Journal of the Korean
Statistical Society
Vol. 22, No. 2, 1993

On Testing Fisher's Linear Discriminant Function When Covariance Matrices Are Unequal

Hea-Jung Kim [1]

ABSTRACT

This paper proposes two test statistics which enable us to proceed the variables selection in Fisher's linear discriminant function for the case of heterogeneous discrimination with equal training sample size. Simultaneous confidence intervals associated with the test are also given. These are exact and approximate results. The latter is based upon an approximation of a linear sum of Wishart distributions with unequal scale matrices. Using simulated sampling experiments, powers of the two tests have been tabulated, and power comparisons have been made between them.

KEYWORDS : Linear discriminant function, Heterogeneous scale matrices, Variables selection, Test statistic, Simulation.

1. INTRODUCTION

In discriminant analysis, when we have large number of variables it may be of interest to find out a smaller number of important variables which are adequate for discrimination. This is known as variables selection problem and is important not only for increasing the ability of discrimination but also for the cost and computational considerations. For homogeneous(equal covariance matrices case) discriminant analysis, various statistics(cf. Dillon and Goldstein, 1984) were suggested and

[1] Department of Statistics, Dongguk University, Seoul, 100-715, Korea

used in the variables selection procedure. However little attention appears to have been paid to the variables selection in heterogeneous(unequal covariance matrices case) discriminant analysis(cf. Fatti *et. al.*, 1982).

Rao(1952) was first to provide the first test for additional discrimination in the two-group case with homogeneous covariance matrices. If $D^2_{(q)}$ and $D^2_{(p)}$ denote respective sample Mahalanobis distance between the two groups based on the q variables in $X_{(1)}$ and all the p variables $X = (X'_{(1)}, X'_{(2)})'$. Then Rao gives the test statistic:

$$F_1 = \frac{N_1 + N_2 - p - 1}{p - q} c(D^2_{(p)} - D^2_{(q)})/(N_1 + N_2 - 2 + cD^2_{(q)}), \qquad (1)$$

where $c = N_1N_2/(N_1 + N_2)$ has an F-distribution with $(p-q)$ and (N_1+N_2-p-1) degrees of freedom under the null hypothesis that the randomly selected $(p-q)$ variables in $X_{(2)}$ provide no extra discrimination between the two groups. Rao(1952) showed that the statistic can be used for testing the hypothesis that the coefficients of the elements of $X_{(2)}$ in the Fisher's linear discriminant function(LDF) are all zero. The LDF can be also used for the two-group heterogeneous discrimination. Investigations by Gilbert(1969) and Marks and Dunn(1974) indicated that the LDF was adequate if the differences among the covariance matrices were not extreme. Robustness of the LDF is also shown by Lachenbruch *et. al.*(1973).

It is interesting to note when training sample sizes are equal($N_1 = N_2 = N$), the estimated LDF is exactly the same for both homogeneous and heterogeneous cases(cf Morrison, 1990). However, in the case of the heterogeneous discrimination, Rao's statistic would not be applicable for testing the coefficients of the LDF. Instead a proper test statistic which takes into account the heterogeneous covariance matrices is needed. However, it has not been seen yet. Our concern in this paper is to propose procedures for testing the hypothesis on the coefficients of the LDF in the case of two-group heterogeneous discrimination with training samples of equal size. The procedures are obtained by generalizing Rao's(1952) criterion. For the test we propose two test statistics. One is obtained from exact sampling distribution. But it loses many degrees of freedom in constructing a test which is independent of heterogeneity and homogeneity of the two covariance matrices. To circumvent this problem the other is proposed by using an approximation to the distribution of a linear sum of independent Wishart matrices obtained by Nel and van del Merwe(1986).

2. DEFINITIONS AND PRELIMINARY RESULTS

Assume that, for individuals from population Π_i, $i = 1, 2$, the p component response vector X is normally distributed with mean μ_i and covariance matrix Σ_i.

Anderson and Bahadur(1962) developed the discriminant functions for the two multivariate normal populations. When the parameters are known, they proposed the discriminant function

$$Y = (\mu_1 - \mu_2)'(t_1\Sigma_1 + t_2\Sigma_2)^{-1}X \tag{2}$$

and a rule assigning an individual with particular response vector value of X to Π_1 if Y $> \delta$ and to Π_2 otherwise. Among various choices of t_1, t_2, and δ, if we set $t_1 = t_2 = 1/2$ and a corresponding value of δ, (2) is called the LDF with equal sample size(cf. Morrison, 1990 and Marks and Dunn ,1974). When parameters are unknown, the method might be extended by replacing the parameters by their estimators. Now consider the LDF with equal sample size:

$$\mathbf{a}'X = 2(\mu_1 - \mu_2)'(\Sigma_1 + \Sigma_2)^{-1}X, \tag{3}$$

where $\mathbf{a}' = (a_1, \ldots, a_p)$, $X' = (x_1, \ldots, x_p)$, for the two heterogeneous populations. If any of the coefficients a_i are zero, then the corresponding variables x_i do not make any contribution for the discrimination between two populations. Therefore, it is of interest to find out as to which of the coefficients are zero.

Let $X = (X'_{(1)}, X'_{(2)})'$ and let $\triangle = \mu_1 - \mu_2$, and $\Omega = \Sigma_1 + \Sigma_2$ be correspondingly partitioned as

$$\triangle = (\triangle'_1, \triangle'_2)', \quad \Omega = \{\Omega_{ij}\}, \quad i,j = 1,2,$$

and $\mathbf{a} = (\mathbf{a}'_1, \mathbf{a}'_2)'$. We set $X_{(1)}, \triangle_1$, and $\mathbf{a}_1$ are of order qx1 and Ω_{11} is of order qxq.

Lemma 1. The hypothesis **H**: $a_{q+1} = \ldots = a_p = 0$ is equivalent to the hypothesis that

$$\triangle_2 - \Omega_{21}\Omega_{11}^{-1}\triangle_1 = 0$$

and this denotes that the mean difference of $X_{(2)}$ in both populations, after eliminating the effect of $X_{(1)}$, is zero.

Proof. Under the hypothesis **H**, *i.e.* $\mathbf{a}_2 = 0$, (3) gives the following equation:

$$-\triangle'_1\Omega_{11}^{-1}\Omega_{12}\Omega_{22.1}^{-1} + \triangle'_2\Omega_{22.1}^{-1} = 0, \quad \text{where } \Omega_{22.1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}.$$

Postmultiplying $\Omega_{22.1}$ and solving the relation gives first two lines of the lemma. This proves the first part of the statement. For the second part of the statement, let $X(i) = (X'_{(1)}(i), X'_{(2)}(i))'$, $i = 1,2$, be the $i-th$ population variates, where $X_{(1)}(i)$ is of order qx1. Then Z $= X(1) - X(2)$ is distributed as multivariate normal with mean vector $\triangle$ and covariance matrix Ω. If we let Z $= (Z'_1, Z'_2)'$, where Z_1 is of order qx1, the conditional distribution of Z_2 given Z_1 is

$$Z_2 \mid Z_1 \sim N_{p-q}(\Delta_2 + \Omega_{21}\Omega_{11}^{-1}(Z_1 - \Delta_1),\ \Omega_{22.1}).$$

Thus

$$E[Z_2 \mid Z_1] = \Delta_2 + \Omega_{21}\Omega_{11}^{-1}(Z_1 - \Delta_1). \tag{4}$$

Upon eliminating the effect of Z_1 in (4) and setting to zero gives the second statement. ♣

Lemma 2. If we define $\mathbf{M} = \Delta'\Omega^{-1}\Delta$ and $\mathbf{M_1} = \Delta_1'\Omega_{11}^{-1}\Delta_1$, the hypothesis **H** is equivalent to the hypothesis that

$$\mathbf{H}_0 : \mathbf{M} = \mathbf{M_1}.$$

Proof. Ω^{-1} may be written as

$$\Omega^{-1} = \begin{bmatrix} \Omega_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} AB^{-1}A' & -AB^{-1} \\ -B^{-1}A' & B^{-1} \end{bmatrix},$$

where $A = \Omega_{11}^{-1}\Omega_{12}$, $B = \Omega_{22.1}$. By decomposing **M**, we have

$$\mathbf{M} = \mathbf{M_1} + \{\Delta_2 - \Omega_{21}\Omega_{11}^{-1}\Delta_1\}'\Omega_{22.1}^{-1}\{\Delta_2 - \Omega_{21}\Omega_{11}^{-1}\Delta_1\}.$$

Substituting the result of Lemma 1 to the left hand side of the equation gives the result. ♣

This can be interpreted as the hypothesis **H** is equivalent to the hypothesis that the Mahalanobis linear distance(**M**) between two heterogeneous populations based on the p variables is equal to the distance ($\mathbf{M_1}$) between the populations on the basis of the first q variables.

3. TEST STATISTICS

In most practical situations the parameters μ_i and Σ_i, $i = 1, 2$ will not be known. So, to test **H** which is equivalent to $\mathbf{M} = \mathbf{M_1}$, we develop tests based upon sampling distributions of estimates of **M** and $\mathbf{M_1}$.

3.1 An Exact Procedure

Let $X_\alpha(1)$ and $X_\alpha(2)$ are the $\alpha - th$ observations from two independent multivariate normal populations $\Pi_i \sim N_p(\mu_i, \Sigma_i)$, $i = 1, 2$, where $\Sigma_1 \neq \Sigma_2$. Assuming the numbering of the observations in two samples of equal size is independent of observations themselves, let $Z_\alpha = X_\alpha(1) - X_\alpha(2)$, $\alpha = 1, \ldots, N$. Then

$$\bar{Z} = \bar{X}(1) - \bar{X}(2) \sim N_p(\triangle, \frac{1}{N}\Omega), \tag{5}$$

$\Omega = \Sigma_1 + \Sigma_2$, $\triangle = \mu_1 - \mu_2$, independently of $(N-1)\hat{\Omega} = \sum_{\alpha=1}^{N}(Z_\alpha - \bar{Z})(Z_\alpha - \bar{Z})'$ which is a Wishart distribution with $N-1$ degrees of freedom with scale matrix Ω:

$$n\hat{\Omega} \sim W_p(n, \Omega), \text{ where } n = N - 1. \tag{6}$$

Let $Z, \triangle, \Omega$, and $\hat{\Omega}$ are partitioned as

$$\bar{Z} = (\bar{Z}_1', \bar{Z}_2')', \ \triangle = (\triangle_1', \triangle_2')', \ \Omega = \{\Omega_{ij}\}, \ \hat{\Omega} = \{\hat{\Omega}_{ij}\}, \ i, j = 1, 2,$$

where $\bar{Z}_1$ and $\triangle_1$ are of order qx1 and Ω_{11} and $\hat{\Omega}_{11}$ are of order qxq. The Hotelling's T^2's(Hotelling, 1931) based upon $\bar{Z}$ and the one based on only $\bar{Z}_1$ are respectively

$$T_p^2 = N\bar{Z}'\hat{\Omega}^{-1}\bar{Z} \text{ and } T_q^2 = N\bar{Z}_1'\hat{\Omega}_{11}^{-1}\bar{Z}_1$$

with noncentrality parameters $\lambda_p = N\triangle'\Omega^{-1}\triangle$ and $\lambda_q = N\triangle_1'\Omega_{11}^{-1}\triangle_1$.

Lemma 3. Conditional on T_q^2 being fixed,

$$T_{22.1}^2 = \frac{(n-q)}{n + T_q^2} N(\bar{Z}_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\bar{Z}_1)'\hat{\Omega}_{22.1}^{-1}(\bar{Z}_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\bar{Z}_1) \tag{7}$$

is Hotelling's T^2 based on $n - q$ degrees of freedom with noncentrality parameter

$$\lambda_{22.1} = \frac{1}{1 + T_q^2/n} N(\triangle_2 - \Omega_{21}\Omega_{11}^{-1}\triangle_1)'\Omega_{22.1}^{-1}(\triangle_2 - \Omega_{21}\Omega_{11}^{-1}\triangle_1). \tag{8}$$

Proof. Under the distributions (5) and (6), it is straightforward to see, from Kaufman(1969) and Kshirsagar(1972), that when T_q^2(*i.e.* $\bar{Z}_1$ and $\hat{\Omega}_{11}$) is fixed

$$(\bar{Z}_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\bar{Z}_1) \sim N_{p-q}(\triangle_2 - \Omega_{21}\Omega_{11}^{-1}\triangle_1, \ \Omega_{22.1}(1 + T_q^2/n)/N), \tag{9}$$

and $n\hat{\Omega}_{22.1} \sim W_{p-q}(n - q, \ \Omega_{22.1})$. Moreover, the two distributions are independent(cf. Kshirsagar, 1972). Constructing Hotelling's T^2 using the two random matrices gives the results. ♣

Noticing that

$$T_p^2 = T_q^2 + N(\bar{Z}_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\bar{Z}_1)'\hat{\Omega}_{22.1}^{-1}(\bar{Z}_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\bar{Z}_1),$$

$$\lambda_p = \lambda_q + N(\triangle_2 - \Omega_{21}\Omega_{11}^{-1}\triangle_1)'\Omega_{22.1}^{-1}(\triangle_2 - \Omega_{21}\Omega_{11}^{-1}\triangle_1),$$

$T_{22.1}^2$ and $\lambda_{22.1}$, defined in Lemma 3, can be expressed in terms of T_p^2, T_q^2, λ_p and λ_q:

$$T_{22.1}^2 = \frac{(n-q)}{n+T_q^2}(T_p^2 - T_q^2), \tag{10}$$

$$\lambda_{22.1} = \frac{1}{1+T_q^2/n}(\lambda_p - \lambda_q). \tag{11}$$

Thus we have following result.

Theorem 4. When $\mathbf{M} = \mathbf{M}_1$, a statistic

$$F_2 = \frac{n-p+1}{(p-q)(n-q)}T_{22.1}^2$$

follows central F-distribution with $(p-q)$ and $(n-p+1)$ degrees of freedom, and it is independent of T_q^2.

Proof. Since $\mathbf{M} - \mathbf{M}_1 = (\lambda_p - \lambda_q)/N$, under the hypothesis, the conditional distribution in (7) becomes Hotelling's T^2 with noncentrality parameter $\lambda_{22.1} = 0$. Moreover, the distributional relation between Hotelling's T^2 and F-distribution(cf. Anderson, 1984) gives the conditional distribution of F_2, when T_q^2 is fixed, as a central F$-$ distribution with $(p-q)$ and $(n-p+1)$ degrees of freedom. However this conditional distribution does not involve T_q^2, and so F_2 is independently distributed of T_q^2. ♣

From Lemma 2 and Theorem 4, we can test the hypothesis $\mathbf{H}$: $\mathbf{a}_2 = 0$ by using the following statistic:

$$F_2 = \frac{N(Z_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}Z_1)'\hat{\Omega}_{22.1}^{-1}(Z_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\bar{Z}_1)(n-p+1)}{(n + N\bar{Z}_1'\hat{\Omega}_{11}^{-1}\bar{Z}_1)(p-q)}. \tag{12}$$

The statistic F_2 is distributed as the central F$-$ distribution with $(p-q)$ and $(n-p+1)$ degrees of freedom when $\mathbf{H}$ is true. Thus the hypothesis is accepted when

$$F_2 \le F_\alpha,$$

where $Pr(F_2 \le F_\alpha \mid \mathbf{H}) = 1-\alpha$.

Corollary 1. (1 - α) simultaneous confidence intervals associated with the above procedure are

$$\begin{aligned}&|\,\gamma'(Z_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}Z_1 - \Delta_2 + \Omega_{21}\Omega_{11}^{-1}\Delta_1)\,| \;\le \\ &\quad \{F_\alpha\, \gamma'\hat{\Omega}_{22.1}\gamma(p-q)(n + N\bar{Z}_1'\hat{\Omega}_{11}^{-1}\bar{Z}_1)/(N(n-p+1))\}^{1/2},\end{aligned} \tag{13}$$

for all nonnull γ.

Proof. Let C = $\bar{Z}_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\bar{Z}_1$ and D = $\Delta_2 - \Omega_{21}\Omega_{11}^{-1}\Delta_1$, then from the distribution (9) and the distribution $n\hat{\Omega}_{22.1} \sim W_{p-q}(n-q,\ \Omega_{22.1})$, we have (1 - α) confidence region for D:

$$(C-D)'\hat{\Omega}_{22.1}^{-1}(C-D) \le F_\alpha \frac{(p-q)(n+T_q^2)}{N(n-p+1)}, \tag{14}$$

where F_α indicates $(1-\alpha)$-th quantile of F-distribution with $(p-q)$ and $(n-p+1)$ degrees of freedom. In addition to the inequality, a generalization of the Cauchy-Schwarz inequality(cf. Anderson, 1984) gives, for all nonnull γ,

$$|\,\gamma'(C-D)\,| \;\le \{\gamma'\hat{\Omega}_{22.1}\gamma\ (C-D)'\hat{\Omega}_{22.1}^{-1}(C-D)\}^{1/2}. \tag{15}$$

Using this inequality to the equation (14) and noticing that $T_q^2 = N\bar{Z}_1'\hat{\Omega}_{11}^{-1}\bar{Z}_1$, we have the result. ♣

It should be observed from (1) that if we had known $\Sigma_1 = \Sigma_2$, we would have used Rao's F-statistic with $(p-q)$ and $(2N-p-1)$ degrees of freedom; thus we have lost $N-1$ degrees of freedom in constructing the test and the simultaneous confidence intervals which are independent of the two covariance matrices. On the other hand, when we know $\Sigma_1 \neq \Sigma_2$, we may need a test statistic which takes into account the heterogeneity of the two covariance matrices, and it would be better to use the test statistic rather than the one independent of the two covariance matrices. The followings derive a test statistic for this consideration.

3.2. An Approximate Procedure

Suppose $\bar{X}(1)$ and $\bar{X}(2)$ are the sample mean vectors and S_1 and S_2 the sample covariance matrices of random sample of equal size $N_1 = N_2 = N$ respectively from two independent multivariate normal populations $\Pi_i \sim N_p(\mu_i, \Sigma_i)$, $i = 1, 2$, where $\Sigma_1 \neq \Sigma_2$. Then

$$\bar{Y} = \bar{X}(1) - \bar{X}(2) \sim N_p(\Delta, \frac{1}{N}\Omega), \tag{16}$$

where $\Omega = \Sigma_1 + \Sigma_2$, and $\Delta = \mu_1 - \mu_2$, independently of $\hat{\Omega}^* = (S_1 + S_2)$ which is a sum of Wishart distributions with different scale matrix. Using the second moment approximation by Nel and van del Merwe(1986), the sum of different scale matrix Wishart distributions have following approximate distribution:

$$f\hat{\Omega}^* \sim W_p(f, \Omega), \tag{17}$$

that is the random matrix $f\hat{\Omega}^*$ is approximately distributed as Wishart distribution with its scale matrix Ω and f degrees of freedom, where

$$f = (N-1)\{tr[\Omega^2] + tr^2[\Omega]\}/\{tr[\Sigma_1^2 + \Sigma_2^2] + tr^2[\Sigma_1] + tr^2[\Sigma_2]\}. \tag{18}$$

In practice we will replace Σ_i by S_i in (18). Let $\bar{Y}, \Delta, \Omega$, and $\hat{\Omega}^*$ are partitioned as

$$\bar{Y} = (\bar{Y}_1', \bar{Y}_2')', \ \Delta = (\Delta_1', \Delta_2')', \ \Omega = \{\Omega_{ij}\}, \ \hat{\Omega}^* = \{\hat{\Omega}_{ij}^*\}, \ i, j = 1, 2,$$

where $\bar{Y}_1$ and Δ_1 are of order qx1 and Ω_{11} and $\hat{\Omega}_{11}^*$ are of order qxq. The approximation of Hotelling's T^2 statistics based upon $\bar{Y}$ and the one based on only $\bar{Y}_1$ are respectively denoted by

$$U_p^2 = N\bar{Y}'\hat{\Omega}^{*-1}\bar{Y} \ \text{ and } \ U_q^2 = N\bar{Y}_1'\hat{\Omega}_{11}^{*-1}\bar{Y}_1$$

with noncentrality parameters $\lambda_p^* = N\Delta'\Omega^{-1}\Delta$ and $\lambda_q^* = N\Delta_1'\Omega_{11}^{-1}\Delta_1$.

Theorem 5. Under the hypothesis **H**: $\mathbf{a}_2 = 0$ the statistic:

$$F_3 = \frac{N(\bar{Y}_2 - \hat{\Omega}_{21}^*\hat{\Omega}_{11}^{*-1}\bar{Y}_1)'\hat{\Omega}_{22.1}^{*-1}(\bar{Y}_2 - \hat{\Omega}_{21}^*\hat{\Omega}_{11}^{*-1}\bar{Y}_1)(f-p+1)}{(f + N\bar{Y}_1'\hat{\Omega}_{11}^{*-1}\bar{Y}_1)(p-q)}. \tag{19}$$

is approximately distributed as the central F- distribution with $(p-q)$ and $(f-p+1)$ degrees of freedom.

Proof. Using the similar proof in Lemma 3, conditional on U_q^2 being fixed, we see that

$$U_{22.1}^2 = \frac{(f-q)}{f+U_q^2} N(\bar{Y}_2 - \hat{\Omega}_{21}^*\hat{\Omega}_{11}^{*-1}\bar{Y}_1)'\hat{\Omega}_{22.1}^{*-1}(\bar{Y}_2 - \hat{\Omega}_{21}^*\hat{\Omega}_{11}^{*-1}\bar{Y}_1) \tag{20}$$

is an approximate Hotelling's T^2 based on $f-q$ degrees of freedom with noncentrality parameter

$$\lambda_{22.1}^* = \frac{1}{1+U_q^2/f} N(\Delta_2 - \Omega_{21}\Omega_{11}^{-1}\Delta_1)'\Omega_{22.1}^{-1}(\Delta_2 - \Omega_{21}\Omega_{11}^{-1}\Delta_1). \tag{21}$$

Using the relations

$$U_p^2 = U_q^2 + N(\bar{Y}_2 - \hat{\Omega}_{21}^*\hat{\Omega}_{11}^{*-1}\bar{Y}_1)'\hat{\Omega}_{22.1}^{*-1}(\bar{Y}_2 - \hat{\Omega}_{21}^*\hat{\Omega}_{11}^{*-1}\bar{Y}_1),$$
$$\lambda_p^* = \lambda_q^* + N(\Delta_2 - \Omega_{21}\Omega_{11}^{-1}\Delta_1)'\Omega_{22.1}^{-1}(\Delta_2 - \Omega_{21}\Omega_{11}^{-1}\Delta_1),$$

and

$$U_{22.1}^2 = \frac{(f-q)}{f+U_q^2}(U_p^2 - U_q^2), \tag{22}$$

$$\lambda_{22.1}^* = \frac{1}{1+U_q^2/f}(\lambda_p^* - \lambda_q^*), \tag{23}$$

we have, when $\mathbf{a_2} = 0$ which is equivalent to $\mathbf{M} = \mathbf{M}_1$, a statistic

$$F_3 = \frac{f-p+1}{(p-q)(f-q)} U_{22.1}^2$$

follows approximate F-distribution with $(p-q)$ and $(f-p+1)$ degrees of freedom, and it is independent of U_q^2. ♣

Thus the hypothesis is accepted when $F_3 \leq F_\alpha$, where $Pr(F_3 \leq F_\alpha \mid \mathbf{H}) = 1-\alpha$. It may be noted that $(f-p+1)$ need not be integer. But this does not cause any difficulty in consulting a table of significant values of the test statistic. Moreover, from the similar proof of Corollary 1, approximate $(1-\alpha)$ simultaneous confidence intervals associated with the above approximate procedure are

$$\begin{aligned} &\mid \gamma'(\bar{Y}_2 - \hat{\Omega}_{21}^*\hat{\Omega}_{11}^{*-1}\bar{Y}_1 - \Delta_2 + \Omega_{21}\Omega_{11}^{-1}\Delta_1) \mid \leq \\ &\{F_\alpha\ \gamma'\hat{\Omega}_{22.1}^*\gamma(p-q)(f + N\bar{Y}_1'\hat{\Omega}_{11}^{*-1}\bar{Y}_1)/(N(f-p+1))\}^{1/2}, \end{aligned} \tag{24}$$

for all nonnull γ.

Noticing that for any symmetric positive definite matrices C and D, $tr[(C+D)^2] > tr[C^2 + D^2]$ and $tr^2[C+D] > tr^2[C] + tr^2[D]$, the degrees of freedom f in (18) always larger than n. Thus, comparing to the exact procedure, the approximate procedure can save the degrees of freedom for testing the hypothesis. As indicated before to calculate the degrees of freedom in (18) we would replace Σ_i by S_i $i = 1, 2$. Thus exact power function of the test based upon F_3 is not available. Such that power comparison of the tests based upon F_2 and F_3 needs a simulation study.

4. PERFORMANCE OF TESTS ON SIMULATED DATA

The statistics F_2 and F_3 are derived to test the null hypothesis($\mathbf{H}$) that discriminant variables $x_{q+1}, \ldots, x_p$ contribute no discriminatory power in heterogeneous discriminant analysis. However, as in the stepwise procedure, an important application of these tests occurs with $q = p-1$, that is when we are testing the importance of one discriminant variable once all the other discriminant variables have been taken into account. For this reason, we confine our simulation study to the case where

$q = p-1$, that is testing whether one discriminant variable may be dropped without affecting the overall discriminatory power.

For tabulating the powers of the tests for **H** based on F_1, F_2 and F_3, the following procedure was adopted. To put the problem into canonical form, we made a transformation(see, Gilbert, 1969) so that $\mu_1 = 0$, $\mu_2 = \nu$, $\Sigma_1 = I_p$, and $\Sigma_2 = \Lambda$, a diagonal matrix. The parameters involved are the p elements of ν; the p elements of Λ; N and p, an uncomfortable total of $2p+2$ parameters. Thus we simplified the parameter structure by setting $\Lambda = \text{diag}\{\lambda, \ldots, \lambda, \lambda+1\}$, and i-th element of ν as $\nu_i = (-1)^{i+1}2$, $i = 1, \ldots, p-1$. The last component ν_p is adjusted according to various quantities of $\xi = \Delta_2 - \Omega_{21}\Omega_{11}^{-1}\Delta_1$(see, Lemma 1). For each set of values N, p, μ_1, μ_2, Σ_1 and Σ_2, 200 different pairs of samples were generated by SAS/IML. Then, using F_1, F_2, and F_3, we tested the null hypothesis **H** : $a_p = 0(i.e.\ \xi = 0)$ which states that the p-th discriminant variable x_p do not make any contribution for the discrimination. Table 1 gives the empirical significance level for each test. All tests were conducted at $\alpha = .05$ level of significance.

Table 1. Empirical Significance Level of The Tests Based Upon F_1, F_2 and F_3 with $\xi = 0$

		N = 10			N = 20			N = 50		
p	λ	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
3	2	.095	.050	.030	.090	.055	.025	.070	.045	.025
	5	.105	.045	.040	.110	.060	.045	.075	.045	.040
	10	.120	.055	.045	.130	.065	.050	.075	.050	.050
5	2	.096	.050	.005	.070	.060	.010	.090	.055	.025
	5	.145	.060	.025	.115	.050	.035	.110	.065	.045
	10	.170	.055	.025	.120	.055	.050	.115	.050	.045
7	2	.115	.085	.005	.060	.040	.010	.070	.055	.015
	5	.175	.095	.020	.110	.050	.015	.120	.055	.020
	10	.330	.185	.030	.125	.035	.015	.110	.050	.030

The following are noted from Table 1: (i) Rao's test based on F_1 has empirical significance level which, in each case, larger than .05, and hence it is not adequate for the variables selection criterion for the LDF in heteroscedastic two group discrimination case. (ii) For the exact test based upon F_2, empirical significance level is around .05, but for the cases with N = 10 and p = 7, even considering for the sampling variations, the empirical significance levels are considerably larger than .05. Thus we can conjecture that the large values are due to inadequate pairing of the independent samples which reduces sizable information obtained by data. (iii) In each case, empirical significance level for the test based on F_3 is less than that based on F_2. Moreover it is less than the significance level $\alpha = .05$. Table 2 is given for comparing the powers of tests based on F_2 and F_3 for $p = 3$.

Table 2. Empirical Powers Of Tests Based Upon F_2 and F_3 For Testing $\mathbf{H} : a_p = 0$ (*i.e.* $\xi = 0$) Against Simple Alternative Hypothesis at $\alpha = .05$

		Values of ξ							
N	λ	0.1	0.5	1.0	1.5	2.0	2.5	3.0	5.0
	2	.060	.070	.130	.235	.330	.440	.585	.915
		(.055)	(.070)	(.175)	(.330)	(.520)	(.715)	(.820)	(.980)
10	5	.055	.065	.120	.195	.285	.370	.510	.845
		(.060)	(.070)	(.115)	(.230)	(.365)	(.500)	(.625)	(.940)
	10	.060	.065	.095	.140	.235	.325	.400	.765
		(.055)	(.065)	(.080)	(.160)	(.265)	(.360)	(.460)	(.830)
	2	.080	.110	.210	.435	.660	.775	.885	.995
		(.075)	(.160)	(.360)	(.635)	(.875)	(.850)	(.965)	(1.00)
20	5	.080	.085	.220	.375	.530	.725	.825	.995
		(.085)	(.130)	(.260)	(.435)	(.695)	(.830)	(.920)	(1.00)
	10	.075	.090	.195	.305	.435	.585	.749	.985
		(.080)	(.100)	(.230)	(.320)	(.490)	(.670)	(.810)	(.990)
	2	.040	.160	.465	.825	.965	.995	1.00	1.00
		(.065)	(.215)	(.700)	(.950)	(.990)	(1.00)	(1.00)	(1.00)
50	5	.045	.115	.405	.745	.935	.980	.995	1.00
		(.045)	(.155)	(.520)	(.840)	(.965)	(.995)	(1.00)	(1.00)
	10	.040	.105	.335	.610	.868	.950	.995	1.00
		(.045)	(.125)	(.355)	(.705)	(.908)	(.955)	(.990)	(1.00)

Note: The values in parentheses denote the powers of the test based upon F_3, and the other values tabulated in the upper space are those based on F_2.

The following observations may be made from Table 2: (i) Generally the test based on F_3 seem to have more power than that based on F_2. (ii) As deviations from the null hypothesis become larger, when the value of ξ is far apart from 0, the powers of the test based on F_3 seem to exceed those of the test based on F_2 and this phenomenon becomes more clear for moderate and large sample sizes. (iii) Finally, as the distance between two populations become larger(in other words, as the value of λ decreases), the two tests tend to have better powers.

5. CONCLUDING REMARKS

In this paper we have given two test statistics which can be used as criteria for the variables selection on Fisher's LDF constructed under two unequal covariance matrices case. It is shown that the exact and an approximate test statistics respectively follow exact and approximate F-distribution under the null hypothesis that some randomly selected coefficients in the LDF are zero. The approximate test statistic may be considered as a generalization of Rao's test statistic in a sense that, when $\Sigma_1 = \Sigma_2$, the proposed test statistic (19) reduces to the Rao's test statistic defined in (1). Limited but informative sampling experiments show that the approximate test statistic generally gives more powerful test than that based upon the exact test statistic. Our study was confined to the case of two group discriminant analysis with equal training sample size. Thus the issues of more developments pertaining to the unequal sample size and the multiple group discriminant analysis are clearly needed and left for continuing study.

ACKNOWLEDGEMENT

The detailed and constructive comments of referees were very useful in making the paper more readable. Their efforts are greatly appreciated.

REFERENCES

(1) Anderson, T. W.(1984). *An Introduction to Multivariate Statistical Analysis.* 2nd ed., John Wiely & Sons, New York.

(2) Anderson, T. W. and Bahadur, R. R.(1962). Classification into Two Multivariate Normal Distributions with Different Covariance Matrices. *Annals of Mathematical Statistics,* Vol. 33, 420-431.

(3) Dillon, W. R. and Goldstein, M.(1984). *Multivariate Analysis: Methods and Applications.* John Wiely & Sons, New York.

(4) Gilbert, E. S.(1969). The Effect of Unequal Variance- Covariance Matrices on Fisher's Linear Discriminant Function. *Biometrics,*Vol. 25, 505-515.

(5) Hotelling, M. (1931). The Generalization of Student's Ratio. *Annals of Mathematical Statistics,* Vol. 2, 360-378.

(6) Kaufman, G. M. (1969). Conditional Prediction and Unbiasedness in Structural Equations. *Econometrics*, Vol. 37, No.1, 44-49.

(7) Kshirsagar, A. M. (1972). *Multivariate Analysis.* Marcel Dekker Inc., New York.

(8) Lachenbruch, P. A., Sneeringly, C. and Redo, L. B.(1973). Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non normality. *Communications in Statistics: Theory and Method,* Vol. 1, No. 1, 39-56.

(9) Marks, S. and Dunn, O. J.(1974). Discriminant Function When Covariance Matrices are Unequal. *Journal of the American Statistical Association,* Vol. 69, 555-559.

(10) Morrison, D. F.(1990). *Multivariate Statistical Methods.* McGraw-Hill Publishing Company, New York.

(11) Nel, D. G. and van del Merwe, C. A.(1986). A Solution to the Multivariate Beheren-Fisher Problem. *Communications in Statistics: Theory and Method,* Vol. 15, No. 12, 3719-3735.

(12) Fatti, L. P., Hawkins D. M. and Raath, E. L.(1982). Discriminant Analysis. In: Hawkins D. M. ed., *Topics in Applied Multivariate Analysis,* 5-65, Cambridge University Press.

(13) Rao, C. R.(1952). *Advanced Statistical Methods in Biometric Research.* John Wiley & Sons, New York.