix}, \ \Sigma_3 = \begin{pmatrix} 5 & 2 \\ 2 & 12 \end{pmatrix}, \text{ and } \Sigma_4 = \begin{pmatrix} 4 & 2.5 \\ 2.5 & 18 \end{pmatrix},$$

Using these data, we want to test

$$H_0 : \mu' = (0, 0) \ \text{ vs. } \ H_a : \mu' \neq (0, 0) \tag{4.1}$$

at the level of significance 0.05.

The summary statistics are:

$$\bar{X}_1 = \begin{pmatrix} 0.7879 \\ 0.2324 \end{pmatrix}, \ \bar{X}_2 = \begin{pmatrix} 0.4781 \\ 1.134 \end{pmatrix}, \ \bar{X}_3 = \begin{pmatrix} 1.125 \\ -2.515 \end{pmatrix}, \ \bar{X}_4 = \begin{pmatrix} 2.013 \\ -0.6159 \end{pmatrix},$$

the sample covariance matrices S_1, S_2, S_3 and S_4 are

$$\begin{pmatrix} 1.770 & 1.063 \\ - & 5.615 \end{pmatrix}, \begin{pmatrix} 1.205 & 1.850 \\ - & 7.264 \end{pmatrix}, \begin{pmatrix} 4.941 & 2.622 \\ - & 12.036 \end{pmatrix}, \text{ and } \begin{pmatrix} 5.863 & 8.204 \\ - & 25.467 \end{pmatrix}$$

respectively.

Table 4. *Simulated Samples; $k = 4, p = 2$ and $n_1 = ... = n_4 = 15$*

Samples

1		2		3		4	
-0.169	-4.433	0.493	2.076	2.094	-7.885	-2.512	-8.773
-0.864	1.377	0.411	0.031	1.958	-0.195	4.654	3.208
0.906	1.573	1.181	-1.809	1.077	-2.936	2.378	-8.794
2.509	2.156	0.482	5.096	3.302	1.544	5.562	1.471
0.572	-0.077	-0.555	1.163	0.100	5.186	5.388	10.345
0.346	2.520	0.806	-0.474	0.458	-1.210	3.637	-1.933
-1.635	0.122	1.161	5.588	3.365	-2.998	2.251	3.671
-0.165	-0.386	1.231	2.298	-0.432	-5.917	2.865	-3.173
2.031	0.047	3.131	5.047	-2.327	-5.038	2.788	1.379
0.258	2.386	-0.087	-0.461	1.550	-5.621	-0.746	-2.642
1.169	0.625	-0.849	0.723	2.415	-2.320	-1.041	-3.417
3.594	3.306	-1.148	-3.374	1.550	-4.096	-0.021	-3.236
1.908	1.088	-0.853	-2.366	-4.374	-5.457	3.425	5.070
0.772	-4.589	0.495	1.482	2.033	-2.469	0.340	-0.334
0.587	-2.229	1.273	1.996	4.100	1.682	1.222	-2.081

The matrices Q_1, Q_2, Q_3 and Q_4 defined in (2.4) are

$$\begin{pmatrix} 7.456 & -1.655 \\ - & 3.201 \end{pmatrix}, \begin{pmatrix} 20.975 & -5.900 \\ - & 3.519 \end{pmatrix}, \begin{pmatrix} 2.556 & 0.0555 \\ - & 0.8554 \end{pmatrix}, \text{ and } \begin{pmatrix} 1.919 & -0.5099 \\ - & 0.7567 \end{pmatrix}$$

respectively. The Hotelling T^2 statistics based on the individual samples are $T_1^2 = 5.436$, $T_2^2 = 3.389$, $T_3^2 = 17.489$ and $T_4^2 = 22.998$. We shall now give the five test procedures given in Section 2.

Fisher's Test

The p-values based on the individual Hotelling T^2 tests are $P_1 = 0.1185$, $P_2 = 0.2444$, $P_3 = 0.005149$ and $P_4 = 0.00181$. The observed value of the test statistic $-2\sum_{i=1}^{4}\ln(P_i) = 30.254$ and the overall p-value is $P(\chi_8^2 > 30.254) = 0.00019$. Thus, this test rejects H_0 in (4.1).

Zhou and Mathew's test with weights based on determinants

The weights are $\gamma_1 = 0.3326$, $\gamma_2 = 0.6142$, $\gamma_3 = 0.0344$ and $\gamma_4 = 0.0188$. The value of $\eta = 0.2795$ and $\alpha(1+\eta) = 0.0640$, and the p-value = 0.1077. Since the p-value is not less than $\alpha(1+\eta)$, this test does not reject H_0.

Zhou and Mathew's test with weights based on traces

The weights are $\gamma_1 = 0.3355$, $\gamma_2 = 0.4667$, $\gamma_3 = 0.1198$ and $\gamma_4 = 0.0780$. The p-value = 0.0243. Since the p-value is less than $\alpha(1+\eta)$, this test rejects H_0.

Zhou and Mathew's test with weights based on eigenvalues

The weights are $\gamma_1 = 0.3244$, $\gamma_2 = 0.4711$, $\gamma_3 = 0.1333$ and $\gamma_4 = 0.0712$. The p-value = 0.0236. Since the p-value is less than $\alpha(1+\eta)$, this test also rejects H_0 in (4.1).

Jordan and Krishnamoorthy's test

The value of $d = 2.353$ and $\nu = 26.50$. The statistic $\sum_{i=1}^{k} c_i T_i^2 = 12.328$, and the p-value = $P(F_{8,26.5} > 12.328/d) = P(F_{8,26.5} > 5.239) = 0.00054$.

Recall that the true mean vector is (0.6,0.2), and hence an efficient test should reject the H_0 in (4.1). We observe from above that all the tests, except the Zhou and Mathew's test based on determinant weights, reject the H_0. Furthermore, among all the tests, the Fisher's test produced the smallest p-value.

ACKNOWLEDGEMENT

The authors are thankful to a referee for providing valuable comments and suggestions.

BIBLIOGRAPHY

(1) Meier, P. Variance of a Weighted Mean. Biometrics, **1953**, 9, 59–73.

(2) Eberhardt, K. R.; Reeve, C. P.; Spiegelman, C. H. A Minimax Approach to Combine Means, with Practical Examples. Chemometrics and Intelligent Laboratory Systems, **1989**, 5, 129–148.

(3) Zhou, L.; Mathew, T. Combining Independent Tests in Multivariate Linear Models. J. Multivariate Anal., **1994**, 51, 265–276.

(4) Graybill, F. A.; Deal, R. B. Combining Unbiased Estimators. Biometrics, **1959**, 15, 543–550.

(5) Cohen, A.; Sackrowitz, H. B. Exact Tests that Recover Interblock Information in Balanced Incomplete Blocks Designs. J. Amer. Statist. Assoc., **1989**, 84, 556–560.

(6) Zhou, L.; Mathew, T. Combining independent tests in linear models. J. Amer. Statist. Assoc., **1993**, 88, 650–655.

(7) Krishnamoorthy, K.; Lu, Y. Inferences on the Common Mean of Several Normal Populations Based on the Generalized variable Method. Biometrics, **2003**, 59, 237–247.

(8) Chiou, W.; Cohen, A. On Estimating a Common Multivariate Normal Mean Vector. Ann. Inst. Statist. Math., **1985**, 37, 499–506.

(9) Loh, W. L. Estimating Common Mean of Two Multivariate Normal Distributions. *Ann. Statist.*, **1991**, 19, 283–296.

(10) Kubokawa, T. Estimating Common Parameters Growth Curve Models Under a Quadratic Loss. Comm. Statist. - Theory and Methods, **1989**, 18, 3149-3155.

(11) George, E. I. Shrinkage Domination in a Multivariate Common Mean Problem. Ann. Statist., **1991**, 19, 952–960.

(12) Jordan, S. J.; Krishnamoorthy, K. Confidence Regions For the Common Mean Vector of Several Multivariate Normal Populations. Canadian J. Statist. **1991**, 23, 283–297.

(13) Smith, W. B.; Hocking, R. R. Wishart Variates Generator (Algorithm AS 53). App. Statist., **1972**, 21, 341–345.