Ann. Inst. Statist. Math.
Vol. 42, No. 4, 769–782 (1990)

A STUDY OF REDUNDANCY OF SOME VARIABLES IN COVARIATE DISCRIMINANT ANALYSIS

YASUNORI FUJIKOSHI[1] AND CHINUBAL G. KHATRI[2]

[1]*Department of Mathematics, Hiroshima University, Hiroshima 730, Japan*
[2]*Center for Multivariate Analysis, Pennsylvania State University, University Park, PA 16802, U.S.A.*

(Received March 1, 1989; revised August 15, 1989)

Abstract. The redundancy of some variables in discriminant analysis and its tests were developed by Rao (1946, *Sankhyā*, **7**, 407–414; 1948, *Biometrika*, **35**, 58–79; 1970, *Essays in Probability and Statistics*, (eds. R. C. Bose *et al.*), 587–602, Univ. of North Carolina Press, Chapel Hill), and were further studied by McKay (1977, *J. Roy. Statist. Soc. Ser. B*, **39**, 371–380) and Fujikoshi (1982, *Ann. Inst. Statist. Math.*, **34**, 523–530). These are now extended to the most general situation which includes redundancy in covariate as well as main variables in discrimination between two or more groups. The likelihood ratio test is derived in a closed form. An alternative test is also suggested.

Key words and phrases: Covariate and discriminate variables, likelihood ratio test, redundancy of variables, test of redundancy, Wilks lambda distribution.

1. Introduction

Rao (1946, 1948, 1970) discussed the redundancy of a given set of variables, for purposes of discrimination between two groups with the same covariance matrix. Rao (1948, 1970), McKay (1977) and Fujikoshi (1982) have extended this idea to the case of multiple groups. Such a question arises in the signal detection theory due to the inclusion of more variables than necessary, and the techniques developed here will be useful in this area as well.

In this paper we consider the redundancy of a given set of variables in covariate discriminant analysis, i.e., in the situation where there are covariate variables as well as discriminate variables. Two-group covariate discriminant analysis was considered by Cochran and Bliss (1948) and Cochran (1964). They proposed using the classification statistic W^*. Memon and Okamoto (1970) obtained an asymptotic expansion of the distribution of W^* and discussed the information gained by assuming that some variables have equal means. Fujikoshi and Kanazawa (1976) considered the ML rule in the covariate case and introduced the classification statistic Z^*. For a summary, see, e.g., Siotani (1982). In Section 2 we give several equivalent statements for the redundancy in the case of two groups. These statements are extended to the case of multiple groups in Section 3. In Section 4 we consider the tests for redundancy. We give a canonical form for the testing problem. It is shown that the LR test can be decomposed as two independent LR tests; some lemmas useful in obtaining the LR test are given in the Appendix. An alternative test is also suggested.

2. The case of two groups

Cochran and Bliss (1948) and Cochran (1964) discussed the classification problem when some variables are known to have the same means in two groups Π_1 and Π_2. Let $\boldsymbol{x}$ and $\boldsymbol{z}$ be the vector variates of p discriminators and q covariates, respectively. It is assumed that

$$
\begin{aligned}
&E\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix}\middle| \Pi_i\right) = \begin{pmatrix} \boldsymbol{\mu}^{(i)} \\ \xi \end{pmatrix}, \\
&\mathrm{Var}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix}\middle| \Pi_i\right) = \Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix}.
\end{aligned} \tag{2.1}
$$

Letting $\Sigma_{xx\cdot z} = \Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$,

$$
\boldsymbol{y} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix}, \quad \boldsymbol{\eta}^{(i)} = \begin{pmatrix} \boldsymbol{\mu}^{(i)} \\ \xi \end{pmatrix} \quad \text{and} \quad \boldsymbol{\delta} = \boldsymbol{\eta}^{(1)} - \boldsymbol{\eta}^{(2)} = \begin{pmatrix} \boldsymbol{\delta}_x \\ 0 \end{pmatrix},
$$

we can write the population linear discriminant function as

$$
\begin{aligned}
&(\boldsymbol{\eta}^{(1)} - \boldsymbol{\eta}^{(2)})'\Sigma^{-1}\left\{\boldsymbol{y} - \frac{1}{2}(\boldsymbol{\eta}^{(1)} + \boldsymbol{\eta}^{(2)})\right\} \\
&\qquad = \boldsymbol{\delta}'\Sigma_{xx\cdot z}^{-1}\left\{\boldsymbol{x} - \Sigma_{xz}\Sigma_{zz}^{-1}(\boldsymbol{z} - \xi) - \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})\right\}.
\end{aligned}
$$

The last expression shows that the discriminant function is one of the random variates $\boldsymbol{x}^* = \boldsymbol{x} - \Sigma_{xz}\Sigma_{zz}^{-1}\boldsymbol{z}$ obtained by subtracting from $\boldsymbol{x}$ its regression on $\boldsymbol{z}$. The vector of coefficients in $\boldsymbol{y}$ is

$$
\begin{aligned}
\gamma &= \Sigma^{-1}(\boldsymbol{\eta}^{(1)} - \boldsymbol{\eta}^{(2)}) \\
&= \begin{pmatrix} I \\ -\Sigma_{zz}^{-1}\Sigma_{zx} \end{pmatrix}\Sigma_{xx\cdot z}^{-1}\boldsymbol{\delta}_x .
\end{aligned} \tag{2.2}
$$

In order to formulate the redundancy of subvectors of $\boldsymbol{x}$ and $\boldsymbol{z}$, we partition $\boldsymbol{x}$ and $\boldsymbol{z}$ as

$$\boldsymbol{x}' = (\boldsymbol{x}_1', \boldsymbol{x}_2'), \quad \boldsymbol{x}_i\colon p_i \times 1, \qquad \boldsymbol{z}' = (\boldsymbol{z}_1', \boldsymbol{z}_2'), \quad \boldsymbol{z}_i\colon q_i \times 1$$

and $\boldsymbol{y}$, $\boldsymbol{\gamma}$, Σ conformably as follows:

$$\text{(2.3)} \qquad \boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \\ \boldsymbol{y}_4 \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \\ \boldsymbol{\gamma}_3 \\ \boldsymbol{\gamma}_4 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} \end{pmatrix}.$$

The Mahalanobis distance between Π_1 and Π_2 based on $\boldsymbol{y}$ is $\Delta_{p+q} = (\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta})^{1/2}$, while the one based on $\boldsymbol{y}_a' = (\boldsymbol{y}_1', \boldsymbol{y}_3')$ is $\Delta_{p_1+q_1}^2 = \boldsymbol{\delta}_a'\Sigma_{aa}^{-1}\boldsymbol{\delta}_a$, where $\boldsymbol{y}_b' = (\boldsymbol{y}_2', \boldsymbol{y}_4')$,

$$\text{(2.4)} \qquad E\begin{pmatrix} \boldsymbol{y}_a \\ \boldsymbol{y}_b \end{pmatrix} = \begin{pmatrix} \boldsymbol{\delta}_a \\ \boldsymbol{\delta}_b \end{pmatrix}, \quad \mathrm{Var}\begin{pmatrix} \boldsymbol{y}_a \\ \boldsymbol{y}_b \end{pmatrix} = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

We note that $\boldsymbol{y}_b$ involves both the subvectors of $\boldsymbol{x}$ and $\boldsymbol{z}$. We consider the following five statements on the redundancy of $\boldsymbol{y}_b$, which are shown to be equivalent.

(1) The random variate $\boldsymbol{y}_b - \Sigma_{ba}\Sigma_{aa}^{-1}\boldsymbol{y}_a$ obtained by subtracting from $\boldsymbol{y}_b$ its regression on $\boldsymbol{y}_a$ has the same expected value for both groups.

(2) The Mahalanobis distances between Π_1 and Π_2 based on $\boldsymbol{y}$ and $\boldsymbol{y}_a$ are the same, i.e.,

$$\text{(2.5)} \qquad \Delta_{p+q} = \Delta_{p_1+q_1}.$$

(3) The coefficients of $\boldsymbol{y}_2$ and $\boldsymbol{y}_4$ in the linear discriminant function based on $\boldsymbol{y}$ are all zero, i.e., $\boldsymbol{\gamma}_2 = 0$, $\boldsymbol{\gamma}_4 = 0$.

(4) The Mahalanobis distances between Π_1 and Π_2 based on the random variates $\boldsymbol{y}_{x\cdot z} = \boldsymbol{x} - \Sigma_{xz}\Sigma_{zz}^{-1}\boldsymbol{z}$ and $\boldsymbol{y}_{1\cdot 3} = \boldsymbol{x}_1 - \Sigma_{13}\Sigma_{33}^{-1}\boldsymbol{z}_1$ are the same, i.e.,

$$\text{(2.6)} \qquad \boldsymbol{\delta}_x'\Sigma_{xx\cdot z}^{-1}\boldsymbol{\delta}_x = \boldsymbol{\delta}_1'\Sigma_{11\cdot 3}^{-1}\boldsymbol{\delta}_1,$$

where $\Sigma_{xx\cdot z} = \Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$ and $\Sigma_{11\cdot 3} = \Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31}$.

(5)

$$\text{(2.7)} \qquad \boldsymbol{\delta}_2 - \Sigma_{21\cdot z}\Sigma_{11\cdot z}^{-1}\boldsymbol{\delta}_1 = 0 \quad \text{and} \quad \Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\boldsymbol{\delta}_1 = 0.$$

The statements (1), (2), (3) in the case when the covariate is not available were given by Rao (1970). His proof for the equivalence of (1)–(3) is applicable to our case. The equivalence (1)–(5) will be shown as a special case of multiple groups.

3. The case of multiple groups

We consider the formulation for redundancy in the case of $k+1$ groups. Suppose that $\boldsymbol{y}' = (\boldsymbol{x}', \boldsymbol{z}')$ has the means

$$E(\boldsymbol{y} \mid \Pi_i) = \boldsymbol{\eta}^{(i)} = \begin{pmatrix} \boldsymbol{\mu}^{(i)} \\ \xi \end{pmatrix}, \qquad i = 1, \ldots, k+1 \tag{3.1}$$

and the same covariance matrix Σ as in (2.1). Following McKay (1977) and Fujikoshi (1982), we extend the statements (1)–(5) to the case of multiple groups. The statement (1) can be extended as

$$\boldsymbol{\eta}_b^{(1)} - \Sigma_{ba}\Sigma_{aa}^{-1}\boldsymbol{\eta}_a^{(1)} = \cdots = \boldsymbol{\eta}_b^{(k+1)} - \Sigma_{ba}\Sigma_{aa}^{-1}\boldsymbol{\eta}_a^{(k+1)} . \tag{3.2}$$

Let Ω be the population between-groups covariance matrix defined by

$$\begin{aligned} \Omega &= \Theta G \Theta' \\ &= \sum_{i=1}^{k+1} g_i(\boldsymbol{\eta}^{(i)} - \bar{\boldsymbol{\eta}})(\boldsymbol{\eta}^{(i)} - \bar{\boldsymbol{\eta}})' = \begin{pmatrix} \Omega_{xx} & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned} \tag{3.3}$$

where g_i's are positive constants such that $\Sigma g_i = 1$, $\bar{\boldsymbol{\eta}} = \Sigma g_i \boldsymbol{\eta}^{(i)}$, $\Theta = (\boldsymbol{\eta}^{(i)} - \bar{\boldsymbol{\eta}}, \ldots, \boldsymbol{\eta}^{(k+1)} - \bar{\boldsymbol{\eta}})$ and $G = \operatorname{diag}(g_1, \ldots, g_{k+1})$. A choice of $\{g_i\}$ may be done, based on the sample sizes from Π_i. The population Fisher's discriminant functions are defined by using the characteristic vectors of Ω with respect to Σ. Let γ_j be the solutions of

$$\Omega\gamma_j = l_j \Sigma \gamma_j, \quad \gamma_j'\Sigma\gamma_j = 1, \quad j = 1, \ldots, m , \tag{3.4}$$

where $m = \operatorname{rank} \Omega \le \min(p, k)$. Let Ω and $\gamma_j' = (\gamma_{1j}', \gamma_{2j}', \gamma_{3j}', \gamma_{4j}')$ partition as in (2.3) and (2.4). Similarly we partion Θ as

$$\Theta = \begin{pmatrix} \Theta_1 \\ \Theta_2 \\ \Theta_3 \\ \Theta_4 \end{pmatrix}, \quad \Theta_2 = \begin{pmatrix} \Theta_1 \\ \Theta_3 \end{pmatrix}, \quad \Theta_b = \begin{pmatrix} \Theta_2 \\ \Theta_4 \end{pmatrix}. \tag{3.5}$$

The statements (2)–(5) can be extended as follows:

$$\text{(3.6)} \qquad \operatorname{tr} \Omega\Sigma^{-1} = \operatorname{tr} \Omega_{aa}\Sigma_{aa}^{-1} ,$$

$$\text{(3.7)} \qquad \gamma_{2j} = 0, \quad \gamma_{4j} = 0, \quad j = 1,\ldots, m ,$$

$$\text{(3.8)} \qquad \operatorname{tr} \Omega_{xx}\Sigma_{xx\cdot z}^{-1} = \operatorname{tr} \Omega_{11}\Sigma_{11\cdot 3}^{-1} ,$$

$$\text{(3.9)} \qquad \Theta_2 - \Sigma_{21\cdot z}\Sigma_{11\cdot z}^{-1}\Theta_1 = 0 \quad \text{and} \quad \Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\Theta_1 = 0 .$$

Statements (3.6) and (3.7) in the case when the covariate is not available have been considered by McKay (1977) and Fujikoshi (1982), respectively.

THEOREM 3.1. *The statements* (3.2), (3.6)–(3.9) *are equivalent.*

PROOF. Though the equivalence of (3.2), (3.6) and (3.7) follows from the proof as in McKay (1977) and Fujikoshi (1982), we give a complete proof. Noting that

$$\text{(3.10)} \quad \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_{aa}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -\Sigma_{aa}^{-1} & \Sigma_{ab} \\ & I \end{pmatrix} \Sigma_{bb\cdot a}^{-1} \begin{pmatrix} -\Sigma_{aa}^{-1} & \Sigma_{ab} \\ & I \end{pmatrix}' ,$$

we obtain

$$\text{(3.11)} \qquad \operatorname{tr} \Omega\Sigma^{-1} = \operatorname{tr} \Omega_{aa}\Sigma_{aa}^{-1} + \operatorname{tr} \Sigma_{bb\cdot a}^{-1}(\Theta_b - \Sigma_{ba}\Sigma_{aa}^{-1}\Theta_a)G(\Theta_b - \Sigma_{ba}\Sigma_{aa}^{-1}\Theta_a)' .$$

Therefore, (3.6) is equivalent to

$$\text{(3.12)} \qquad \Theta_b - \Sigma_{ba}\Sigma_{aa}^{-1}\Theta_a = 0 ,$$

which is equivalent to (3.2). We can write (3.4) as

$$\text{(3.13)} \qquad \begin{pmatrix} \Omega_{aa} & \Omega_{ab} \\ \Omega_{ba} & \Omega_{bb} \end{pmatrix} \begin{pmatrix} \gamma_{aj} \\ \gamma_{bj} \end{pmatrix} = l_j \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \begin{pmatrix} \gamma_{aj} \\ \gamma_{bj} \end{pmatrix}, \qquad j = 1,\ldots, m .$$

Suppose that (3.12) is true. Then, premultiplying both sides of (3.13) by $(-\Sigma_{ba}\Sigma_{aa}^{-1}, I)$ we have

$$0 = l_j(0, \Sigma_{bb\cdot a}) \begin{pmatrix} \gamma_{aj} \\ \gamma_{bj} \end{pmatrix} .$$

Hence, $\gamma_{bj} = 0$, $j = 1,\ldots, m$. Conversely, suppose that (3.7) is true, i.e., $\gamma_{bj} = 0$. Then from (3.13) we have

$$\Omega_{aa}\gamma_{aj} = l_j\Sigma_{aa}\gamma_{aj}, \qquad j = 1,\ldots, m$$

which implies (3.6), i.e., $\operatorname{tr}\,\Omega_{aa}\Sigma_{aa}^{-1} = \sum_{j=1}^{m} l_j = \operatorname{tr}\,\Omega\Sigma^{-1}$. The equivalence of (3.8) and (3.9) follows from

$$\begin{aligned}\operatorname{tr}\,\Omega_{xx}\Sigma_{xx\cdot z}^{-1} = \operatorname{tr}\,&\Omega_{11}\Sigma_{11\cdot 3}^{-1}\\ &+ \operatorname{tr}\,(\Sigma_{22\cdot z} - \Sigma_{21\cdot z}\Sigma_{11\cdot z}^{-1}\Sigma_{12\cdot z})^{-1}(\Theta_2 - \Sigma_{21\cdot z}\Sigma_{11\cdot z}^{-1}\Theta_1)\\ &\quad\cdot G(\Theta_2 - \Sigma_{11\cdot z}^{-1}\Sigma_{11\cdot 3}^{-1}\Theta_1)'\\ &+ \operatorname{tr}\,(\Sigma_{44\cdot 3} - \Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\Sigma_{14\cdot 3})^{-1}(\Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\Theta_1)\\ &\quad\cdot G(\Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\Theta_1)'\end{aligned}$$

which is obtained in the same way as in (3.11). Substituting

$$\Sigma_{aa}^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{33}^{-1} \end{pmatrix} + \begin{pmatrix} I \\ -\Sigma_{33}^{-1}\Sigma_{31} \end{pmatrix}\Sigma_{11\cdot 3}^{-1}\begin{pmatrix} I \\ -\Sigma_{33}^{-1}\Sigma_{31} \end{pmatrix}'$$

into (3.12), we can see that (3.12) is equivalent to

$$\Theta_2 - \Sigma_{21\cdot 3}\Sigma_{11\cdot 3}^{-1}\Theta_1 = 0 \quad \text{and} \quad \Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\Theta_1 = 0\,. \tag{3.14}$$

The equivalence of (3.12) and (3.14) is shown by using

$$\Sigma_{11\cdot z}^{-1} = \Sigma_{11\cdot 3}^{-1} + \Sigma_{11\cdot 3}^{-1}\Sigma_{14\cdot 3}(\Sigma_{44\cdot 3} - \Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\Sigma_{14\cdot 3})^{-1}\Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\,.$$

This completes the proof.

4. Tests for redundancy

4.1 *A canonical form*

We consider the problem of testing the hypothesis that $\boldsymbol{y}_b = (\boldsymbol{y}_2' \quad \boldsymbol{y}_4')'$ is redundant, based on random samples $\boldsymbol{y}_1^{(i)},\ldots,\boldsymbol{y}_{N_i}^{(i)}$ of sizes N_i from Π_i, $i = 1,\ldots,k+1$. It is assumed that the samples are multivariate normal. Let W, B and $T = W + B$ be the matrices of sums of squares and products due to within-groups, between-groups and the total variation, respectively, i.e.,

$$B = \sum_{i=1}^{k+1} N_i(\bar{\boldsymbol{y}}^{(i)} - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}^{(i)} - \bar{\boldsymbol{y}})'\,,$$

$$W = \sum_{i=1}^{k+1}\sum_{j=1}^{N_i} (\boldsymbol{y}_j^{(i)} - \bar{\boldsymbol{y}}^{(i)})(\boldsymbol{y}_j^{(i)} - \bar{\boldsymbol{y}}^{(i)})'\,,$$

where $\bar{\boldsymbol{y}}^{(i)} = (1/N_i)\sum_{j=1}^{N_i} \boldsymbol{y}_j^{(i)}$ and $\bar{\boldsymbol{y}} = (1/N)\sum_{i=1}^{k+1} N_i\bar{\boldsymbol{y}}^{(i)}$ with $N = N_1 + \cdots + N_{k+1}$. Then W and B are independently distributed as a central Wishart distribution $W_{p+q}(\Sigma, n)$ and a noncentral Wishart distribution $W_{p+q}(\Sigma, k; N\Omega)$, where $n = N - k - 1$ and Ω is defined by (3.3) with $g_i = N_i/N$. We can write

$$B = XX', \qquad W = UU',$$

where the columns of $(X \quad U)$: $(p+q) \times (k+n)$ are independently distributed as $N_{p+q}(\cdot, \Sigma)$ with

$$E(X \quad U) = (\nu \quad 0).$$

Here ν satisfies $\nu\nu' = N\Omega$. Using the same partitions for X, U and ν as in (3.5), we can write the hypothesis on the redundancy of y_b as

$$H\colon \nu_b - \Sigma_{ba}\Sigma_{aa}^{-1}\nu_a = 0. \tag{4.1}$$

Here we note that $\nu_3 = 0$ and $\nu_4 = 0$. From Theorem 3.1 the hypothesis (4.1) can be decomposed as

$$H_1\colon \nu_2 - \Sigma_{21\cdot z}\Sigma_{11\cdot z}^{-1}\nu_1 = 0 \quad \text{and} \quad H_2\colon \Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\nu_1 = 0. \tag{4.2}$$

We shall obtain the LR test for H by using a conditional approach. It is easily seen that

$$E\left[(X_2 \quad U_2) \middle| \begin{pmatrix} X_1 & U_1 \\ X_z & U_z \end{pmatrix}\right] = (\nu_2^* \quad 0) + (\Gamma_1 \quad \Gamma_2)\begin{pmatrix} X_1 & U_1 \\ X_z & U_z \end{pmatrix} \tag{4.3}$$

and

$$E\left[(X_4 \quad U_4) \middle| \begin{pmatrix} X_1 & U_1 \\ X_3 & U_3 \end{pmatrix}\right] = (\beta_1 \quad \beta_3)\begin{pmatrix} X_1 - \nu_1 & U_1 \\ X_3 & U_3 \end{pmatrix}, \tag{4.4}$$

where $\nu_2^* = \nu_2 - \Gamma_1\nu_1$,

$$(\Gamma_1 \quad \Gamma_2) = (\Sigma_{21} \quad \Sigma_{2z})\begin{pmatrix} \Sigma_{11} & \Sigma_{1z} \\ \Sigma_{z1} & \Sigma_{zz} \end{pmatrix}^{-1},$$

$$(\beta_1 \quad \beta_3) = (\Sigma_{41} \quad \Sigma_{43})\begin{pmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{pmatrix}^{-1}.$$

Noting that $\Gamma_1 = \Sigma_{21\cdot z}\Sigma_{11\cdot z}^{-1}$ and $\beta_1 = \Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}$, we can write (4.2) as

$$H_1\colon \nu_2^* = 0 \quad \text{and} \quad H_2\colon \beta_1\nu_1 = 0. \tag{4.5}$$

Therefore, the LR criterion λ for H can be decomposed as

$$\lambda = \lambda_1 \lambda_2 , \tag{4.6}$$

where

λ_1 = the LR criterion for H_1 based on the conditional density

$$f\left(X_2, U_2 \middle| \begin{pmatrix} X_1 & U_1 \\ X_z & U_z \end{pmatrix} \right)$$

and

λ_2 = the LR criterion for H_2 based on the conditional density

$$f\left(\begin{pmatrix} X_1 & U_1 \\ X_4 & U_4 \end{pmatrix} \middle| X_3, U_3 \right) = f\left(X_4, U_4 \middle| \begin{pmatrix} X_1 & U_1 \\ X_3 & U_3 \end{pmatrix} \right) f(X_1, U_1 | X_3, U_3) .$$

The explicit formulas for λ_1 and λ_2 are given in Sections 4.2 and 4.3, respectively.

4.2 *The LR Criterion λ_1*

The problem of testing H_1: $v^* = 0$ in the conditional model of $(X_2 \quad U_2)$ given $\begin{pmatrix} X_1 & U_1 \\ X_z & U_z \end{pmatrix}$ is the one of testing a linear hypothesis in a multivariate linear model. Therefore, we obtain

$$\Lambda_1 = \lambda_1^{2/N} = \frac{|W_{22\cdot 1z}|}{|T_{22\cdot 1z}|} \tag{4.7}$$

where $W_{22\cdot 1z} = W_{22} - W_{2(1z)} W_{(1z)(1z)}^{-1} W_{(1z)2}$ and $T_{22\cdot 1z} = T_{22} - T_{2(1z)} T_{(1z)(1z)}^{-1} T_{(1z)2}$. Here we use the same partitions for W and T as in (2.1) and (2.3). The conditional null distribution of Λ_1 is the Wilks lambda distribution $\Lambda_{p_2}(k, n - p_1 - q)$ which does not depend on $(X_1 \quad U_1)$ and $(X_z \quad U_z)$. Therefore, the null distribution of Λ_1 is $\Lambda_{p_2}(k, n - p_1 - q)$ and is independent of λ_2. The limiting null distribution of $-\log \Lambda_1$ is a chi-squared distribution with kp_2 degrees of freedom.

4.3 *The LR criterion λ_2*

First we consider the maximum of $f\left(\begin{pmatrix} X_1 & U_1 \\ X_4 & U_4 \end{pmatrix} \middle| X_3, U_3 \right)$ when the parameters are unrestricted. The conditional density can be written as

$$
\begin{aligned}
(4.8)\quad & f(X_1, U_1 | X_z, U_z) f(X_4, U_4 | X_3, U_3) \\
&= (2\pi)^{-(p_1+p_2)N/2} |\Sigma_{11\cdot z}|^{-N/2} \\
&\quad \times \operatorname{etr}\Big\{ -\frac{1}{2} \Sigma_{11\cdot z}^{-1} [(X_1 - \nu_1 \quad U_1) \\
&\qquad - \Sigma_{1z}\Sigma_{zz}^{-1}(X_z \quad U_z)][(X_1 - \nu_1 \quad U_1) \\
&\qquad - \Sigma_{1z}\Sigma_{zz}^{-1}(X_z \quad U_z)]' \Big\} \\
&\quad \times |\Sigma_{44\cdot 3}|^{-N/2} \operatorname{etr}\Big\{ -\frac{1}{2} \Sigma_{44\cdot 3}^{-1} [(X_4 \quad U_4) - \Sigma_{43}\Sigma_{33}^{-1}(X_3 \quad U_3)] \\
&\qquad \times [(X_4 \quad U_4) - \Sigma_{43}\Sigma_{33}^{-1}(X_3 \quad U_3)]' \Big\}.
\end{aligned}
$$

This expression implies that

$$
(4.9)\quad \max f\left(\begin{pmatrix} X_1 & U_1 \\ X_4 & U_4 \end{pmatrix} \middle| X_3, U_3 \right) = \text{const.} \times \{ |W_{11\cdot z}| \, |T_{44\cdot 3}| \}^{-N/2}.
$$

Next we consider the maximum of $f\left(\begin{pmatrix} X_1 & U_1 \\ X_4 & U_4 \end{pmatrix} \middle| X_3, U_3 \right)$ when H_2 is true. Under H_2 we can write the conditional density as

$$
\begin{aligned}
(4.10)\quad & f\left(X_4, U_4 \middle| \begin{pmatrix} X_1 & U_1 \\ X_3 & U_3 \end{pmatrix} \right) f(X_1, U_1 | X_3, U_3) \\
&= (2\pi)^{-(p_1+p_2)N/2} |\Sigma_{44\cdot 13}|^{-N/2} \\
&\quad \times \operatorname{etr}\Big\{ -\frac{1}{2} \Sigma_{44\cdot 13}^{-1} \Big[(X_4 \quad U_4) - (\beta_1 \quad \beta_3) \begin{pmatrix} X_1 & U_1 \\ X_3 & U_3 \end{pmatrix} \Big] \\
&\qquad \times \Big[(X_4 \quad U_4) - (\beta_1 \quad \beta_3) \begin{pmatrix} X_1 & U_1 \\ X_3 & U_3 \end{pmatrix} \Big]' \Big\} \\
&\quad \times |\Sigma_{11\cdot 3}|^{-N/2} \operatorname{etr}\Big\{ -\frac{1}{2} \Sigma_{11\cdot 3}^{-1} [(X_1 - \nu \quad U_1) \\
&\qquad - \Sigma_{13}\Sigma_{33}^{-1}(X_3 \quad U_3)] \\
&\qquad \times [(X_4 - \nu_1 \quad U_1) - \Sigma_{13}\Sigma_{33}^{-1}(X_3 \quad U_3)]' \Big\}.
\end{aligned}
$$

Considering the maximization of (4.10) with respect to $\Sigma_{44\cdot 13}$ and $\Sigma_{11\cdot 3}$, it is easily seen that

$$\Lambda_2 = \lambda_2^{2/N} = \frac{|W_{11\cdot z}||T_{44\cdot 3}|}{\underset{\beta_1\nu_1=0}{\operatorname{Min}} g(\beta_1, \beta_3, \nu_1, \zeta)} \tag{4.11}$$

where $\zeta = \Sigma_{13}\Sigma_{33}^{-1}$,

$$\begin{aligned} g(\beta_1, \beta_3, \nu_1, \zeta) = &|\{(X_4 \quad U_4) - \beta_1(X_1 \quad U_1) - \beta_3(X_3 \quad U_3)\} \\ &\times \{(X_4 \quad U_4) - \beta_1(X_1 \quad U_1) - \beta_3(X_3 \quad U_3)\}'| \\ &\times |\{(X_1 - \nu_1 \quad U_1) - \zeta(X_3 \quad U_3)\} \\ &\times \{(X_1 - \nu_1 \quad U_1) - \zeta(X_3 \quad U_3)\}'| \,. \end{aligned} \tag{4.12}$$

Considering the minimization of (4.12) with respect to β_3 and ζ, we have

$$\begin{aligned} &\underset{\beta_1\nu_1=0}{\operatorname{Min}} g(\beta_1, \beta_3, \nu_1, \zeta) \\ &= \underset{\beta_1\nu_1=0}{\operatorname{Min}} |T_{44\cdot 3} - \beta_1 T_{14\cdot 3} - T_{41\cdot 3}\beta_1' + \beta_1 T_{11\cdot 3}\beta_1'| \\ &\quad \times |T_{11\cdot 3} - \nu_1(X_1 - T_{13}T_{33}^{-1}X_3)' - (X_1 - T_{13}T_{33}^{-1}X_3)\nu_1' \\ &\qquad + \nu_1(I - X_3'T_{33}^{-1}X_3)\nu_1'| \,. \end{aligned} \tag{4.13}$$

The problem of minimizing (4.13) is discussed in the Appendix. Using Lemma A.3 in the Appendix and noting that $W_{11\cdot 3} = T_{11\cdot 3} - (X_1 - T_{13}T_{33}^{-1}X_3)(I - X_3'T_{33}^{-1}X_3)^{-1}(I - X_3'T_{33}^{-1}X_3)'$, we obtain

$$\underset{\beta_1\nu_1=0}{\operatorname{Min}} g(\beta_1, \beta_3, \nu_1, \zeta) = |W_{11\cdot 3}||T_{44\cdot 3}| \left(\prod_{i=p_1-s_0+1}^{p_1} l_i\right), \tag{4.14}$$

where $l_1 \geq \cdots \geq l_{p_1-s_0} > 1 \geq l_{p_1-s_0+1} \geq \cdots \geq l_{p_1} > 0$ are the eigenvalues of $T_{11\cdot 34}W_{11\cdot 3}^{-1}$. Hence

$$\Lambda_2 = \frac{|W_{11\cdot 34}||T_{44\cdot 3}|}{|W_{11\cdot 3}||T_{44\cdot 3}| \prod_{i=p_1-s_0+1}^{p_1} l_i} = \frac{\prod_{i=1}^{p_1}(1-\rho_i^2)}{\prod_{i=p_1-s_0+1}^{p_1} l_i}, \tag{4.15}$$

where $\rho_1^2 \geq \cdots \geq \rho_{p_1}^2$ are the eigenvalues of $W_{14\cdot 3}W_{44\cdot 3}^{-1}W_{41\cdot 3}W_{11\cdot 3}^{-1}$. Unfortunately, the distribution of Λ_2 appears to be extremely complicated even under the null hypothesis H_2.

For practical purposes we consider an approximate test. Starting the conditional model (4.4), we can use the test statistic

$$\Lambda_{(2)} = |W_{44\cdot 13}|/|T_{44\cdot 13}| \tag{4.16}$$

instead of (4.15). It is easily seen that the null distribution of $\Lambda_{(2)}$ is the Wilks lambda distribution $\Lambda_{q_2}(k, n - p_1 - q_1)$. This test will not be efficient because it ignores the information on the mean value of X_4 and the conditional density of $[X_1, U_1]$ given $[X_3, U_3]$. An approximate test for H is to use the statistic $-n \log \Lambda_1\Lambda_{(2)}$ whose asymptotic null distribution is a chi-squared distribution with $k(p_2 + q_2)$ degrees of freedom.

We shall rewrite (4.15) in an alternative form to point out the connection with (4.16). For this, we note that

$$
\begin{aligned}
|T_{11\cdot 34}|/|W_{11\cdot 3}| &= \prod_{i=1}^{p_1} l_i \\
&= \begin{vmatrix} T_{11\cdot 3} & T_{14\cdot 3} \\ T_{41\cdot 3} & T_{44\cdot 3} \end{vmatrix} \Big/ \{|W_{11\cdot 3}T_{44\cdot 3}|\} = \frac{|T_{11\cdot 3}||T_{44\cdot 13}|}{|T_{11\cdot 3}||T_{44\cdot 3}|} \\
&= \left[\prod_{i=1}^{p_1} (1 - \rho_i^2)\right] (|W_{44\cdot 3}|/|T_{44\cdot 3}|) \\
&\quad \times \left(\begin{vmatrix} T_{11\cdot 3} & T_{14\cdot 3} \\ T_{41\cdot 3} & T_{44\cdot 3} \end{vmatrix} \Big/ \begin{vmatrix} W_{11\cdot 3} & W_{14\cdot 3} \\ W_{41\cdot 3} & W_{44\cdot 3} \end{vmatrix} \right).
\end{aligned}
$$

Hence, we can rewrite (4.15) as

$$
\Lambda_2 = \left(\prod_{i=1}^{p_1 - s_0} l_i \right) \left(\frac{|T_{44\cdot 3}|}{|W_{44\cdot 3}|} \right) \left(\frac{|W_{11\cdot 3}|}{|T_{11\cdot 3}|} \right) \Lambda_{(2)} = \zeta \Lambda_{(2)}, \qquad \text{say}\,.
$$

Note that the quantity ζ depends on (X_4, U_4) and hence ζ and $\Lambda_{(2)}$ are dependent variables. It is not clear whether the distribution of ζ will depend on the unknown parameters under H_2 or not. For this, we require simulation study which will be done at a later stage.

Acknowledgements

The authors would like to express their deep thanks to Professor C. R. Rao, Pennsylvania State University for his encouragement and helpful discussion during this investigation. Thanks are also due to the referees for their useful comments.

Appendix

We consider the minimization problem which is related to (4.13). Let Q and T be the positive definite matrices of orders $m + p_1$ and $p_1 + p_2$, respectively, and let

$$Q=\begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \qquad Q_{11}\colon m\times m, \qquad T=\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}, \qquad T_{11}\colon p_1\times p_1 .$$

LEMMA A.1. *Let β and v be $p_2\times p_1$ and $p_1\times m$ matrices such that $\beta v=0$. The minimum value of $|Q_{22}-vQ_{12}-Q_{21}v'+vQ_{11}v'|$ over v subject to $\beta v=0$ for a given β is*

$$|Q_{22\cdot 1}|\,|\beta_0 Q_{22}\beta_0'|/|\beta_0 Q_{22\cdot 1}\beta_0'| ,$$

where $\beta=\beta_1\beta_0$, β_1 and β_0 are $p_2\times s$ and $s\times p_1$ matrices of ranks s, $s=\operatorname{rank}\beta$, and $Q_{22\cdot 1}=Q_{22}-Q_{21}Q_{11}^{-1}Q_{12}$.

PROOF. Note that $\beta v=0 \Leftrightarrow \beta_0 v=0$. Using Lagrangian multipliers, we have to minimize

$$\phi=\log|Q_{22}-vQ_{12}-Q_{21}v'+vQ_{11}v'|+2\operatorname{tr}\Lambda\beta_0 v .$$

Differentiating ϕ with respect to v and Λ, we get

$$\begin{aligned}&(Q_{22}-vQ_{12}-Q_{21}v'+vQ_{11}v')^{-1}(Q_{21}-vQ_{11})=\beta_0'\Lambda' ,\\ &\beta_0 v=0\end{aligned}$$

which are equivalent to

$$Q_{21}-vQ_{11}=(Q_{22}-vQ_{12})\beta_0'\Lambda', \qquad \beta_0 v=0 . \tag{A.1}$$

Premultiplying both sides of the first equation of (A.1) by β_0, we have $\Lambda'=(\beta_0 Q_{22}\beta_0')^{-1}\beta_0 Q_{21}$. Hence (A.1) gives

$$\begin{aligned} v&=[I-Q_{22}\beta_0'(\beta_0 Q_{22}\beta_0')^{-1}\beta_0]Q_{21}[Q_{11}-Q_{12}\beta_0'(\beta_0 Q_{22}\beta_1')^{-1}\beta_0 Q_{21}]^{-1}\\ &=[I-Q_{22\cdot 1}\beta_0'(\beta_0 Q_{22\cdot 1}\beta_0')^{-1}\beta_0]Q_{21}Q_{11}^{-1}\end{aligned}$$

because $Q_{21}[Q_{11}-Q_{12}\beta_0'(\beta_0 Q_{22}\beta_0')^{-1}\beta_0 Q_{21}]^{-1}=Q_{21}Q_{11}^{-1}+Q_{21}Q_{11}^{-1}Q_{12}\beta_0'\cdot(\beta_0 Q_{22\cdot 1}\beta_0')^{-1}\beta_0 Q_{21}Q_{11}^{-1}$. Substituting this into $|Q_{22}-vQ_{12}-Q_{21}v'+vQ_{11}v'|$, we have

$$\begin{aligned} &\operatorname*{Min}_{\beta_0 v=0}|Q_{22}-vQ_{12}Q_{21}v'+vQ_{11}v'|\\ &\quad=|Q_{22\cdot 1}+Q_{22\cdot 1}\beta_0'(\beta_0 Q_{22\cdot 1}\beta_0')^{-1}\beta_0(Q_{21}Q_{11}^{-1}Q_{12})\\ &\qquad\qquad\times\beta_0'(\beta_0 Q_{22\cdot 1}\beta_0')^{-1}\beta_0 Q_{22\cdot 1}|\\ &\quad=|Q_{22\cdot 1}|\,|I+(\beta_0 Q_{22\cdot 1}\beta_0')^{-1}\beta_0 Q_{21}Q_{11}^{-1}Q_{12}\beta_0'|\end{aligned}$$

$$= |Q_{22\cdot 1}||\beta_0 Q_{22}\beta_0'| / |\beta_0 Q_{22\cdot 1}\beta_0'|$$

which proves Lemma A.1.

LEMMA A.2. *Let $\beta = \beta_1\beta_0$, β_1 and β_0 be matrices of orders $p_2 \times p_1$, $p_2 \times s$, $s \times p_1$ and of ranks s. Then, given β_0, the minimum of $|T_{22} - \beta T_{12} - T_{21}\beta' + \beta T_{11}\beta'|$ is $|T_{22} - T_{21}\beta_0'(\beta_0 T_{11}\beta_0')^{-1}\beta_0 T_{12}| = |T_{22}||\beta_0 T_{11\cdot 2}\beta_0'| / |\beta_0 T_{11}\beta_0'|$ with $T_{11\cdot 2} = T_{11} - T_{12}T_{22}^{-1}T_{21}$.*

PROOF. The result follows from the identity

$$\begin{aligned} &T_{22} - \beta T_{12} - T_{21}\beta' + \beta T_{11}\beta' \\ &\quad = \{T_{21}\beta_0'(\beta_0 T_{11}\beta_0')^{-1} - \beta_1\}(\beta_0 T_{11}\beta_0')\{T_{21}\beta_0'(\beta_0 T_{11}\beta_0')^{-1} - \beta_1\}' \\ &\qquad + T_{22} - T_{21}\beta_0'(\beta_0 T_{11}\beta_0')^{-1}\beta_0 T_{12}\,. \end{aligned}$$

LEMMA A.3. *Suppose that $Q_{22} = T_{11}$. Then,*

$$\text{(A.2)} \quad \begin{aligned} &\operatorname*{Min}_{\beta\nu=0} |T_{22} - \beta T_{12} - T_{21}\beta' + \beta T_{11}\beta'||Q_{22} - \nu Q_{12} - Q_{21}\nu' + \nu Q_{11}\nu| \\ &\quad = |T_{22}||Q_{22\cdot 1}| \prod_{i=p_1-s_0+1}^{p_1} l_i\,, \end{aligned}$$

where $l_1 \geq \cdots \geq l_{p_1-s_0} > 1 \geq l_{p_1-s_0+1} \geq \cdots \geq l_{p_1} > 0$ are the eigenvalues of $T_{11\cdot 2}Q_{22\cdot 1}^{-1}$.

PROOF. Let $s = \operatorname{rank}\beta$. Then, using Lemmas A.1 and A.2, we can write the left-hand side of (A.2) as

$$\begin{aligned} &\operatorname*{Min}_{s,\beta_0} \frac{|T_{22}||\beta_0 T_{11\cdot 2}\beta_0'|}{|\beta_0 T_{11}\beta_0'|} \cdot \frac{|Q_{22\cdot 1}||\beta_0 Q_{22}\beta_0'|}{|\beta_0 Q_{22\cdot 1}\beta_0'|} \\ &\quad = \operatorname*{Min}_{s} |T_{22}||Q_{22\cdot 1}| \prod_{i=p_1-s+1}^{p_1} l_i \end{aligned}$$

which is equal to the right-hand side of (A.2).

REFERENCES

Cochran, W. G. (1964). Comparison of two methods of handling covariates in discriminatory analysis, *Ann. Inst. Statist. Math.*, **16**, 43–53.

Cochran, W. G. and Bliss, C. I. (1948). Discriminant functions with covariance, *Ann. Math. Statist.*, **19**, 151–176.

Fujikoshi, Y. (1982). A test for additional information in canonical correlation analysis, *Ann. Inst. Statist. Math.*, **34**, 523–530.

Fujikoshi, Y. and Kanazawa, M. (1976). The ML classification statistic in covariate discriminant analysis and its asymptotic expansions, *Essays in Probability and Statistics*, 305–320, Shinko-Tsusho, Tokyo.

McKay, R. J. (1977). Variable selection in multivariate regression: An application of simultaneous test procedures, *J. Roy. Statist. Soc. Ser. B*, **39**, 371–380.

Memon, A. Z. and Okamoto, M. (1970). The classification statistic W^* in covariate discriminant function, *Ann. Math. Statist.*, **41**, 1491–1499.

Rao, C. R. (1946). Tests with discriminant functions in multivariate analysis, *Sankhyā*, **7**, 407–414.

Rao, C. R. (1948). Tests of significance in multivariate analysis, *Biometrika*, **35**, 58–79.

Rao, C. R. (1970). Inference on discriminant function coefficients, *Essays in Probability and Statistics*, (eds. R. C. Bose and others), 587–602, Univ. of NorthCarolina Press, Chapel Hill, North Carolina.

Siotani, M. (1982). Large sample approximations and asymptotic expansions of classification statistics, *Handbook of Statistics*, Vol. 2, (eds. P. R. Krishnaiah and L. N. Kanal), 61–100, North-Holland, Amsterdam.