RESEARCH REPORT
Institute of Mathematics
University of Oslo

No 3
March 1977

MAXIMIN SELECTION AND MULTIPLE COMPARISON PROCEDURES IN MULTIVARIATE NORMAL MODELS

by

Erik Bølviken

SUMMARY

Optimum multiple tests are found for a family of multiple hypothesis testing problems in multivariate normal models. The optimum criteria are directed towards the simultaneous rather than the individual performance of the tests. Applications are made to selection and multiple comparison problems.

CONTENTS

1. INTRODUCTION

Let X_{ij}, $j=1,\ldots,n_i$, $i=1,\ldots,m$ be p-dimensional random vectors, independently and normally distributed with common covariance matrix $\underset{\sim}{\Sigma}$ (*) and expectations $EX_{ij} = \xi_i$. For given m-dimensional column vectors $b_1,\ldots,b_a$ introduce p-dimensional linear forms $v_i = b_i'\underset{\sim}{\xi}$, where $\underset{\sim}{\xi}$ is the m×p-matrix with rows $\xi_1,\ldots,\xi_m$. Consider the multiple hypothesis testing problem

$$H_i : v_i = 0 \quad \text{against} \quad K_i : v_i \neq 0 , \quad i=1,\ldots,a .$$

The purpose of the paper is to derive optimum multiple tests. Attention is directed towards maximizing the simultaneous rather than the individual power of the tests. The optimum criteria which are of the maximin type, are due to Lehmann [5].

Applications are made to the following examples:

1. The selection of ξ_i different from some known standard
2. Testing the equality of successive ξ_i
3. Multiple comparison of $\xi_1,\ldots,\xi_m$
4. The selection of non-zero regression coefficients.

The paper is based on a part of the author's Cand.Real. thesis at the University of Oslo.

(*) The symbol ~ below the letter will be used throughout to denote matrices.

2. FORMULATION OF THE OPTIMUM PROBLEM

The formulation of the optimum problem is based upon an idea due to Lehmann [5]. As in ordinary Neyman-Pearson theory, there are two sources of error. We may reject H_i while $v_i = 0$, or we may fail to state K_i while $v_i \neq 0$. The latter will be regarded as an error only if v_i is outside some neighbourhood of the origin. The area $0 < v_i \Sigma v_i' < \Delta_i$, for some specified number $\Delta_i > 0$, will be taken as an indifference zone where it does not matter whether H_i is rejected or not.

As a measure of the performance of a procedure with respect to false rejections we shall take either

(2.1 a) The expected number of false rejections

or

(2.1 b) The expected proportion of true H_i rejected, that is the quantity (2.1a) divided by the total number of true H_i.

In the opinion of the author these criteria usually should be preferred to

(2.1 c) The probability of at least one false rejection

which has the disadvantage of measuring only how often false rejections occur, not how many we are likely to make each time.

Introduce

$$\mathfrak{V} = \{v_i \mid v_i \Sigma^{-1} v_i' \geq \Delta_i\} . \tag{2.2}$$

To see how well a test carries out its task of identifying non-zero v_i, we shall use one of the following four quantities:

(2.3 a) The expected number of v_i in $\mathcal{V}$ identified as non-zero

(2.3 b) The expected proportion of v_i in $\mathcal{V}$ recognized as non-zero, that is quantity (2.3 a) divided by the total number of elements in $\mathcal{V}$.

(2.3 c) The probability of identifying at least one non-zero v_i in $\mathcal{V}$.

(2.3 d) The probability of stating $v_M \neq 0$ where M is given by $v_M \underset{\sim}{\Sigma}^{-1} v_M' = \max_{v_i \in \mathcal{V}} v_i \underset{\sim}{\Sigma}^{-1} v_i'$.

(2.3 a) and (2.3 b) are appropriate if it is desired to identify as many non-zero v_i as possible, while (2.3 c) or (2.3 d) may be preferred if the decision is only a part of a scheme aiming at a final selection of a single v_i .

A fifth possibility is

(2.3 e) The probability of identifying as non-zero all $v_i \in \mathcal{V}$.

However, this criterion seems appropriate only in rare cases (see [5].)

As generic notations for (2.1 a) - (2.1 b) and (2.3 a) - (2.3 d) we shall use respectively $R(\underset{\sim}{\xi};\varphi)^{(*)}$ and $S(\underset{\sim}{\xi};\varphi)^{(*)}$ where φ is the procedure in question.

Clearly it is desirable to have $S(\underset{\sim}{\xi};\varphi)$ as large as possible and $R(\underset{\sim}{\xi};\varphi)$ as small as possible. In this paper we shall regard a procedure as optimum if it subject to

(*) if $\underset{\sim}{\Sigma}$ is known. In section 5 where $\underset{\sim}{\Sigma}$ is unknown, we shall write $R(\underset{\sim}{\xi},\underset{\sim}{\Sigma};\varphi)$ and $S(\underset{\sim}{\xi},\underset{\sim}{\Sigma};\varphi)$.

$$(2.4) \qquad \sup_{\underset{\sim}{\xi}} R(\underset{\sim}{\xi};\varphi) \leq \epsilon$$

maximizes

$$(2.5) \qquad \inf_{\underset{\sim}{\xi}\in\Omega^*} S(\underset{\sim}{\xi};\varphi)$$

where $\Omega^* = \{\underset{\sim}{\xi} \mid v_i \underset{\sim}{\Sigma}^{-1} v_i' \geq \Delta_i$ for at least one $i\}$.

A dual problem is to minimize $\sup R(\underset{\sim}{\xi};\varphi)$ subject to $\inf S(\underset{\sim}{\xi};\varphi) \geq \epsilon'$. This problem will not be dealt with here although its solution is analogous to the solution of (2.5).

For optimum criteria aiming at the maximization of the individual power of the tests, the reader may consult Spjøtvoll [7] . Studies from a decision theoretical viewpoint are undertaken, for example, in [4] . None of these references treat multivariate problems.

3. A FUNDAMENTAL THEOREM

Suppose the procedures are to be based on some statistic Z . (In sections 4 and 5 Z will be derived from sufficiency and invariance arguments.) Let $f(z;\underset{\sim}{\xi})$ be the density of Z . Introduce

$$(3.1) \qquad \theta_i(\underset{\sim}{\xi}) = v_i \underset{\sim}{\Sigma}^{-1} v_i' = (b_i'\underset{\sim}{\xi}) \underset{\sim}{\Sigma}^{-1} (b_i'\underset{\sim}{\xi})' \ .$$

Suppose there exist parameter points $\underset{\sim}{\xi}_o, \underset{\sim}{\xi}_1, \ldots, \underset{\sim}{\xi}_a$ satisfying the following conditions:

C_1 : $\theta_i(\underset{\sim}{\xi}_o) = 0$, $i=1,\ldots,a$

C_2 : $\theta_i(\underset{\sim}{\xi}_i) = \Delta_i$, $\theta_j(\underset{\sim}{\xi}_i) < \Delta_j$ for any $j \neq i$, $i=1,\ldots,a$

C_3 : $\left\{ \frac{f(z;\underset{\sim}{\xi}_i)}{f(z;\underset{\sim}{\xi}_o)} = h_i(U_i(z)) \right.$ where h_i is a strictly increasing function and $U_i = U_i(Z)$ some statistic.

C_4 : $\left\{\right.$ The distribution of U_i depends on $\underset{\sim}{\xi}$ through $\theta_i(\underset{\sim}{\xi})$, and U_i is stochastically increasing in θ_i .

Introduce the following multiple test (*) :

$$(3.2)\qquad \hat{\varphi}_i = \begin{cases} 1 , & \text{if } U_i > c_i \\ 0 , & \text{otherwise} \end{cases} , \quad i=1,\ldots,a .$$

Suppose the constants $c_1,\ldots,c_a$ are determined to satisfy the equations

$$(3.3)\qquad \begin{cases} R(\underset{\sim}{\xi}_o;\hat{\varphi}) = \epsilon \\ S(\underset{\sim}{\xi}_i;\hat{\varphi}) = S(\underset{\sim}{\xi}_1;\hat{\varphi}) \quad , \quad i=2,\ldots,a \end{cases}$$

where R and S respectively are one of the quantities (2.1 a) - (2.1 b) and (2.3 a) - (2.3 d) . Then $\hat{\varphi}$ is maximin. This is the conclusion of the following theorem which is only a restatement of a theorem in [5] for the problem at hand.

<u>Theorem 3.1.</u> Suppose the conditions C_1 - C_4 hold. If the constants $c_1,\ldots,c_a$ are determined to satisfy (3.3), then $\hat{\varphi}$, among all procedures based on Z , maximizes $\inf_{\Omega^*} S(\underset{\sim}{\xi};\varphi)$ subject to $\sup R(\underset{\sim}{\xi};\varphi) \leq \epsilon$.

Proof: Conditions C_2 and C_4 easily yield

$$\inf S(\underset{\sim}{\xi};\hat{\varphi}) = S(\underset{\sim}{\xi}_i;\hat{\varphi}) \quad , \quad i=1,\ldots,a .$$

Hence, for an arbitrary procedure φ ,

(*) A multiple test will in general be represented by a vector-valued function $\varphi(z) = (\varphi_1(z),\ldots,\varphi_a(z))$ where $\varphi_i(z)$ has the familiar interpretation as the probability of rejecting H_i given $Z = z$.

$$\inf S(\underset{\sim}{\xi};\hat{\varphi}) - \inf S(\underset{\sim}{\xi};\varphi) \geq S(\underset{\sim}{\xi}_i;\hat{\varphi}) - S(\underset{\sim}{\xi}_i;\varphi) .$$

For arbitrary positive numbers $\pi_1,\dots,\pi_a$ satisfying $\sum_i \pi_i = 1$, this implies

$$\inf S(\underset{\sim}{\xi};\hat{\varphi}) - \inf S(\underset{\sim}{\xi};\varphi) \geq \sum_i \pi_i[S(\underset{\sim}{\xi}_i;\hat{\varphi})-S(\underset{\sim}{\xi}_i;\varphi)] .$$

Now, suppose $\sup R(\underset{\sim}{\xi};\varphi) \leq \epsilon$. Then, in particular, $R(\underset{\sim}{\xi}_o;\varphi) \leq \epsilon$. Hence, for any $\alpha > 0$

$$\inf S(\underset{\sim}{\xi};\hat{\varphi}) - \inf S(\underset{\sim}{\xi};\varphi) \geq \sum_i \pi_i[S(\underset{\sim}{\xi}_i;\hat{\varphi})-S(\underset{\sim}{\xi}_i;\varphi)] - \alpha[R(\underset{\sim}{\xi}_o;\hat{\varphi})-R(\underset{\sim}{\xi}_o;\varphi)].$$

Introduce $f_i(z) = f(z;\underset{\sim}{\xi}_i)$, $i=0,1,\dots,a$. Clearly $S(\underset{\sim}{\xi}_i;\varphi) = \int \varphi_i f_i$ and $R(\underset{\sim}{\xi}_o;\varphi) = \sum_i \int \varphi_i f_o$. Thus the last inequality may be rewritten

$$(3.4) \qquad \inf S(\underset{\sim}{\xi};\hat{\varphi}) - \inf S(\underset{\sim}{\xi};\varphi) \geq \int \sum_i (\pi_i f_i - \alpha f_o)(\hat{\varphi}_i-\varphi_i) .$$

Choose in particular

$$\alpha^{-1} = \sum_i \frac{1}{h_i(c_i)}$$

$$\pi_i = \frac{\alpha}{h_i(c_i)} , \quad i=1,\dots,a .$$

Then (3.4) becomes

$$\inf S(\underset{\sim}{\xi};\hat{\varphi}) - \inf S(\underset{\sim}{\xi};\varphi) \geq \alpha \int \sum_i \frac{1}{h_i(c_i)}(f_i-h_i(c_i)f_o)(\hat{\varphi}_i-\varphi_i)$$

From condition C_3 it is easily seen that the integrand is everywhere non-negative.

Hence

$$\inf S(\underset{\sim}{\xi};\hat{\varphi}) \geq \inf S(\underset{\sim}{\xi};\varphi)$$

which completes the proof of the theorem.

4. $\underset{\sim}{\Sigma}$ KNOWN

4.1. The general maximin procedure

The maximin procedure may be based on the sufficient $m\times p$-matrix $\bar{\underset{\sim}{X}}$ with rows

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad , \quad i=1,\ldots,m \; .$$

Furthermore, let $\underset{\sim}{D}$ be a fixed non-singular $p\times p$ matrix satisfying

$$\underset{\sim}{D}' \underset{\sim}{\Sigma} \underset{\sim}{D} = \text{the } p\times p \text{ identity matrix} \; .$$

It is easy to show that the decision problem is invariant under the following group of transformations

$$G_1 : \bar{\underset{\sim}{X}} \to \bar{\underset{\sim}{X}} \underset{\sim}{D} \underset{\sim}{g}_1 \underset{\sim}{D}^{-1}, \; \underset{\sim}{g}_1 \text{ any } p\times p \text{ ortonormal matrix.}$$

It follows from the Hunt-Stein theorem that we may look for an optimum procedure among invariant procedures.

Let Z be a maximal invariant function. Explicit knowledge of the distribution of Z will not be needed. Below we shall, for each $i=1,2,\ldots,a$, construct a set ω_i associated with the i'th problem. The distribution of Z will be characterized on each ω_i by means of a sufficient statistic U_i . Eventually parameter points $\underset{\sim}{\xi}_i$ satisfying conditions $C_1 - C_4$ of section 3, will be selected from the ω_i's .

Observe that the density of $\bar{\underset{\sim}{X}} = \{\bar{X}_{k\mu}\}$ may be written

$$r(\bar{\underset{\sim}{x}};\underset{\sim}{\xi}) = r(\bar{\underset{\sim}{x}};\underset{\sim}{O})\exp\left\{\sum_{\nu=1}^{p} \sum_{\mu=1}^{p} A_{\mu\nu} \sum_{k=1}^{m} n_k \bar{x}_{k\mu} \xi_{k\nu} - \tfrac{1}{2} \sum_{k=1}^{m} \xi_k \underset{\sim}{A} \xi_k'\right\}$$

where $\underset{\sim}{\xi} = \{\xi_{k\nu}\}$ and $\underset{\sim}{A} = \{A_{\mu\nu}\} = \underset{\sim}{\Sigma}^{-1}$. Hence, it is easy to see that $Y_i = b_i'\bar{\underset{\sim}{X}}$ is a sufficient statistic for $\bar{\underset{\sim}{X}}$ on the set

(4.1) $$\omega_i = \{\underset{\sim}{\xi} \mid \xi_k = \frac{b_{ik}}{n_k}\eta \quad , \; k=1,\ldots,m \quad , \quad \eta \in \mathbb{R}^p\}$$

(b_{ik} is the k'th element of b_i). Note that G_1 induces a corresponding group in the Y_i-space:

$$G_1^* : Y_i \rightarrow Y_i \underset{\sim}{D} \underset{\sim}{g}_1 \underset{\sim}{D}^{-1} .$$

By straightforward calculations it may be proved that

(4.2) $$U_i = Y_i \underset{\sim}{\Sigma}^{-1} Y_i' = (b_i' \underset{\sim}{\bar{X}}) \underset{\sim}{\Sigma}^{-1} (b_i' \underset{\sim}{\bar{X}})'$$

is a maximal invariant for G_1^* .

Observe that U_i was derived by first applying ω_i-sufficiency (going from $\underset{\sim}{\bar{X}}$ to Y_i) and then invariance. It is the statement of a general theorem (*) that we should arrive at the same result by first applying invariance (going from $\underset{\sim}{\bar{X}}$ to Z) and then sufficiency. Hence we have proved that U_i is sufficient for Z if $\underset{\sim}{\xi} \in \omega_i$. This means that the density of Z may be written

(*) Theorem 3.1 in [3]. It is the statement of this theorem that we under weak conditions on the transformation group may interchange the sequence of applications of the invariance and sufficiency principles without affecting the final result. The conditions are:

(i) $g\omega_i = \omega_i$ for all transformations g of the group (satisfied in our case).

(ii) Any almost invariant function is equivalent to an invariant function (satisfied also). (For definitions of the concepts "almost invariant" and "equivalent to an invariant function", see [6].)

$$f(z\,;\underset{\sim}{\xi}) = \gamma_i(u_i;\theta_i)\, f_i(z|u_i) \quad , \quad \underset{\sim}{\xi} \in \omega_i \tag{4.3}$$

where $\gamma_i(u_i;\theta_i)$ is the density of U_i, (depending on $\underset{\sim}{\xi}$ through $\theta_i = \theta_i(\underset{\sim}{\xi})$), and $f_i(z|u_i)$ is the conditional density of Z given U_i (not depending on $\underset{\sim}{\xi} \in \omega_i$) .

The next step is to choose $\underset{\sim}{\xi}_o, \underset{\sim}{\xi}_1, \dots, \underset{\sim}{\xi}_a$ and verify the conditions C_1-C_4. First, let $\underset{\sim}{\xi}_o = 0$ which trivially satisfies C_1. Secondly, suppose $\underset{\sim}{\xi} \in \omega_i$. Then $\xi_k = \frac{b_{ik}}{n_k}\eta$, $k=1,\dots,m$. For this $\underset{\sim}{\xi}$ straightforward calculations yield

$$\theta_j(\underset{\sim}{\xi}) = B_{ij}^2 \circ \eta\, \underset{\sim}{\Sigma}^{-1}\, \eta' \tag{4.4}$$

where

$$B_{ij} = \sum_{k=1}^{m} \frac{b_{ik}b_{jk}}{n_k} . \tag{4.5}$$

Hence, choose as $\underset{\sim}{\xi}_i$ any member of ω_i corresponding to an η satisfying

$$\eta\, \underset{\sim}{\Sigma}^{-1}\, \eta' = \frac{\Delta_i}{B_{ii}^2} .$$

Then $\theta_i(\underset{\sim}{\xi}_i) = \Delta_i$ and condition C_2 is satisfied if and only if

$$\left(\frac{B_{ij}}{B_{ii}}\right)^2 \Delta_i < \Delta_j \quad \text{for any} \quad j \neq i.$$

As an immediate consequence of (4.3)

$$\frac{f(z;\underset{\sim}{\xi}_i)}{f(z;\underset{\sim}{\xi}_o)} = \frac{\gamma_i(u_i;\Delta_i)}{\gamma_i(u_i;0)} .$$

Furthermore, it is for any $\underset{\sim}{\xi}$ clear that

$$\frac{U_i}{B_{ii}} \sim \chi_p^2\left(\frac{\theta_i}{B_{ii}}\right)^{(*)} .$$

Hence conditions C_3 and C_4 follow from well-known properties of the non-central χ^2-distribution.

Applying Theorem 3.1, we can now write down the maximin procedure:

$$\hat{\varphi}_i = \begin{cases} 1 , & \text{if } (b_i'\bar{X}) \Sigma^{-1}(b_i'\bar{X})' > c_i \\ 0 , & \text{otherwise} \end{cases} \tag{4.6}$$

$i=1,2,\ldots,a$.

Observe that the power functions of the individual tests are given by

$$E\,\hat{\varphi}_i = 1 - \Gamma_p\left(\frac{c_i}{B_{ii}} ; \frac{\theta_i}{B_{ii}}\right) .$$

Hence, the constants $c_1,\ldots,c_a$ should, according to (3.3), be determined to satisfy

$$\begin{cases} \sum_{i=1}^{a} \left[1 - \Gamma_p\left(\frac{c_i}{B_{ii}}\right)\right] = \epsilon \\ \Gamma_p\left(\frac{c_i}{B_{ii}} ; \frac{\Delta_i}{B_{ii}}\right) = \Gamma_p\left(\frac{c_1}{B_{11}} ; \frac{\Delta_1}{B_{11}}\right) , \quad i=2,\ldots,a . \end{cases} \tag{4.7}$$

(If R is given by (2.1b) the right hand side of the first equation should be $a\epsilon$ rather than ϵ.)

(*) $\chi_p(\lambda)$ denotes the non-central χ^2-distribution with p degrees of freedom and eccentricity λ . The corresponding cumulative distribution function is written $\Gamma_p(\circ;\lambda)$,or if $\lambda=0$, $\Gamma_p(\circ)$.

The results of this section are summarized in the following theorem.

Theorem 4.1. Suppose

$$\max_{i \neq j} [\Delta_j - \Delta_i (\frac{B_{ij}}{B_{ii}})^2] > 0 \tag{4.8}$$

where $B_{ij} = \sum_{k=1}^{m} \frac{b_{ik} b_{jk}}{n_k}$. Then $\hat{\varphi}$, given by (4.6) and (4.7), is maximin, i.e. maximizes $\inf S(\underset{\sim}{\xi};\varphi)$ among all φ satisfying $\sup R(\underset{\sim}{\xi};\varphi) \leq \epsilon$.

4.2. Remarks:

1) (4.8) was clearly essential for the proof of Theorem 4.1. (Without this condition C_2 could not have been established.) The author conjectures that (4.8) is not only sufficient, but also necessary for $\hat{\varphi}$ to be maximin.

Based on this conjecture the theorem has the following interesting interpretation. Suppose the vectors b_i are in some sense normalized (either to the same length or to make B_{ii} independent of i). Then $B_{ij} = \sum_k \frac{b_{ik} b_{jk}}{n_k}$ may be taken as a measure of how strongly the i'th and the j'th problem is related. In this sense (4.8) states that the stronger the relation between the individual problems, the more uniformly $\Delta_1, \ldots, \Delta_a$ have to be selected to make $\hat{\varphi}$ maximin. In the extreme case $B_{ij} = 0$ for all $i \neq j$ (the problems being completely "unrelated") there are no conditions on the Δ_i's at all.

2) Suppose the vectors $b_1, \ldots, b_a$ are normalized to have the same length. Then one reasonable choice of $\Delta_1, \ldots, \Delta_a$ appears to be

(4.9) $$\Delta_1 = \ldots = \Delta_a = \Delta \ .$$

Condition (4.8) reduces to

(4.10) $$\max_{i \neq j} \left[\sum_{k=1}^{m} \frac{b_{ij}^2}{n_k} - \left| \sum_{k=1}^{m} \frac{b_{ik} b_{jk}}{n_k} \right| \right] > 0 \ .$$

In particular, this inequality is an immediate consequence of the Cauchy-Schwarz inequality if $n_1 = \ldots n_m$. For most problems of interest (4.10) is satisfied whatever the values of $n_1, \ldots, n_m$.

Another reasonable choice of the Δ_i's is

(4.11) $$\Delta_i = \Delta \circ \sum_{k=1}^{m} \frac{b_{ik}^2}{n_k} \ .$$

Then (4.8) is universally satisfied (again as a consequence of the Cauchy-Schwarz inequality). The system of equalities (4.7) simplifies to

(4.12) $$\begin{cases} c_i = c \displaystyle\sum_{k=1}^{m} \frac{b_{ik}^2}{n_k} \\ c \ = \Gamma_p^{-1}(1-\varepsilon/a) \end{cases} .$$

Thus the maximin procedure may be rewritten

(4.13) $$\hat{\varphi}_i = \begin{cases} 1 \ , & \text{if} \quad \dfrac{(b_i' \underset{\sim}{\bar{X}}) \underset{\sim}{\Sigma}^{-1} (b_i' \underset{\sim}{\bar{X}})'}{\sum_{k=1}^{m} b_{ik}^2 / n_k} > c \\ 0 \ , & \text{otherwise.} \end{cases}$$

This choice of $\Delta_1, \ldots, \Delta_a$ is suggested by the fact that the test statistics in (4.13) are normalized in the sense that they all have the same variance.

4.3. Applications

Example 1. Comparison with some known standard.

Suppose we want to compare each ξ_i with some known standard which we without loss of generality take to be zero. Note that this problem corresponds to $b_i' = (0,..0,1,0..,0)$. The indifference zone is for each ξ_i defined to be the area $0 < \xi_i \underset{\sim}{\Sigma}^{-1} \xi_i' < \Delta_i$. The maximin procedure is to state $\xi_i \neq 0$ if

$$(4.14) \qquad \bar{X}_i \underset{\sim}{\Sigma}^{-1} \bar{X}_i' > c_i$$

where $c_1, \dots, c_m$ are determined from

$$\sum_{i=1}^{m} [1-\Gamma_p(n_i c_i)] = \varepsilon$$

$$\Gamma_p(n_i c_i ; n_i \Delta_i) = \Gamma_p(n_1 c_1 ; n_1 \Delta_1) \quad , \quad i=2,\dots,a \; .$$

Observe that condition (4.8) is always satisfied. Hence (4.14) is maximin whatever the values of $\Delta_1, \dots, \Delta_a$.

If we in particular choose $\Delta_i = \frac{\Delta}{n_i}$, then $n_i c_i = c = \Gamma_p^{-1}(1-\varepsilon/m)$ and (4.14) becomes

$$n_i \bar{X}_i \underset{\sim}{\Sigma}^{-1} \bar{X}_i' > c \; .$$

Note that with this choice of Δ_i the indifference zones, $0 < \xi_i \underset{\sim}{\Sigma}^{-1} \xi_i' < \Delta_i$, become smaller the larger the value of n_i . This is very reasonable if the unequal sample sizes $n_1, \dots, n_m$ originally were motivated from unequal priorities on $\xi_1, \dots, \xi_m$.

Example 2. Testing the equality of successive ξ_i .

Let ξ_i represent the value of a quantity at some time t_i $(t_1 < t_2 < \dots < t_m)$. Suppose it is desired to find out at which points in the process the quantity has undergone changes. This means that each ξ_i is to be compared to ξ_{i+1} , $i=1,\dots,m-1$.

Observe that $b_i' = (0,..0,-1,1,0,..0)$, the non-zero elements being no. i and $i+1$. The indifference zone is for each difference $\xi_{i+1}-\xi_i$ the area $0 < (\xi_{i+1}-\xi_i) \underset{\sim}{\Sigma}^{-1} (\xi_{i+1}-\xi_i)' < \Delta_i$. Condition (4.8) is not satisfied in general. However, it is certainly satisfied when $\Delta_1 = \ldots = \Delta_{m-1} = \Delta$. Hence, under this choice of the Δ_i's a maximin procedure is to state $\xi_{i+1} \neq \xi_i$ if

$$(4.15) \qquad (\bar{X}_{i+1}-\bar{X}_i) \underset{\sim}{\Sigma}^{-1} (\bar{X}_{i+1}-\bar{X}_i)' > c_i$$

where $c_1,\ldots,c_{m-1}$ are determined from the conditions

$$\sum_{i=1}^{m} [1-\Gamma_p(\frac{c_i}{n_i^{-1}+n_{i+1}^{-1}})] = \epsilon$$

$$\Gamma_p(\frac{c_i}{n_i^{-1}+n_{i+1}^{-1}} ; \frac{\Delta}{n_i^{-1}+n_{i+1}^{-1}}) \quad \text{independent of } i .$$

<u>Example 3. Complete comparison of $\xi_1,\ldots,\xi_m$.</u>

For each pair (ξ_i,ξ_j) we are to decide whether $\xi_i \neq \xi_j$. Note that $a = \frac{1}{2}m(m-1)$. The indifference zone is for each pair (ξ_i,ξ_j) the area $0 < (\xi_i-\xi_j) \underset{\sim}{\Sigma}^{-1} (\xi_i-\xi_j)' < \Delta_{ij}$. Condition (4.8) does not hold in general, but as in the preceding example it is clearly satisfied if $\Delta_{ij} = \Delta$. Hence, under this choice of Δ_{ij} a maximin decision rule is to state $\xi_i \neq \xi_j$ if

$$(4.16) \qquad (\bar{X}_i-\bar{X}_j) \underset{\sim}{\Sigma}^{-1} (\bar{X}_i-\bar{X}_j)' > c_{ij}$$

where the constants c_{ij} are determined to satisfy

$$\sum_{i<j} [1-\Gamma_p(\frac{c_{ij}}{n_i^{-1}+n_j^{-1}})] = \epsilon$$

$$\Gamma_p(\frac{c_{ij}}{n_i^{-1}+n_j^{-1}} ; \frac{\Delta}{n_i^{-1}+n_j^{-1}}) \quad \text{independent of } i \text{ and } j .$$

Example 4. Selection of non-zero regression coefficients.

Suppose a multivariate regression model is given, i.e. let $\underset{\sim}{X}$ be a $m \times p$ random matrix of the form

(4.17) $$\underset{\sim}{X} = \underset{\sim}{y}\,\underset{\sim}{\beta} + \underset{\sim}{U}$$

where $\underset{\sim}{y}$ is a known $m \times r$ matrix $(r \leq m)$ of full rank, $\underset{\sim}{\beta}$ an unknown $r \times p$ matrix and $\underset{\sim}{U}$ an unobservable random matrix with independently and normally distributed rows $U_1, \ldots, U_m$ with common known covariance matrix $\underset{\sim}{\Sigma}$ and expectations zero. Suppose the selection of non-zero regression coefficients is desired, that is the non-zeros among the rows $\beta_1, \ldots, \beta_r$ of $\underset{\sim}{\beta}$.

Writing $\underset{\sim}{\xi} = E\,\underset{\sim}{X}$ it is clear that $\underset{\sim}{\beta} = \underset{\sim}{d}'\,\underset{\sim}{\xi}$, where $\underset{\sim}{d} = \underset{\sim}{y}(\underset{\sim}{y}'\underset{\sim}{y})^{-1}$. Hence the selection problem is contained in the general formulation of the paper. Note that $b_i = d_i$ (= the i'th row of $\underset{\sim}{d}$). According to the general procedure (4.6) we should state $\beta_i \neq 0$ if $(d_i'\underset{\sim}{X})\,\underset{\sim}{\Sigma}^{-1}\,(d_i'\underset{\sim}{X})' > c_i$, or equivalently if

(4.18) $$\hat{\beta}_i\,\underset{\sim}{\Sigma}^{-1}\,\hat{\beta}_i' > c_i$$

where $\hat{\beta}_i = d_i'\,\underset{\sim}{X}$ is the least-squares estimate of β_i. Introducing the area $0 < \beta_i\,\underset{\sim}{\Sigma}^{-1}\,\beta_i' < \Delta_i$ as the indifference zone, the constants $c_1, \ldots, c_r$ are to be determined to satisfy

$$\begin{cases} \sum_{i=1}^{r}\left[1-\Gamma_p\left(\frac{c_i}{q_{ii}}\right)\right] = \epsilon \\ \Gamma_p\left(\frac{c_i}{q_{ii}}\,;\,\frac{\Delta_i}{q_{ii}}\right) = \Gamma_p\left(\frac{c_1}{q_{11}}\,;\,\frac{\Delta_1}{q_{11}}\right) \end{cases}$$

where q_{ii} is the diagonal elements of the matrix $\underset{\sim}{q} = (\underset{\sim}{y}'\underset{\sim}{y})^{-1}$. This selection rule is maximin provided

$$(4.19) \qquad \max_{i \neq j} \left[\Delta_j - \Delta_i \left(\frac{q_{ij}}{q_{ii}}\right)^2 \right] > 0$$

(q_{ij} is the general element of $\underset{\sim}{q}$). Observe that if $\underset{\sim}{y}$ represents an ortogonal design, this condition is fullfilled whatever the values of $\Delta_1, \ldots, \Delta_a$.

One reasonable choice of Δ_i is to make it proportional to the variance of $\hat{\underset{\sim}{\beta}}_i$, that is to let $\Delta_i = q_{ii} \cdot \Delta$. Then (4.19) is always satisfied and (4.18) reduces to

$$\frac{1}{q_{ii}} \hat{\beta}_i \underset{\sim}{\Sigma}^{-1} \hat{\beta}_i' > c$$

where $c = \Gamma_p^{-1} (1 - \frac{\varepsilon}{r})$.

5. $\underset{\sim}{\Sigma}$ UNKNOWN

The natural extension of the general test procedure of the preceding section is

$$(5.1) \qquad \hat{\psi}_i = \begin{cases} 1 , & \text{if } T_i > c_i \\ 0 , & \text{otherwise} \end{cases} \qquad , i=1,\ldots,a$$

where

$$(5.2) \qquad T_i = (b_i' \underset{\sim}{\bar{X}}) \hat{\underset{\sim}{\Sigma}}^{-1} (b_i' \underset{\sim}{\bar{X}})'$$

$$(5.3) \qquad \hat{\underset{\sim}{\Sigma}} = \frac{1}{N-m} \sum_{k=1}^{m} \sum_{j=1}^{n_k} (X_{kj} - \bar{X}_k)'(X_{kj} - \bar{X}_k)$$

and $N = \sum_k n_k$. The constants $c_1, \ldots, c_a$ are to be determined from a system of equations analogously to (4.7).

Let $\mathcal{F}_{\mu,\nu}(\delta)$ denote the non-central F-distribution with μ and ν degrees of freedom and eccentricity δ . The corresponding cumulative distribution function is written $F_{\mu,\nu}(\cdot;\delta)$ or if $\delta=0$, $F_{\mu,\nu}(\cdot)$. It is well-known that

$$\frac{T_i}{B_{ii}} \circ \frac{N'}{(N-m)p} \sim \mathcal{F}_{p,N-m-p+1}\left(\frac{\theta_i}{B_{ii}}\right)$$

where $N' = N-m-p+1$ and $\theta_i = \theta_i(\underset{\sim}{\xi},\underset{\sim}{\Sigma}) = (b_i'\underset{\sim}{\xi})\ \underset{\sim}{\Sigma}^{-1}(b_i'\underset{\sim}{\xi})'$. Suppose the constants $c_1,\ldots,c_a$ are determined to satisfy

$$(5.4)\qquad \begin{cases} \sum_{i=1}^{a}\left[1-F_{p,N'}\left(\frac{c_i}{B_{ii}} \circ \frac{N'}{(N-m)p}\right)\right] = \epsilon \\ \\ F_{p,N'}\left(\frac{c_i}{B_{ii}} \circ \frac{N'}{(N-m)p}\ ;\ \frac{\Delta_i}{B_{ii}}\right) \quad \text{independent of} \quad i=1,2,\ldots,a\ . \end{cases}$$

Then, from well-known results concerning the non-central F-distribution, it follows that $\hat{\psi}$ has the following two properties:

1) $\sup R(\underset{\sim}{\xi},\underset{\sim}{\Sigma};\hat{\psi}) = R(0,\underset{\sim}{\Sigma};\hat{\psi}) = \epsilon$

2) $S(\underset{\sim}{\xi},\underset{\sim}{\Sigma};\hat{\psi})$ attains its minimum value (*) in points $(\underset{\sim}{\xi},\underset{\sim}{\Sigma})$ for which $\theta_i(\underset{\sim}{\xi},\underset{\sim}{\Sigma}) = \Delta_i$ for exactly one i and $\theta_j(\underset{\sim}{\xi},\underset{\sim}{\Sigma}) < \Delta_j$ for all $j \neq i$.

As in the preceding section the solution of (5.4) is particularilarily simple if we choose $\Delta_i = \Delta \circ B_{ii}$. In this case (5.1) may be rewritten

$$(5.5)\qquad \hat{\psi}_i = \begin{cases} 1\ , & \text{if}\quad T_i > c \circ B_{ii} \\ 0\ , & \text{otherwise} \end{cases}$$

where $c = \frac{(N-m)p}{N'} F^{-1}_{p,N'}\ (1-\frac{\epsilon}{a})$.

(*) Over the set of all $(\underset{\sim}{\xi},\underset{\sim}{\Sigma})$ for which $\theta_i(\underset{\sim}{\xi},\underset{\sim}{\Sigma}) \geq \Delta_i$ for at least one i .

It is shown in [5] in a special case (for $p=1$) that $\hat{\psi}$ is not maximin. However, we now proceed to show that if condition (4.8) is satisfied, it usually comes close.

Note that $(\bar{\underset{\sim}{X}}, \hat{\underset{\sim}{\Sigma}})$ are sufficient statistics. Furthermore, the decision problem is clearly invariant under the following group of transformations:

$$G_2 : \begin{cases} \bar{\underset{\sim}{X}} \to \bar{\underset{\sim}{X}}\, \underset{\sim}{g}_2 & g_2 \text{ any non-singular } p\times p \text{ matrix} \\ \hat{\underset{\sim}{\Sigma}} \to \underset{\sim}{g}_2'\, \hat{\underset{\sim}{\Sigma}}\, \underset{\sim}{g}_2 . \end{cases}$$

However, as the Hunt-Stein theorem does not apply for this transformation group (see [6], page 338-339), it can not be concluded that there exists an invariant maximin procedure. Instead we shall require the procedures themselves to be invariant.

Let Z^* be a maximal invariant function for G_2. Let

(5.6) $$\omega_i^* = \{(\underset{\sim}{\xi}, \underset{\sim}{\Sigma}) \mid \xi_k = \frac{b_{ik}}{n_k}\eta \ , \ k=1,\dots,m \ , \ \eta \in \mathbb{R}^p\} .$$

Introduce

$$s(\psi) = \inf_{\underset{\sim}{\xi},\underset{\sim}{\Sigma}} S(\underset{\sim}{\xi},\underset{\sim}{\Sigma};\psi)$$

$$s^*(\psi) = \inf_i \inf_{\omega_i^*} S(\underset{\sim}{\xi},\underset{\sim}{\Sigma};\psi).$$

Trivially $s(\psi) \leq s^*(\psi)$ for any ψ. Thus

$$\sup_\psi s(\psi) \leq \sup_\psi s^*(\psi)$$

where the suprema are taken over all invariant ψ satisfying $\sup R(\underset{\sim}{\xi},\underset{\sim}{\Sigma};\psi) \leq \epsilon$. Hence, if we have found a procedure ψ^* maximizing $s^*(\psi)$, we may judge how close $\hat{\psi}$ comes to being maximin by comparing $s(\hat{\psi})$ with $s^*(\psi^*)$.

To derive $\hat{\psi}$, note that it can easily be proved that $(\bar{\underset{\sim}{X}},\hat{\underset{\sim}{\Sigma}})$ under ω_i^* has a sufficient set of statistics

$$W_i = (b_i' \bar{\underset{\sim}{X}} \,,\; (N-m)\hat{\underset{\sim}{\Sigma}} + \sum_{k=1}^{m-1} V_k' V_k)$$

where $V_1,\ldots,V_{m-1}$ are p-dimensional row vectors independently distributed from $b_i'\bar{\underset{\sim}{X}}$ and $\hat{\underset{\sim}{\Sigma}}$, and themselves mutually independent and normal with expectations zero and covariance matrix $\underset{\sim}{\Sigma}$. Furthermore, it can be seen that G_2 induces the following transformation group in the W_i-space:

$$G_2^* : \begin{cases} b_i'\bar{\underset{\sim}{X}} \rightarrow b_i'\bar{\underset{\sim}{X}}\, g_2 \\ (N-m)\hat{\underset{\sim}{\Sigma}} + \sum_{k=1}^{m-1} V_k'V_k \rightarrow g_2[(N-m)\hat{\underset{\sim}{\Sigma}} + \sum_{k=1}^{m-1} V_k'V_k]\, g_2' \end{cases} .$$

Maximal invariant is known to be

$$T_i^* = (N-1)(b_i'\bar{\underset{\sim}{X}})[(N-m)\hat{\underset{\sim}{\Sigma}} + \sum_{k=1}^{m-1} V_k'V_k]^{-1}(b_i'\bar{\underset{\sim}{X}})' .$$

It follows, as in section 4.1, that T_i^* is sufficient for Z^* under ω_i^* . Hence the density of Z^* may be written

$$h^*(z^*;\underset{\sim}{\xi},\underset{\sim}{\Sigma}) = f_i^*(t_i^*;\theta_i)h_i^*(z^*|t_i^*) \quad , \quad \underset{\sim}{\xi} \in \omega_i^*$$

where f_i^* and h_i^* are respectively the density of T_i^* and the conditional density of Z^* given $T_i^* = t_i^*$. Note that

$$\frac{T_i^*}{B_{ii}} \; \frac{N-p}{(N-1)p} \sim \mathcal{F}_{p,N-1}\left(\frac{\theta_i}{B_{ii}}\right) .$$

As in section 4.1 it follows that a version of ψ^* is given by

(5.7) $$\psi_i^* = \begin{cases} 1 , & \text{if } T_i^* > c_i^* \\ 0 , & \text{otherwise} \end{cases}$$

provided (4.8) holds, that is

$$\max_{i \neq j}[\Delta_j - \Delta_i \left(\frac{B_{ij}}{B_{ii}}\right)^2] > 0 .$$

The constants $c_1^*, \ldots, c_a^*$ are to be determined to satisfy

(5.8) $$\begin{cases} \sum_{i=1}^{a} [1 - F_{p,N-p}(\frac{c_i^*}{B_{ii}} \circ \frac{N-p}{(N-1)p})] = \epsilon \\ F_{p,N-p}(\frac{c_i^*}{B_{ii}} \quad \frac{N-p}{(N-1)p} ; \frac{\Delta_i}{B_{ii}}) \qquad \text{independent of } i . \end{cases}$$

It is clear that the possible improvement of the maximin value of $\hat{\psi}$ is bounded by the quantity

$$Q = F_{p,N'}(\frac{c_i}{B_{ii}} \circ \frac{N'}{(N-m)p} ; \frac{\Delta_i}{B_{ii}}) - F_{p,N-p}(\frac{c_i^*}{B_{ii}} \circ \frac{N-p}{(N-1)p} ; \frac{\Delta_i}{B_{ii}}).$$

(The right hand side is independent of i).

Recall that $N' = N-m-p+1$. Since $N-p$ usually is considerably larger than m , it is clear that Q typically is small. (It is easy to prove that the difference between the two functions $F_{p,N'}(\circ ; \frac{\Delta_i}{B_{ii}})$ and $F_{p,N-p}(\circ ; \frac{\Delta_i}{B_{ii}})$ more than accounts for the value of Q).

References

[1] Anderson, T.W. (1958). Introduction to multivariate statistical analysis. John Wiley & Sons.

[2] Bølviken, E. (1975). Noen desisjonsproblemer i lineær-normale modeller. Cand.Real. thesis at the University of Oslo.

[3] Hall, W.J., Wijsman, R.A. and Ghosh, J.K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. Ann.Math.Stat.,36, 575-614.

[4] Lehmann, E. (1975). A theory of some multiple decision procedures I. Ann.Math.Stat., 28, 1-25.

[5] Lehmann, E. (1961). Some model I problems of selection. Ann.Math.Stat., 32, 990-1012.

[6] Lehmann, E. (1959). Testing statistical hypothesis. John Wiley & Sons.

[7] Spjøtvoll, E. (1971). On the optimality of multiple comparison problems. Ann.Math.Stat., 42, 398-411.