Ann. Inst. Statist. Math.
34 (1982), Part A, 523-524

# A TEST FOR ADDITIONAL INFORMATION IN CANONICAL CORRELATION ANALYSIS

YASUNORI FUJIKOSHI

(Received July 16, 1981; revised Jan. 25, 1982)

## Summary

In canonical correlation analysis a hypothesis concerning the relevance of a subset of variables from each of the two given variable sets is formulated. The likelihood ratio statistic for the hypothesis and an asymptotic expansion for its null distribution are obtained. In discriminant analysis various alternative forms of a hypothesis concerning the relevance of a specified variable subset are also discussed.

## 1. Introduction

In canonical correlation analysis McKay [9] has formulated a hypothesis concerning the relevance of a subset of variables from one of the two given variable sets. The test of the hypothesis is designed to examine whether or not a subset of variables provides additional information about the relationships between the two variable sets. Such a hypothesis is useful in the problem of variable selection. In this paper we formulate a hypothesis concerning the relevance of a subset of variables from each of the two given variable sets, resulting in a generalization of the hypothesis formulated by McKay [9]. We obtain various alternative forms of the hypothesis. A form of the hypothesis is given in terms of the coefficient vectors of the canonical variates. Using a characterization of the hypothesis in a conditional set-up we shall obtain the likelihood ratio criterion for testing the hypothesis. Its null distribution is studied and an asymptotic expansion for the distribution is obtained.

In discriminant analysis Rao's [12] additional information hypothesis is known as a hypothesis concerning the relevance of a specified variable subset. McKay [8] has given an alternative form of the hypoth-

AMS 1980 subject classifications: Primary 62H15; secondary 62H10.
Key words and phrases: Canonical correlation analysis, additional information, likelihood ratio statistic, null distribution, discriminant analysis.

esis and proposed simultaneous procedures for variable selection, based on the hypothesis. In this paper we give a new interpretation of the hypothesis in terms of the coefficient vectors of the canonical variates due to Fisher [3].

## 2. The hypotheses for additional information

Let $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ be two vectors of $p_1$ and $p_2$ ($p_1 \leqq p_2$) components, respectively, with means $\boldsymbol{\mu}_u$ and $\boldsymbol{\mu}_v$, respectively, and a nonsingular covariance matrix

$$\Sigma=\begin{pmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{pmatrix} . \tag{2.1}$$

When we are concerned to summarize the relationships between $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ in terms of only a few variables, the canonical variates $(\boldsymbol{a}'_{uj}\boldsymbol{x}_u, \boldsymbol{a}'_{vj}\boldsymbol{x}_v)$ due to Hotelling [6] are used. The coefficient vectors $\boldsymbol{a}_{uj}$ and $\boldsymbol{a}_{vj}$ are defined by

$$\Sigma_{uv}\Sigma_{vv}^{-1}\Sigma_{vu}\boldsymbol{a}_{uj}=\rho_j^2\Sigma_{uu}\boldsymbol{a}_{uj} , \qquad \boldsymbol{a}'_{ui}\Sigma_{uu}\boldsymbol{a}_{uj}=\delta_{ij} , \tag{2.2}$$

$$\Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{uv}\boldsymbol{a}_{vj}=\rho_j^2\Sigma_{vv}\boldsymbol{a}_{vj} , \qquad \boldsymbol{a}'_{vi}\Sigma_{vv}\boldsymbol{a}_{vj}=\delta_{ij} \tag{2.3}$$

where $\rho_j^2$ are the characteristic roots of $\Sigma_{uv}\Sigma_{vv}^{-1}\Sigma_{vu}\Sigma_{uu}^{-1}$, $\rho_1 \geqq \cdots \geqq \rho_{p_1} \geqq 0$ and $\delta_{ij}=1$ for $i=j$, 0 for $i \neq j$. Let $m$ be the number of non-zero canonical correlations $\rho_j$. Then $m=\operatorname{rank}(\Sigma_{uv}) \leqq p_1$ and the relationships between $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ can be summarized in terms of the first $m$ canonical variates $(\boldsymbol{a}'_{uj}\boldsymbol{x}_u, \boldsymbol{a}'_{vj}\boldsymbol{x}_v)$, $j=1,\cdots, m$. In order to formulate a hypothesis concerning the relevance of a subset of variables from each of $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ we partition $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ as $\boldsymbol{x}_u=(\boldsymbol{x}'_1, \boldsymbol{x}'_2)'$, $\boldsymbol{x}_1: r_1 \times 1$, $\boldsymbol{x}_v=(\boldsymbol{x}'_3, \boldsymbol{x}'_4)'$, $\boldsymbol{x}_3: r_2 \times 1$, and $\boldsymbol{a}_{uj}$, $\boldsymbol{a}_{vj}$, $\boldsymbol{\mu}_u$, $\boldsymbol{\mu}_v$, $\Sigma$ conformably:

$$\begin{pmatrix} \boldsymbol{a}_{uj} \\ \boldsymbol{a}_{vj} \end{pmatrix}=\begin{pmatrix} \boldsymbol{a}_{1j} \\ \boldsymbol{a}_{2j} \\ \boldsymbol{a}_{3j} \\ \boldsymbol{a}_{4j} \end{pmatrix}, \quad \begin{pmatrix} \boldsymbol{\mu}_{uj} \\ \boldsymbol{\mu}_{vj} \end{pmatrix}=\begin{pmatrix} \boldsymbol{\mu}_{1j} \\ \boldsymbol{\mu}_{2j} \\ \boldsymbol{\mu}_{3j} \\ \boldsymbol{\mu}_{4j} \end{pmatrix}, \quad \Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} \end{pmatrix} . \tag{2.4}$$

We use the notation $\Sigma_{23\cdot1}=\Sigma_{23}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13}$, $\Sigma_{4u}=(\Sigma_{41}, \Sigma_{42})$, etc. Then it is natural to consider that in the interpretation of the relationships between $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ the variates $\boldsymbol{x}_2$ and $\boldsymbol{x}_4$ are irrelevant, in the presence of $\boldsymbol{x}_1$ and $\boldsymbol{x}_3$ if and only if

$$H_1: \boldsymbol{a}_{2j}=0 , \quad \boldsymbol{a}_{4j}=0 \qquad (j=1,\cdots, m) . \tag{2.5}$$

If $H_1$ is true, we can say that the subset $(\boldsymbol{x}'_1, \boldsymbol{x}'_3)$ has as much information about the relationships between $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ as the full set $(\boldsymbol{x}'_u, \boldsymbol{x}'_v)$.

Following Rao's terminology as in a multivariate linear model, the subset $(\boldsymbol{x}_2', \boldsymbol{x}_4')$ may be said to supply no additional information about the relationships between $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$, independently of $(\boldsymbol{x}_1', \boldsymbol{x}_3')$. It may be noted that the effect on the canonical correlations by adding extra variates to one of the original two sets of variates has been discussed in Laha [7], Siotani [13], McKay [9]. Let $\rho_1^* \geqq \cdots \geqq \rho_{m^*}^*$ be the non-zero population canonical correlations between $\boldsymbol{x}_1$ and $\boldsymbol{x}_3$, i.e., the positive square roots of the non-zero characteristic roots of $\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33}^{-1}$. Then it holds that

$$\rho_j \geqq \rho_j^* , \qquad j=1,\cdots,m^*, \text{ and } m \geqq m^* . \tag{2.6}$$

This follows by applying a Lemma in Gabriel [4] to

$$\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33}^{-1} = D_2\{\Sigma_{vu}D_1'(D_1\Sigma_{uu}D_1')^{-1}D_1\Sigma_{uv}\}D_2'(D_2\Sigma_{vv}D_2')^{-1}$$

where $D_1=(I_{r_1}, 0)$ and $D_2=(I_{r_2}, 0)$. Intuitively, the condition

$$\rho_j = \rho_j^* , \qquad j=1,\cdots,m^*, \text{ and } m = m^* , \tag{2.7}$$

also means that the subset $(\boldsymbol{x}_1', \boldsymbol{x}_3')$ has as much information as the full set. In the following Theorem 1 we show that (2.5) and (2.7) are equivalent, and (2.5) is also equivalent to (2.8) and (2.9):

$$\operatorname{tr} \Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{uv}\Sigma_{vv}^{-1} = \operatorname{tr} \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33}^{-1} , \tag{2.8}$$

$$\Sigma_{4u\cdot 3}=0 , \qquad \Sigma_{2v\cdot 1}=0 \text{ (or } \Sigma_{23\cdot 1}=0) . \tag{2.9}$$

THEOREM 1. *The four statements* (2.5), (2.7), (2.8) *and* (2.9) *are equivalent.*

PROOF. The equivalence of (2.7) and (2.8) follows from (2.6). It is easily seen that

$$\operatorname{tr} \Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{uv}\Sigma_{vv}^{-1} = \operatorname{tr} \Sigma_{3u}\Sigma_{uu}^{-1}\Sigma_{u3}\Sigma_{33}^{-1} + \operatorname{tr} \Sigma_{4u\cdot 3}\Sigma_{uu}^{-1}\Sigma_{u4\cdot 3}\Sigma_{44\cdot 3}^{-1} . \tag{2.10}$$

and

$$\operatorname{tr} \Sigma_{3u}\Sigma_{uu}^{-1}\Sigma_{u3}\Sigma_{33}^{-1} = \operatorname{tr} \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33}^{-1} + \operatorname{tr} \Sigma_{32\cdot 1}\Sigma_{22\cdot 1}^{-1}\Sigma_{23\cdot 1}\Sigma_{33}^{-1} . \tag{2.11}$$

Using (2.10) and (2.11) we have the equivalence of (2.8) and (2.9). To complete the proof we need only show that (2.5)$\Longleftrightarrow$(2.9). If $\boldsymbol{a}_{4j}=0$ ($j=1,\cdots,m$), then from (2.3) we have $\Sigma_{3u}\Sigma_{uu}^{-1}\Sigma_{u3}\boldsymbol{a}_{3j}=\rho_j^2\Sigma_{33}\boldsymbol{a}_{3j}$, and hence $\operatorname{tr} \Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{uv}\Sigma_{vv}^{-1} = \operatorname{tr} \Sigma_{3u}\Sigma_{uu}^{-1}\Sigma_{u3}\Sigma_{33}^{-1}$. The last equality and (2.10) imply $\Sigma_{4u\cdot 3}=0$. Conversely, assume that $\Sigma_{4u\cdot 3}=0$. Then premultiplying both sides of (2.3) by $(-\Sigma_{43}\Sigma_{33}^{-1}, I_{p_2-r_2})$, we obtain $\rho_j^2\Sigma_{44\cdot 3}\boldsymbol{a}_{4j}=0$, i.e., $\boldsymbol{a}_{4j}=0$ ($j=1,\cdots,m$). Similarly we can show the equivalence of $\boldsymbol{a}_{2j}=0$ ($j=1,\cdots,m$) and $\Sigma_{2v\cdot 1}=0$. This implies the equivalence of (2.5) and (2.9).

As a special case of $H_1$ the hypothesis that the subset $(\boldsymbol{x}_u', \boldsymbol{x}_3')$ has as much information as the full set $(\boldsymbol{x}_u', \boldsymbol{x}_v')$ can be expressed as

$$H_2: \boldsymbol{a}_{4j}=0 \qquad (j=1,\cdots,m)\,. \tag{2.12}$$

From Theorem 1 we obtain the following alternative forms of $H_2$:

$$\operatorname{tr} \Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{uv}\Sigma_{vv}^{-1}=\operatorname{tr} \Sigma_{3u}\Sigma_{uu}^{-1}\Sigma_{u3}\Sigma_{33}^{-1}\,, \tag{2.13}$$

$$\Sigma_{4u\cdot 3}=0\,. \tag{2.14}$$

The equivalence of (2.13) and (2.14) has been established by McKay [9].

## 3. The likelihood ratio test

In this section we assume that $\boldsymbol{x}=(\boldsymbol{x}_u', \boldsymbol{x}_v')'$ is normally distributed. Then the conditional distribution of $(\boldsymbol{x}_2', \boldsymbol{x}_4')$ given $(\boldsymbol{x}_1', \boldsymbol{x}_3')$ is a $(p-r)$-variate normal distribution with mean vector

$$\mathrm{E}\left[\begin{pmatrix}\boldsymbol{x}_2\\ \boldsymbol{x}_4\end{pmatrix}\middle|\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_3\end{pmatrix}\right]=\begin{pmatrix}\boldsymbol{\mu}_2\\ \boldsymbol{\mu}_4\end{pmatrix}+\begin{pmatrix}B_{21} & B_{23}\\ B_{41} & B_{43}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_1-\boldsymbol{\mu}_1\\ \boldsymbol{x}_3-\boldsymbol{\mu}_3\end{pmatrix} \tag{3.1}$$

and covariance matrix

$$\mathrm{V}\left[\begin{pmatrix}\boldsymbol{x}_2\\ \boldsymbol{x}_4\end{pmatrix}\middle|\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_3\end{pmatrix}\right]=\begin{pmatrix}\Sigma_{22\cdot 13} & \Sigma_{24\cdot 13}\\ \Sigma_{42\cdot 13} & \Sigma_{44\cdot 13}\end{pmatrix} \tag{3.2}$$

where $p=p_1+p_2$, $r=r_1+r_2$,

$$\begin{pmatrix}B_{21} & B_{23}\\ B_{41} & B_{43}\end{pmatrix}=\begin{pmatrix}\Sigma_{21} & \Sigma_{23}\\ \Sigma_{41} & \Sigma_{43}\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{13}\\ \Sigma_{31} & \Sigma_{33}\end{pmatrix}^{-1},$$

$$\begin{pmatrix}\Sigma_{22\cdot 13} & \Sigma_{24\cdot 13}\\ \Sigma_{42\cdot 13} & \Sigma_{44\cdot 13}\end{pmatrix}=\begin{pmatrix}\Sigma_{22} & \Sigma_{24}\\ \Sigma_{42} & \Sigma_{44}\end{pmatrix}-\begin{pmatrix}\Sigma_{21} & \Sigma_{23}\\ \Sigma_{41} & \Sigma_{43}\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{13}\\ \Sigma_{31} & \Sigma_{33}\end{pmatrix}^{-1}\begin{pmatrix}\Sigma_{12} & \Sigma_{14}\\ \Sigma_{32} & \Sigma_{34}\end{pmatrix}.$$

The hypothesis $H_1$ can be expressed in terms of the conditional set-up as

$$B_{23}=0\,, \qquad B_{41}=0\,, \qquad \Sigma_{24\cdot 13}=0\,. \tag{3.3}$$

This result follows from (2.9) and the identities

$$B_{23}=\Sigma_{23\cdot 1}\Sigma_{33\cdot 1}^{-1}\,, \qquad B_{41}=\Sigma_{41\cdot 3}\Sigma_{11\cdot 3}^{-1}\,,$$

$$\Sigma_{24\cdot 13}=\Sigma_{24\cdot 3}-\Sigma_{21\cdot 3}\Sigma_{11\cdot 3}^{-1}\Sigma_{14\cdot 3}\,.$$

Suppose that a random sample of size $N=n+1$ $(>p)$ of $\boldsymbol{x}$ is available. Let $n^{-1}S$ be the sample covariance matrix formed from the sample. Then, using the conditional distribution of $(\boldsymbol{x}_2', \boldsymbol{x}_4')$ given $(\boldsymbol{x}_1', \boldsymbol{x}_3')$ and (3.3) it can be shown that the likelihood ratio criterion for $H_1$ is

$$\lambda_1^{N/2}=\Lambda_1=\begin{vmatrix}S_{22\cdot 13} & S_{24\cdot 13}\\ S_{42\cdot 13} & S_{44\cdot 13}\end{vmatrix}\Big/\{|S_{22\cdot 1}||S_{44\cdot 3}|\} \tag{3.4}$$

$$= \{|S_{22\cdot 13} - S_{24\cdot 13} S_{44\cdot 13}^{-1} S_{42\cdot 13}| / |S_{22\cdot 1}|\} \{|S_{44\cdot 13}| / |S_{44\cdot 3}|\}$$
$$= \Lambda_{11} \times \Lambda_{12}$$

where the method used previously to describe submatrices of $\Sigma$ is used here to describe submatrices of $S$.

THEOREM 2. *If $H_1$ is true, the statistics $\Lambda_{11}$ and $\Lambda_{12}$ in* (3.4) *are independently distributed according to the $\Lambda$ distributions $\Lambda(p_1-r_1, p_2, n-r_1-p_2)$ and $\Lambda(p_2-r_2, r_1, n-r_1-r_2)$, respectively. Here we denote the distribution of $\Lambda=|A|/|A+B|$ by $\Lambda(p, q, n)$, where $A$ and $B$ are independently distributed according to the Wishart distributions $W_p(n, \Sigma)$ and $W_p(q, \Sigma)$, respectively.*

PROOF. We can write $S=Y'Y$, where each row of $Y: n\times(p_1+p_2)$ is independently distributed according to $N(0, \Sigma)$. Let $Y$ partition as $Y=(Y_1, Y_2, Y_3, Y_4)$ with $r_1, p_1-r_1, r_2, p_2-r_2$ columns, respectively. Then we can write $\Lambda_{11}$ as

$$\Lambda_{11}=|Y_2'F_1Y_2|/|Y_2'F_1Y_2+Y_2'F_2Y_2|$$

where $F_1=I_n-P_{1\cdot 3}-(I_n-P_{1\cdot 3})Y_4\{Y_4'(I_n-P_{1\cdot 3})Y_4\}^{-1}Y_4'(I_n-P_{1\cdot 3})$, $F_2=I_n-P_1-F_1$, $P_{i\cdot j}=(Y_i, Y_j)\{(Y_i, Y_j)'(Y_i, Y_j)\}^{-1}(Y_i, Y_j)'$ and $P_i=Y_i(Y_i'Y_i)^{-1}Y_i'$. It is easily seen that $F_1^2=F_1$, $F_2^2=F_2$ and $F_1F_2=0$. Noting that the conditional distribution of $Y_2$ given $(Y_1, Y_3, Y_4)$ is normal with means $Y_1B_{21}'$ when $H_1$ is true we obtain that the conditional distribution of $\Lambda_{11}$ is $\Lambda(p_1-r_1, p_2, n-r_1-p_2)$. Therefore, $\Lambda_{11}\sim\Lambda(p_1-r_1, p_2, n-r_1-p_2)$, and $\Lambda_{11}$ is independent of $(Y_1, Y_3, Y_4)$ and hence of $\Lambda_{12}=|Y_4'G_1Y_4|/|Y_4'G_1Y_4+Y_4'G_2Y_4|$, where $G_1=I_n-P_{1\cdot 3}$ and $G_2=I-P_3-G_1$. Similarly, noting that the conditional distribution of $Y_4$ given $(Y_1, Y_3)$ is normal with means $Y_3B_{43}'$ when $H_1$ is true, we obtain that $\Lambda_{12}\sim\Lambda(p_2-r_2, r_1, n-r_1-r_2)$.

From Theorem 2 we can express the characteristic function of $-n\log\Lambda_1$ in terms of $\Gamma$-functions (cf. Anderson [1], p. 193). By expanding the characteristic function as in Box [2] and Anderson [1] we can obtain the approximation to the null distribution of $\Lambda_1$ given in the following Theorem 3:

THEOREM 3. *The null distribution of the likelihood ratio criterion* (3.4) *for $H_1$ can be approximated asymptotically up to the order $m^{-2}$ by*

$$\text{(3.5)}\quad \mathrm{P}(-m\log\Lambda_1\leqq x)=\mathrm{P}(\chi_\phi^2\leqq x)+\frac{\beta}{m^2}\{\mathrm{P}(\chi_{\phi+4}^2\leqq x)-\mathrm{P}(\chi_\phi^2\leqq x)\}+O(m^{-3}),$$

*where $p=p_1+p_2$, $r=r_1+r_2$, $\phi=p_1p_2-r_1r_2$,*

$$m=n-\frac{1}{2}(p+1)-\frac{1}{2}r_1r_2(p-r)/\phi,$$

$$\beta=\frac{1}{48}[\{r_1^2+(p_1-r_1)^2+r_2^2+(p_2-r_2)^2-5\}\phi+2r_1^2(p_1-r_1)p_2 \\ +2p_1r_2^2(p_2-r_2)+2(p_1-r_1)(p_2-r_2)\{r_1(p_1-r_1) \\ +r_2(p_2-r_2)-3r_1r_2\}-3(r_1r_2)^2(p-r)^2/\phi] .$$

As a special result of the above we obtain that the likelihood ratio criterion $\lambda_2$ for $H_2$ is

$$\lambda_2^{N/2}=\Lambda_2=|S_{44\cdot u3}|/|S_{44\cdot 3}| \tag{3.6}$$

and the null distribution of $\Lambda_2$ is the $\Lambda$ distribution $\Lambda(p_2-r_2, p_1, n-p_1-r_2)$. An approximation to the null distribution of $\Lambda_2$ is given by the formula (3.5) with $p_1=r_1$. The statistic $\Lambda_2$ has been also obtained by McKay [9] and in the complex normal case by Hannan [5], p. 300.

## 4. The case of discriminant analysis

Consider $q+1$ $p$-variate populations $\Pi_i$ $(i=1,\cdots,q+1)$ with means $\boldsymbol{\mu}_i$ and the same covariance matrix $\Sigma$. Let $\boldsymbol{x}'=(x_1,\cdots,x_p)'$ be the column vector of the $p$ variables. Suppose that $N_i$ samples from $\Pi_i$ are available. Let $\Omega$ be the population between-groups covariance matrix defined by

$$\Omega=\sum_{i=1}^{q+1}(N_i/N)(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}})' \tag{4.1}$$

where $N=N_1+\cdots+N_{q+1}$ and $\bar{\boldsymbol{\mu}}=(1/N)\sum_{i=1}^{q+1}N_i\boldsymbol{\mu}_i$. We are interested in a hypothesis indicating that a specified variable set has as much information about the differences between the populations as the full set.

Now, consider the partitions $\boldsymbol{x}'=(\boldsymbol{x}_1', \boldsymbol{x}_2')$, $\boldsymbol{x}_1=(x_1,\cdots,x_k)'$, and

$$\boldsymbol{\mu}_j=\begin{pmatrix}\boldsymbol{\mu}_{1i}\\ \boldsymbol{\mu}_{2i}\end{pmatrix}, \quad \Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}, \quad \Omega=\begin{pmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{pmatrix} \tag{4.2}$$

conforming with the partition of $\boldsymbol{x}$. Considering the conditional distribution of $\boldsymbol{x}_2$ given $\boldsymbol{x}_1$, Rao [12] formulated the hypothesis that $\boldsymbol{x}_2$ provides no additional information about departures from nullity of the hypothesis of equality of $\boldsymbol{\mu}_i$ $(i=1,\cdots,q+1)$ independently of $\boldsymbol{x}_1$. The hypothesis is defined by

$$H_3: \boldsymbol{\mu}_{21}-\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_{11}=\cdots=\boldsymbol{\mu}_{2\cdot(q+1)}-\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_{1\cdot(q+1)} . \tag{4.3}$$

McKay [9] proved that the hypothesis $H_3$ is equivalent to each of the following statements:

$$\theta_j=\theta_j^*, \quad j=1,\cdots,m^* \text{ and } m=m^*, \tag{4.4}$$

$$\text{(4.5)} \qquad \operatorname{tr} \Sigma^{-1}\Omega = \operatorname{tr} \Sigma_{11}^{-1}\Omega_{11}$$

where $\theta_1 \geqq \cdots \geqq \theta_m$ and $\theta_1^* \geqq \cdots \geqq \theta_{m^*}^*$ are the non-zero characteristic roots of $\Sigma^{-1}\Omega$ and $\Sigma_{11}^{-1}\Omega_{11}$, respectively.

In this section we give an alternative form of $H_3$ in terms of the canonical variates due to Fisher [3]. The canonical variates method is used to summarize the differences between the populations in terms of only a few transformed variates. If $\operatorname{rank}(\Omega) = m$, then the differences between the populations can be expressed in terms of the first $m$ canonical variates $\boldsymbol{a}_j'\boldsymbol{x}$, $j=1, \cdots, m$, where $\boldsymbol{a}_j$ are the solutions of

$$\text{(4.6)} \qquad \Omega \boldsymbol{a}_j = \theta_j \Sigma \boldsymbol{a}_j , \qquad \boldsymbol{a}_i' \Sigma \boldsymbol{a}_j = \delta_{ij} .$$

THEOREM 4. *Each of the two statements* (4.3) *and* (4.5) *is equivalent to*

$$\text{(4.7)} \qquad \boldsymbol{a}_{2j} = 0 , \qquad j=1, \cdots, m$$

*where* $\boldsymbol{a}_j = (\boldsymbol{a}_{1j}', \boldsymbol{a}_{2j}')'$, $\boldsymbol{a}_{1j}$ : $k \times 1$ *and* $m = \operatorname{rank}(\Omega)$.

PROOF. It is sufficient to show the equivalence of (4.5) and (4.7) since the equivalence of (4.3) and (4.5) has been established by McKay [9]. We use the identity

$$\text{(4.8)} \qquad \operatorname{tr} \Omega\Sigma^{-1} = \operatorname{tr} \Omega_{11}\Sigma_{11}^{-1} + \operatorname{tr} (-\Sigma_{21}\Sigma_{11}^{-1}, I_{p-k})\Omega(-\Sigma_{21}\Sigma_{11}^{-1}, I_{p-k})'\Sigma_{22\cdot 1}$$

where $\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. If (4.5) holds, from (4.8) we have $(-\Sigma_{21}\Sigma_{11}^{-1}, I_{p-k})\Omega = 0$. Using this result and premultiplying both sides of the first equation of (4.6) by $(-\Sigma_{21}\Sigma_{11}^{-1}, I_{p-k})$, we obtain $\theta_j \Sigma_{22\cdot 1}\boldsymbol{a}_j = 0$, $j=1, \cdots, m$ and hence (4.7). Conversely, if (4.7) is true, from (4.6) we have $\Omega_{11}\boldsymbol{a}_{1j} = \theta_j \Sigma_{11}\boldsymbol{a}_{1j}$, $j=1, \cdots, m$ and hence (4.5) holds. This completes the proof.

The statement (4.7) means that for the description of the differences between the populations $\boldsymbol{x}_2$ may be considered irrelevant, in the presence of $\boldsymbol{x}_1$. It is known (Rao [11]) that in the case of $q=1$ the statements (4.3), (4.5) and (4.7) are equivalent. The likelihood ratio test of $H_3$ has been used by Rao [10].

## Acknowledgment

The author wishes to thank the referee for his valuable comments.

HIROSHIMA UNIVERSITY

## REFERENCES

[1] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*, John Wiley, New York.

[2] Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria, *Biometrika*, **36**, 317-346.

[3] Fisher, R. A. (1938). The statistical utilization of multiple measurements, *Ann. Eugen.*, 8, 376-386.

[4] Gabriel, K. R. (1968). Simultaneous test procedures in multivariate analysis of variance, *Biometrika*, **55**, 489-504.

[5] Hannan, E. J. (1970). *Multiple Time Series*, John Wiley, New York.

[6] Hotelling, H. (1936). Relations between two sets of variates, *Biometrika*, 28, 321-377.

[7] Laha, R. G. (1954). On some problems in canonical correlations, *Sankhyā*, **14**, 61-66.

[8] McKay, R. J. (1977). Simultaneous procedures for variable selection in multiple discriminant analysis, *Biometrika*, **64**, 283-290.

[9] McKay, R. J. (1977). Variable selection in multivariate regression: an application of simultaneous test procedures, *J. R. Statist. Soc.*, B, **39**, 371-380.

[10] Rao, C. R. (1949). Tests of significance in multivariate analysis, *Biometrika*, **35**, 58-79.

[11] Rao, C. R. (1970). Inference on discriminant function coefficients, *Essays in Prob. Statist.* (ed. R. C. Bose), 587-602.

[12] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, John Wiley, New York.

[13] Siotani, M. (1957). Effect of the additional variates on the canonical correlation coefficients, *Proc. Inst. Statist. Math.*, **5**, 52-57.