Sankhyā: The Indian Journal of Statistics
1995, Volume 57, Series A, Pt. 2, pp. 256–266

A DECISION RULE FOR DIMENSION IN THE CONTEXT OF MANOVA

By ANINDYA K. DE
Purdue University
and
JAYANTA K. GHOSH
Indian Statistical Institute and *Purdue University*

SUMMARY. An ad hoc rule is proposed to determine the dimension of the space generated by the mean vectors centered at the origin. Under normality, the proposed ad hoc rule is close to being Bayes for a prior which seems to distribute the total mass in a reasonable way. The ad hoc rule is then slightly modified to incorporate some of the finer features of the Bayes rule without losing good frequentist features. Simulation studies show that the rule nearly attains the Bayes risk and so is nearly admissible. It has other attractive frequentist features.

1. INTRODUCTION

In the context of Multivariate Analysis of Variance (MANOVA) it is of interest to know the dimension of the space generated by the mean vectors centered at zero. This often provides insight by identifying a structural relationship among the parameters if there is any.

An interesting application is discussed in Cochran (1943) where the problem is to compare different methods or *scales* of measurement. For example, in comparing two different methods of assaying vitamins with independent experiments one likes to know if the scales are linearly related, (see Cochran (1943) for details). If they do lie closely around a straight line then the slope of the line is meaningful.

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 $\boldsymbol{\Omega}$ 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 $\boldsymbol{\mu} = p^{-1} \sum_{i=1}^{p} \boldsymbol{\mu}_i$. We have a multiple decision problem with possible decisions (or actions): $a_0, a_1, \ldots, a_d$ where a_i denotes the decision or action and rank of $\mathbf{M}$ is i. For simplicity we will focus on the case $d = 2$ and assume available decisions are a_1 and a_2.

Our goal in this paper is to obtain a multiple decision rule which has some optimality properties from frequentist point of view and is also attractive to a

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

Bayesian. We propose an ad hoc rule first, then show that it is close to being a proper Bayes rule under suitable choice of priors and finally refine the ad hoc rule in the light of the Bayes rule. We have attempted to find the ad hoc rule and the prior parameters/hyper-parameters in an iterative way. Starting with an attractive ad hoc rule, a suitable prior is found with respect to which the ad hoc rule is approximately Bayes. Then the Bayes rule and the ad hoc rule are perturbed until a satisfactory pattern of the error probabilities is found and at the same time the prior weights are distributed on the parameter space in a 'reasonable' way (see §§2.2 and table 1). Care is taken to ensure that the perturbed or the refined ad hoc rule is both approximately Bayes with respect to the chosen prior and has good frequentist properties. We shall also study its risk function at various parametric points and check that no substantial improvement is possible.

Apart from being simple, ad hoc rule can be extended very easily to the following cases where a) the common covariance matrix is unknown, b) the covariance matrices are different but known and c) the covariance matrices are different and unknown. Simulation studies showed that the performance is quite satisfactory under these situations also, but deriving the analogous Bayes rule appears to be much more difficult. A sample of simulation results for the case (c), as stated above, is given in section 3 of this paper.

There is a substantial literature on this subject, e.g. Fisher (1938), Anderson (1951), Rao (1973, 1985), Fujikoshi (1974). The concern in these papers is with the derivation of the likelihood ratio test or its asymptotic distribution. In Baiqi (1992) the concern is with consistency of an estimate of the rank of the regression coefficient matrix. Somewhat more similar in spirit to that of ours is an interesting paper of Shen and Sinha (1991), where a proper prior and the corresponding Bayes solution are exhibited. However since the prior puts weight on certain particular lower dimensional spaces, the Bayes solution, though admissible, is unattractive for practical use.

In section 2 we introduce the ad hoc rule, construct the prior and derive the Bayes rule. A simulation study is presented in section 3.

2. Decision rules

2.1. Ad hoc rule. Let $X_{i1}, X_{i2}, \ldots, X_{in}$ be the set of d dimensional vectors observed from population $i, i = 1, 2, \ldots, p$. Without loss of generality we will assume the dispersion matrix to be $\mathbf{I}$. Between sum of squares and products matrix $\mathbf{H} = n\sum_{i=1}^{p}(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})' = n\sum_{i=1}^{p} Y_i Y_i'$ and the within sum of squares and products matrix $\mathbf{E} = \sum_{i=1}^{p}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$ where $\bar{X}_i = n^{-1}\sum_{j=1}^{n} X_{ij}$, $\bar{X} = p^{-1}\sum_{i=1}^{p}\bar{X}_i$, $Y_i = \bar{X}_i - \bar{X}$. We define $\boldsymbol{\Delta} = p^{-1}\sum_{i=1}^{p}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})' = p^{-1}\sum_{i=1}^{p}\boldsymbol{\psi}_i\boldsymbol{\psi}_i'$, where $\boldsymbol{\psi}_i = \boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}$. We know

$$E\left(\frac{\mathbf{H}}{p-1}\right) = \mathbf{I} + \frac{np}{p-1}\boldsymbol{\Delta} \qquad \ldots (1)$$

Let d^* be the number of eigenvalues of $\mathbf{H}/(p-1)$ which are greater than 1. Then d^* may be taken as an estimate of the rank of $\mathbf{\Delta}$ which is in turn equal to the dimension of the basis of the column space of M. In this vein a decision rule of the following form is proposed. Let H_k be the hypothesis that the rank of $\mathbf{\Delta}$ is k i.e. the p mean vectors lie in a k dimensional subspace. The rule is:

Choose the hypothesis H_k if

$$\lambda_r < c(d,p,r) \text{ for } r = k+1,\ldots,d \text{ and } \lambda_k > c(d,p,k) \qquad \ldots (2)$$

where c's are constants to be chosen suitably. If $c(d,p,r) = 1$ for all r then the rule exactly matches with the above approach but some allowance must be made for the sampling variation of the eigenvalues. So instead of choosing d^* we shall permit ourselves more flexibility in the choice of c's.

It may be noted that we may determine c's under the assumption of multinormality of X's because even under non-normality $\sqrt{n}\bar{X}_i$ will behave like a normal for large n, on which our rule is based.

2.2. Bayes rule. As explained in the introduction, for the sake of simplicity, we assume that the populations are bivariate. Let $X'_{ij} = (X^{(1)}_{ij}, X^{(2)}_{ij})$ be the jth observation from the ith population, $i = 1,2,\ldots,p; j = 1,2,\ldots,n$. We assume $X_{ij} \sim N_2(\boldsymbol{\mu}_i, \mathbf{I}_2)$ where $\boldsymbol{\mu}'_i = (\mu^{(1)}_i, \mu^{(2)}_i)$. X_{ij} and $X_{i'j'}$ are independent if $(i,j) \neq (i',j')$. The joint density of the sufficient statistics $(\bar{X}_1,\ldots,\bar{X}_p)$ which we will refer to as X' for convenience, is multivariate normal and the density is given by

$$C(Y,\bar{X})\, \mathbf{e}^{-\frac{n}{2}\mathbf{tr}\left\{-2\sum \boldsymbol{\psi}_i Y'_i + \sum \boldsymbol{\psi}_i \boldsymbol{\psi}'_i - 2p\bar{\boldsymbol{\mu}}\bar{X}' + p\bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}'\right\}} \qquad \ldots (3)$$

where $Y' = (Y_1,\ldots,Y_p)$. The factor C will get cancelled from the numerator and the denominator of the Bayes factor.

We wish to introduce a prior given H_k is true, $k = 1,2$, such that the part depending on $\bar{X}$ cancels in the numerator and the denominator of the Bayes factor. The way this can be done is to put the same prior on $\bar{\boldsymbol{\mu}}$ under H_1 and H_2. On the other hand the prior for $\boldsymbol{\psi}$'s will reflect the difference between H_1 and H_2. *These goals cannot be achieved with identically and independently distributed normal $\boldsymbol{\mu}_i$'s under H_1 and H_2 for the simple reason that $\bar{\boldsymbol{\mu}}$ will then have different distributions under H_1 and H_2.* So we follow a different route. We specify separately a prior for $\bar{\boldsymbol{\mu}}$, same under H_1 and H_2 and a prior for $\boldsymbol{\Psi}' = (\boldsymbol{\psi}_1,\ldots,\boldsymbol{\psi}_p)$, different under H_1 and H_2. Under this specification $\boldsymbol{\mu}_i$'s will then be exchangeable. Details are now given below.

We assume $\bar{\boldsymbol{\mu}}$ and $\boldsymbol{\Psi}$ are independent and both have normal distributions. The distribution of $\bar{\boldsymbol{\mu}}$ is assumed to be the same under the hypotheses H_1, H_2. The Bayes factor would thus be free of $\bar{X}$ and would involve only Y_i's.

The prior for $\boldsymbol{\Psi}$ is chosen from the following consideration. If $\mathbf{e}^{-\frac{n}{2}\mathbf{tr}\{-2\sum \boldsymbol{\psi}_i Y'_i\}}$ were the only factor involving $\boldsymbol{\Psi}$ in the expression (3), that is, if the other factor

involving Ψ namely, $e^{-\frac{n}{2}tr\{\sum \psi_i \psi_i'\}}$ were absent, then (3) could be easily integrated with respect to a normal prior for Ψ. Moreover, under an appropriate prior the result would depend only on the eigenvalues of the between sum of squares and products matrix, the calculation being similar to that for deriving the moment generating function of a multivariate normal distribution. So to take care of the factor $e^{-\frac{n}{2}tr\{\sum \psi_i \psi_i'\}}$ we choose a prior which is proportional to $e^{\frac{n}{2}tr\{\sum \psi_i \psi_i'\}}$ times a normal which, as we would see, is again a normal with the same support. The prior for $\bar{\mu}$ would also be chosen from similar considerations. We need to introduce some notations. Let $\pi_1(\cdot)$ and $\pi_2(\cdot)$ denote the prior distributions under H_1 and H_2. As indicated $\pi_k(\cdot)$ would be chosen as $e^{\frac{n}{2}tr\{\sum \psi_i \psi_i'\}}$ times a normal which we would call as $\pi_k^*(\cdot)$, that is

$$\pi_k(\Psi) = K_\psi e^{\frac{n}{2}tr\{\sum \psi_i \psi_i'\}} \pi_k^*(\Psi) \qquad \ldots (4)$$

$$\pi_k(\bar{\mu}) = K_\mu e^{\frac{np}{2}tr\{\sum \bar{\mu}\bar{\mu}'\}} \pi_k^*(\bar{\mu}) \qquad \ldots (5)$$

where K's are constants. While choosing $\pi_k^*(\Psi)$ we incorporate the condition $\mathbf{1}'\Psi = 0$, where $\mathbf{1}'$ is a vector and its elements are all equal to 1, in the following manner.

Let $W_i \overset{iid}{\sim} N_2(O, \Sigma_k^*)$, where Σ_k^* is a 2×2 matrix of rank k. We assume $\psi|\Sigma_k^* \sim W_i - \bar{W}$ under $\pi_k^*(\cdot)$, $k = 1, 2$. We put a prior on the eigenvectors of Σ_k^* as follows. Let the spectral decomposition of Σ_k^* be $\mathbf{U}\mathbf{D}_k^*\mathbf{U}'$ where $\mathbf{U}$ is an orthogonal matrix and $\mathbf{D}_k^*$ is a diagonal matrix, let $\mathbf{D}_k^* = \text{diag}(\sigma_{11}^*, \sigma_{22}^*)$, $\sigma_{11}^* \geq \sigma_{22}^*$. For $\mathbf{D}_1^*$, $\sigma_{22}^* = 0$ and for $\mathbf{D}_2^*$, $\sigma_{22}^* > 0$. The σ_{kk}^*'s must be such that the corresponding σ_{kk}'s are positive, vide Proposition 2.1. We take $\mathbf{D}_k^*$ to be fixed and let the probability law of $\mathbf{U}$ be the Haar measure in the space of 2×2 orthogonal matrices. The 2×2 orthogonal matrix $\mathbf{U}$ can be parametrized by a single parameter θ, $\mathbf{U} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ and our assumption of Haar measure on $\mathbf{U}$ translates to uniform distribution over $[0, 2\pi]$ for θ, (see Muirhead (1982)). We may point out here that the direction in which the mean vectors lie under H_1 and the direction of the major axis of the elliptical spread under H_2 get a uniform prior. Unlike Shen and Sinha (1991) no particular direction gets a positive mass under H_1 which should be the case in almost all real life situations where there is no prior knowledge in favour of some special directions. This construction leads to prior distributions for Ψ and μ as given in the following proposition which we state without proof.

Proposition 1. *Under H_1, the prior distribution of Ψ given $\Sigma_1^* = \mathbf{U}diag(\sigma_{11}^*, 0)\mathbf{U}'$ is normal with mean $\mathbf{0}$ and covariance $\mathbf{P} \otimes \Sigma_1$ where $\Sigma_1 = \mathbf{U}diag(\sigma_{11}, 0)\mathbf{U}'$ and $\sigma_{11} = (\sigma_{11}^{*^{-1}} - n)^{-1}$.*

Under H_2, the prior distribution of Ψ given $\Sigma_2^* = \mathbf{U}diag(\sigma_{11}^*, \sigma_{22}^*)\mathbf{U}'$ is normal with mean $\mathbf{0}$ and covariance $\mathbf{P} \otimes \Sigma_2$ where $\Sigma_2 = \mathbf{U}diag(\sigma_{11}, \sigma_{22})\mathbf{U}'$ and $\sigma_{kk} = (\sigma_{kk}^{*^{-1}} - n)^{-1}$, $k = 1, 2$.

Under both H_1 and H_2, the prior distribution of $\underline{\mu}$ given a positive definite matrix $\Sigma^*_{\underline{\mu}} = \mathbf{U}'$ diag$(\sigma^*_{\mu 11}, \sigma^*_{\mu 22})$ $\mathbf{U}$ is normal with mean $\mathbf{O}$ and covariance $\Sigma_{\underline{\mu}}$ where $\Sigma_{\underline{\mu}} = \mathbf{U}'$ diag$(\sigma_{\mu 11}, \sigma_{\mu 22})\mathbf{U}$ and $\sigma_{\mu kk} = (\frac{1}{\sigma^*_{\mu kk}} - np)^{-1}$.

The Bayes factor under the priors $\pi_k(\Psi)$ and $\pi_k(\mu)$ for Ψ and $\underline{\mu}$ respectively, assuming $\sigma^*_{22} < \sigma^*_{11}$, is given by

$$\frac{(1-n\sigma^*_{22})^{\frac{p-1}{2}} \int_{-\frac{\pi}{2}}^{\frac{3\pi}{4}} \mathbf{e}^{\frac{n(p-1)}{2}\tau_2(\theta)} d\theta}{\int_{-\frac{\pi}{2}}^{\frac{3\pi}{4}} \mathbf{e}^{\frac{n(p-1)}{2}\tau_1(\theta)} d\theta} \qquad \ldots (6)$$

where $\tau_1(\theta) = \sigma^*_{11}\lambda_1 \cos^2\theta + \sigma^*_{11}\lambda_2 \sin^2\theta$ and $\tau_2(\theta) = \sigma^*_{11}\lambda_1 \cos^2\theta + \sigma^*_{22}\lambda_1 \sin^2\theta + \sigma^*_{11}\lambda_2 \sin^2\theta + \sigma^*_{22}\lambda_2 \cos^2\theta$. The Bayes factor ($BF$) may be reduced to a simplified form, as given below, by applying Laplace's method to approximate the integrals, and the approximation was checked to be quite good.

$$BF \approx \frac{(1-n\sigma^*_{22})^{\frac{p-1}{2}} \sqrt{\sigma^*_{22}}\, \mathbf{e}^{\frac{n(p-1)}{2}\sigma^*_{22}\lambda_2}}{\sqrt{\sigma^*_{11} - \sigma^*_{22}}}. \qquad \ldots (7)$$

If $\sigma^*_{11} = \sigma^*_{22}$ then BF has a very simple form, namely,

$$BF \approx \left\{\frac{1}{2}\pi(p-1)n\sigma^*_{11}(\lambda_1 - \lambda_2)\right\}^{\frac{1}{2}} (1-n\sigma^*_{11})^{\frac{p-1}{2}} \mathbf{e}^{\frac{n(p-1)}{2}\sigma^*_{11}\lambda_2}. \qquad \ldots (8)$$

Let Π_1 be the prior probability of H_1 and Π_2 be the prior probability of H_2. Under the condition $\sigma^*_{11} = \sigma^*_{22}$, Bayes rule leads to the acceptance of H_2 if

$$\left\{\frac{1}{2}\pi(p-1)\sigma^*_{11}(\lambda_1 - \lambda_2)\right\}^{\frac{1}{2}} (1-n\sigma^*_{11})^{\frac{p-1}{2}} \mathbf{e}^{\frac{n(p-1)}{2}\sigma^*_{11}\lambda_2} > \frac{\Pi_1}{\Pi_2}$$

which may be expressed as

$$\lambda_2 > \frac{2}{n(p-1)\sigma^*_{11}} \ln \frac{\Pi_1}{\Pi_2(1-n\sigma^*_{11})^{\frac{p-1}{2}}} - \frac{2}{n(p-1)\sigma^*_{11}} \ln\left\{\frac{1}{2}\pi(p-1)\sigma^*_{11}(\lambda_1 - \lambda_2)\right\} \qquad \ldots (9)$$

We write in this way to show the basic form of the rule. The expression involving $(\lambda_1 - \lambda_2)$ has been kept on the right hand side with other constants because it is not possible to solve the inequality for λ_2 analytically. Nevertheless this form is convenient for discussion. The term involving $(\lambda_1 - \lambda_2)$ does not influence the rule very much if n or p or both are moderate.

It is evident that the rule depends on the data through the eigenvalues of the between sum of squares and products matrix. Moreover the rule depends on the

smallest eigenvalue λ_2 for large n or p while the difference $(\lambda_1 - \lambda_2)$ plays a role when λ_2 is near the first term of the right hand side of the inequality (9). Thus if λ_2^0 is the first term on the right hand side of (9), typically large (or small) values of λ_2 relative to λ_2^0 will lead to acceptance of H_2 (or H_1) but when λ_2 is near the boundary λ_2^0, the difference $(\lambda_1 - \lambda_2)$ also comes into play. Now we present a simulation study to show how the parameters are chosen, how the Bayes rule is found iteratively, how the ad hoc rule is refined and how the resulting rules perform.

3. A SIMULATION STUDY

We first decide on the values of some of the parameters. We have considered four populations i.e., $p = 4$, for detailed study. For $p = 3$ detailed simulations were not carried out for the following reason. With only three populations it is not possible to distinguish well between H_1 and H_2 with acceptably low error probabilities. This is reflected in the relatively high overall Bayes risk of about 0.3 (see table 2). Also the case of three populations is not of much practical interest in the sense that the efficiency that would be gained by recognizing a structural relationship, would not be much. Under these circumstances it did not seem warranted to get a detailed picture for the case $p = 3$. For $p \geq 5$ the performance of the rules get better (see table 2) otherwise the same pattern as observed in the case $p = 4$ follows and we do not give the detailed tables. We chose a sample of size n=10 from each population, which is neither too small nor too large. In any case it is easy to calculate from table 3 and table 4 which give the pointwise risk, what the effect of changing n would be on the risk of the ad hoc rule at different parametric points. For example if we keep the ad hoc rule fixed and change n from 10 to 100, then the risk at $(e_1, e_2, n = 100)$ is same as at $(10e_1, 10e_2, n = 10)$, where e_i's are the eigenvalues of the matrix $\frac{np}{p-1}\Delta$. Here it should be pointed out that our priors depend on n and so a given ad hoc rule need not be close to being Bayes for different values of n.

We chose $\sigma_{11}^* = 1/11$ and $\sigma_\mu^* = 1/41$ which imply $\sigma_{11} = 1$ and $\sigma_\mu = 1$ and assume $\Sigma_2^* = \sigma_{11}^* \mathbf{I}$ and Σ_1^* =diag$(\sigma_{11}^*, 0)$, giving a moderate variation to the μ_i's.

From the tables of simulated joint distributions of $(\lambda_1 - \lambda_2)$ and λ_2 (see De (1993)) under H_1 and H_2, we derived approximate likelihood ratios at various cells. We found that the likelihood ratios are mostly greater than one for $\lambda_2 > b$ and mostly less than one for $\lambda_2 < a$ where b is a number near 1.5 and a is a number near 0.5.

As a starting point we chose $\Pi_1 = \Pi_2 = 0.5$ and substituted in the inequality (9). The Bayes rule is then compared with the likelihood rule based on simulated joint distribution tables and cut off point one. The two matched well as expected and on examining the likelihood ratio table, it was found that the ad hoc rule that seemed to be close to the Bayes rule has a cut off point approximately at 1.44 for λ_2. We rounded it off to 1.5 and checked through some simulations that its frequentist performance seemed acceptable. Then Π_1 and Π_2 were adjusted a little to make the Bayes rule resemble more the ad hoc rule with the cut off for λ_2 at 1.5. New values

are $\Pi_1 = 0.523$ and $\Pi_2 = 1 - 0.523$. At this point the Bayes rule is finalized. Again the table of the likelihood ratio is studied together with the Bayes factor table and the following refinement is suggested for the ad hoc rule.

$$\begin{gathered}\text{Accept } H_1 \text{ if } \lambda_2 > a \\ \text{and} \\ \text{accept } H_2 \text{ if } \lambda_2 > b. \\ \text{If } a \leq \lambda_2 \leq b, \\ \text{accept } H_1 \text{ for large values of } \lambda_1.\end{gathered}$$

We specify the rule completely after trying quite a few combinations of (a, b) where a is in the neighbourhood of 0.5 and b is in the neighbourhood of 1.5 and trying few straight line boundaries separating H_1 and H_2 when $a \leq \lambda_2 \leq b$. Final ad hoc rule is

$$\begin{gathered}\text{Accept } H_1 \text{ if } \lambda_2 > 0.5, \\ \text{Accept } H_2 \text{ if } \lambda_2 > 1.5. \\ \text{When } 0.5 \leq \lambda_2 \leq 1.5, \\ \text{accept } H_2 \text{ if } 52.5 - 35\lambda_2 < \lambda_1 < 35, \\ \text{otherwise accept } H_1.\end{gathered}$$

This brings in a slight refinement and improvement as far as the average risk goes but more importantly, when $a \leq \lambda_2 \leq b$, the rule chooses H_1 for large $(\lambda_1 - \lambda_2)$. For λ_2 neither too large to indicate H_2 nor small enough to indicate H_1, we have perturbed the Bayes rule to get small error probability for points in H_1 with a large value of e_1, at the cost of points in H_2 with small value of e_2. We felt that the points in H_2 referred to above are close to being one dimensional and therefore are less important than the points in H_1 with large e_1. These latter points would appear to be one dimensional and so one should have less error probability $P(H_2|H_1)$. This is not what the Bayes rule does but the ad hoc rule is made to do so to have good frequentist properties, with negligible increase in the Bayes risk (see figure 1). The performance of the refined ad hoc rule under H_2 is shown in figure 2. Detailed tables are given in De (1993).

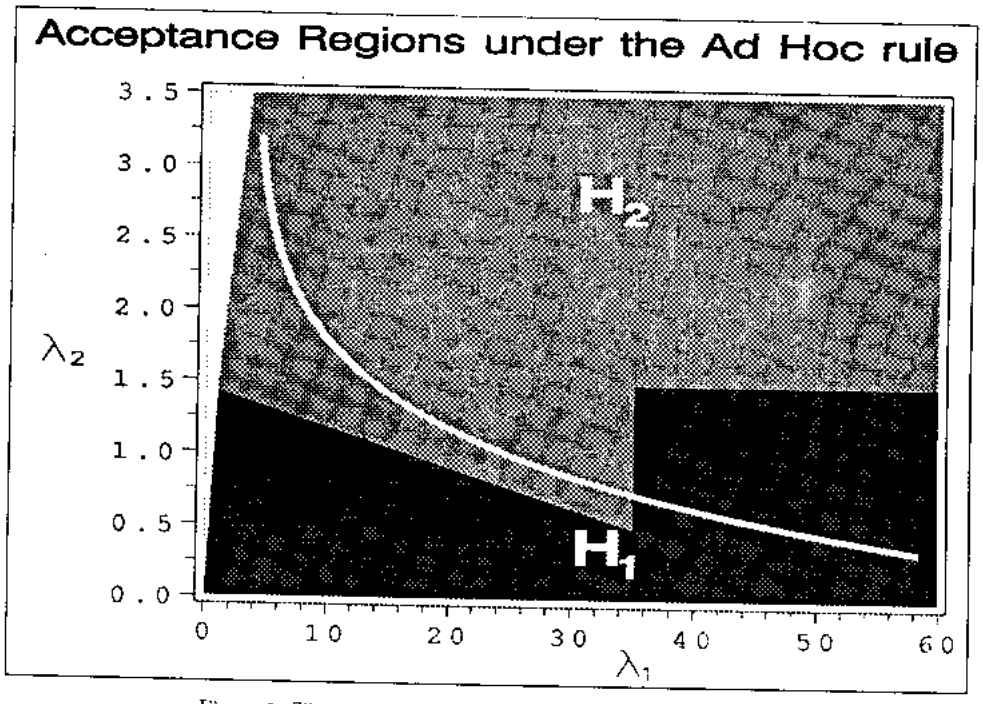

Figure 1. The refined ad hoc rule and superimposed Bayes rule is shown by the white curve.

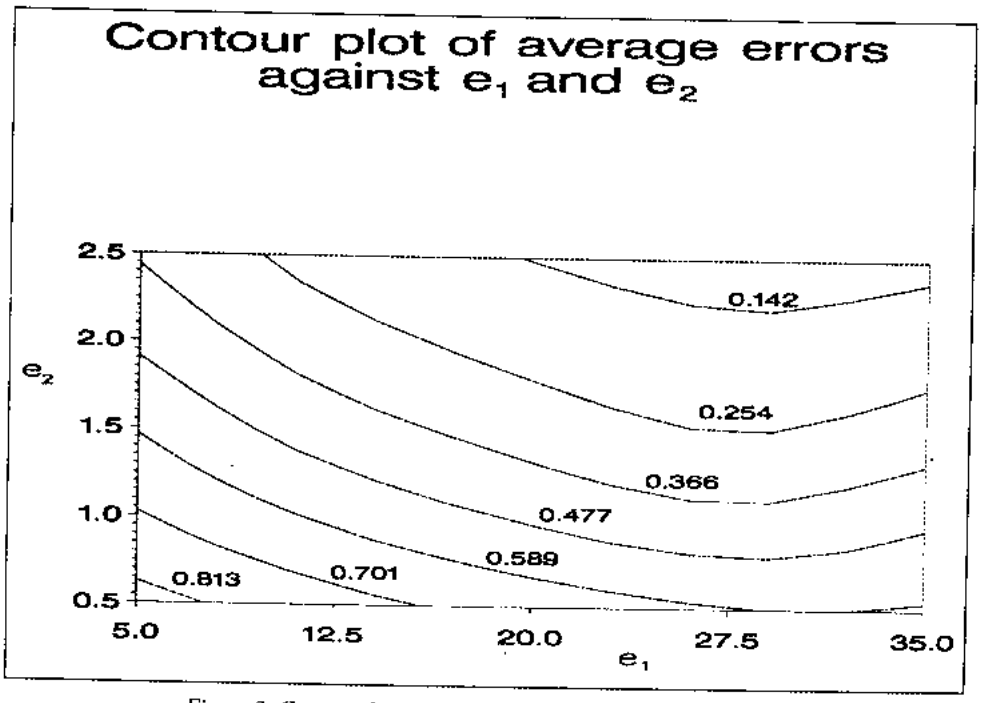

Figure 2. Contour diagram of error under H_2 for the ad hoc rule.

Table 1. JOINT PROBABILITY DISTRIBUTION OF THE EIGENVALUES $e_1 \geq e_2$ OF $\frac{np}{p-1}\Delta$ FROM SIMULATION (size=80, 000).

	Under H_2							Under H_1
	e_2							e_2
e_1	0–0.25	0.25–0.75	0.75–1.25	1.25–1.75	1.75–2.25	2.25–2.75	2.75–∞	0
0–2.5	.003	.006	.002	.001	.000	.000	.000	.143
2.5–7.5	.020	.033	.025	.022	.016	.007	.031	.340
7.5–12.5	.019	.034	.029	.025	.023	.009	.099	.230
12.5–17.5	.014	.024	.020	.019	.018	.007	.113	.135
17.5–22.5	.009	.015	.013	.014	.011	.005	.087	.073
22.5–27.5	.005	.009	.008	.008	.006	.003	.057	.039
27.5–32.5	.003	.005	.005	.005	.004	.002	.035	.020
32.5–37.5	.001	.003	.003	.003	.002	.001	.022	.010
37.5–42.5	.000	.001	.001	.001	.001	.000	.007	.003
42.5–∞	.001	.002	.002	.002	.002	.001	.019	.007

Table 2 shows the performance of the new ad hoc rule and the Bayes rule for various values of p. 1000 sets of mean vectors were generated and for each set rules were tested 100 times. Difference between ad hoc and Bayes was also noted. For the case $p = 4$ we also examined the case where Ω is unknown. Results are given under the column 4(u) using unbiased estimator of Ω.

Table 2. PERFORMANCE OF AD HOC AND BAYES RULES FOR VARIOUS VALUES OF p.

	p				
	3	4	4(u)	5	6
$\Pi(H_1)$	.546	.523	.523	.377	.287
Overall error ad hoc	.283	.198	.205	.157	.120
Overall error Bayes	.284	.191	.200	.154	.115
Overall difference	.071	.043	.054	.024	.024
Error under H_1 ad hoc	.100	.145	.104	.168	.180
Error under H_1 Bayes	.152	.109	.130	.141	.130
Difference under H_1	.066	.045	.051	.034	.040
Error under H_2 ad hoc	.527	.261	.315	.150	.093
Error under H_2 Bayes	.459	.290	.276	.162	.106
Difference under H_2	.078	.041	.057	.017	.014

Now we shall try to examine whether the ad hoc rule is really performing well at various points in the parameter space locally. For this we choose a point $\theta_1 \in \Theta_1$ where ad-hoc rule has not done well and a point $\theta_2 \in \Theta_2$ close to θ_1 where it has done reasonably. Is it possible to improve the performance at θ_1 without sacrificing the same at θ_2? We shall construct an approximately most powerful test (based only on λ_1, λ_2) from the frequency table of λ_1 and λ_2 generated by simulation at θ_1 and at θ_2 respectively. We choose $\theta_1 = (e_1^{(1)}, 0)$ and $\theta_2 = (e_1^{(2)}, e_2^{(2)})$ and take $e_1^{(1)} = e_1^{(2)}$ to make the problem difficult. We avoid choosing $e_2^{(2)}$ close to zero because it represents Θ_1 for practical purposes.

Table 3. COMPARISON OF THE RULE WITH AN APPROXIMATELY MOST POWERFUL(AMP) RULE AT VARIOUS CHOICES OF (e_1, e_2) UNDER H_2 AND $(e_1, 0)$ UNDER H_1

(e_1, e_2)	$P(H_1\|H_2)$		$P(H_2\|H_1)$		average error	
	ad-hoc	AMP	ad-hoc	AMP	ad-hoc	AMP
(18, 1.7)	0.235	0.245	0.223	0.217	0.229	0.231
(24, 1.7)	0.284	0.287	0.185	0.183	0.234	0.235
(30, 1.7)	0.290	0.255	0.191	0.205	0.241	0.230
(34, 1.7)	0.242	0.198	0.226	0.239	0.234	0.219

Table 3 shows there is not much scope for improvement for the ad hoc and table 4 gives a similar picture even when the covariance matrices of the four populations are unknown and different. For table 4, sets of four different 2×2 covariance matrices, $\mathbf{\Omega}_i, i = 1, \ldots, 4$, are chosen with 1 in the diagonals and off-diagonals as given below. Here the ad hoc rule uses $p^{-1} \sum \hat{\mathbf{\Omega}}_i$ in place of $\mathbf{\Omega}$, where $\hat{\mathbf{\Omega}}_i$'s are estimated covariances.

Table 4. SIMILAR TO TABLE 3 BUT HERE THE COVARIANCE MATRICES ARE UNEQUAL AND UNKNOWN

Variances are 1 and covariances are .2, - .9, .0 & 0 respectively

(e_1, e_2)	$P(H_1\|H_2)$		$P(H_2\|H_1)$		average error	
	ad-hoc	AMP	ad-hoc	AMP	ad-hoc	AMP
(18, 1.7)	0.242	0.240	0.222	0.225	0.232	0.233
(24, 1.7)	0.266	0.279	0.209	0.195	0.237	0.237
(30, 1.7)	0.251	0.229	0.242	0.241	0.247	0.235
(34, 1.7)	0.220	0.174	0.270	0.279	0.245	0.227

Acknowledgements. The authors are thankful to Dr. Bimal Sinha and Dr. G. McCabe for their useful comments and suggestions.

REFERENCES

ANDERSON, T. W. (1951). Estimating linear restrictions on coefficients for multivariate normal distributions. *Ann. Math. Statist*, **22**, 327-351. Correction, *Ann. Statist*, (1980), **8**, 1400.

BAIQI, M. (1992). The estimate of the rank for regression coefficient matrix in a median regression model. *Chin. Ann. Math.*, **13B**, 205-214.

COCHRAN, W. G. (1943). The comparison of different scales of measurement for experimental results. *Ann. Math. Statist*, **14**, 205-216.

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.

FISHER, R. A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics*, **8**, 376-386.

FUJIKOSHI, Y. (1974). The likelihood ratio tests for the dimensionality of regression coefficients. *Jour. Multivariate Anal.*, **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 Anal.*, **16**, 173-184.

SHEN, W. and SINHA, B. K. (1991). Statistical tests for structural relationship. *Statist. and Probab. Lett.*, **12**, 545-559.

SCHOOL OF NURSING
EMORY UNIVERSITY
531 ASBURY CIRCLE
ATLANTA GA 30322

STAT-MATH UNIT
INDIAN STATISTICAL INSTITUTE
203, B.T. ROAD
CALCUTTA - 35, INDIA