Sankhyā: The Indian Journal of Statistics
1997, Volume 59, Series A, Pt. 1, pp. 74-87

A GENERAL DECISION RULE FOR DIMENSION IN THE CONTEXT OF MANOVA

By ANINDYA K. DE
Purdue University, West Lafayette
and
JAYANTA K. GHOSH
Indian Statistical Institute, Calcutta

SUMMARY. It is often of interest to identify a linear structural relationship among the mean vectors in the context of Multivariate Analysis of Variance(MANOVA). In this paper we have generalized the results of De and Ghosh (1994) for the bivariate case to a d-variate case for $d > 2$. We have exhibited a Bayes solution for populations of arbitrary dimension and also derived a useful approximation which could be used in practice. As in the bivariate case, the Bayes rule closely resembles the ad hoc rule introduced in De and Ghosh (1994) which was shown to have good frequentist properties. A simulation study was carried out to compare the performance of the proposed rules.

1. Introduction

In a Multivariate Analysis of Variance (MANOVA) set up (for examples and references see De (1993) and De and Ghosh (1994)) it is important to know if the mean vectors have a linear structural relationship among themselves. Let us state the problem formally. We have p populations each of dimension d, $p > d$. Let the mean vectors be $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_p$ and Ω be the common known dispersion matrix. Let $\mathbf{M} \equiv (\boldsymbol{\mu}_1 - \bar{\boldsymbol{\mu}}, \boldsymbol{\mu}_2 - \bar{\boldsymbol{\mu}}, \ldots \boldsymbol{\mu}_p - \bar{\boldsymbol{\mu}})$, where $\bar{\boldsymbol{\mu}} = p^{-1} \sum_{i=1}^{p} \boldsymbol{\mu}_i$. We have a multiple decision problem with $d+1$ possible decisions (or actions): $a_0, a_1, \ldots, a_d$ where a_k denotes the decision or action – rank of $\mathbf{M}$ is k i.e. the mean vectors lie in a k dimensional subspace. In this paper we will generalize to arbitrary d the result of De and Ghosh (1994) for $d = 2$. We will find a Bayes rule for the multiple decision theoretic problem in d dimension which resembles the ad hoc rule of De and Ghosh (1994) based on the eigenvalues of the between sum of squares and product matrix. We will then examine the performance of

Paper received. October 1994; revised July 1995.
AMS (1991) *subject classification.* 62 C10, 62 F15.
Key words and phrases. MANOVA; decision rule for dimension; approximate Bayes.

the ad hoc rule and that of the Bayes rule for the cases $p = 5, 6, 7$ and 8 for $d = 3$, using simulation.

See also Rao (1973, 1985), Fujikoshi (1974) and Shen and Sinha (1991) which are some other relevant references in this area. However, they deal with tests rather than multiple decision procedures.

2. Prior distributions

Prior distributions are chosen from considerations similar to those of De and Ghosh (1994). Let X_{ij} represent a d dimensional vector m-th element of which is $X_{ij}^{(m)}$. The subscript (ij) stands for the j-th observation from the i-th population, $i = 1, \ldots, p$ and $j = 1, \ldots, n$. We assume $X_{ij} \sim N_d(\boldsymbol{\mu}_i, \mathbf{I})$ where $\boldsymbol{\mu}_i' = (\mu_i^{(1)}, \ldots, \mu_i^{(d)})$. X_{ij} and $X_{i'j'}$ are independent if $(ij) \neq (i'j')$. Let $\bar{X}_i = n^{-1} \sum_{j=1}^n X_{ij}$. We also assume that the common covariance matrix is a known positive definite matrix and without loss of generality, we may take it as $\mathbf{I}$.

The prior distributions of Ψ and $\bar{\boldsymbol{\mu}}$ are assumed to be independent and the same distribution for $\bar{\boldsymbol{\mu}}$ is assumed under all $d + 1$ hypotheses. Let $\pi_m(\cdot)$ represent the prior distribution under the hypothesis H_m. Then $\pi_m(\Psi, \bar{\boldsymbol{\mu}}) = \pi_m(\Psi)\pi_m(\bar{\boldsymbol{\mu}}), m = 0, 1, \ldots, p$. Now $\pi_m(\cdot)$ is defined in terms of another distribution $\pi_m^*(\cdot)$, (see De and Ghosh (1994) for explanation) and they are related as

$$\pi_m(\Psi) = K_m e^{\frac{n}{2}\mathbf{tr} \sum \psi_i \psi_i'} \pi_m^*(\Psi) \qquad \ldots (1)$$

$$\pi_m(\bar{\boldsymbol{\mu}}) = K_\mu e^{\frac{n}{2} p \bar{\boldsymbol{\mu}} \bar{\boldsymbol{\mu}}'} \pi_m^*(\bar{\boldsymbol{\mu}}) \qquad \ldots (2)$$

where K's are constants. $\pi_m^*(\bar{\boldsymbol{\mu}})$ and thus $\pi_m(\bar{\boldsymbol{\mu}})$ would be chosen to be the same distribution under each H_m.

Under π_m^*, $\bar{\boldsymbol{\mu}}|\Sigma_{\bar{\mu}} \sim N_d(O, \Sigma_{\bar{\mu}})$ and $\Psi|\Sigma_m^* \sim N_{dp}(O, \mathbf{P} \otimes \Sigma_m^*)$, $\mathbf{P} = \mathbf{I}_p - p^{-1}\mathbf{1}\mathbf{1}'$. Let $\mathbf{U}\mathbf{D}_m^*\mathbf{U}'$ be the spectral decomposition of Σ_m^* where $\mathbf{U}$ is a d dimensional orthogonal matrix and $\mathbf{D}_m^*$ is a diagonal matrix. We assume the prior probability law of $\mathbf{U}$ to be the Haar measure in the space of d dimensional orthogonal matrices. $\mathbf{D}_m^* = \text{diag}(\sigma_{11}^*, \ldots, \sigma_{mm}^*, 0, \ldots, 0)$, where $\sigma_{11}^* \geq \cdots \geq \sigma_{mm}^* > 0$ are fixed and are to be chosen.

THEOREM 2.1. *Distribution of* Ψ *under the prior* π_m *for a given* $\Sigma_m^* = \mathbf{U}\mathbf{D}_m^*\mathbf{U}', \mathbf{D}_m^* = \text{diag}\ (\sigma_{11}^* \ldots, \sigma_{mm}^*, 0, \ldots, 0)$ *is normal with mean* O *and covariance* $\mathbf{P} \otimes \Sigma_m$ *where* $\Sigma_m = \mathbf{U}\mathbf{D}_m\mathbf{U}'$ *and* $\mathbf{D}_m = diag(\sigma_{11} \ldots, \sigma_{mm}, 0, \ldots, 0)$ *and* $\sigma_{rr} = (\frac{1}{\sigma_{rr}^*} - n)^{-1}$. *The constant* K_m *in the expression (1) is* $\prod_{r=1}^m (1 - n\sigma_{rr}^*)^{\frac{p-1}{2}}$.

PROOF. Let $\gamma_i = \mathbf{U}'\psi_i$, $i = 1, \ldots, p$ and $\gamma_i' = (\gamma_i^{(1)}, \ldots, \gamma_i^{(d)})$ and $\Gamma' = (\gamma_1', \ldots, \gamma_p') = (\psi_1'\mathbf{U}, \ldots, \psi_p'\mathbf{U})$. So $\Gamma = \mathbf{I}_p \otimes \mathbf{U}'\Psi$ and $\Gamma_p = \mathbf{I}_{p-1} \otimes \mathbf{U}'\Psi_p$ where $\Gamma_p' = (\gamma_1', \ldots, \gamma_{p-1}')$.

Γ has a multinormal distribution with mean $\mathbf{0}$ and covariance $\mathbf{P}\otimes\mathbf{D}_m^*$ under H_m. So Γ_p has a multinormal distribution with mean $\mathbf{0}$ and covariance $\mathbf{P}_p\otimes\mathbf{D}_m^*$ under H_m. Note γ_i^r is identically equal to zero because its mean and variance are zero, for $m < r \le d$ and for all i. Let $\Gamma_p^{(m)\prime} = (\,(\gamma_1^{(1)},\ldots,\gamma_1^{(m)}),\ldots,(\gamma_{p-1}^{(1)},\ldots,\gamma_{p-1}^{(m)})\,)$. Then

$$\begin{aligned}\pi_m^*(d\Gamma_p^{(m)}) &= \frac{1}{(2\pi)^{m(p-1)/2}|\mathbf{P}_p\otimes\mathbf{D}_{nz}^*|^{1/2}}\mathbf{e}^{-\frac{1}{2}\Gamma_p^{(m)\prime}(\mathbf{P}_p\otimes\mathbf{D}_{nz}^*)^{-1}\Gamma_p^{(m)}}d(\Gamma_p^{(m)})\\ \text{where } \mathbf{D}_m^* &= \left(\begin{array}{c|c}\mathbf{D}_{nz}^* & \mathbf{O}\\ \hline \mathbf{O} & \mathbf{O}\end{array}\right)\end{aligned}$$

and $\mathbf{D}_{nz}^*$ is the $m\times m$ diagonal matrix with positive diagonal elements.

Now,

$$\begin{aligned}\Psi'\Psi &= \Psi_p'\Psi_p + \psi_p'\psi_p\\ &= \textstyle\sum_{i=1}^{p-1}\psi_i'\psi_i + (-\sum_{i=1}^{p-1}\psi_i')(-\sum_{i=1}^{p-1}\psi_i)\\ &= 2\textstyle\sum_{i=1}^{p-1}\psi_i'\psi_i + \sum_{i=1}^{p-1}\sum_{j\neq i}^{p-1}\psi_i'\psi_j\\ &= 2\textstyle\sum_{i=1}^{p-1}\gamma_i'\gamma_i + \sum_{i=1}^{p-1}\sum_{j\neq i}^{p-1}\gamma_i'\gamma_j\\ &= 2\textstyle\sum_{i=1}^{p-1}\sum_{r=1}^{m}\gamma_i^{r^2} + \sum_{i=1}^{p-1}\sum_{j\neq i}^{p-1}\sum_{r=1}^{m}\gamma_i^r\gamma_j^r\\ &= \Gamma_p^{(m)\prime}\,(\mathbf{P}_p\otimes\mathbf{I}_m)^{-1}\,\Gamma_p^{(m)}\end{aligned}$$

Therefore, the density of $\Gamma_p^{(m)}$ under π_m is given as

$$\begin{aligned}\pi_m(d\Gamma_p^{(m)}) &= \frac{K_m}{(2\pi)^{m(p-1)/2}|\mathbf{P}_p\otimes\mathbf{D}_{nz}^*|^{1/2}}\mathbf{e}^{\frac{n}{2}\Gamma_p^{(m)\prime}(\mathbf{P}_p\otimes\mathbf{I}_m)^{-1}\Gamma_p^{(m)}-\frac{1}{2}\Gamma_p^{(m)\prime}(\mathbf{P}_p\otimes\mathbf{D}_{nz}^*)^{-1}\Gamma_p^{(m)}}d(\Gamma_p^{(m)})\\ &= \frac{K_m}{(2\pi)^{m(p-1)/2}|\mathbf{P}_p\otimes\mathbf{D}_{nz}^*|^{1/2}}\mathbf{e}^{-\frac{1}{2}\Gamma_p^{(m)\prime}\left(\mathbf{P}_p^{-1}\otimes(\mathbf{D}_{nz}^{*\,-1}-n\mathbf{I}_m)\right)^{-1}\Gamma_p^{(m)}}d(\Gamma_p^{(m)})\end{aligned} \quad \ldots(3)$$

Hence under π_m given Σ_m^*, $E(\Gamma_p^{(m)}) = \mathbf{0}$ and $Cov(\Gamma_p^{(m)}) = [\mathbf{P}_p^{-1}\otimes(\mathbf{D}_{nz}^{*\,-1} - n\mathbf{I}_m)]^{-1} = \mathbf{P}_p\otimes(\mathbf{D}_{nz}^{*\,-1} - n\mathbf{I}_m)^{-1}$ imply

$$\begin{aligned}&E(\Gamma) = 0 \quad\text{and}\quad E(\Psi) = O,\\ &Cov(\Gamma) = \mathbf{P}\otimes\left(\begin{array}{c|c}(\mathbf{D}_{nz}^{*\,-1} - n\mathbf{I}_m)^{-1} & \mathbf{O}\\ \hline \mathbf{O} & \mathbf{O}\end{array}\right) = \mathbf{P}\otimes\mathbf{D}_m\\ &\text{and}\quad Cov(\Psi) = \mathbf{P}\otimes\mathbf{U}\mathbf{D}_m\mathbf{U}'\end{aligned}$$

Also it is clear from (3) that the distribution of Ψ is normal. On integrating both sides of (3) we solve for K_m.

$$K_m^{-1} = \frac{|\mathbf{P}_p\otimes\mathbf{D}_{nz}^*|^{-1/2}}{|\mathbf{P}_p^{-1}\otimes(\mathbf{D}_{nz}^{*\,-1} - n\mathbf{I}_m)|^{1/2}}$$

$$K_m^{-2} = \frac{|\mathbf{D}_{nz}^{*^{-1}}|^{p-1}}{|\mathbf{D}_{nz}^{*^{-1}} - n\mathbf{I}_m|^{p-1}} = \frac{1}{|\mathbf{I}_m - \mathbf{D}_{nz}^{*}|^{p-1}}$$

$$\text{Therefore,} \qquad K_m = \prod_{r=1}^{m} (1 - n\sigma_{rr}^{*})^{\frac{p-1}{2}}$$

▱

3. Bayes rule

We first prove a few lemmas and then use them to derive the Bayes rule (vide Theorem 3.1). We then prove a useful approximation to simplify the Bayes rule.

LEMMA 3.1. *Let* $\mathbf{Q}$ *be a* $d \times d$ *orthogonal matrix and* $\mathbf{Q}_1$ *be the* $d \times m$ *matrix consisting of first* m *columns of* $\mathbf{Q}$, $m \leq d$. *Let* $\mathbf{A}$ *be a* $m \times m$ *diagonal matrix,* $\mathbf{A} = \text{diag}\,(a_1, \ldots, a_m)$ *say, such that* $a_i > a_j$ *if* $i < j$. *Let* $\mathbf{B}$ *be a* $d \times d$ *diagonal matrix,* $\mathbf{B} = \text{diag}\,(b_1, \ldots, b_d)$ *say, such that* $b_i > b_j$ *if* $i < j$. *Let* $f(\mathbf{Q}_1) = \mathbf{tr}\mathbf{A}\mathbf{Q}_1'\mathbf{B}\mathbf{Q}_1$. *Then the maximum of* f *is* $\sum_{i=1}^{m} a_i b_i$ *and it is attained at* 2^m *values of* $\mathbf{Q}_1$ *of the form*

$$\begin{pmatrix} \pm 1 & 0 & \cdots & 0 \\ 0 & \pm 1 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \pm 1 \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \qquad \ldots (4)$$

PROOF. The (ij)-th element of $\mathbf{Q}_1'\mathbf{B}\mathbf{Q}_1$ is

$$\sum_{l=1}^{d}\sum_{k=1}^{d} q_{li} b_{lk} q_{kj} = \sum_{l=1}^{d} q_{li} b_l q_{lj}$$

as $\mathbf{B}$ is diagonal and $b_{ll} = b_l$ and $b_{lk} = 0$ for $l \neq k$. So (ii)-th element of $\mathbf{Q}_1'\mathbf{B}\mathbf{Q}_1$ is $\sum_{l=1}^{d} b_l q_{li}^2$. Hence,

$$\mathbf{tr}(\mathbf{A}\mathbf{Q}_1'\mathbf{B}\mathbf{Q}_1) = \sum_{i=1}^{m}\sum_{l=1}^{d} a_i b_l q_{li}^2$$

Let $a_i = \sum_{k=i}^{m} \alpha_k,\ \alpha_k > 0$ for all k and $b_i = \sum_{k=i}^{d} \beta_k,\ \beta_k > 0$ for all k. Then

$$\begin{aligned}
\mathbf{tr}(\mathbf{AQ}_1'\mathbf{BQ}_1) &= \sum_{i=1}^{m}\sum_{l=1}^{d}\left(\sum_{t=l}^{d}\beta_t\right) q_{li}^2 \left(\sum_{r=i}^{m}\alpha_r\right) \\
&= \sum_{t=1}^{d}\sum_{r=1}^{m}\alpha_r\beta_t \sum_{l=1}^{t}\sum_{i=1}^{r} q_{li}^2 \\
&\leq \sum_{t=1}^{d}\sum_{r=1}^{m}\alpha_r\beta_t \min(t,r)
\end{aligned}$$

because $((q_{ij}^2))$ is doubly stochastic hence both $\sum_{l=1}^{t} q_{li}^2$ and $\sum_{i=1}^{r} q_{li}^2$ are less than equal to 1. So it follows that

$$\begin{aligned}
\mathbf{tr}(\mathbf{AQ}_1'\mathbf{BQ}_1) &\leq \sum_{r=1}^{m}\alpha_r\left\{\left(\sum_{j=1}^{r} j\beta_j\right) + \left(r\sum_{j=r+1}^{d}\beta_j\right)\right\} \\
&= \sum_{r=1}^{m}\alpha_r\sum_{i=1}^{r}\sum_{l=i}^{d}\beta_l \\
&= \sum_{i=1}^{m}\sum_{r=i}^{m}\alpha_r\sum_{l=i}^{d}\beta_l \\
&= \sum_{i=1}^{m} a_i b_i.
\end{aligned}$$

Now we claim that the equality holds iff $\mathbf{Q}_1$ is of the form as given in equation 4. We prove that if

$$\sum_{l=1}^{t}\sum_{i=1}^{r} q_{li}^2 = \min(t,r) \quad \text{for } t = 1,\ldots,d; r = 1,\ldots,m;$$

then $\mathbf{Q}_1$ must be of the form as given in equation 4. The other part of the claim is trivial.

Let $(t,r) = (1,1)$ then $q_{11}^2 = 1$ which implies $q_{1j}^2 = 0$ for $j > 1$ and $q_{i1}^2 = 0$ for $i > 1$.

Take $(t,r) = (2,2)$ then

$$\begin{aligned}
& q_{11}^2 + q_{12}^2 + q_{21}^2 + q_{22}^2 = 2 \\
\text{or}\quad & q_{22}^2 = 1 \ \Rightarrow q_{2j}^2 = 0,\ q_{i2}^2 = 0 \text{ for } i,j \neq 2.
\end{aligned}$$

Next we consider $(t,r) = (3,3)$ and by similar argument it follows that $q_{3j}^2 = 0,\ q_{i3}^2 = 0$ for $i,j \neq 2$. After $s-1$ steps we know the form of first $s-1$ rows and those many columns of $\mathbf{Q}_1$ and they are as given in equation 4. Now take $(t,r) = (s,s)$ and $\sum_{l=1}^{s}\sum_{i=1}^{s} q_{li}^2 = s$ implies $q_{ss}^2 = 1$ then $q_{sj}^2 = 0$, $q_{is}^2 = 0$ for $i, j \neq s$. This completes the proof of the claim and that of the lemma. ▱

LEMMA 3.2. *Let f be as defined in the previous lemma. Let* **H** *be such that the columns of*

$$\left(\frac{\mathbf{I}}{\mathbf{O}}\right) + \mathbf{H} \qquad \ldots(5)$$

are of norm 1 and mutually orthogonal and the elements of **H** *be infinitesimally small (so that 3rd and higher order products are negligible compared to the second order ones). Then*

$$f\left(\left(\frac{\mathbf{I}}{\mathbf{O}}\right) + \mathbf{H}\right) - f\left(\frac{\mathbf{I}}{\mathbf{O}}\right)$$

$$\approx - \sum_{1 \le i < j \le m} (a_i - a_j)(b_i - b_j =)h_{ij}^2 \; - \sum_{1 \le i \le m,\ (m+1) \le j \le d} a_i(b_i - b_j)h_{ij}^2 \qquad \ldots(6)$$

PROOF. From the condition that columns are of norm 1 we have the following equations

$$(1 + h_{jj})^2 + \sum_{i \neq j}^{d} h_{ij}^2 = 1, \quad j = 1, \ldots, m.$$

$$\Rightarrow \quad h_{jj} = -\frac{1}{2}\sum_{i=1}^{d} h_{ij}^2, \quad j = 1, \ldots, m.$$

Setting the cross product of j-th and the k-th columns equal to zero we get

$$\sum_{i \neq j,k} h_{ij}h_{ik} + (1 + h_{jj})h_{jk} + (1 + h_{kk})h_{kj} = 0$$

Substituting

$$h_{jj} = -\frac{1}{2}\sum_{i=1}^{d} h_{ij}^2 \text{ and } h_{kk} = -\frac{1}{2}\sum_{i=1}^{d} h_{ik}^2$$

in the above equation we get

$$\begin{aligned} h_{jk} &= -h_{kj} - \sum_{i \neq j,k} h_{ij}h_{ik} \\ \Rightarrow \quad h_{jk}^2 &\approx h_{kj}^2 \end{aligned}$$

Now

$$
\begin{aligned}
&f\left(\begin{pmatrix} \mathrm{I} \\ \hline \mathrm{O} \end{pmatrix} + \mathrm{H}\right) - f\begin{pmatrix} \mathrm{I} \\ \hline \mathrm{O} \end{pmatrix} \\
&= a_1\left\{b_1(h_{11}+1)^2 + b_2h_{21}^2 + \cdots + b_dh_{d1}^2\right\} \\
&\quad + a_2\left\{b_1h_{12}^2 + b_2(h_{22}+1)^2 + \cdots + b_dh_{d2}^2\right\} \\
&\quad \vdots \\
&\quad + a_m\left\{b_1h_{1m}^2 + b_2h_{2m}^2 + \cdots + b_m(h_{mm}+1)^2 + \cdots + b_dh_{dm}^2\right\} \\
&\quad - \sum a_ib_i
\end{aligned}
$$

Writing $\sum a_ib_i$ as $\sum a_ib_i\left\{(h_{ii}+1)^2 + \sum_{j\neq i} h_{ij}^2\right\}$ the above expression is equal to

$$
\begin{aligned}
&a_1h_{21}^2(b_2-b_1) + a_1h_{31}^2(b_3-b_1) + \cdots + a_1h_{d1}^2(b_d-b_1) \quad + \\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&+ \quad a_mh_{1m}^2(b_1-b_m) + a_mh_{2m}^2(b_2-b_m) + \cdots + a_mh_{dm}^2(b_d-b_m)
\end{aligned}
$$

Using $h_{jk}^2 \approx h_{kj}^2$ for $j < k$, the above is approximately equal to

$$
-\sum_{1\le i<j\le m} (a_i-a_j)(b_i-b_j)h_{ij}^2 \quad - \sum_{1\le i\le m,\ (m+1)\le j\le d} a_i(b_i-b_j)h_{ij}^2
$$

▱

The following lemma and the subsequent theorem use the notations of Muirhead (1982). The symbol $(d\mathbf{X})$ denotes the exterior product of the elements of $\mathbf{X}$. Exterior product is denoted by the symbol $\bigwedge$. It is a non-commutative binary operation like the ordinary product but has the property $a \bigwedge b = -b \bigwedge a$, in particular $a \bigwedge a = 0$.

LEMMA 3.3. *In the neighborhood of* $\mathbf{Q}_1 = \begin{pmatrix} \mathrm{I} \\ \hline \mathrm{O} \end{pmatrix}$, *the following is true*

$$
(\mathbf{Q}_1'd\mathbf{Q}_1) = \prod_{i=2}^{m}\prod_{j<i}^{m-1} dq_{ij} \prod_{i=m+1}^{d}\prod_{j=1}^{m} dq_{ij} \qquad \ldots(7)
$$

PROOF. By definition

$$
\begin{aligned}
(\mathbf{Q}_1'd\mathbf{Q}_1) &= \bigwedge_{i=1}^{m}\bigwedge_{j=i+1}^{d} \mathbf{q}_j'd\mathbf{q}_i \\
&= \bigwedge_{i=1}^{m}\bigwedge_{j=i+1}^{d}\sum_{k=1}^{d} q_{kj}dq_{ki}.
\end{aligned}
$$

In the neighborhood of $\left(\frac{\mathbf{I}}{\mathbf{0}}\right)$ the above expression reduces to

$$dq_{21}\bigwedge dq_{31}\bigwedge\cdots\bigwedge q_{m1}\bigwedge\Big(\sum_{k=m+1}^{d} q_{k\,m+1}dq_{k1}\Big)\cdots\bigwedge\Big(\sum_{k=m+1}^{d} q_{kd}dq_{k1}\Big)\bigwedge$$
$$dq_{32}\bigwedge\cdots\bigwedge q_{m2}\bigwedge\Big(\sum_{k=m+1}^{d} q_{k\,m+1}dq_{k2}\Big)\cdots\bigwedge\Big(\sum_{k=m+1}^{d} q_{kd}dq_{k2}\Big)\bigwedge$$
$$\ddots$$
$$dq_{m\,m-1}\bigwedge\Big(\sum_{k=m+1}^{d} q_{k\,m+1}dq_{k\,m}\Big)\cdots\bigwedge\Big(\sum_{k=m+1}^{d} q_{kd}dq_{k\,m}\Big)$$

$$= \bigwedge_{i=2}^{m}\bigwedge_{j<i}^{m-1} dq_{ij}\bigwedge\Big(\sum_{k=m+1}^{d} q_{k\,m+1}dq_{k1}\Big)\cdots\bigwedge\Big(\sum_{k=m+1} q_{kd}dq_{k1}\Big)\bigwedge$$
$$\Big(\sum_{k=m+1}^{d} q_{k\,m+1}dq_{k2}\Big)\cdots\bigwedge\Big(\sum_{k=m+1}^{d} q_{kd}dq_{k2}\Big)\bigwedge$$
$$\vdots$$
$$\Big(\sum_{k=m+1}^{d} q_{k\,m+1}dq_{k\,m}\Big)\cdots\bigwedge\Big(\sum_{k=m+1}^{d} q_{kd}dq_{k\,m}\Big)$$
$$= \bigwedge_{i=2}^{m}\bigwedge_{j<i}^{m-1} dq_{ij}\bigwedge\Big(|\mathbf{Q}_{22}|\bigwedge_{i=m+1}^{d} dq_{i1}\Big)\bigwedge\Big(|\mathbf{Q}_{22}|\bigwedge_{i=m+1}^{d} dq_{i2}\Big)\bigwedge\cdots\Big(|\mathbf{Q}_{22}|\bigwedge_{i=m+1}^{d} dq_{im}\Big)$$
$$= \bigwedge_{i=2}^{m}\bigwedge_{j<i}^{m-1} dq_{ij}\bigwedge\Big(|\mathbf{Q}_{22}|^{m}\bigwedge_{i=m+1}^{d}\bigwedge_{j=1}^{m} dq_{ij}\Big)$$

Now the determinant of $\mathbf{Q}_{22}$ is 1 because it is $(d-m)\times(d-m)$ orthogonal matrix. Hence

$$(\mathbf{Q}_1' d\mathbf{Q}_1) = \prod_{i=2}^{m}\prod_{j<i}^{m-1} dq_{ij}\prod_{i=m+1}^{d}\prod_{j=1}^{m} dq_{ij}.$$

$\square$

We now state the main results of this section.

THEOREM 3.1. *The marginal density of X, where $X' = (\bar{X}'_1,\ldots,\bar{X}'_p)$, under H_m for a given $\mathbf{D}^*_m$ may be expressed as*

$$p(X|H_m) = C(X)\prod_{k=1}^{m}(1-n\sigma^*_{kk})^{\frac{p-1}{2}}\int_{\mathbf{Q}\in O(d)} \mathbf{e}^{\frac{n(p-1)}{2}\mathbf{tr}(\mathbf{D}^*_m\mathbf{Q}'\mathbf{L}\mathbf{Q})}\,(d\mathbf{Q}) \quad \ldots(8)$$

where $O(d)$ is the space of orthogonal matrices of dimension d, $(d\mathbf{Q})$ denotes normalized Haar measure and $C(X)$ is a factor common to $p(X|H_m)$ for $m = 0, \dots, d$.

THEOREM 3.2.

$$p(X|H_m) \sim \qquad \dots (9)$$

$$C(X)\frac{2^m \mathrm{Vol}O(d-m)}{\mathrm{Vol}O(d)} \prod_{k=1}^{m} (1 - n\sigma_{kk}^*)^{\frac{p-1}{2}} \mathbf{e}^{\frac{n(p-1)}{2}\sum_1^m \sigma_{kk}^* \lambda_k} \times \qquad \dots (10)$$

$$\prod_{1 \le i < j \le m} \left\{\tfrac{n(p-1)}{2\pi}(\sigma_{kk}^* - \sigma_{ll}^*)(\lambda_k - \lambda_l)\right\}^{-\frac{1}{2}} \times \prod_{1 \le i \le m,\ (m+1) \le j \le d} \left\{\tfrac{n(p-1)}{2\pi}\sigma_{kk}^*(\lambda_k - \lambda_l)\right\}^{-\frac{1}{2}}$$

PROOF OF THEOREM 3.1. For fixed $\Sigma_{\bar{\mu}}, \Sigma_m^*$

$$p(X|H_m) \propto \int \cdots \int \prod_{k=1}^{m} (1 - n\sigma_{kk}^*)^{\frac{p-1}{2}} \mathbf{e}^{\frac{n}{2}\mathbf{tr}\{2\sum \psi_i Y_i' + 2p\bar{\mu}\bar{X}'\}} \pi_m^*(d\Psi)\pi_m^*(d\bar{\mu})$$

$$= \prod_{k=1}^{m} (1 - n\sigma_{kk}^*)^{\frac{p-1}{2}} \mathbf{e}^{\frac{(np)^2}{2}\bar{X}'\ \Sigma_{\bar{\mu}}\bar{X} + \frac{(n)^2}{2} Y'\mathbf{P}\otimes\Sigma_m^* Y}$$

where $Y' = (Y_1, \dots, Y_p)$. Now

$$\begin{aligned}
& Y'\mathbf{P} \otimes \Sigma_m^* Y \\
= \quad & (1 - p^{-1}) \sum Y_i' \Sigma_m^* Y_i - p^{-1} \sum \sum_{i \ne j} Y_i' \Sigma_m^* Y_j \\
= \quad & (1 - p^{-1})\mathbf{tr}\Sigma_m^* \sum Y_i Y_i' - p^{-1}\mathbf{tr} \sum Y_i' \Sigma_m^* \sum_{i \ne j} Y_j \\
= \quad & (1 - p^{-1})\mathbf{tr}\Sigma_m^* \sum Y_i Y_i' - p^{-1}\mathbf{tr} \sum Y_i' \Sigma_m^* (-Y_i) \\
= \quad & (1 - p^{-1})\mathbf{tr}\Sigma_m^* \sum Y_i Y_i' + p^{-1}\mathbf{tr} \sum \sum \Sigma_m^* Y_i Y_i' \\
= \quad & \mathbf{t} = \mathbf{r}\Sigma_m^* \sum Y_i Y_i'.
\end{aligned}$$

So for a fixed Σ_m^*,

$$p(X|H_m) = C(X) \prod_{k=1}^{m} (1 - n\sigma_{kk}^*)^{\frac{p-1}{2}} \mathbf{e}^{\frac{n^2}{2}\mathbf{tr}\Sigma_m^* \sum Y_i Y_i'}$$

$$= C(X) \prod_{k=1}^{m} (1 - n\sigma_{kk}^*)^{\frac{p-1}{2}} \mathbf{e}^{\frac{n^2}{2}\mathbf{tr}\Sigma_m^*(\frac{p-1}{n}\mathbf{ZLZ}')}$$

$$= C(X) \prod_{k=1}^{m} (1 - n\sigma_{kk}^*)^{\frac{p-1}{2}} \mathbf{e}^{\frac{n(p-1)}{2}\mathbf{trUD}_m^*\mathbf{U}'\mathbf{ZLZ}'}$$

$$= C(X) \prod_{k=1}^{m} (1 - n\sigma_{kk}^*)^{\frac{p-1}{2}} \mathbf{e}^{\frac{n(p-1)}{2}\mathbf{trQD}_m^*\mathbf{Q}'\mathbf{L}}.$$

Integrating with respect to the normalized Haar measure $(d\mathbf{Q})$

$$p(X|H_m, \mathbf{D}_m^*) = C(X)\prod_{k=1}^{m}(1-n\sigma_{kk}^*)^{\frac{p-1}{2}} \int_{\mathbf{Q}\in O(d)} e^{\frac{n(p-1)}{2}\mathbf{trQD}_m^*\mathbf{Q}'\mathbf{L}} \quad (d\mathbf{Q}).$$

▱

PROOF OF THEOREM 3.2. Let

$$\mathbf{D}_m^* = \left(\begin{array}{c|c} \mathbf{D}_{nz}^* & \mathbf{O} \\ \hline \mathbf{O} & \mathbf{O} \end{array}\right).$$

Then

$$\mathbf{trQD}_m^*\mathbf{Q}'\mathbf{L} = \mathbf{trD}_{nz}^*\mathbf{Q}_1'\mathbf{LQ}_1.$$

By lemma 9.5.3 (Muirhead 1982) we can write

$$\begin{aligned} &\int_{\mathbf{Q}\in O(d)} e^{\frac{n(p-1)}{2}\mathbf{trQD}_m^*\mathbf{Q}'\mathbf{L}} \quad (d\mathbf{Q}) \\ =\ & \frac{1}{\mathrm{Vol}O(d)}\int_{\mathbf{Q}\in O(d)} e^{\frac{n(p-1)}{2}\mathbf{trQD}_m^*\mathbf{Q}'\mathbf{L}}(\mathbf{Q}'d\mathbf{Q}) \\ =\ & \frac{1}{\mathrm{Vol}O(d)}\int_{\mathbf{Q}_1\in V_{m,d}}\int_{\mathbf{K}\in O(d-m)} e^{\frac{n(p-1)}{2}\mathbf{trD}_{nz}^*\mathbf{Q}_1'\mathbf{LQ}_1}(\mathbf{K}'d\mathbf{K})(\mathbf{Q}_1'd\mathbf{Q}_1) \\ =\ & \frac{\mathrm{Vol}O(d-m)}{\mathrm{Vol}O(d)}\int_{\mathbf{Q}_1\in V_{m,d}} e^{\frac{n(p-1)}{2}\mathbf{trD}_{nz}^*\mathbf{Q}_1'\mathbf{LQ}_1}(\mathbf{Q}_1'd\mathbf{Q}_1) \end{aligned}$$

where $\mathrm{Vol}O(d)$ is the volume of the space of d dimensional orthogonal matrices and is equal to (Murihead 1982)

$$\mathrm{Vol}O(d) = \frac{2^d\pi^{\frac{d^2}{2}}}{\Gamma_d(\frac{d}{2})}. \qquad \ldots(11)$$

Let $f(\mathbf{Q}_1) = \mathbf{trD}_{nz}^*\mathbf{Q}_1'\mathbf{LQ}_1$. Then from lemma 3.1 it follows that maximum of $f(\mathbf{Q}_1)$ is $\sum_{i=1}^{m}\sigma_{ii}^*\lambda_i$ and it is attained at each of the 2^m values of $\mathbf{Q}_1$ of the form as given in (4). So we apply the Laplace's method by considering neighborhoods around the 2^m points of maximum of f. Using symmetry the above integral is approximated by

$$\frac{2^m\mathrm{Vol}O(d-m)}{\mathrm{Vol}O(d)}\int_{\mathbf{Q}_1\in N(*)} e^{\frac{n(p-1)}{2}\mathbf{trD}_{nz}^*\mathbf{Q}_1'\mathbf{LQ}_1}(\mathbf{Q}_1'd\mathbf{Q}_1)$$

where $N(*)$ denotes a neighborhood around the matrix given in expression (5) in lemma 3.2.

From lemma 3.2 f can be represented locally in terms of $m^2 + m(d-m)$ independent variables and using lemma 3.3 we can write the above as

$$\frac{2^m \mathrm{Vol}O(d-m)}{\mathrm{Vol}O(d)} e^{\frac{n(p-1)}{2}\sum \sigma^*_{kk}\lambda_k} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{\frac{-n(p-1)}{2}(\sigma^*_{kk}-\sigma^*_{ll})(\lambda_k-\lambda_l)q_{kl}^2} \prod_{1\le k<l\le m} dq_{kl} \times$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{\frac{-n(p-1)}{2}(\sigma^*_{kk}-\sigma^*_{ll})(\lambda_k-\lambda_l)q_{kl}^2} \prod_{1\le k\le m,\ (m+1)\le l\le d} q_{kl}$$

$$= \frac{2^m \mathrm{Vol}O(d-m)}{\mathrm{Vol}O(d)} e^{\frac{n(p-1)}{2}\sum \sigma^*_{kk}\lambda_k} \prod_{1\le k<l\le m} \left\{\tfrac{n(p-1)}{2\pi}(\sigma^*_{kk}-\sigma^*_{ll})(\lambda_k-\lambda_l)\right\}^{-\frac{1}{2}} \times$$

$$\prod_{1\le k\le m,\ (m+1)\le l\le d} \left\{\tfrac{n(p-1)}{2\pi}\sigma^*_{kk}(\lambda_k-\lambda_l)\right\}^{-\frac{1}{2}}.$$

Using the previous theorem the assertion (9) follows. ▱

REMARK 3.1. The approximation (local) is good for large $n(p-1)$. It is necessary to assume that the σ^*_{kk}'s are all different so that there are finite number of points of maximum. Laplace's method is then applicable in the neighborhood of finite number of peaks of f. If the σ^*_{kk}'s are not different then there is no sharp peak and the integral can not be approximated well by this method.

REMARK 3.2. One can also prove theorem (3.2) by using proof of theorem (9.5.4) of Muirhead (1982) and some of the results it refers to. However, our calculations seem new and simpler, especially in the handling of the local maxima and integration with respect to the Haar measure locally. Also the application in this context is new.

So the Bayes rule may be expressed as follows

$$\text{accept } H_r \text{ over } H_s \text{ if} \frac{p(X|H_r, \mathbf{D}^*_r)}{p(X|H_s, \mathbf{D}^*_s)} = > \frac{\Pi_s}{\Pi_r} \qquad \ldots(12)$$

where Π_k is the prior probability of H_k.

For moderately large $n(p-1)$ the left hand side of the above inequality is well approximated by

$$\frac{2^{(r-s)}\mathrm{Vol}O(d-r)}{\mathrm{Vol}O(d-s)} \prod_{k=s+1}^{r} (1-n\sigma^*_{kk})^{\frac{p-1}{2}} e^{\frac{n(p-1)}{2}\sum_{i=s+1}^{r}\sigma^*_{kk}\lambda_k} \text{ if } r > s. \qquad \ldots(13)$$

For the three dimensional case the marginal probabilities are as follows

$p(X|H_3) =$

$$\frac{C(X)8(2\pi)^{3/2}\{(1-n\sigma^*_{11})(1-n\sigma^*_{22})(1-n\sigma^*_{33})\}^{\frac{p-1}{2}} e^{\frac{n(p-1)}{2}(\sigma^*_{11}\lambda_1+\sigma^*_{22}\lambda_2+\sigma^*_{33}\lambda_3)}}{\mathrm{Vol}O(3)\{n(p-1)\}^{3/2}\{(\sigma^*_{11}-\sigma^*_{22})(\lambda_1-\lambda_2)(\sigma^*_{11}-\sigma^*_{33})(\lambda_1-\lambda_3)(\sigma^*_{22}-\sigma^*_{33})(\lambda_2-\lambda_3)\}^{\frac{1}{2}}} \qquad \ldots(14)$$

$p(X|H_2) =$

$$\frac{C(X)8(2\pi)^{3/2}\left\{(1-n\sigma_{11}^*)(1-n\sigma_{22}^*)\right\}^{\frac{p-1}{2}} e^{\frac{n(p-1)}{2}(\sigma_{11}^*\lambda_1+\sigma_{22}^*\lambda_2)}}{\mathrm{Vol}O(3)\left\{n(p-1)\right\}^{3/2}\left\{(\sigma_{11}^*-\sigma_{22}^*)(\lambda_1-\lambda_2)(\sigma_{11}^*)(\lambda_1-\lambda_3)(\sigma_{22}^*)(\lambda_2-\lambda_3)\right\}^{\frac{1}{2}}}. \quad \ldots(15)$$

$$p(X|H_1) = \frac{C(X)2(2\pi)(4\pi)\left\{(1-n\sigma_{11}^*)\right\}^{\frac{p-1}{2}} e^{\frac{n(p-1)}{2}(\sigma_{11}^*\lambda_1)}}{\mathrm{Vol}O(3)n(p-1)\left\{(\sigma_{11}^*)(\lambda_1-\lambda_2)(\sigma_{11}^*)(\lambda_1-\lambda_3)\right\}^{\frac{1}{2}}}. \quad \ldots(16)$$

4. Simulation study for the trivariate case

For detailed study of the trivariate case we choose $p = 6$. We avoid the values 4 or 5 because a preliminary study showed the error probabilities under different hypotheses are too high to be of much practical use. We keep the sample size n= 10. The covariance matrix for μ is taken as **I** under all the 4 hypotheses which of course, as mentioned earlier, does not affect the rule.

Unlike the case of $d = 2$ in De and Ghosh (1994), here the values for σ_{kk}^*'s are chosen in conjunction with those for $\Pi(H_k)$'s . We want the σ_{kk}^*'s to be near 1 but all different. We restricted the choice to the set { 1.5 1.4 1.3 1.2 1.1 1.0 0.9 0.8 }. As the joint distribution of $\lambda_1, \lambda_2, \lambda_3$ is a bit unwieldy, we look at their marginal distributions under each hypothesis for various combinations of σ_{kk}^*'s. A rule of the following form emerged after studying the marginal distributions

$$\begin{aligned} \text{accept } H_3 &\text{ if } \lambda_3 > a \\ \text{accept } H_2 &\text{ if } \lambda_2 > b \text{ and } \lambda_3 < a \\ \text{accept } H_1 &\text{ if } \lambda_1 > c \text{ and } \lambda_2 < b \text{ and } \lambda_3 < a \\ \text{accept } H_0 &\text{ otherwise} \end{aligned}$$

where a, b and c are in the neighborhoods of 0.9, 1.9 and 3.5 respectively. The Bayes rule that follows from the previous section may be written as

accept H_3 over H_2 if

$$\lambda_3 > \ln\left(\frac{\Pi_2}{\Pi_3 c_{32}}\right)\frac{2}{n(p-1)\sigma_{33}^*}$$

$$\text{where } c_{32} = \frac{(1-n\sigma_{33}^*)^{\frac{p-1}{2}}}{\left\{\left(1-\frac{\sigma_{33}^*}{\sigma_{11}^*}\right)\left(1-\frac{\sigma_{33}^*}{\sigma_{22}^*}\right)\right\}^{1/2}};$$

accept H_2 over H_1 if

$$\lambda_2 > \ln\left(\frac{\Pi_1}{\Pi_2 c_{21}}\right)\frac{2}{n(p-1)\sigma_{22}^*}$$

$$\text{where } c_{21} = \sqrt{\frac{2}{\pi}}\sqrt{\frac{\sigma_{11}^*}{p-1}}\frac{(1-n\sigma_{22}^*)^{\frac{p-1}{2}}}{\{n(\sigma_{11}^*-\sigma_{22}^*)\sigma_{22}^*(\lambda_2-\lambda_3)\}^{1/2}};$$

accept H_1 over H_0 if

$$\lambda_1 > \ln\left(\frac{\Pi_0}{\Pi_1 c_{10}}\right)\frac{2}{n(p-1)\sigma_{11}^*}$$

$$\text{where } c_{10} = \frac{(1-n\sigma_{11}^*)^{\frac{p-1}{2}}}{n(p-1)\left\{\sigma_{11}^{*2}(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)\right\}^{1/2}}.$$

As in the bivariate case in De and Ghosh (1994) some proxy values for $(\lambda_k - \lambda_l)$ on the right hand side of the above equations are used. Since $p = 6$ is moderately large, it turns out that the factors involving the differences $(\lambda_k - \lambda_l)$ has little influence on the Bayes ratio. This implies that we would not need to refine the ad hoc rule as was done in the bivariate case through the $(\lambda_k - \lambda_l)$'s.

Different sets of values for σ_{kk}^*'s and Π_l's are used to determine the Bayes rule and their performance is studied. Combinations that lead to either too high or too low error probability under one hypothesis are rejected. After repeated trials and adjustments the following combination appeared to be satisfactory.

$$n\sigma_{11}^* = \frac{12}{13}; \quad n\sigma_{22}^* = \frac{11}{12}; \quad n\sigma_{33}^* = \frac{9}{10}; \qquad \ldots(17)$$

$$\Pi_0 = 0.14; \quad \Pi_1 = 0.25; \quad \Pi_2 = 0.37; \quad \Pi_3 = 0.24; \qquad \ldots(18)$$

and the ad hoc rule is finalized as

$$\text{accept } H_3 \text{ if } \lambda_3 \geq 1; \qquad \ldots(19)$$

$$\text{accept } H_2 \text{ if } \lambda_2 \geq 2.1 \text{ and } \lambda_3 < 1;$$

$$\text{accept } H_1 \text{ if } \lambda_1 \geq 3.5 \text{ and } \lambda_2 < 2.1 \text{ and } \lambda_3 < 1;$$

$$\text{accept } H_0 \text{ otherwise.}$$

We do not give the joint distribution of the eigenvalues of $\frac{np}{p-1}\mathbf{\Delta}$ and that of $\frac{\mathbf{H}}{p-1}$ and also omit the table of Bayes ratio because of their voluminous size. Table-1 shows the performance of the ad hoc and the Bayes rules for various values of p.

TABLE 1. PERFORMANCE OF AD-HOC AND BAYES RULES FOR VARIOUS VALUES OF p.

	p			
	5	6	7	8
Overall error ad-hoc	0.28	0.18	0.15	0.13
Overall error Bayes	0.29	0.18	0.13	0.10
Overall difference	0.12	0.03	0.08	0.11

The two rules perform equally well for $p = 6$ because the ad hoc rule was constructed to mimic the Bayes rule for $p = 6$. The performance of the Bayes rule improves very fast as p increases. The ad hoc rule also shows improvement in error probabilities with the increase in p but the gap between the ad hoc rule and the Bayes rule widens. So for each value of p a new set of cut-off values may be determined to reduce this difference.

Frequentist performance was checked in some detail, in the same manner as in De and Ghosh (1994). Not only the average errors at various (e_1, e_2, e_3) of the ad hoc rule and the Bayes rule are comparable but the two rules agree widely (mostly, more than 94% of the times).

Also it was found that under H_3, except where either e_2 or e_3 or both are small, probability of a wrong decision favouring H_1 or H_0 is negligible. Under H_2, except where e_2 is small, errors are mainly due to the choice of the hypothesis H_3. Most of the wrong decisions under H_1 lead to H_2. These provide evidence that the rules behave consistently in the sense that mostly adjacent hypotheses are chosen in case of errors.

The performance of the ad hoc rule seemed satisfactory when it was compared with the most powerful test (based on eigenvalues of the between sum of squares and product matrix) derived from the simulated probability tables. For the various sets of (e_1, e_2, e_3) we tried, most of the times the ad hoc rule performed at least 90% as efficiently as the most powerful test.

References

DE, A.K. (1993). A decision theoretic approach to the dimensionality problem in the context of multivariate analysis of variance. *Ph.D. Thesis*, Purdue University.

DE, A.K. and GHOSH, J.K. (1994). A decision rule for dimension in the context of MANOVA. To appear in *Sankhyā*.

FUJIKOSHI, Y. (1974). The likelihood ratio tests for the dimensionality of regression coefficients. *Jour. Multivariate Analysis*, **4**, 327-340.

MUIRHEAD, R.J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

RAO, C.R. (1973). *Linear Statistical Inference and Its Applications.* Second edition, Wiley, New York.

---- (1985). Tests for dimensionality and interactions of mean vectors under general and reducible covariance structures. *Jour. Multivariate Statistics*, **16**, 173-184.

SHEN, W. and SINHA, B.K. (1991). Statistical tests for structural relationship. *Statistics and Probability Letters*, **12**, 545-559.

SCHOOL OF NURSING
EMORY UNIVERSITY
531 ASBURY CIRCLE
ATLANTA GA 30322
U.S.A.

INDIAN STATISTICAL INSTITUTE
203, B.T. ROAD
CALCUTTA - 700 035
INDIA