AD-766 469

STATISTICAL MULTIPLE-DECISION PROCEDURES FOR SOME MULTIVARIATE SELECTION PROBLEMS

Ricardo M. Frischtak

Cornell University

Prepared for:

Office of Naval Research
Army Research Office-Durham

July 1973

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

DEPARTMENT OF OPERATIONS RESEARCH
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK

TECHNICAL REPORT NO. 187

July 1973

AD 766469

STATISTICAL MULTIPLE-DECISION PROCEDURES FOR SOME MULTIVARIATE SELECTION PROBLEMS

by

Ricardo M. Frischtak

SEP 1973

Prepared under Contracts

DA-31-124-ARO-D-474, U.S. Army Research Office-Durham

and

N00014-67-A-0077-0020, Office of Naval Research

Approved for Public Release; Distribution Unlimited.

THE FINDINGS IN THIS REPORT ARE NOT TO BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, UNLESS SO DESIGNATED BY OTHER AUTHORIZED DOCUMENTS.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
HISTORICAL REMARKS	vi
STATEMENT OF PROBLEMS	x
CHAPTER 1 - Selection of the Variate with the Largest Population Mean From a Single Multivariate Normal Population with Common Known Variances	1
1.0. Introduction	1
1.1. Preliminaries	2
1.2. Case of Equal Correlations	6
1.3. Case $k = 2$	7
1.4. Case $k = 3$	8
1.5. Case $k > 3$	14
1.6. A Conservative Approximation to the Sample Size	17
1.7. A Sequential Procedure	18
CHAPTER 2 - Selection of the Variate with the Smallest Population Variance From a Single Multivariate Normal Population	22
2.0. Introduction	22
2.1. Formulation of the Problem	23
2.2. Case $k = 2$	25
2.3. Case $k \geq 3$	29
2.4. A Conservative Approximation to the Sample Size	32

Page

2.5. Large-sample Theory 32

CHAPTER 3 - Selection of a Subclass of Variates with the Smallest Population Generalized Variance From a Single Multivariate Normal Population (Asymptotic Theory) 44

3.0. Introduction 44

3.1. Selecting the Smallest Population Generalized Variance (Disjoint Subclasses) 44

3.2. Selecting the Smallest Population Generalized Variance (Intersecting Subclasses) 52

CHAPTER 4 - Selection of Subclasses of Variates or of Populations Based on Measures of Association Between Two Subclasses of Variates (Asymptotic Theory) 60

4.0. Introduction 60

4.1. Preliminaries 61

4.2. Selecting the Best Out of k Populations with Respect to the Population Coefficients of Alienation . . . 64

4.3. Selecting the Best Subclass of Predictors (Single Population) 71

BIBLIOGRAPHY 82

ABSTRACT

In this thesis we are concerned with multiple-decision problems involving the selection of a variate, or of a set of variates, corresponding to the "best" (in a specified sense) parameter of interest, in a multivariate statistical context, in the presence of nuisance parameters. Our main concern is with the rational choice of sample size, when single-stage procedures are employed; all problems are treated using the indifference-zone and subset approaches. We require of these procedures that they guarantee a stipulated probability requirement. In order to determine the sample size necessary to achieve this objective using a single-stage procedure, it is first necessary to minimize the probability of a correct selection associated with the procedure, with respect to the parameters of interest (in a specified region of the parameter space) and the nuisance parameters (for all possible values of these parameters).

Our objective at the outset of research in the present thesis was to provide a solution to the problem of selecting the best subclass of predictors for a specified subclass of variates. (This is accomplished in Chapter 4.) We soon realized that this problem is intimately connected with other selection problems involving covariance matrices

of multivariate normal distributions. Therefore, Chapters 2, 3 and 4 are very closely related, while Chapter 1, although related to these chapters, treats a different topic.

In Chapter 1, we consider the problem of selecting the variate associated with the largest population mean, in a multivariate normal population, with unknown population means, known (unknown) population variances, and unknown population correlations.

In Chapter 2, we consider the problem of selecting the component associated with the smallest population variance, in a multivariate normal population, with totally unknown parameters.

The results of Chapter 2 are extended in Chapter 3 to some selection problems concerning generalized variances in multivariate normal populations. The results of this chapter involve large-sample (asymptotic) theory.

Finally, in Chapter 4, we solve (using asymptotic theory) two problems which have aroused recent interest in the literature. The first is that of selecting the multivariate normal population (among independent populations), with the smallest vector coefficient of alienation between two sets of components. Gupta and Panchapakesan (1969) and Rizvi and Solomon (1973) give different formulations and solutions for this problem.

Secondly, and perhaps more importantly from the viewpoint of applications, we consider the problem of selecting the best subclass of predictors for a fixed subclass of variates, each of the contending subclasses being correlated with the subclass previously specified. This problem is treated in a multivariate normal context, and a quite general asymptotic solution is displayed. The vector coefficient

of alienation is used as a measure of association. Ramberg (1969) and Arvensen (1971) obtained partial solutions for related problems. All asymptotic results of Chapters 2 - 4 are valid under quite general families of multivariate distributions, although, for simplicity, we have stated them under normality assumptions.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Department of Operations Research College of Engineering, Cornell University Ithaca, New York 14850	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

STATISTICAL MULTIPLE-DECISION PROCEDURES FOR SOME MULTIVARIATE SELECTION PROBLEMS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, July 1973

5. AUTHOR(S) (First name, middle initial, last name)

Frischtak, Ricardo M.

6. REPORT DATE July 1973	7a. TOTAL NO. OF PAGES ~~85~~ 102	7b. NO. OF REFS 50
8a. CONTRACT OR GRANT NO DA-31-124-ARO-D-474 b. ~~XXXXXXXXXX~~ N00014-67-A-0077-0020 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 187 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. ~~XXXXXXXXXXXXXXXXX~~ Sponsoring Military Activity U.S. Army Research Office Durham, N.C. 27706	12. SPONSORING MILITARY ACTIVITY Logistics and Mathematical Statistics Branch, Office of Naval Research Washington, D.C. 20360

13. ABSTRACT

The following statistical multiple-decision problems are considered for a multivariate normal distribution with unknown (or partially known) covariance matrix, using the indifference-zone and subset approaches: a) selecting the variate with the largest population mean; b) selecting the variate with the smallest population variance; c) selecting the subclass of variates with the smallest population generalized variance; d) selecting the population with the smallest vector coefficient of alienation between two subclasses of variates; e) selecting the best subclass of predictors for a specified subclass of variates. Small-sample theory is employed in a) and b), while large-sample theory is used in b), c), d) and e).

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

1b

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-0811

Unclassified
Security Classification

A-31408

14 KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
generalized variances						
indifference-zone approach						
mathematical statistics						
multivariate prediction						
ranking procedures						
selection procedures						
statistical multinle-decision						
subset approach						
vector coefficient of alienation						

HISTORICAL REMARKS

The birth and development of the idea of treating certain statistical problems as decision problems is generally credited to A. Wald. His work culminated with the publication of the book Statistical Decision Functions (see Wald (1950)).

The first instances of multiple-decision problems, with some bearing on the present thesis, may be traced back to this period. In particular, we should mention the work of Paulson (1949, 1952a, 1952b) who treated classification schemes, comparison with a control and the "slippage" problem. Bahadur (1950) and Bahadur and Goodman (see also Lehmann (1957, 1961, 1966) and Eaton (1967a)), proved strong optimality properties for "natural" selection procedures, when the experimenter is interested in selecting the "best" population.

Bechhofer (1954) wrote a pioneering paper in which he defined precisely several possible ranking and selection goals as alternatives to classical tests of homogeneity. In this paper, the idea of planning the sample size using an indifference-zone approach with the purpose of guaranteeing a specified probability of a correct selection or ranking was set forth.

Somerville (1954) considered a selection problem, with explicit reference to the use of the category selected after the decision process. In planning the initial experiment, he considered loss functions which "take into consideration the amount of use to be made of

the result, the cost of making a wrong decision and the cost of sampling". A minimax criterion was used.

W. J. Hall (1958, 1959) introduced the notion of most economical multiple decision rules (roughly, rules which require the smallest sample sizes to achieve a certain objective). He then proved the most economical character of some of Bechhofer's rules.

Dunnett (1960) proposed selection procedures for normal means, introducing prior distributions on the means, and assuming a known and particular covariance matrix. After a rather complete analysis without loss functions, he introduced linear loss functions and invoked a minimax criterion, as in Somerville (1954), and other criteria, such as minimizing the maximum regret.

Gupta (1956) introduced the subset selection approach, in which the experimenter's goal is to select a subset of variates, including the best one. In many practical situations, these may be regarded as screening procedures, to be used in the presence of a large number of variates, before one demands the selection of a best one.

Much of the literature on multiple-decision (selection and ranking) procedures since then has been concerned with the indifference-zone and subset approaches. The most important development using indifference-zone ideas is perhaps the monograph Sequential Identification and Ranking Procedures by Bechhofer, Kiefer and Sobel (1968), in which sequential procedures for ranking parameters of Koopman-Darmois populations are treated. This book also contains a rather complete survey of the field. The reader may consult it for references to practically all of the literature up to 1968.

The following papers, using the indifference-zone approach, are particularly relevant to the present thesis:

Bechhofer and Sobel (1954) considered the problem of ranking population variances for independent normal variates;

Bechhofer (1968) studied ranking problems arising in connection with multiply-classified variances and a multiplicative model for these variances;

Paulson (1964) gave a closed fully sequential procedure, which eliminates noncontending populations, for the problem of selecting the normal population with the largest population mean, when the common population variance is known or unknown;

Ramberg (1969) considered the problem of finding a best set of predictors for a specified variate, in a multivariate normal context;

Rizvi and Solomon (1973) considered the problem of selecting the population with the largest population multiple correlation coefficient between a specified variate and a set of variates.

In the area of subset selection procedures, the reader is referred to the papers of Gupta (1965) and Gupta and Panchapakesan (1972) wherein there are given rather broad surveys of the main results, and many of the important references.

The following papers, using the subset approach, are important to this thesis:

Gupta and Sobel (1962) considered the problem of selecting a subset of normal variates containing the variate with the smallest population variance;

Gupta and Panchapakensan (1969) considered problems of selection in terms of multiple correlation coefficients and conditional generalized variances;

Arvensen (1971) considered the problem of selecting a subset of subclasses of variates containing the best predictor subclass, and used a Bayesian approach.

Finally, there are several papers which employ different formulations for selection and ranking problems. Among these, we mention Fabian (1962) and Mahamunulu (1966, 1967). Recently, Gupta and Santner (1972) proposed a multiple-decision procedure which selects a subset of size not exceeding a specified upper-bound; their procedure bridges the indifference-zone and subset approaches.

STATEMENT OF PROBLEMS

In this section we formulate the problems of interest to us in a general enough framework for our purposes. Let $X = (X_1,\ldots,X_k)$ be a random vector with distribution function $F_X(\cdot|\Theta,\phi)$, where $\theta = (\theta_1,\ldots,\theta_k)$ and $\phi = (\phi_1,\ldots,\phi_s)$, each θ_i and ϕ_j being unknown scalars. Our major interest is in the θ_i while the ϕ_j are regarded as nuisance parameters. Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ be the ranked values of the elements of the vector θ . We will say that X_i is associated with θ_i if the marginal distribution of X_i depends on θ_i and not on $\{\theta_j, j \neq i\}$. It is assumed that no prior knowledge exists concerning the pairing of the $\theta_{[j]}$ with the X_i $(1 \leq i,j \leq k)$.

Indifference-zone formulation

Our goal, when using the indifference-zone approach, will be to select the variate X_i associated with $\theta_{[k]}$. For this goal, we permit only k possible decisions, namely "X_i $(1 \leq i \leq k)$ is associated with $\theta_{[k]}$." There are many other ranking goals treated in the literature, but we will consider only this one in the present thesis. Here correct selection means selection of the variate associated with $\theta_{[k]}$ (or of any one of $\theta_{[q]},\theta_{[q+1]},\ldots,\theta_{[k]}$ if

$\theta_{[q]} = \theta_{[k]}$).

The probability requirement associated with this goal is not completely formulated until a "distance" function $\psi(\theta_i,\theta_j)$, between the marginal distributions of X_i and X_j , is adopted. We assume ψ to satisfy:

$\psi(a,b) \geq 0$ for all pairs (a,b) ;

$\psi(a,b) = 0$ iff $a = b$;

$\psi(a,b) = \psi(b,a)$;

$\psi(a,b)$ is strictly increasing in a for fixed b , and strictly decreasing in b for fixed a , if $a \geq b$.

The specification of this distance function is fundamental when using the indifference-zone approach. Bechhofer, Kiefer and Sobel (1968) showed that, in certain problems, the adoption of a particular distance function implies the nonexistence of a single or multi-stage procedure which will guarantee the probability requirement (to be defined shortly).

The experimenter specifies real constants $\{\delta^*,P^*\}$, $\delta^* > 0$, $1/k < P^* < 1$, prior to experimentation. For example, if θ_i are location parameters in the marginal distribution of X_i $(1 \leq i \leq k)$), we may take $\psi(a,b) = a - b$. If the θ_i are scale parameters, we may use $\psi(a,b) = \log(a/b)$.

When there exists a decision Rule R which guarantees the probability requirement,

$$\inf_{\Omega} P_{\theta,\phi}(\text{Correct selection using } \mathcal{R}) \geq P^* ,$$

where

$$\Omega = \{(\theta,\phi) | \psi(\theta_{[k]},\theta_{[k-1]}) \geq \delta^*\} ,$$

we say that $\mathcal{R}$ provides a solution to the selection problem relative to the distance function ψ . Ω is called the <u>preference zone</u>, and all parameter points not in Ω are said to be in the <u>indifference-zone</u>. When the experimenter adopts this approach he states in effect that, for all parameter points not in Ω , he is indifferent as to which decision is made. Any point (θ,ϕ) for which the infimum is attained is called a least favorable configuration of the parameters.

Usually, we define $\mathcal{R} = \mathcal{R}(N)$, a function of the sample size N . Then we determine the smallest N necessary to guarantee the above probability requirement when $\mathcal{R}(N)$ is employed.

<u>Subset formulation</u>

Another possible goal is to select a subset of variates X_i $(1 \leq i \leq k)$ containing a variate associated with $\theta_{[k]}$. There are $2^k - 1$ possible decisions, namely all nonempty subsets of $(X_1,\ldots,X_k)$. When using the so-called subset approach there is no need to consider distance functions; instead, the experimenter specifies $\{P^*\}$, $1/k < P^* < 1$ before experimentation starts. Then, if correct selection means selection of a subset of variates containing a variate associated with $\theta_{[k]}$, Rule $\mathcal{R}$ is said to provide a solution to the selection

problem if it guarantees the probability requirement,

$$\inf_{\theta,\phi} P_{\theta,\phi}(\text{Correct selection using } R) \geq P^* .$$

In the problem we consider, $R = R_{d^*}(N)$ is a function of the sample size N, and of d^*, which is a specified "yardstick." Our method will be to fix d^*, and then find the smallest N such that the probability requirement is guaranteed, when $R_{d^*}(N)$ is employed. This is in contrast with the usual formulation of such problems using the subset approach, where N is fixed and d^* is found to guarantee the same probability requirement. It will be seen that the mathematical problems are equivalent, and our approach is taken just as a matter of convenience.

A few words about notation. Correct selection will always mean a selection for which the goal under consideration is achieved. PCS denotes probability of a correct selection. a.d. stands for asymptotic distribution. PCS_a denotes PCS, and E_a the operator expectation, when an a.d. theory is employed.

CHAPTER 1

SELECTION OF THE NORMAL VARIATE WITH THE LARGEST POPULATION MEAN FROM A SINGLE MULTIVARIATE NORMAL POPULATION WITH COMMON KNOWN VARIANCES

1.0. Introduction

In most of the present chapter we consider a k-variate normal population and propose single-stage procedures for selecting the component with the largest population mean. We assume throughout that the population variances are common and known.

Section 1.1 gives certain preliminaries including a statement of an indifference-zone and a subset formulation of the problem, which we later treat simultaneously. In Section 1.2 we consider, for $k \geq 3$, the simple special case of equal but unknown population correlations. The case $k = 2$ is treated in Section 1.3. For $k = 3$, we show in Section 1.4 that the theory is quite involved, but still tractable; exact small-sample results are obtained. However, for $k > 3$, only tentative results are available; these are given in Section 1.5. In Section 1.6 we use Bonferroni's inequality to determine a conservative approximation to the sample size required to guarantee the probability requirement for the general $k \geq 3$ case. Finally, in Section 1.7, we show that Paulson's (1964) sequential procedure can be modified slightly to apply to the indifference-zone formulation of the problem described in this chapter.

The most interesting results of the present chapter, when single-stage procedures are used, are the following: a) The fact

that the least favorable configuration of the correlation matrix depends on the sample size; b) Using "natural" procedures (i.e., the same procedures, based only on sample means, that have been used for independent components), the probability of a correct selection can attain values less than $1/k$, when the sample size is small; therefore, these "natural" procedures are not minimax when this situation obtains.

1.1. Preliminaries

Consider a k-variate normal population $X^t = (X_1,\ldots,X_k)$ with population mean vector $\mu^t = (\mu_1,\ldots,\mu_k)$ and population covariance matrix $\sigma^2 R$. We assume that σ^2 is the common known population variance, while $R = (\rho_{ij})$ is the unknown population correlation matrix. Let $\mu_{[1]} \leq \cdots \leq \mu_{[k]}$ be the ranked values of the μ_i. We assume no prior knowledge concerning the values of the μ_i, or of the pairing of the $\mu_{[i]}$ with the variates X_j $(1 \leq i,j \leq k)$.

Indifference-zone formulation

The experimenter's goal is to select the variate associated with $\mu_{[k]}$. The experimenter specifies constants $\{\delta^*,P^*\}$, $\delta^* > 0$, $1/k < P^* < 1$, prior to the start of experimentation. Let $PCS_{\mathcal{R}}(\mu,R)$ denote the probability of a correct selection using decision procedure $\mathcal{R}$, when μ and R are the unknown set of parameters. We limit consideration to decision procedures $\mathcal{R}$ which guarantee the probability requirement:

$$\inf_{\Omega} PCS_R(\mu,R) \geq P^*$$

where

$$\Omega = \{(\mu,R) \mid \mu_{[k]} - \mu_{[k-1]} \geq \delta^* \text{ , R a correlation matrix}\} .$$

Most of the present chapter will be concerned with single-stage procedures. For such procedures, the experimenter takes a sample of N independent vector observations, $X_\alpha^t = (X_{1\alpha},\ldots,X_{k\alpha})$ $(\alpha = 1,\ldots,N)$. The following decision rule has been proposed for this <u>indifference-zone</u> formulation of the problem:

<u>Rule B</u>: Let $\overline{X}_j = \sum_{\alpha=1}^{N} X_{j\alpha}/N$ $(1 \leq j \leq k)$. Then assert that the variate associated with $\overline{X}_{[k]} = \max\{\overline{X}_1,\ldots,\overline{X}_k\}$ has the largest population mean.

The problem is to determine the smallest value of the integer N for which the probability requirement is guaranteed if Rule B is employed.

Bechhofer (1954) introduced the indifference-zone philosophy when solving the above problem for the case where $R = I_k$, i.e., when all components of X are mutually independent. Our objective is to generalize his result in the multivariate setting.

<u>Subset formulation</u>

If the experimenter's goal is to select a subset of components of X which will include the component associate with $\mu_{[k]}$, he specifies $\{P^*\}$, $1/k < P^* < 1$, prior to experimentation. Letting

$PCS_R(\mu,R)$ be defined as above, we limit consideration to decision procedures R which guarantee the probability requirement:

$$\inf_{\mu,R} PCS_R(\mu,R) \geq P^* .$$

The following decision rule has been proposed for this subset formulation of the problem:

Rule G: Include the component associated with $\overline{X}_j$ in the selected subset if $\overline{X}_j \geq \overline{X}_{[k]} - d^*$, where $d^* > 0$ is specified in the units of the problem.

Our task is then to determine the smallest integer N for which the probability requirement is guaranteed when Rule G is used.

This rule was introduced by Gupta (1956), where the subset approach was first proposed. The problem solved by Gupta (1956) assumed $R = I_k$ (independent components). Our objective is to generalize his result in the multivariate setting.

In order to obtain solutions to these problems we will first derive some preliminary results which will be used throughout the present chapter. We assume, without loss of generality, that $\mu_k \geq \mu_j$ $(j \neq k)$.

Lemma 1.1. Let

$$Y_j^i = \frac{\overline{X}_i - \overline{X}_j - (\mu_i - \mu_j)}{\sigma\left(\frac{2}{N}\right)^{1/2}(1-\rho_{ij})^{1/2}} \quad (j \neq i) .$$

Then, for each fixed i , the $\{Y_j^i , j \neq i\}$ have a standard multivariate normal distribution with

$$\mathrm{corr}(Y_j^i, Y_{j'}^i) \equiv \gamma_{jj'}^i = \frac{1-\rho_{ij}-\rho_{ij'}+\rho_{jj'}}{2(1-\rho_{ij})^{1/2}(1-\rho_{ij'})^{1/2}} .$$

Proof. The result follows at once from the above definitions.

For simplicity of notation, we now let

$$Y_j \equiv Y_j^k , \quad \gamma_{ij} \equiv \gamma_{ij}^k \quad (1 \le i,j \le k-1) .$$

Lemma 1.2. Let the $\{Y_j , j \neq k\}$ be as in Lemma 1.1.

(a) If Rule B is used, then in Ω we have

$$\text{(1.1)} \qquad PCS \ge P(Y_j > -a(N)(1-\rho_{jk})^{-1/2} , j \neq k)$$

where $a(N) = (\delta^*/\sigma)(N/2)^{1/2}$.

(b) If Rule G is used, then we have

$$\text{(1.2)} \qquad PCS \ge P(Y_j > -a(N)(1-\rho_{jk})^{-1/2} , j \neq k)$$

where $a(N) = (d^*/\sigma)(N/2)^{1/2}$.

Proof. We use Lemma 1.1 and notice that, in (a)

$$PCS = P(\overline{X}_k > \overline{X}_j , j \neq k) = P(Y_j > -\frac{(\mu_k-\mu_j)(N/2)^{1/2}}{\sigma(1-\rho_{jk})^{1/2}} , j \neq k)$$

while in (b),

$$PCS = P(\overline{X}_k > \overline{X}_j - d^* , j \neq k) = P(Y_j > -\frac{(\mu_k-\mu_j+d^*)(N/2)^{1/2}}{\sigma(1-\rho_{jk})^{1/2}} , j \neq k) . \quad \text{QED}$$

Our task for most of the present chapter is to minimize the right-hand sides of (1.1) and (1.2) with respect to R. Formally, these are identical problems, and thus we will not make a distinction, as far as the minimization is concerned, between the indifference-zone and the subset approach. The expressions (1.1) and (1.2) depend on δ^*/σ or d^*/σ, which may be specified, instead of σ alone.

Lemma 1.3. Let S be the size of the selected subset associated with Rule G. Then

$$\text{(a)} \quad E(S|\mu,R) = \sum_{i=1}^{k} P\left(Y_j^i > -\frac{(\mu_i-\mu_j+d^*)(N/2)^{1/2}}{\sigma(1-\rho_{ij})^{1/2}},\ i \neq j\right)$$

where the $\{Y_j^i,\ j \neq i\}$ are as in Lemma 1.1.

(b) $\sup_{\mu,R} E(S|\mu,R) = k$, which occurs when $\mu_1 = \ldots = \mu_k$, and all elements of R are equal to unity.

Proof. This result is a consequence of previous developments.

1.2. Case of equal correlations

When the off-diagonal elements of R are known to be equal to a common unknown ρ $(-1/(k-1) \leq \rho \leq 1)$, the minimization of (1.1) and (1.2) simplifies considerably. In this case, $\gamma_{ij} = 1/2$ $(i \neq j)$, and the minimum occurs when $\rho = -1/(k-1)$, in which case the k-variate distribution of X is degenerate, being concentrated in a linear subspace of $k - 1$ dimensions. However, the distribution of the $\{Y_j,\ j \neq k\}$ is not degenerate. Therefore,

one obtains for either (1.1) or (1.2),

$$\text{(1.3)} \qquad \inf PCS = P(Y_j > -a(N)((k-1)/k)^{1/2}, \; j \neq k)$$

where the $\{Y_j, j \neq k\}$ are as in Lemma 1.1, with $\gamma_{ij} = 1/2 \quad (i \neq j)$.

The infimum in (1.3) was known to Milton (1963) and Gupta (1963). They have provided tables for the distribution of the $(Y_j, j \neq k)$, for several values of k. Using these tables, an experimenter determines $h = h(k,P^*) > 0$, such that $P(Y_j > -h, j \neq k) = P^*$, and upon equating

$$a(N)((k-1)/k)^{1/2} = h ,$$

a value of N_ℓ then follows; the experimenter employs the smallest integer $\geq N_\ell$.

It should be mentioned that Rule B has many optimum properties when the correlations are equal. For a large class of "natural" loss functions, the rule has uniformly smallest risk function among all symmetrical (invariant under permutation of components) procedures, being minimax and admissible (cf. Eaton (1967a), Lehmann (1966), Hall (1959)).

1.3. Case $k = 2$

Although this is a particular case of the preceding section, we state the result explicitly, so that it may be compared easily with the results of section 1.4.

Here, since $k = 2$, (1.1) and (1.2) reduce to a univariate normal integral, the minimum of which clearly occurs when $\rho_{12} = -1$. Therefore,

$$\inf \mathrm{PCS} = P(Y_1 > -a(N)2^{-1/2}) , \tag{1.4}$$

where Y_1 is a standard univariate normal variate.

1.4. Case $k = 3$

Here the problem is considerably more complicated than for $k = 2$. We wish to minimize the right-hand side of (1.1) and (1.2),

$$\overline{\mathrm{PCS}} = P(Y_1 > -a(N)(1-\rho_{13})^{-1/2} , Y_2 > -a(N)(1-\rho_{23})^{-1/2})$$

over all permissible values of $\rho_{12},\rho_{13},\rho_{23}$, where the $\{Y_1,Y_2\}$ have a standard bivariate normal distribution with

$$\mathrm{corr}(Y_1,Y_2) = \gamma_{12} = \frac{1-\rho_{13}-\rho_{23}+\rho_{12}}{2(1-\rho_{12})^{1/2}(1-\rho_{23})^{1/2}} .$$

The region of Euclidean 3-space where R is positive semi-definite is given by $\det R \geq 0$, $\rho_{ij}^2 \leq 1$ $(i \neq j)$. The region $\det R \geq 0$ is the ellipsoid

$$1 + 2\rho_{12}\rho_{13}\rho_{23}-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2 \geq 0 .$$

Lemma 1.4.

$$\frac{\partial}{\partial\rho_{12}} \overline{\mathrm{PCS}} > 0 \quad \text{for} \quad \rho_{13} \neq 1 \quad \text{and} \quad \rho_{23} \neq 1 .$$

Proof. Let $f_{\gamma_{12}}(y_1,y_2)$ be the p.d.f. of $\{Y_1,Y_2\}$. According to the known relation (cf., for example, Plackett (1954)),

$$\frac{\partial}{\partial\gamma_{12}} f_{\gamma_{12}}(y_1,y_2) = \frac{\partial^2}{\partial y_1 \partial y_2} f_{\gamma_{12}}(y_1,y_2) ,$$

$$\frac{\partial}{\partial\rho_{12}}\overline{PCS} = \int_{\frac{-a(N)}{(1-\rho_{13})^{1/2}}}^{\infty} \int_{\frac{-a(N)}{(1-\rho_{23})^{1/2}}}^{\infty} \frac{\partial^2}{\partial y_1 \partial y_2} f_{\gamma_{12}}(y_1,y_2)\,dy_1 dy_2 \frac{\partial\gamma_{12}}{\partial\rho_{12}}$$

$$= f_{\gamma_{12}}\left(\frac{-a(N)}{(1-\rho_{13})^{1/2}}, \frac{-a(N)}{(1-\rho_{23})^{1/2}}\right)(1/2)(1-\rho_{13})^{-1/2}(1-\rho_{23})^{-1/2} > 0 .$$

QED

Some of the ideas underlying many proofs in this thesis, including the one above, derive from a basic paper of Slepian (1962).

It is easy to check that the inf of $\overline{PCS}$ does not occur when either ρ_{13} or ρ_{23} equals unity. Hence, this case is excluded in the following discussion.

Lemma 1.5. inf $\overline{PCS}$ occurs when $\det R = 0$.

Proof. Suppose we fix ρ_{13} and ρ_{23}. By the previous lemma, we would set ρ_{12} at its smallest possible value, which is the smallest root of the quadratic equation $\det R = 0$; thus we obtain

$$\rho_{12} = \rho_{13}\rho_{23} - (1-\rho_{13}^2)^{1/2}(1-\rho_{23}^2)^{1/2} \geq -1 . \qquad \text{QED}$$

We proceed directly to the minimization of $\overline{PCS}$. Let us define the following Lagrangean function,

$$F = \overline{PCS} + \tilde{\lambda} \det R .$$

The parameter point R, which leads to an infimum of $\overline{PCS}$, subject to the restriction $\det R = 0$, must satisfy the following equations:

$$(1.5)\quad \frac{\partial F}{\partial \rho_{12}} = f_{\gamma_{12}}\left(\frac{-a(N)}{(1-\rho_{13})^{1/2}}, \frac{-a(N)}{(1-\rho_{23})^{1/2}}\right) \frac{1}{2(1-\rho_{13})^{1/2}(1-\rho_{23})^{1/2}} + 2\tilde{\lambda}(\rho_{13}\rho_{23} - \rho_{12}) = 0 .$$

$$(1.6)\quad \frac{\partial F}{\partial \rho_{13}} = f_{\gamma_{12}}\left(\frac{-a(N)}{(1-\rho_{13})^{1/2}}, \frac{-a(N)}{(1-\rho_{23})^{1/2}}\right) \frac{\rho_{13}+\rho_{12}-\rho_{23}-1}{4(1-\rho_{23})^{1/2}(1-\rho_{13})^{3/2}} + \frac{a(N)}{2(1-\rho_{13})^{3/2}} \int_{\frac{-a(N)}{(1-\rho_{23})^{1/2}}}^{\infty} f_{\gamma_{12}}\left(\frac{-a(N)}{(1-\rho_{13})^{1/2}}, y_2\right) dy_2 + 2\tilde{\lambda}(\rho_{23}\rho_{12} - \rho_{13}) = 0 .$$

$$(1.7)\quad \frac{\partial F}{\partial \rho_{23}} = f_{\gamma_{12}}\left(\frac{-a(N)}{(1-\rho_{13})^{1/2}}, \frac{-a(N)}{(1-\rho_{23})^{1/2}}\right) \frac{\rho_{23}+\rho_{12}-\rho_{13}-1}{4(1-\rho_{13})^{1/2}(1-\rho_{23})^{3/2}} + \frac{a(N)}{2(1-\rho_{23})^{3/2}} \int_{\frac{-a(N)}{(1-\rho_{13})^{1/2}}}^{\infty} f_{\gamma_{12}}\left(y_1, \frac{-a(N)}{(1-\rho_{23})^{1/2}}\right) dy_1 + 2\tilde{\lambda}(\rho_{12}\rho_{13} - \rho_{23}) = 0 .$$

$$(1.8)\quad \frac{\partial F}{\partial \tilde{\lambda}} = \det R = 0 .$$

By the symmetry of equations (1.6) and (1.7) with respect to ρ_{13} and ρ_{23}, one is led to study a solution of the form

$$\rho_{13} = \rho_{23} = \tau$$

which consequently implies by (1.8) and Lemma 1.4, that

$$\rho_{12} = 2\tau^2 - 1 .$$

Moreover, by substitution, we have $\gamma_{12} = -\tau$. With such a solution, equations (1.6) and (1.7) become identical, and in order to find we must eliminate the Lagrange multiplier $\tilde{\lambda}$ between equations (1.5) and (1.6). After simplifications, we arrive at,

$$(1.9) \quad \int_{\frac{-a(N)}{(1-\tau)^{1/2}}}^{\infty} f_{-\tau}\left(\frac{-a(N)}{(1-\tau)^{1/2}}, y\right) dy = \frac{(1-\tau)^{3/2}}{a(N)} f_{-\tau}\left(\frac{-a(N)}{(1-\tau)^{1/2}}, \frac{-a(N)}{(1-\tau)^{1/2}}\right) .$$

Using the factorization $f(x,y) = f(y|x)f(x)$ for the density inside the integral, and simplifying further still, we obtain,

$$(1.10) \quad \int_{-b}^{\infty} (2\pi)^{-1/2} \exp(-y^2/2) dy = (2\pi)^{-1/2} (1/b) \exp(-b^2/2)$$

where

$$b = a(N)(1+\tau)^{1/2}/(1-\tau) .$$

Equation (1.10) has a unique solution, $b \cong .5$, which gives

$$a(N) \cong .5(1 - \tau)(1 + \tau)^{-1/2}$$

and

$$\overline{PCS} = P(Y_i > \frac{-a(N)}{(1-\tau)^{1/2}}, \ i = 1,2)$$

$$\cong P(Y_i > -\frac{.5(1-\tau)^{1/2}}{(1+\tau)^{1/2}}, \ i = 1,2)$$

where $\mathrm{corr}(Y_1, Y_2) = -\tau$.

For numerical evaluations of $\overline{PCS}$ it is convenient to start with a fixed value of τ $(-1 < \tau < 1)$, and then obtain $a(N)$ and $\overline{PCS}$. Some rough numerical calculations are given in Table 1.1. The purpose of this table is to illustrate the variation of $\overline{PCS}$ and $a(N)$ with τ , rather than to provide the reader with a working device. Table 1.1 was computed using the National Bureau of Standards (1959) tables of the bivariate normal integral.

One notes that as $a(N)$ increases so does $\overline{PCS}$, as is to be expected; but as $a(N) \to 0$, $\overline{PCS}$ attains values less than $1/3$. In other words, for small values of $a(N)$ one does better by simply selecting one of the three components at random rather than by using Rules **B** or **G**. Therefore, for small values of $a(N)$, these rules are not minimax (with respect to simple $0 - 1$ loss functions).

Another curious fact is that, for small $a(N)$, the least favorable configuration of R is very close to a correlation matrix all entries of which are equal to unity. However, this is also the most favorable configuration of R , since then $\overline{PCS} = 1$. In other words, for $a(N)$ close to zero, the least favorable configuration of R is "close" to the most favorable configuration of R . One may interpret this as happening when σ is large compared to δ^* or d^* , in which case our intuition fails.

We have not been able to prove analytically that the solution (1.10) of equations (1.5), (1.6), (1.7) and (1.8) which we selected is indeed the one which leads to the global minimum of $\overline{PCS}$. However, some limited numerical results do indicate that this is in fact the global minimum. We recommend that more extensive numerical

TABLE 1.1

Values of the Infimum of the Probability of a Correct Selection as a Function of τ $(k = 3)$

τ	a(N)	$\overline{PCS}$
-.9	3.00	.98
-.7	1.55	.83
-.5	1.06	.69
-.2	.67	.55
0	.50	.48
.2	.37	.41
.3	.31	.37
.4	.26	.33
.5	.21	.30
.6	.16	.27
.7	.12	.23
.8	.07	.19
.9	.04	.13
.99	$\cong$.00	.04

computations be carried out in the future, and hope to do so ourselves.

Table 1.1 shows that when $a(N) \to \infty$, which may also be thought of as $N \to \infty$, the least favorable configuration is near $\rho_{13} = \rho_{23} = -1$, $\rho_{12} = 1$. Another way to see this is to notice that as $a(N) \to \infty$, equations (1.5), (1.6), (1.7) and (1.8) become,

$$2\tilde{\lambda}(\rho_{13}\rho_{23} - \rho_{12}) = 0$$

$$2\tilde{\lambda}(\rho_{23}\rho_{12} - \rho_{13}) = 0$$

$$2\tilde{\lambda}(\rho_{12}\rho_{13} - \rho_{23}) = 0 \qquad \tilde{\lambda} \neq 0$$

$$\det R = 0 .$$

The only solutions of these equations are $\rho_{13} = \rho_{23} = \rho_{12} = 1$, the most favorable configuration, and $\rho_{13} = \rho_{23} = -1$, $\rho_{12} = 1$, the least favorable configuration.

1.5. Case $k > 3$

In this section our results are more tentative than the results of the previous section, since we have not made any numerical computations to verify that what we obtain is indeed a least favorable configuration. The present section could be written in parallel with the previous one, the basic ideas being the same, except for the much more involved algebra. Instead, we simply give below the main results, without proofs. Let

$$\overline{PCS} \equiv P(Y_j > -a(N)(1 - \rho_{jk})^{-1/2} , j \neq k)$$

where the $\{Y_j , j \neq k\}$ are as in Lemma 1.1.

Lemma 1.6.

$$\frac{\partial \overline{PCS}}{\partial \rho_{ij}} > 0 \quad \text{for} \quad (1 \leq i < j \leq k - 1) \quad \text{if} \quad \rho_{ik} \neq 1 \quad (1 \leq i \leq k - 1) .$$

Lemma 1.7. $\inf \overline{PCS}$ occurs when $\det R = 0$.

Consider the Lagrangean function

$$F = \overline{PCS} + \tilde{\lambda} \det R .$$

It can be shown that the equations

$$\frac{\partial F}{\partial \rho_{ij}} = 0 \quad (1 \leq i < j \leq k) , \quad \frac{\partial F}{\partial \tilde{\lambda}} = 0 ,$$

admit a solution of the form

$$\rho_{1k} = \dots = \rho_{k-1,k} = \tau \quad (-1 < \tau < 1) ,$$

$$\rho_{12} = \dots = \rho_{k-2,k-1} = \{(k-1)\tau^2 - 1\}/(k-2) .$$

Moreover, by substitution,

$$\rho \equiv \gamma_{ij} = \{(k - 3) - (k - 1)\tau\}/(2(k-2)) .$$

In the present context, equation (1.10) is a particular case of (1.11), when $k = 3$.

$$(1.11) \quad \int \dots \int f_{\Gamma_1}(z_1, \dots, z_{k-2}) dz_1 \dots dz_{k-2}$$

$$= \frac{(k-1)(1-\tau)^{3/2}}{2(2\pi)^{1/2} a(N)(1-\rho^2)^{1/2}} \exp\left\{ - \frac{1}{2} \frac{a^2(N)(1-\rho)}{(1-\tau)(1+\rho)} \right\}$$

$$\cdot \int \dots \int f_{\Gamma_2}(w_1, \dots, w_{k-3}) dw_1 \dots dw_{k-3} ,$$

where the limits of integration in the left-hand size are from $-a(N)(1-\rho)^{1/2}(1-\tau)^{-1/2}(1+\rho)^{-1/2}$ to ∞, while the right-hand size limits of integration are from $-a(N)(1-2\rho)(1+\rho)^{1/2}(1-\tau)^{-1/2}(1-\rho)^{-1/2}(2\rho+1)^{-1/2}$ to ∞. Moreover, f_{Γ_i} $(i = 1,2)$ are the p.d.f.'s of standard multivariate normal distributi .s with correlation matrices Γ_i, where Γ_1 has all its off-diagonal elements equal to $\rho/(1 + \rho)$, while Γ_2 has all its off-diagonal elements equal to $\rho/(2\rho+1)$.

(1.11) does not lend itself to an easy solution as did (1.10) where we found b and consequently computed Table 1.1. Although we have not pursued numerical computations for $k > 3$, we recommend that (1.11) be used as follows: for fixed values of τ $(-1 < \tau < 1)$, (1.11) gives a unique value of $a(N)$; then, with τ and $a(N)$, one computes $\overline{PCS}$. As τ varies from 1 to -1, $a(N)$ ranges from 0 to ∞, and $\overline{PCS}$ from 0 to 1.

Again for $k > 3$, the $\overline{PCS}$ may attain values less than $1/k$, if $a(N)$ is sufficiently small. For example, if we take $\tau = (k - 3)/(k - 1)$, implying $\rho = 0$, then

$$\overline{PCS} = \left\{ \int_{\frac{-a(N)}{(1-\tau)^{1/2}}}^{\infty} f(z)dz \right\}^{k-1} ,$$

where $f(z)$ is a standard univariate normal density. For $a(N)$ very small,

$$\overline{PCS} \cong 2^{-(k-1)} < 1/k .$$

While for $k = 3$ we were able to show computationally, in a few cases, that the minimum obtained is indeed a global minimum, for $k > 3$ these computational results are very difficult to obtain because of the unavailability of tables of general multivariate normal integrals of dimension greater than 2 . It may be possible that a proof exists for the uniqueness of the minimum, but we were unable to provide it.

1.6. <u>A conservative approximation to the sample size when $k \geq 3$</u>

While expressions such as (1.11) seem to be unmanageable, a lower bound on $\overline{PCS}$ may be obtained using Bonferroni's inequality as given in Feller (1968). Indeed, for a collection of p events $A_1, \ldots, A_p$,

$$P(\bigcap_{i=1}^{p} A_i) = 1 - P(\bigcup_{i=1}^{p} A_i^c) \geq 1 - \sum_{i=1}^{p} P(A_i^c) = \sum_{i=1}^{p} P(A_i) - (p - 1) ,$$

where A_i^c is the complement of A_i , and Boole's inequality has been used.

Therefore, since we know the minimum when $k = 2$, if we take any $k \geq 3$,

$$\begin{aligned}
\overline{PCS} &= P(Y_i > - a(N)(1 - \rho_{ik})^{-1/2} , i \neq k) \\
&\geq \sum_{i=1}^{k-1} P(Y_i > - a(N)(1 - \rho_{ik})^{-1/2}) - (k - 2) \\
&\geq (k - 1)P(Y_i > - a(N)2^{-1/2}) - (k - 2) .
\end{aligned}$$

Setting the right-hand side equal to P^* , one may easily solve for N using tables of the standard univariate normal distribution.

It is also possible to use the results we have for $k = 3$, possibly in conjunction with results for $k = 2$, to obtain a Bonferroni approximation. For example, suppose that $k = 5$. Then,

$$\begin{aligned}\overline{PCS} &\geq P(Y_1 > -a(N)(1-\rho_{15})^{-1/2}, Y_2 > -a(N)(1-\rho_{25})^{-1/2}) \\ &\quad + P(Y_3 > -a(N)(1-\rho_{35})^{-1/2}, Y_4 > -a(N)(1-\rho_{45})^{-1/2}) - 1 \\ &\geq 2P(Y_i > -.5(1-\tau)^{1/2}(1+\tau)^{-1/2}, i = 1,2) - 1 .\end{aligned}$$

Setting the right-hand side equal to P^* , with the aid of Table 1.1, one determines N .

1.7. A sequential procedure

Paulson (1964) devised a sequential procedure for the problem of selecting the normal population with the largest population mean, when the variances are known and equal. This procedure is fully sequential and truncated, in the sense that populations are eliminated as sampling proceeds and there is a predetermined upper bound on the total number of stages. In this section we show how Paulson's procedure can be slightly modified to handle the problem of correlated variates, when the variances are known, but not necessarily equal. Since the proof that this procedure guarantees the PCS over the preference region parallels Paulson's proof, we prove only what is strictly necessary and refer the reader to Paulson's paper for the

remaining details. In what follows, we will use, as far as possible, Paulson's notation.

Let $(X_{1s},\ldots,X_{ks})$ $s = 1,2,\ldots$ be a sequence of independent vectors each with a multivariate normal distribution with unknown population means $(\mu_1,\ldots,\mu_k)$, known population variances $(\sigma_1^2,\ldots,\sigma_k^2)$, and unknown population correlations $\rho_{ij} \equiv \mathrm{corr}(X_{is},X_{js})$. Our objective is to select, with probability at least P^*, the component with the largest mean, whenever $\mu_{[k]} - \mu_{[k-1]} \geq \delta^* > 0$. Let $0 < \lambda < \delta^*$ be an arbitrary fixed number, and set $\bar{\sigma}^2 = \max_{i \neq j} (\sigma_i + \sigma_j)^2$. Next define,

$$a_\lambda = [\bar{\sigma}^2/2(\delta^* - \lambda)] \log ((k - 1)/(1 - P^*)) ,$$

and W_λ = the largest integer less than a_λ/λ. (Note: Our definition of a_λ is different from Paulson's.) Then Paulson describes his

<u>Rule P_λ</u>: "At the first stage of the experiment we take one observation from each variate, obtaining ... $(X_{11},X_{21},\ldots,X_{k1})$. Then we eliminate from further consideration any variate j for which

$$X_{j1} < \max \{X_{11},X_{21},\ldots,X_{k1}\} - a_\lambda + \lambda .$$

If all but one variate are eliminated after the first stage of the experiment, we stop the experiment and select the remaining variate as the best one. Otherwise we go on to the second stage of the experiment and take one observation on each variate not eliminated

after the first stage. Proceeding by induction, at the rth stage of the experiment $(r = 2,3,\ldots,W_\lambda)$ we take one observation on each variate not eliminated after the $(r - 1)$ stage, and then eliminate any remaining variate j for which

$$\sum_{s=1}^{r} X_{js} < \max_\nu \left\{ \sum_{s=1}^{r} X_{\nu s} \right\} - a_\lambda + r\lambda ,$$

where the max is taken over all variates left after the $(r - 1)$ stage. If only one variate is left after the rth stage, the experiment is terminated and the remaining variate is selected, otherwise we go on to the $(r + 1)$ stage. If more than one variate remains after the W_λ stage, the experiment is terminated at the $(W_\lambda + 1)$ stage by selecting the remaining variate for which the sum of the $(W_\lambda + 1)$ observations is a maximum."

<u>Lemma 1.8.</u> For each $0 < \lambda < \delta^*$, Rule P_λ guarantees the probability requirement

$$\inf_\Omega PCS_R(\mu, R) \geq P^*$$

where

$$\Omega = \{(\mu,R) \mid \mu_{[k]} - \mu_{[k-1]} \geq \delta^* , \ R \text{ is a correlation matrix.}\}$$

<u>Proof.</u> It follows from the lines at the bottom of p. 176 of Paulson's paper that in Ω,

$$P(\text{incorrect selection}) \leq \sum_{\nu=1}^{k-1} P\left(\sum_{s=1}^{n} X_{ks} < \sum_{s=1}^{n} X_{\nu s} - a_\lambda + n\lambda \text{ for some } n < \infty \right)$$

and,

$$P\left(\sum_{s=1}^{n} (X_{\nu s}-X_{ks}+\lambda) > a_\lambda \quad \text{for some} \quad n < \infty \right)$$

$$\leq \exp \frac{2(\mu_\nu-\mu_k+\lambda)a_\lambda}{\sigma_\nu^2+\sigma_k^2-2\sigma_\nu\sigma_k\rho_{\nu k}} \leq \exp \frac{-2(\delta^*-\lambda)a_\lambda}{\sigma_\nu^2+\sigma_k^2-2\sigma_\nu\sigma_k\rho_{\nu k}}$$

$$\leq \exp \frac{-2(\delta^*-\lambda)a_\lambda}{(\sigma_\nu+\sigma_k)^2} \leq \exp \frac{-2(\delta^*-\lambda)a_\lambda}{\bar{\sigma}^2} = \frac{1-P^*}{1-k}$$

Therefore,

$$P(\text{incorrect solution}) \leq 1 - P^* \quad \text{and} \quad PCS \geq P^* .$$

In the first inequality above we have used the fact that the equation

$$0 = Ee^{t(X_{\nu s}-X_{ks}+\lambda)} = \exp\left\{t(\mu_\nu-\mu_k+\lambda) + \frac{t^2}{2}(\sigma_\nu^2+\sigma_k^2-2\sigma_\nu\sigma_k\rho_{\nu k})\right\}$$

has the unique nonzero root

$$t_0 = -2(\mu_\nu-\mu_k+\lambda)/(\sigma_\nu^2+\sigma_k^2-2\sigma_\nu\sigma_k\rho_{\nu k}) .$$

QED

CHAPTER 2

SELECTION OF THE VARIATE WITH THE SMALLEST POPULATION VARIANCE FROM A SINGLE MULTIVARIATE NORMAL POPULATION

2.0. Introduction

The problem studied in the present chapter was motivated by the problem posed in Section 4.3 of Chapter 4. The asymptotic solution provided by Theorem 2.3 will be crucial to the developments of Chapters 3 and 4.

In this chapter we study single-stage procedures for selecting the variate with the smallest population variance from a single k-variate normal distribution. We formulate the general problem in Section 2.1. In Section 2.2 we obtain exact small-sample results for $k = 2$. However, when $k > 2$, it does not seem possible to extend the analysis for $k = 2$, as we point out in Section 2.3. In Section 2.4 we show how a conservative approximation to the single-stage sample size can be obtained. In Section 2.5 we develop a large-sample solution for the general case $k \geq 3$. For $k \geq 3$, and arbitrary correlation matrix, it turns out (perhaps surprisingly) that the least favorable configuration of the correlation matrix depends on N, the single-stage sample size, in a very complicated way. This is reminiscent of the results of Chapter 1. The large-sample results of the present chapter are special cases of the results of Section 3.1 of Chapter 3. These large-sample results, although stated in a normal framework, are valid for large classes of multivariate distributions, for which Lemma 2.8 is also true.

2.1. Formulation of the problem

We consider a k-variate normal population with population means $(\mu_1,\ldots,\mu_k)$, population variances $(\sigma_1^2,\ldots,\sigma_k^2)$ and population correlations ρ_{ij} $(1 \leq i < j \leq k)$. We denote the covariance matrix by $\Sigma = \{\sigma_{ij}\} \equiv \tilde{\sigma}\, R\, \tilde{\sigma}$, where $\tilde{\sigma} = \mathrm{diag}(\sigma_1,\ldots,\sigma_k)$ and $R = \{\rho_{ij}\}$ are $k \times k$ matrices. Therefore, $\sigma_{ii} \equiv \sigma_i^2$ are the variances. Let the ranked values of the σ_i^2 be $\sigma_{[1]}^2 \leq \ldots \leq \sigma_{[k]}^2$. The experimenter does not have any prior knowledge concerning the values of the parameters of this multivariate normal population, or of the pairing of the $\sigma_{[i]}^2$ with the variates.

Indifference-zone formulation

The experimenter's goal is to select the variate associated with $\sigma_{[1]}^2$, the smallest population variance. Two constants $\{\theta^*,P^*\}$, $\theta^* > 1$, $1/k < P^* < 1$, are specified prior to experimentation. We denote the probability of a correct selection when decision procedure $\mathcal{R}$ is used by $PCS_{\mathcal{R}}(\tilde{\sigma},R)$, and restrict consideration to decision procedures which guarantee the probability requirement:

$$\inf_{\Omega} PCS_{\mathcal{R}}(\tilde{\sigma},R) \geq P^* \tag{2.1}$$

where

$$\Omega = \{(\tilde{\sigma},R) \,|\, \sigma_{[2]}^2 \geq \theta^*\sigma_{[1]}^2 \text{ , } R \text{ a correlation matrix}\} .$$

Bechhofer and Sobel (1954) proposed the following decision

procedure, when considering this problem for the case $R = I_k$. A sample of N independent vector observations, $(X_{1\alpha},\ldots,X_{k\alpha})$ $(1 \leq \alpha \leq N)$, is taken and one computes,

$$a_{ii} = \sum_{\alpha=1}^{N} (X_{i\alpha} - \overline{X}_i)^2 \text{ , where } \overline{X}_i = \sum_{\alpha=1}^{n} X_{i\alpha}/N \quad (1 \leq i \leq k) \text{ .}$$

Rule BS: Assert that the component associated with $a_{[11]} \equiv \min\{a_{11},\ldots,a_{kk}\}$ has population variance $\sigma^2_{[1]}$.

Our task is to determine the smallest sample size N necessary to guarantee the probability requirement (2.1) when Rule BS is used and R is an unknown correlation matrix.

Subset formulation

In certain situations, the experimenter may be interested in the selection of a subset of variates, which includes the variate with the smallest variance. A constant $\{P^*\}$, $1/k < P^* < 1$, is specified prior to experimentation. Letting $PCS_R(\tilde{\sigma},R)$ be defined as above, we restrict consideration to decision procedures which guarantee the probability requirement:

$$\inf_{\tilde{\sigma},R} PCS_R(\tilde{\sigma},R) \geq P^* \text{ .} \tag{2.2}$$

The following decision procedure, proposed by Gupta and Sobel (1954), when considering this problem for the case $R = I_k$, will be used:

Rule GS: Include the variate associated with a_{ii} in the selected subset if $a_{ii} \leq d^* a_{[11]}$, where $d^* > 1$ is a specified constant.

Our objective is to find the smallest sample size N which will guarantee the probability requirement (2.2) when Rule GS is employed and R is an unknown correlation matrix.

Throughout this chapter, we assume, without loss of generality, that $\sigma_1^2 < \sigma_j^2$ $(j \neq 1)$. No consideration will be given to the population means, since their configuration is irrelevant for our purposes.

2.2. Case $k = 2$

In this section we consider the case $k = 2$, i.e., the parent population is bivariate normal. Writing $\rho_{12} = \rho$ and $\Sigma^{-1} = (\sigma^{ij})$, we have

Lemma 2.1. The joint p.d.f. of a_{11} and a_{22} is

$$(2.3) \quad p_{a_{11},a_{22}}(y_1,y_2) = \sum_{j=0}^{\infty} c_j(\rho) \prod_{i=1}^{2} \frac{(\sigma^{ii})^{j+\frac{n}{2}} y_i^{j+\frac{n}{2}-1} \exp(-\sigma^{ii} y_i/2)}{2^{\frac{n}{2}+j} \Gamma(\frac{n}{2}+j)},$$

$$y_i > 0,$$

where

$$c_j(\rho) = (1-\rho^2)^{n/2} \rho^{2j} \frac{\Gamma(\frac{n}{2}+j)}{\Gamma(\frac{n}{2})\, j!} \geq 0, \quad \sum_{j=0}^{\infty} c_j(\rho) = 1.$$

Proof. Let $A = (a_{ij})$, $a_{ij} = \sum_{\alpha=1}^{N} (X_{i\alpha} - \bar{X}_i)(X_{j\alpha} - \bar{X}_j)$. Then A has a Wishart density. Make the transformation of variables $a_{11} = a_{11}$,

$a_{22} = a_{22}$, $r_{12} = a_{12}a_{11}^{-1/2}a_{22}^{-1/2}$, and then obtain (2.3) as the marginal p.d.f. of (a_{11},a_{22}) . Note that the joint p.d.f. of (a_{11},a_{22}) is a weighted sum of products of gamma densities. QED

Lemma 2.2. Let

$$v = \frac{a_{22}\sigma^{22}}{a_{11}\sigma^{11}} = \frac{a_{22}\sigma_{11}}{a_{11}\sigma_{22}} .$$

Then the p.d.f. of v is

$$(2.4) \qquad p_v(z) = \sum_{j=0}^{\infty} c_j(\rho) \frac{\Gamma(2j+n)}{(\Gamma(j+n/2))^2} z^{j+\frac{n}{2}-1}(1+z)^{-(n+2j)}$$

$$\equiv \sum_{j=0}^{\infty} c_j(\rho) f_j(z) \quad , \quad z > 0 .$$

Proof. In (2.3) make the transformation

$$a_{11} = a_{11} , \quad v = \frac{a_{22}\sigma_{11}}{a_{11}\sigma_{22}} ,$$

then integrate out y_1 , obtaining (2.4) as a final result. The p.d.f. of v is a weighted sum of central F densities. QED

Lemma 2.3. Define

$$b_j = \int_{1/\theta^*}^{\infty} f_j(z)dz .$$

Then $b_o \leq b_1 \leq b_2 \leq \cdots$.

Proof. For $j \geq 1$,

$$(2.5)\quad b_j - b_{j-1} = \frac{\Gamma(n+2j)}{(\Gamma(j+\frac{n}{2}))^2}\int_{1/\theta^*}^{\infty} z^{\frac{n}{2}+j-1}(1+z)^{-(n+2j)}dz$$

$$- \frac{\Gamma(n+2j-2)}{(\Gamma(\frac{n}{2}+j-1))^2}\int_{1/\theta^*}^{\infty} z^{\frac{n}{2}+j-2}(1+z)^{-(n+2j-2)}dz$$

It is easy to show that, if we integrate by parts the first integral in (2.5), and then twice integrate by parts the second integral in (2.5), we obtain,

$$b_j - b_{j-1} = \frac{\Gamma(n+2j-2)}{\Gamma(\frac{n}{2}+j-1)\Gamma(\frac{n}{2}+j)}(1/\theta^*)^{j+\frac{n}{2}-1}(1+1/\theta^*)^{-n-2j+1}(1-1/\theta^*) \geq 0 .$$

QED

If the experimenter uses Rule BS, we obtain,

<u>Theorem 2.1.</u> The least favorable configuration of the relevant parameters is $\sigma^2_{[2]} = \theta^*\sigma^2_{[1]}$, $\rho = 0$, yielding,

$$(2.6)\qquad \inf_{\Omega} PCS(\tilde{\sigma},\rho) = \int_{1/\theta^*}^{\infty} \frac{\Gamma(n)}{(\Gamma(\frac{n}{2})^2} z^{\frac{n}{2}-1}(1+z)^{-n}dz .$$

<u>Proof.</u> If $\rho = \pm 1$, we have $PCS(\tilde{\sigma},\rho) = 1$. Indeed, in this case,

$$X_{2\alpha} - \mu_2 = b(X_{1\alpha} - \mu_1) \quad \text{a.e. } (1 \leq \alpha \leq N) ,$$

where $b = \rho\sigma_2/\sigma_1$. Hence,

$$a_{22} = \sum_{\alpha=1}^{N} (X_{2\alpha} - \overline{X}_2)^2 = b^2 a_{11} = \frac{\sigma_2^2}{\sigma_1^2} a_{11} \quad \text{a.e.}$$

resulting in

$$PCS(\tilde{\sigma},\rho) = P(a_{11} \leq a_{22}) = P(\sigma_1^2 \leq \sigma_2^2) = 1 .$$

For other values of ρ , and $\sigma_2^2 \geq \theta^*\sigma_1^2$,

$$PCS(\tilde{\sigma},\rho) = P(a_{11} \leq a_{22}) = P(v \geq \sigma_1^2/\sigma_2^2) \geq P(v \geq 1/\theta^*)$$

$$= \sum_{j=0}^{\infty} c_j(\rho)b_j .$$

Since $b_0 = \inf_{j\geq 0} b_j$, and $c_0(0) = 1$, it follows that

$$\inf_{\Omega} PCS(\tilde{\sigma},\rho) = b_0 . \qquad \text{QED}$$

Bechhofer and Sobel (1954) provide a table of values of the integral on the right-hand side of (2.6). For θ^* and P^* specified, the experimenter uses the table to determine $N = n + 1$.

<u>Lemma 2.4.</u> Consider a loss function $L_i(\tilde{\sigma},\rho)$ = loss when component i is selected and $(\tilde{\sigma},\rho)$ are the parameters, such that,

(i) $L_i(\tilde{\sigma},\rho) \leq L_j(\tilde{\sigma},\rho)$ when $\sigma_j^2 > \sigma_i^2$;

(ii) $0 \leq L_i(\tilde{\sigma},\rho) = L_{\pi i}(\pi\tilde{\sigma},\rho)$, where $\pi(\sigma_1^2,\sigma_2^2) = (\pi\sigma_1^2,\pi\sigma_2^2)$ is any permutation of (σ_1^2,σ_2^2) .

Then Rule BS is minimax and admissible, uniformly minimizing the risk function among all invariant (under permutations of components) procedures.

<u>Proof.</u> Since $c_j(\rho) \geq 0$ for all ρ , and since the gamma densities appearing in (2.3) have monotone likelihood ratio, invoking a result of Eaton (1967a) (a generalization of a theorem of Bahadur and Goodman (1952)), the conclusion follows at once. QED

If the experimenter uses Rule GS, we have,

Theorem 2.2. The least favorable configuration of the relevant parameters is

$$\sigma^2_{[1]} = \sigma^2_{[2]} \ , \quad \rho = 0 \ , \text{ yielding,}$$

$$\text{(2.7)} \qquad \inf \mathrm{PCS}(\tilde{\sigma},\rho) = \int_{1/d^*}^{\infty} \frac{\Gamma(n)}{(\Gamma(\frac{n}{2}))^2} z^{\frac{n}{2}-1} (1+z)^{-n} dz \ .$$

Proof. The proof parallels that of Theorem 2.1.

Lemma 2.5. If S denotes the size of the selected subset when Rule GS is employed, then

(a) $E(S|\tilde{\sigma},\rho) = P(v_1 \geq \frac{\sigma_{11}}{d^*\sigma_{22}}) + P(v_2 \geq \frac{\sigma_{22}}{d^*\sigma_{11}})$

where v_1 and v_2 are both distributed as in (2.4).

(b) $\sup_{\tilde{\sigma},\rho} E(S|\tilde{\sigma},\rho) = 2$, when $\sigma_1^2 = \sigma_2^2$, $\rho = 1$.

Proof. The result follows easily from previous developments.

2.3. Case $k \geq 3$.

In this section we develop some preliminary results for the case $k \geq 3$, and outline some of the difficulties encountered. We have not been able to obtain definitive general small-sample results when $k \geq 3$. Unfortunately, the method employed for $k = 2$ in Section 2.2 fails here. In particular, it is easy to develop similar results to those given as Lemmas 2.1 and 2.2, but there is very strong evidence that the least favorable configuration of R depends on N

and θ^*(or d^*). This will be seen in Section 2.5, where we develop complete asymptotic $(N \to \infty)$ results.

If Rule BS is used, without loss of generality, we assume $\sigma_1^2 = 1$, $\sigma_j^2 = \theta^*$, $j \neq 1$, since this is a least favorable configuration of the variances. Indeed, in Ω, we have,

$$\begin{aligned} PCS &= P(a_{11} < a_{jj}, j \neq 1) = P(\sigma_1^2 \chi_n^2(1) < \sigma_j^2 \chi_n^2(j)) \\ &\geq P(\chi_n^2(1) < \theta^* \chi_n^2(j), j \neq 1) \end{aligned}$$

where $\chi_n^2(j)$ $(1 \leq j \leq k)$ are the diagonal elements of a Wishart matrix with mean nR.

We define

$$R = \begin{pmatrix} 1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \sum_{\alpha=1}^{N} (X_\alpha - \overline{X})(X_\alpha - \overline{X})^t ,$$

where Σ_{22} and A_{22} are $(k-1) \times (k-1)$ symmetric positive definite matrices. Then, the following lemma is stated in a slightly different form in Johnson and Kotz (1972), p. 223. It provides a convenient representation for the distribution function of the diagonal elements of A.

<u>Lemma 2.6.</u> The conditional distribution of a_{11} given A_{22} is noncentral χ_n^2, with noncentrality parameter

$$\lambda = \frac{\Sigma_{12}\Sigma_{22}^{-1}A_{22}\Sigma_{22}^{-1}\Sigma_{21}}{2\theta^*(1-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})} .$$

In other words,

$$p_{a_{11}|A_{22}}(\cdot) = \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} p_{\chi^2_{\frac{n}{2}+j}}(\cdot) .$$

If $p_{A_{22}}(\cdot)$ denotes the density of the Wishart matrix A_{22}, we have

$$(2.8) \quad PCS = \int_{W>0} P(a_{11} < a_{jj}, j\neq 1 | A_{22}) p_{A_{22}}(W) dW$$

$$= \int_{W>0} p_{A_{22}}(W) \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \int_0^{\frac{\min_{j\neq 1} a_{jj}}{1-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}} p_{\chi^2_{\frac{n}{2}+k}}(u) du \, dW ,$$

where $W > 0$ means W symmetric positive definite.

Using (2.8), a tedious but straightforward computation shows that

$$\frac{\partial PCS}{\partial \rho_{ij}} = 0 \quad (i \neq j), \text{ at } R = I_k .$$

One might conjecture, in view of this last result, and the results of the previous section, that $R = I_k$ is a least favorable configuration of R. However, we dot believe this to be the case for $k > 2$. In fact, we shall prove in Section 2.5, using asymptotic $(N \to \infty)$ distribution theory, that $R = I_k$ can be a saddle-point of the PCS. It approaches a global minimum when $c_{\theta^*}(n) = (1/2)n^{1/2}\log \theta^* \to \infty$. When the

experimenter knows that the off-diagonal elements of R are equal, then we show in Section 2.5, using asymptotic theory, that $R = I_k$ is a least favorable configuration, which does not depend on N and θ^*. In other words, we are facing a situation similar to the one encountered in Chapter 1, where the least favorable configuration varies with the sample size.

The same remarks are valid when Rule GS is used.

2.4. A conservative approximation to the sample size

In view of the difficulty of determining a least favorable configuration of R for $k \geq 3$, the following Bonferroni approximation (cf. Section 1.6 of Chapter 1) can be used to determine a value of N, which will be larger than the minimum N required to guarantee the probability requirement.

Lemma 2.7. If Rule BS is used,

$$(2.9) \qquad \inf_{\Omega} PCS(\tilde{\sigma},R) > (k-1) \int_{1/\theta^*}^{\infty} \frac{\Gamma(n)}{(\Gamma(\frac{n}{2}))^2} z^{\frac{n}{2}-1} (1+z)^{-n} dz - (k-2) .$$

Hence Bechhofer and Sobel's (1954) table may be used to determine a conservative value of $N = n + 1$.

If Rule GS is used, a similar approximation is available, replacing θ^* by d in (2.9).

2.5. Large-sample theory

In this section we develop a large-sample theory for the problems considered in Section 2.1. One of the results obtained (Theorem 2.3) will be used in the next two chapters as an important

tool for obtaining large-sample results. We start with a version of the Central Limit Theorem, stated and proved in Anderson (1958), p. 75.

Lemma 2.8. Let X_α , $\alpha = 1,2,\ldots$, be a sequence of independent k-dimensional normal vectors, each with mean vector μ and covariance matrix $\Sigma = (\sigma_{ij})$. Let

$$B(n) = (b_{ij}(n)) = n^{\frac{1}{2}}\{(1/n) \sum_{\alpha=1}^{N} (X_\alpha - \overline{X}_N)(X_\alpha - \overline{X}_N)^t - \Sigma\} ,$$

where

$$\overline{X}_N = \sum_{\alpha=1}^{N} X_\alpha/N , \quad n = N - 1 .$$

Then the asymptotic $(N \to \infty)$ distribution (a.d.) of $B(n)$ if multivariate normal, with zero means, and covariances

$$E\{b_{ij}(n) \cdot b_{k\ell}(n)\} = \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk} .$$

Another tool that will be used extensively, is given below as a lemma, the proof of which may be found, for example, in Rao (1968), Chapter 6.

Lemma 2.9. Let $(Y_{1n},\ldots,Y_{kn})$, $n = 1,2,\ldots$, be a sequence of not necessarily independent vector variates, such that,

$$n^{1/2}(Y_{1n} - \theta_1^0,\ldots,Y_{kn} - \theta_k^0)$$

has multivariate normal asymptotic distribution with zero means and covariance matrix Σ . Let $g_1,\ldots,g_r$ be real functions defined on

E^k , the k dimensional Euclidean space, which are differentiable in a neighborhood of $\theta^0 = (\theta_1^0,\ldots,\theta_k^0)$. Then the a.d. of $n^{1/2}(g_j(Y_{1n},\ldots,Y_{kn}) - g_j(\theta_1^0,\ldots,\theta_k^0))$ $(1 \le j \le r)$, is multivariate normal, with zero means, and covariance matrix,

$$\Sigma(\theta^0) = (\nabla^t g_i \Sigma \nabla g_j) ,$$

where

$$\nabla^t g_j = \left(\frac{\partial g_j}{\partial \theta_1}, \ldots, \frac{\partial g_j}{\partial \theta_k}\right)\Bigg|_{\theta=\theta^0} \quad (1 \le j \le r) ,$$

if $\Sigma(\theta^0)$ is nonsingular.

We shall use the notation introduced in Section 2.1, and assume, without loss of generality, that $\sigma_1^2 < \sigma_j^2$ $(j \ne 1)$.

<u>Lemma 2.10.</u> The a.d. of

$$(2.10) \quad Y_j = n^{1/2}\left\{\frac{\log(a_{11}/a_{jj})-\log(\sigma_1^2/\sigma_j^2)}{2(1-\rho_{1j}^2)^{1/2}}\right\} \quad (j \ne 1) ,$$

is standard multivariate normal, with correlations,

$$\mathrm{corr}(Y_i,Y_j) \equiv \gamma_{ij} = \frac{1-\rho_{1i}^2-\rho_{ij}^2+\rho_{ij}^2}{2(1-\rho_{1i}^2)^{1/2}(1-\rho_{1j}^2)^{1/2}} \quad (i \ne j) .$$

<u>Proof.</u> From Lemma 2.8, the a.d. of

$$n^{1/2}(a_{ii}/n - \sigma_i^2) \quad (1 \le i \le k) ,$$

is multivariate normal with zero means, variances equal to $2\sigma_i^2$, and covariances $2\sigma_{ij}^2$. Therefore, using a variance stabilizing transformation (cf. Bartlett and Kendall (1946) and Lemma 2.9), we have that

$$n^{1/2}(\log(a_{ii}/n) - \log \sigma_i^2) \quad (1 \leq i \leq k) ,$$

has a multivariate normal a.d. with zero means, variances equal to 2, and covariances equal to $2\rho_{ij}^2$. Finally, (2.10) is obtained using Lemma 2.9 once again. QED

The proof of the above lemma is essentially contained in Ramberg (1969). Note that (2.10) resembles the distributions of the previous chapter (cf. Lemma 1.1).

Let PCS_a denote probability of a correct selection when an asymptotic $(N \to \infty)$ distribution function is used. The following is an important result for our purposes.

<u>Theorem 2.3.</u> If the experimenter uses Rule BS, the asymptotic $(N \to \infty)$ least favorable configuration of the relevant parameters is

$$\theta^*\sigma_{[1]}^2 = \sigma_{[2]}^2 = \ldots = \sigma_{[k]}^2 , \quad \rho_{ij} = 0 \quad (i \neq j) .$$

Therefore,

$$(2.11) \quad \inf_{\Omega} PCS_a(\tilde{\sigma},R) = P(Y_j \leq n^{1/2}(1/2)\log \theta^* , \ j \neq 1) ,$$

where the $\{Y_j, j \neq 1\}$ are distributed as in (2.10) with $\gamma_{ij} = 1/2 \quad (i \neq j)$.

<u>Proof.</u> In Ω , if $\sigma_1^2 < \sigma_j^2$ $(j \neq 1)$,

$$PCS(\tilde{\sigma},R) = P(a_{11} < a_{jj}, j \neq 1) = P(Y_j < \frac{n^{1/2}\log(\sigma_j^2/\sigma_1^2)}{2(1-\rho_{1j}^2)^{1/2}} , j\neq 1)$$

$$\geq P(Y_j < n^{1/2}(1/2)(\log \theta^*)(1 - \rho_{1j}^2)^{-1/2} , j \neq 1)$$

$$= \Phi_{(\gamma_{ij})}(c_{\theta^*}(n)(1 - \rho_{12}^2)^{-1/2},\ldots,c_{\theta^*}(n)(1 - \rho_{1k}^2)^{-1/2})$$

where,

$$c_{\theta^*}(n) = n^{1/2}(1/2)\log \theta^* ,$$

and

$$\Phi_{(\gamma_{ij})} \equiv \Phi_{(\gamma_{ij})}(c_2(n),\ldots,c_k(n))$$

$$= \int_{-\infty}^{c_2(n)} \cdots \int_{-\infty}^{c_k(n)} f_{(\gamma_{ij})}(y_2,\ldots,y_k)dy_2\ldots dy_k ,$$

$$c_j(n) = c_{\theta^*}(n)(1 - \rho_{1j}^2)^{-1/2} \quad (j \neq 1)$$

and $f_{(\gamma_{ij})}(y_2,\ldots,y_k)$ is the p.d.f. of the $\{Y_j , j \neq 1\}$. Since,

$$\frac{\partial\Phi_{(\gamma_{ij})}}{\partial\rho_{k\ell}} = 2\rho_{k\ell}\frac{\partial\Phi_{(\gamma_{ij})}}{\partial\rho_{k\ell}^2} = 0$$

if all $\rho_{k\ell} = 0$, it follows that the correlation matrix $R = I_k$ is a stationary point of $\Phi_{(\gamma_{ij})}$. We must show that it is a point of global minimum as $c_{\theta^*}(n) \to \infty$.

First assume $\rho_{1j}^2 \neq 1 \quad (j \neq 1)$. We will prove that

$$\frac{\partial \Phi_{(\gamma_{ij})}}{\partial \rho_{\ell m}^2} > 0 \quad (\ell, m > 1) .$$

Without loss of generality, consider $\ell = 2$, $m = 3$.

$$(2.12) \quad \frac{\partial \Phi_{(\gamma_{ij})}}{\partial \rho_{23}^2} = \frac{\partial \gamma_{23}}{\partial \rho_{23}^2} \int_{-\infty}^{c_4(n)} \cdots \int_{-\infty}^{c_k(n)} f_{(\gamma_{ij})}(c_2(n), c_3(n), y_4, \ldots, y_k) dy_4 \ldots dy_k$$

$$> 0$$

since

$$\frac{\partial \gamma_{23}}{\partial \rho_{23}^2} = (1/2)(1 - \rho_{12}^2)^{-1/2}(1 - \rho_{13}^2)^{-1/2} > 0 .$$

Next, we show that, as $c_{\theta *}(n) \to \infty$,

$$(2.13) \qquad \frac{\partial \Phi_{(\gamma_{ij})}}{\partial \rho_{1p}^2} > 0 \quad (p \neq 1) .$$

Without loss of generality, take $p = 2$. Since ρ_{12}^2 appears in the expressions of $\gamma_{23}, \ldots, \gamma_{2k}$, we have

$$(2.14)\quad \frac{\partial\Phi(\gamma_{ij})}{\partial\rho_{12}^2} = \sum_{j=3}^{k} \left.\frac{\partial\Phi(\gamma_{ij})}{\partial\gamma_{2j}}\right|_{\rho_{12}^2 \text{ fixed}} \cdot \frac{\partial\gamma_{2j}}{\partial\rho_{12}^2} + \left.\frac{\partial\Phi(\gamma_{ij})}{\partial\rho_{12}^2}\right|_{\gamma_{ij} \text{ fixed}}$$

$$= \sum_{j=3}^{k} \frac{\partial\gamma_{2j}}{\partial\rho_{12}^2} \int_{-\infty}^{c_3(n)} \cdots \int_{-\infty}^{c_{j-1}(n)} \int_{-\infty}^{c_{j+1}(n)} \cdots \int_{-\infty}^{c_k(n)} f_{(\gamma_{ij})}(c_2(n), y_3, \ldots, y_{j-1}, c_j(n),$$

$$y_{j+1}, \ldots, y_k)\, dy_3 \ldots dy_k$$

$$+ \frac{\partial c_2(n)}{\partial\rho_{12}^2} \int_{-\infty}^{c_3(n)} \cdots \int_{-\infty}^{c_k(n)} f_{(\gamma_{ij})}(c_2(n), y_3, \ldots, y_k)\, dy_3 \ldots dy_k \ .$$

$$\equiv \sum_{j=3}^{k} \frac{\partial\gamma_{2j}}{\partial\rho_{12}^2} M_{2j} + \frac{\partial c_2(n)}{\partial\rho_{12}^2} Q_2 \ .$$

where

$$(2.15)\quad \begin{aligned} \frac{\partial c_2(n)}{\partial\rho_{12}^2} &= c_{\theta*}(n)(1/2)(1-\rho_{12}^2)^{-3/2} \to \infty \quad \text{as} \quad c_{\theta*}(n) \to \infty \ ; \\ \frac{\partial\gamma_{2j}}{\partial\rho_{12}^2} &= \frac{\rho_{12}^2+\rho_{2j}^2-\rho_{1j}^2-1}{4(1-\rho_{1j}^2)^{1/2}(1-\rho_{12}^2)^{3/2}} \equiv \frac{\lambda_{12j}}{4(1-\rho_{1j}^2)^{1/2}(1-\rho_{12}^2)^{3/2}} \end{aligned}$$

Since $M_{2j} > 0$ $(j \geq 3)$ and $Q_2 > 0$, we only have to consider situations where $\frac{\partial\gamma_{2j}}{\partial\rho_{12}^2} < 0$ for some j. Suppose, for instance, that $\frac{\partial\gamma_{23}}{\partial\rho_{12}^2} < 0$. This is equivalent to $\lambda_{123} < 0$. We will show that $Q_2 - M_{23} > 0$, which proves (2.13), in view of (2.15).

A straightforward computation leads to

$$Q_2 - M_{23} = f(c_2(n)) \int_{-\infty}^{c_4(n)-\gamma_{24}c_2(n)} \cdots \int_{-\infty}^{c_k(n)-\gamma_{2k}c_2(n)}$$

$$\left\{ \int_{-\infty}^{c_3(n)-\gamma_{23}c_2(n)} \tilde{f}(z_3,\ldots,z_k)dz_3 - \tilde{f}(c_3(n)-\gamma_{23}c_2(n), z_4,\ldots,z_k) \right\}$$

$$dz_4 \ldots dz_k ,$$

where

$$f(c_2(n)) = \frac{1}{(2\pi)^{1/2}} \exp\left\{ - \frac{c_2^2(n)}{2} \right\}$$

and $\tilde{f}$ is the density of a multivariate normal distribution with zero means and covariance matrix which depends only on the γ_{ij} $(i \neq j)$.

Since $\lambda_{123} < 0$, we have

$$c_3(n) - \gamma_{23}c_2(n) = \frac{c_{\theta*}(n)(-\lambda_{123})}{2(1-\rho_{13}^2)^{1/2}(1-\rho_{12}^2)} \to \infty \quad \text{as} \quad c_{\theta*}(n) \to \infty ,$$

showing that $Q_2 - M_{23} > 0$. This, in turn, implies (2.13).

Equations (2.12) and (2.13) imply that

$$\Phi_{(\gamma_{ij})}(c_2(n),\ldots,c_k(n)) \geq \Phi_{(1/2)}(c_{\theta^*}(n),\ldots,c_{\theta^*}(n)) \ .$$

This last lower bound is achieved when $\rho_{ij} = 0 \quad (i \neq j)$, proving (2.11) when $\rho_{1j}^2 \neq 1 \quad (j \neq 1)$.

Let $J = \{j_1,\ldots,j_\ell\}$, $1 \notin J$. Assume that $\rho_{1j}^2 = 1$, $j \in J$. Using an argument similar to the one employed in the proof of Theorem 2.1, we have,

$$a_{11} = (\sigma_1^2/\sigma_j^2)a_{jj} \quad \text{a.e.} \quad (j \in J) \ .$$

Therefore, for such an R , if $\sigma_1^2 < \sigma_j^2 \quad (j \neq 1)$,

$$\begin{aligned} PCS &= P(\bigcap_{j>1}(a_{11}<a_{jj})) = P(\bigcap_{j\in J}(\sigma_1^2<\sigma_j^2) \ , \bigcap_{j\notin J}(a_{11}<a_{jj})) \\ &= P(\bigcap_{j\notin J}(a_{11}<a_{jj})) \end{aligned}$$

Since, for any R ,

$$P(\bigcap_{j\notin J}(a_{11} < a_{jj})) \geq P(\bigcap_{j>1}(a_{11} < a_{jj}))$$

it is seen that $\rho_{1j}^2 = 1$, $j \in J$ does not lead to the infimum. QED

Theorem 2.4. If the experimenter uses Rule GS, the asymptotic $(N \to \infty)$ least favorable configuration of the relevant parameters is

$$\sigma_1^2 = \ldots = \sigma_k^2 , \quad \rho_{ij} = 0 \quad (i \neq j) .$$

Therefore,

$$(2.16) \quad \inf PCS_a(\tilde{\sigma},R) = P(Y_j \leq n^{1/2}(1/2)\log d^* , \ j \neq 1) ,$$

where the $\{Y_j , j \neq 1\}$ are as in Theorem 2.3.

Proof. The proof is similar to the proof of Theorem 2.3.

Lemma 2.11. If S denotes the size of the selected subset when Rule GS is employed, we have asymptotically $(N \to \infty)$,

$$(a) \quad E_a(S|\tilde{\sigma},R) = \sum_{i=1}^{k} P(Y_j^i < n^{1/2}(1/2)(1-\rho_{ij}^2)^{-1/2}\log(d^*\sigma_j^2/\sigma_i^2) , \ i \neq j)$$

where the $\{Y_j^i , j \neq i\}$ are distributed as in (2.10) with i in place of 1 .

$$(b) \qquad \sup E_a(S|\tilde{\sigma},R) = k \quad \text{when} \quad \sigma_i^2 = \sigma_j^2 \ (i \neq j) , \ R = \begin{pmatrix} 1 & \ldots & 1 \\ & \ldots & \\ 1 & \ldots & 1 \end{pmatrix} .$$

Proof. Consequence of previous developments.

Theorem 2.5. Suppose that $\rho_{ij} = \rho$ $(i \neq j)$, where ρ is unknown. Then,

(a) If Rule BS is used, an asymptotic $(N \to \infty)$ least favorable

configuration of the relevant parameters is

$$\theta^*\sigma^2_{[1]} = \sigma^2_{[2]} = \ldots = \sigma^2_{[k]} , \quad \rho = 0 ,$$

(2.11) being pertinent;

(b) If Rule GS is used, an asymptotic $(N \to \infty)$ least favorable configuration of the relevant parameters is

$$\sigma^2_1 = \ldots = \sigma^2_k , \quad \rho = 0 ,$$

(2.16) being true.

Proof. This follows directly from Lemma 2.10, without the need for further arguments.

There is evidence that the approximation used in Lemma 2.10 is very good, even for small values of N. The reader may consult Bechhofer and Sobel's (1954) tables where some comparisons are given. Hence, we would conjecture that Theorem 2.5 provides an excellent approximation to N, even for relatively small values of N. As for Theorems 2.3 and 2.4, we have used the fact that N is large in a stronger manner, but still it is expected that moderate values of N would provide a very good approximation to the small sample results. We would expect that the approximation will be an excellent one if P^* and θ^* are close to unity. Values of N may be determined using formulae (2.11) and (2.16) in conjunction with the tables of Gupta (1963) or Milton (1963).

We next explore the behavior of $\Phi_{(\gamma_{ij})}$ in the vicinity of $R = I_k$. It is easy to compute, from (2.12) and (2.14), that at $R = I_k$,

$$\left.\frac{\partial\Phi_{(\gamma_{ij})}}{\partial\rho^2_{\ell m}}\right|_{R=I_k} > 0 \qquad (\ell,m > 1)$$

$$\left.\frac{\partial\Phi_{(\gamma_{ij})}}{\partial\rho^2_{1j}}\right|_{R=I_k} = -\frac{(k-2)}{4}\int_{-\infty}^{c_{\theta*}(n)}\cdots\int_{-\infty}^{c_{\theta*}(n)} f_{(1/2)}(c_{\theta*}(n),c_{\theta*}(n),y_4,\ldots,y_k)\,dy_4\ldots dy_k$$

$$+\frac{c_{\theta*}(n)}{2}\int_{-\infty}^{c_{\theta*}(n)}\cdots\int_{-\infty}^{c_{\theta*}(n)} f_{(1/2)}(c_{\theta*}(n),y_3,\ldots,y_k)\,dy_3\ldots dy_k$$

$$(j > 1)\ .$$

Therefore,

$$\left.\frac{\partial\Phi_{(\gamma_{ij})}}{\partial\rho^2_{1j}}\right|_{R=I_k} < 0 \quad \text{if} \quad c_{\theta*}(n) \cong 0\ ,$$

and

$$> 0 \quad \text{if} \quad c_{\theta*}(n) \to \infty\ .$$

In other words, if $c_{\theta*}(n)$ is small enough, $\Phi_{(\gamma_{ij})}$ has a saddle-point at $R = I_k$, while as $c_{\theta*}(n)$ increases it will have a local minimum there, and eventually a global minimum.

Finally, we show that PCS_a can be less than $1/k$, as was also the case in Chapter 1. Note that $1/k$ is the lowest possible value for the PCS when $R = I_k$. Take $k = 2$, $\rho^2_{12} = \rho^2_{13} = 1/2$, $\rho_{23} = 0$. Then, $\gamma_{23} = 0$ and

$$PCS_a = \int_{-\infty}^{c_2(n)}\int_{-\infty}^{c_3(n)} f_{(0)}(y_2,y_3)\,dy_2dy_3 \cong 1/4 < 1/3$$

if $c_{\theta*}(n) \cong 0$.

CHAPTER 3

SELECTION OF A SUBCLASS OF VARIATES WITH THE SMALLEST POPULATION GENERALIZED VARIANCE FROM A SINGLE MULTIVARIATE NORMAL POPULATION (ASYMPTOTIC THEORY)

3.0. Introduction

In this chapter we study selection procedures in terms of population generalized variances associated with subclasses of variates from a single multivariate normal population. In Section 3.1 we consider disjoint subclasses and the results obtained are extensions of the results of Section 2.5 of Chapter 2. In Section 3.2 we consider intersecting subclasses. Many other selection problems in terms of generalized variances may be treated using the ideas of the present chapter. We decided to restrict consideration to these two particular problems, since they illustrate well the methods we propose. For instance, it is easy to extend these results to selection problems involving subclasses of different sizes. Throughout the entire chapter, the theory developed is asymptotic (large-sample), and could be stated in a more general framework than normality.

3.1. Selecting the smallest population generalized variance (disjoint subclasses).

Consider a kp-variate normal population, with unknown population mean vector and unknown population covariance matrix,

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_2 & \cdots & \Sigma_{2k} \\ \cdots & & & \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & \Sigma_k \end{pmatrix}$$

where Σ_j $(1 \leq j \leq k)$ are $p \times p$ symmetric positive definite matrices. The quantity $\det \Sigma_i$ is referred to as the population generalized variance associated with the ith subclass of variates $(1 \leq i \leq k)$. Let $\det \Sigma_{[1]} \leq \ldots \leq \det \Sigma_{[k]}$ be the ranked values of the $\det \Sigma_i$ $(1 \leq i \leq k)$. It is assumed that no prior knowledge exists concerning the values of $\det \Sigma_j$ $(1 \leq j \leq k)$, or of the pairing of $\det \Sigma_{[j]}$ with the subclasses of variates.

Indifference-zone formulation

The experimenter's goal is to select the subclass of variates associated with $\det \Sigma_{[1]}$. He specifies $\{\theta^*, P^*\}$, $\theta^* > 1$, $1/k < P^* < 1$, prior to the start of experimentation. If $PCS_R(\Sigma)$ denotes the probability of a correct selection when decision procedure R is employed, we restrict consideration to procedures R which satisfy the probability requirement:

$$\inf_{\Omega} PCS_R(\Sigma) \geq P^* ,$$

where

$$\Omega = \{\Sigma | \det \Sigma_j \geq \theta^* \det \Sigma_{[1]} , j \neq [1]\} .$$

In this chapter we propose "natural" single-stage selection

procedures, which associate sample quantities with the corresponding population parameters.

A sample of N independent kp-vector observations, $X_\alpha^t = (X_{1\alpha}, \ldots, X_{kp,\alpha})$ $(1 \leq \alpha \leq N)$, is taken, and the sufficient statistics $(\bar{X}_N, S)$, $\bar{X}_N = \sum_{\alpha=1}^{N} X_\alpha / N$, $S = \sum_{\alpha=1}^{N} (X_\alpha - \bar{X}_N)(X_\alpha - \bar{X}_N)^t / n$, $n = N - 1$, are obtained. Let S be partitioned according to Σ, in such a way that S_j corresponds to Σ_j $(1 \leq j \leq k)$, and S_{ij} to Σ_{ij} $(i \neq j)$. For this <u>indifference-zone</u> goal, we adopt the following decision rule:

<u>Rule R_{GV1}</u>: Assert that the subclass associated with $\det S_{[1]} \equiv \min_j \det S_j$, has the smallest population generalized variance, $\det \Sigma_{[1]}$.

Our objective is to determine the smallest sample size N such that R_{GV1} will guarantee the probability requirement.

When $\Sigma_{ij} = \Sigma^*$ $(i \neq j)$, R_{GV1} is minimax, and also has uniformly smallest risk for a class of natural (invariant) decision procedures and loss functions (cf. Eaton (1967b)).

<u>Subset formulation</u>

If the experimenter wishes to select a subset of subclasses containing the subclass associated with $\det \Sigma_{[1]}$, he specified $\{P^*\}$, $1/k < P^* < 1$, prior to the start of experimentation. If $PCS_R(\Sigma)$ has the same meaning as above, we restrict consideration

to decision procedures R which satisfy the probability requirement:

$$\inf_{\Sigma} PCS_R(\Sigma) \geq P^* .$$

We propose the following decision procedure for this <u>subset</u> goal:

<u>Rule R_{GV2}</u>: Include the subclass of variates associated with S_j in the selected subset if $\det S_j \leq d^* \det S_{[1]}$, where $d^* > 1$ is a specified constant.

Our objective is to find the smallest sample size N which will guarantee the probability requirement when R_{GV2} is employed.

Gnanadesikan and Gupta (1970) studied R_{GV2} for the case where $\Sigma_{ij} = 0 \quad (i \neq j)$.

We disregard the population means in what follows, since they are irrelevant in our problems. We assume, without loss of generality, that $\det \Sigma_1 < \det \Sigma_j \quad (j \neq 1)$.

The following linearization result, proved in Siotani and Hayakawa (1964), and which goes back to Olkin and Siotani (1964), will be used extensively in this and the following chapter.

<u>Lemma 3.1.</u> Let S be as above, $\Sigma = (\sigma_{\alpha\beta})$, and $f_j(S)$, $j \in J$, J a finite set, be real valued functions of S , not algebraically dependent, having first and second derivatives in a neighborhood of Σ (in the topology inherited from $E^{\frac{pk(1+pk)}{2}}$). Then, the a.d. of

$$n^{1/2}(f_j(S) - f_j(\Sigma)) \quad (j \in J)$$

is multivariate normal with zero means, variances equal to $2(f_j(\Sigma))^2 \operatorname{tr} (\phi_j(\Sigma)\Sigma)^2 \quad (j \in J)$, and covariances equal to $2f_j(\Sigma)f_i(\Sigma) \operatorname{tr} \phi_j(\Sigma)\Sigma\phi_i(\Sigma)\Sigma \quad (i,j \in J)$, where

$$\phi_j(\Sigma) = (\phi_{j\alpha\beta}(\Sigma)) \ , \quad \phi_{j\alpha\beta}(\Sigma) = \partial_{\alpha\beta} \log f_j(\Sigma) \ ,$$

$$\partial_{\alpha\beta} = (1/2)(1 + \delta_{\alpha\beta})\frac{\partial}{\partial\sigma_{\alpha\beta}} \ ,$$

$$\delta_{\alpha\beta} = 1 \text{ if } \alpha = \beta \ , = 0 \text{ if } \alpha \neq \beta .$$

<u>Lemma 3.2.</u> The a.d. of

$$n^{1/2}(\det S_1 - \det \Sigma_1, \ldots, \det S_k - \det \Sigma_k)$$

is multivariate normal, with zero means, variances equal to $2p(\det \Sigma_j)^2$, and covariances $2 \det \Sigma_i \det \Sigma_j \operatorname{tr} \Sigma_i^{-1}\Sigma_{ij}\Sigma_j^{-1}\Sigma_{ji}$.

<u>Proof.</u> This lemma is a consequence of Lemma 3.1. Using the notation peculiar to that lemma, let $f_j(\Sigma) = \det \Sigma_j$. Then, it is known (cf. Anderson (1958), p. 347) that,

$$\phi_1(\Sigma) = \begin{bmatrix} \Sigma_1^{-1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} , \quad \phi_2(\Sigma) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma_2^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix} .$$

Noticing that

$$\operatorname{tr}(\phi_1(\Sigma)\Sigma)^2 = \operatorname{tr}\begin{bmatrix} I_p & \Sigma_1^{-1}\Sigma_{12} & \cdots & \Sigma_1^{-1}\Sigma_{1k} \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix}^2 = p$$

$$\operatorname{tr}\ \phi_1(\Sigma)\Sigma\phi_2(\Sigma)\Sigma = \operatorname{tr}\begin{bmatrix} I_p & \Sigma_1^{-1}\Sigma_{12} & \cdots & \Sigma_1^{-1}\Sigma_{1k} \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & \cdots & 0 \\ \Sigma_2^{-1}\Sigma_{21} & I_p & \cdots & \Sigma_2^{-1}\Sigma_{2k} \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$= \operatorname{tr}\ \Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{21}\ ,$$

the present lemma follows at once. QED

Hooper (1959) defined

$$\rho_{ij}^2 = (1/p)\ \operatorname{tr}\ \Sigma_i^{-1}\Sigma_{ij}\Sigma_j^{-1}\Sigma_{ji}\ ,$$

as the squared trace correlation coefficient between subclasses i and j. If $\nu_{1ij}^2,\ldots,\nu_{pij}^2$ are the canonical correlations (cf. Anderson (1958)) between the two subclasses, it can be shown that

$$\rho_{ij}^2 = \sum_{\ell=1}^{p} \nu_{\ell ij}^2/p\ ,$$

which implies

$$0 \le \rho_{ij}^2 \le 1\ .$$

<u>Lemma 3.3.</u> The a.d. of

$$(3.1)\quad Y_j^1 = n^{1/2}\left\{ \frac{\log(\det S_1/\det S_j)-\log(\det \Sigma_1/\det \Sigma_j)}{2p^{1/2}(1-\rho_{1j}^2)^{1/2}} \right\} \quad (j \neq 1)\ ,$$

is standard multivariate normal, with

$$\mathrm{corr}(Y_i^1, Y_j^1) \equiv \gamma_{ij} = \frac{1-\rho_{1i}^2-\rho_{1j}^2+\rho_{ij}^2}{2(1-\rho_{1i}^2)^{1/2}(1-\rho_{1j}^2)^{1/2}} \quad (i \neq j) .$$

Proof. This follows using Lemma 2.9 of Chapter 2 and Lemma 3.2 above.

Theorem 3.1. If the experimenter uses Rule R_{GV1} , an asymptotic $(N \to \infty)$ least favorable configuration of the relevant parameters is

$$\det \Sigma_j / \det \Sigma_1 = \theta^* \quad (j \neq 1) , \quad \Sigma_{ij} = 0 \quad (i \neq j) .$$

Therefore,

$$\text{(3.2)} \qquad \inf_{\Omega} PCS_a(\Sigma) = P(Y_j^1 \leq n^{1/2}(1/2)p^{-1/2}\log \theta^* , \ j \neq 1)$$

where the $\{Y_j^1 , j \neq 1\}$ are distributed as in (3.1) with $\gamma_{ij} = 1/2 \quad (i \neq j)$.

Proof. Using Lemma 3.3, we have for $\det \Sigma_j \geq \theta^* \det \Sigma_1 \quad (j \neq 1)$,

$$\begin{aligned} PCS_a(\Sigma) &= P(\det S_1 < \det S_j , \ j \neq 1) \\ &= P(Y_j < n^{1/2}(1/2)p^{-1/2}(1-\rho_{1j}^2)^{-1/2}\log(\det \Sigma_j/\det \Sigma_1) , \ j \neq 1) \\ &\geq p(Y_j < n^{1/2}(1/2)p^{-1/2}(1-\rho_{1j}^2)^{-1/2}\log \theta^* , \ j \neq 1) . \end{aligned}$$

We can now use Theorem 2.3 of Chapter 2, and set $\rho_{ij}^2 = 0 \quad (i \neq j)$, to obtain a lower bound on $PCS_a(\Sigma)$. Since the parameter configuration $\det \Sigma_j = \theta^* \det \Sigma_1 \quad (j \neq 1)$, $\Sigma_{ij} = 0 \quad (i \neq j)$, leads to this

lower bound, it is a least favorable configuration.

It is easy to show that $\Sigma_{ij} = 0 \quad (i \neq j)$, is also necessary for an asymptotic least favorable configuration. Indeed, if $\rho_{ij}^2 = 0$, then $\Sigma_{ij} = 0$ necessarily, from the definition of squared trace correlation coefficient. QED

Theorem 3.2. If the experimenter uses Rule R_{GV2}, an asymptotic $(N \to \infty)$ least favorable configuration of the relevant parameters is $\Sigma_1 = \ldots = \Sigma_k$, $\Sigma_{ij} = 0 \quad (i \neq j)$. Therefore,

$$(3.3) \qquad \inf_{\Sigma} PCS_a(\Sigma) = P(Y_j^1 \leq n^{1/2}(1/2)p^{-1/2}\log d^* \ , \ j \neq 1)$$

where the $\{Y_j^1 \ , \ j \neq 1\}$ are as in Theorem 3.1.

Proof. The result follows immediately from Theorem 3.1.

Lemma 3.4. If $\tilde{S}$ denotes the size of the selected subset of subclasses of variates when R_{GV2} is employed, we have

$$(a) \quad E_a(\tilde{S}|\Sigma) = \sum_{i=1}^{k} P(Y_j^i \leq n^{1/2}(1/2)p^{-1/2}(1-\rho_{ij}^2)^{-1/2}\log(d^*\det \Sigma_j/\det \Sigma_i) \ , \quad i \neq j)$$

where the $\{Y_j^i \ , \ i \neq j\}$ are distributed as in (3.1) with 1 replaced by i .

$$(b) \quad \sup_{\Sigma} E_a(\tilde{S}|\Sigma) = k \quad \text{which occurs when} \quad \Sigma = \begin{pmatrix} I_p & \cdots & I_p \\ \vdots & & \\ I_p & \cdots & I_p \end{pmatrix} .$$

Proof. The result is a consequence of previous developments.

It is interesting to note that Theorems 3.1 and 3.2, and Lemma 3.4 reduce to Theorems 2.3 and 2.4, and Lemma 2.11 of Chapter 2, respectively, when $p = 1$.

3.2. Selecting the smallest population generalized variance (intersecting subclasses).

The last problem of the present chapter is a problem of intersecting subclasses of variates, where some variates belong to more that a single subclass. We start by proving a lemma, which will be basic to what follows.

Lemma 3.5. Let $X_\alpha^t = (X_{1\alpha}^t, X_{2\alpha}^t, X_{3\alpha}^t)$ $(1 \leq \alpha \leq N)$, be independent normally distributed $(p_1+p_2+p_3)$-vectors, with unknown population means and unknown population covariance matrix

$$\Sigma = \begin{pmatrix} A & D & E \\ D^t & B & F \\ E^t & F^t & C \end{pmatrix}$$

where $A(p_1 \times p_1)$, $B(p_2 \times p_2)$ and $C(p_3 \times p_3)$. Define,

$$X_{12,\alpha}^t = (X_{1\alpha}^t, X_{2\alpha}^t) \ , \quad X_{23,\alpha}^t = (X_{2\alpha}^t, X_{3\alpha}^t) \quad (1 \leq \alpha \leq N) \ ,$$

$$\overline{X}_{12} = \sum_{\alpha=1}^{N} X_{12,\alpha}/N \ , \quad \overline{X}_{23} = \sum_{\alpha=1}^{N} X_{23,\alpha}/N \ ,$$

$$S_{12} = \sum_{\alpha=1}^{N} (X_{12,\alpha} - \overline{X}_{12})(X_{12,\alpha} - \overline{X}_{12})^t/n \ , \quad n = N - 1 \ ,$$

$$S_{23} = \sum_{\alpha=1}^{N} (X_{23,\alpha} - \overline{X}_{23})(X_{23,\alpha} - \overline{X}_{23})^t/n \ ,$$

$$\Sigma_{12} = \begin{pmatrix} A & D \\ D^t & B \end{pmatrix} , \quad \Sigma_{23} = \begin{pmatrix} B & F \\ F^t & C \end{pmatrix} .$$

Then, the a.d. of

$$n^{1/2}(\det S_{12} - \det \Sigma_{12}, \det S_{23} - \det \Sigma_{23})$$

is multivariate normal, with zero means, variances equal to $2(p_1+p_2)(\det\Sigma_{12})^2$ and $2(p_2+p_3)(\det\Sigma_{23})^2$, and covariance $2(p_2+\lambda)\det \Sigma_{12}\det \Sigma_{23}$, where $\lambda \geq 0$ is defined in the course of the proof.

<u>Proof.</u> Using the notation of Lemma 3.1, we obtain, $f_{ij}(\Sigma) = \det \Sigma_{ij}$ $(1 \leq i < j \leq 3)$,

$$\phi_{12}(\Sigma) = \begin{pmatrix} \Sigma_{12}^{-1} & 0 \\ 0 & 0 \end{pmatrix}, \quad \phi_{23}(\Sigma) = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{23}^{-1} \end{pmatrix}.$$

The expressions for the variances follow immediately. Now we define,

$$\Sigma_{12}^{-1} = \begin{bmatrix} \Sigma_{12}^{-1}(u) \\ \Sigma_{12}^{-1}(\ell) \end{bmatrix} = \begin{bmatrix} (A-DB^{-1}D^t)^{-1} & -(A-DB^{-1}D^t)^{-1}DB^{-1} \\ \Sigma_{12}^{-1}(\ell) & \end{bmatrix}$$

$$\Sigma_{23}^{-1} = \begin{bmatrix} \Sigma_{23}^{-1}(u) \\ \Sigma_{23}^{-1}(\ell) \end{bmatrix} = \begin{bmatrix} \Sigma_{23}^{-1}(u) & \\ -(C-F^tB^{-1}F)^{-1}F^tB^{-1} & (C-F^tB^{-1}F)^{-1} \end{bmatrix}.$$

In order to compute the covariance we must evaluate

$$\operatorname{tr} \phi_{12}(\Sigma)\Sigma\phi_{23}(\Sigma)\Sigma$$

$$= \operatorname{tr} \begin{pmatrix} \Sigma_{12}^{-1} & & 0 \\ & & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \Sigma_{12} & & E \\ & & F \\ E^t & F^t & C \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & & \\ 0 & \Sigma_{23}^{-1} & \end{pmatrix} \begin{pmatrix} A & D & E \\ D^t & & \\ E^t & & \Sigma_{23} \end{pmatrix}$$

$$= \operatorname{tr} \begin{pmatrix} I_{p_1} & 0 & \Sigma_{12}^{-1}(u)\binom{E}{F} \\ 0 & I_{p_2} & \Sigma_{12}^{-1}(\ell)\binom{E}{F} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ \Sigma_{23}^{-1}(u)\begin{pmatrix} D^t \\ E^t \end{pmatrix} & I_{p_2} & 0 \\ \Sigma_{23}^{-1}(\ell)\begin{pmatrix} D^t \\ E^t \end{pmatrix} & 0 & I_{p_3} \end{pmatrix}$$

$$= p_2 + \operatorname{tr} \Sigma_{12}^{-1}(u)\binom{E}{F}\Sigma_{23}^{-1}(\ell)\begin{pmatrix} D^t \\ E^t \end{pmatrix}$$

$$= p_2 + \operatorname{tr}(A-DB^{-1}D^t)^{-1}(E-DB^{-1}F)(C-F^tB^{-1}F)^{-1}(E^t-F^tB^{-1}D^t)$$

$$= p_2 + \lambda \ .$$

Hooper (1962) defined

$$\rho^2_{13.2} = \lambda/p_1 \ ,$$

as the squared partial trace correlation coefficient between subclasses X_1 and X_3 conditional on X_2. If $p_1 < p_3$ (say) and $\nu_1^2, \ldots, \nu_{p_1}^2$ are the canonical correlations between X_1 and X_3 in the conditional distribution of X_1 and X_3 given X_2, it can be shown that,

$$\rho^2_{13.2} = \sum_{\ell=1}^{p_1} \nu_\ell^2/p_1 \ ,$$

which implies,

$$0 \leq \rho^2_{13.2} \leq 1 .$$

Hence, $0 \leq \lambda \leq p_1$. QED

We note in passing that $\lambda = 0$ if D , E and F are zero matrices.

In this section we consider $X^t = (X_1,\ldots,X_k)$ a k-variate normal population, with unknown population mean vector, and unknown population covariance matrix Σ. We assume $k \geq 3$. Consider all possible subclasses of specified size t $(t < k)$ of X , whose total number is $U = \binom{k}{t}$. Let $\Sigma_1,\ldots,\Sigma_U$ be the covariance matrices (submatrices of Σ) corresponding to these U subclasses, and let $\det \Sigma_{[1]} \leq \ldots \leq \det \Sigma_{[U]}$ be the ranked values of $\det \Sigma_j$ $(1 \leq j \leq U)$. It is assumed that the experimenter has no prior knowledge concerning the values of $\det \Sigma_j$ $(1 \leq j \leq U)$, or of the pairing of the $\det \Sigma_{[i]}$ with the subclasses of variates.

Indifference-zone formulation

The experimenter's goal is to select a subclass of t variates out of the k-variate population, with the smallest population generalized variance, $\det \Sigma_{[1]}$. He specifies $\{\theta^*,P^*\}$, $\theta^* > 1$, $1/U < P^* < 1$, before experimentation starts. If $PCS_R(\Sigma)$ is as defined in Section 3.1, we restrict consideration to decision procedures R which guarantee the probability requirement:

$$\inf_{\Omega} PCS_R(\Sigma) \geq P^*$$

where

$$\Omega = \{\Sigma | \theta^* \det \Sigma_{[1]} \leq \det \Sigma_j \ , \quad j \neq [1]\} \ .$$

We propose the following decision procedure for this <u>indifference-zone</u> formulation of the problem:

<u>Rule R_{GV3}</u>: Let S be the sample covariance matrix computed using a sample of size N from the above population, as in Section 3.1. Let S_i $(1 \leq i \leq U)$ be submatrices of S corresponding to Σ_i $(1 \leq i \leq U)$. Then assert that the subclass of t variates associated with $\det S_{[1]} \equiv \min \det S_j$ has the smallest population generalized variance, $\det \Sigma_{[1]}$.

Our task is to determine N which will guarantee the probability requirement when R_{GV3} is used.

<u>Subset formulation</u>

If the experimenter is interested in selecting a subset of subclasses of variates, which includes the subclass associated with $\det \Sigma_{[1]}$, he must specify $\{P^*\}$, $1/U < P^* < 1$, before experimentation starts. If $PCS_R(\Sigma)$ has the same meaning as above, we restrict consideration to decision procedures which guarantee the probability requirement:

$$\inf_{\Sigma} PCS_R(\Sigma) \geq P^* \ .$$

For this <u>subset</u> formulation, we propose,

<u>Rule R_{GV4}</u>: Include the subclass associated with S_j $(1 \leq j \leq U)$, in the selected subset of subclasses of variates if $\det S_j \leq d^* \det S_{[1]}$, where $d^* > 1$ is a specified constant.

Our task is to determine the smallest sample size N which will guarantee the probability requirement when R_{GV4} is employed.

Due to the symmetry of the present problem, we may assume, without loss of generality, that $\det \Sigma_1 = \det \Sigma_{[1]}$. Moreover, we give no consideration to population means in what follows, since they are irrelevant for our problems.

<u>Lemma 3.6.</u> The a.d. of

$$n^{1/2}(\det S_1 - \det \Sigma_1, \ldots, \det S_U - \det \Sigma_U)$$

is multivariate normal with zero means, variances equal to $2t(\det \Sigma_j)^2$ $(1 \leq j \leq U)$, and covariances $2(t_{ij} + \lambda_{ij})\det \Sigma_i \det \Sigma_j$ $(1 \leq i < j \leq U)$, where $\lambda_{ij} \geq 0$ (defined in Lemma 3.5), and t_{ij} is the number of common variates of subclasses i and j (corresponding to S_i and S_j).

<u>Proof.</u> This result is a consequence of Lemma 3.5.

We define,

$$\rho_{ij}^2 \equiv (t_{ij} + \lambda_{ij})/t \quad (i \neq j) \ , \quad t_{ij}/t \leq \rho_{ij}^2 \leq 1 \ .$$

<u>Lemma 3.7.</u> The a.d. of

$$(3.4) \quad Y_j^1 = n^{1/2}\left\{ \frac{\log(\det S_1/\det S_j) - \log(\det \Sigma_1/\det \Sigma_j)}{2t^{1/2}(1-\rho_{1j}^2)^{1/2}} \right\} \quad (j \neq 1) \ ,$$

is standard multivariate normal with

$$\mathrm{corr}(Y_i^1, Y_j^1) \equiv \gamma_{ij} = \frac{1-\rho_{1i}^2-\rho_{1j}^2+\rho_{ij}^2}{2(1-\rho_{1i}^2)^{1/2}(1-\rho_{1j}^2)^{1/2}} \quad (i \neq j) .$$

Proof. The result follows using Lemma 3.6 and Lemma 2.9 of the previous chapter.

Theorem 3.3. When one employs R_{GV3} , an asymptotic $(N \to \infty)$ lower bound on the PCS_a is given by:

$$(3.5) \qquad \inf_{\Omega} PCS_a(\Sigma) \geq P(Y_j^1 \leq n^{1/2}(1/2)t^{-1/2}\log \theta^* \ , \ j \neq 1)$$

where the $\{Y_j^1 \ , \ j \neq 1\}$ are distributed as in (3.4) with $\gamma_{ij} = 1/2$ $(i \neq j)$. When $t = 1$ (resp. $t = k - 1$) lower bound (3.5) is sharp, and an asymptotic least favorable configuration of the relevant parameters is $\Sigma = \mathrm{diag}(1,\theta^*,\ldots,\theta^*)$ (resp. $\Sigma = \mathrm{diag}(\theta^*,1,\ldots,1)$) .

Proof. The proof of this theorem is similar to the proof of Theorem 2.3 of the previous chapter.

Theorem 3.4. When one employs R_{GV4} , an asymptotic $(N \to \infty)$ lower bound on the PCS_a is given by:

$$(3.6) \qquad \inf_{\Sigma} PCS_a(\Sigma) \geq P(Y_j^1 \leq n^{1/2}(1/2)t^{-1/2}\log d^* \ , \ j \neq 1)$$

where the $\{Y_j^1 \ , \ j \neq 1\}$ are as in Theorem 3.3.

When $t = 1$ or $t = k - 1$, lower bound (3.6) is sharp, and as asymptotic least favorable configuration of the relevant parameters is $\Sigma = \mathrm{diag}(1,\ldots,1)$.

Proof. The proof of this theorem is similar to the proof of Theorem 2.3.

Lemma 3.8. If $\tilde{S}$ denotes the size of the selected subset of subclasses when R_{GV4} is used,

(a) $$E_a(\tilde{S}|\Sigma) = \sum_{i=1}^{k} P\left(Y_j^i \le \frac{n^{1/2}\log(d^*\det \Sigma_j/\det \Sigma_i)}{2t^{1/2}(1-\rho_{ij}^2)^{1/2}},\ i \ne j\right)$$

where the $\{Y_j^i,\ i \ne j\}$ are as in (3.4) with i in place of 1.

(b) $$\sup_{\Sigma} E_a(\tilde{S}|\Sigma) = k \quad \text{which occurs when} \quad \Sigma = \begin{pmatrix} 1 & \dots & 1 \\ & \dots & \\ 1 & \dots & 1 \end{pmatrix}.$$

Proof. The result is a consequence of Lemma 3.7.

CHAPTER 4

SELECTION OF SUBCLASSES OF VARIATES OR OF POPULATIONS BASED ON MEASURES OF ASSOCIATION BETWEEN TWO SUBCLASSES OF VARIATES (ASYMPTOTIC THEORY)

4.0. Introduction

In the present chapter we consider two problems which have been studied recently by several investigators. We provide solutions to these problems using asymptotic theory.

Section 4.1 contains certain preliminaries and definitions employed in the later sections. In particular, we define a measure of association known as the vector coefficient of alienation between two classes of components. Then, in Section 4.2, we consider the problem of selecting a multivariate normal population (among independent populations) with the smallest vector coefficient of alienation between two classes of components. Gupta and Panchapakesan (1969) and Rizvi and Solomon (1973) give different formulations for this problem.

In Section 4.3, we consider the important problem of selecting the best subclass of predictors for a fixed subclass of variates, each of the contending subclasses being correlated with the subclass previously specified. A quite general asymptotic solution is displayed. The vector coefficient of alienation is used as a measure of association. Ramberg (1969) and Arvensen (1971) obtained partial results for related problems.

Although the problems are formulated in a multivariate normal framework, the same asymptotic results are valid for a very general

class of multivariate distribution functions.

4.1. Preliminaries

In this section we describe a few properties of certain measures of association between two sets of variates. For further details the reader is referred to Hotelling (1936) and Hooper (1959, 1962).

Let (Y,X) be a $(q + p)$-dimensional random variable with covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{pmatrix} .$$

We assume that $q \leq p$ and let $\nu_1^2,\ldots,\nu_q^2$ be the canonical correlations (cf. Anderson (1958)) associated with Y and X.

The <u>conditional generalized variance of Y given X</u> is

$$\det \Sigma_{y\cdot x} = \frac{\det \begin{pmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{pmatrix}}{\det \Sigma_x} = \det(\Sigma_y - \Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}) .$$

It can be shown that, if X, Y and Z are three vectors of variates,

$$\det \Sigma_{y\cdot x} \leq \det \Sigma_y ,$$

$$\det \Sigma_{y\cdot xz} = \det(\Sigma_{y\cdot x} - \Sigma_{yz\cdot x}\Sigma_{z\cdot x}^{-1}\Sigma_{zy\cdot x})$$

$$\leq \det \Sigma_{y\cdot x} .$$

No single measure of association is sufficient to fully describe the relation between two sets of variates. A complete

description would be based on the set of canonical correlations. However, as we need in the present development, a single number to describe such a relation, we shall restrict consideration to real functions of the canonical correlations. The following are a few of the measures of association which have been proposed in the literature:

The <u>vector coefficient of alienation between Y and X</u> is γ_{yx}, where

$$\gamma_{yx}^2 = \frac{\det \Sigma_{y\cdot x}}{\det \Sigma_y} = \frac{\det \Sigma}{\det \Sigma_y \det \Sigma_x} .$$

It can be shown that,

(i) $\gamma_{yx}^2 = (1 - \nu_1^2)\ldots(1 - \nu_q^2)$, $0 \leq \gamma_{yx}^2 \leq 1$.

(ii) $\gamma_{yx}^2 = 0$ iff $\nu_\ell^2 = 1$ for some ℓ .

$\gamma_{yx}^2 = 1$ iff $\nu_\ell^2 = 0$ for all ℓ, i.e., $\Sigma_{yx} = 0$.

The <u>vector multiple correlation coefficient between Y and X</u> is R_{yx}, where

$$R_{yx}^2 = \frac{\det \Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}}{\det \Sigma_y} = \frac{\det \begin{pmatrix} 0 & -\Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{pmatrix}}{\det \Sigma_y \det \Sigma_x} .$$

It can be shown that

(i) $R^2_{yx} = \nu_1^2 \ldots \nu_q^2 , \quad 0 \leq R^2_{yx} \leq 1 .$

(ii) $R^2_{yx} = 0$ iff $\nu_\ell^2 = 0$ for some ℓ .

$R^2_{yx} = 1$ iff $\nu_\ell^2 = 1$ for all ℓ , i.e., $Y = BX$ a.e.

(iii) $R^2_{yx} + \gamma^2_{yx} = \prod_{\ell=1}^{q} \nu_\ell^2 + \prod_{\ell=1}^{q} (1 - \nu_\ell^2) \leq 1 ,$

and, in general, inequality holds, except when $q = 1$.

The <u>trace correlation coefficient between Y and X</u> is ρ_{yx} , where

$$\rho^2_{yx} = (1/q) \text{ tr } \Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1} .$$

It can be shown that

(i) $\rho^2_{yx} = (1/q)(\nu_1^2 + \ldots + \nu_q^2) , \quad 0 \leq \rho^2_{yx} \leq 1 .$

(ii) $\rho^2_{yx} = 0$ iff $\nu_\ell^2 = 0$ for all ℓ , i.e., $\Sigma_{yx} = 0$.

$\rho^2_{yx} = 1$ iff $\nu_\ell^2 = 1$ for all ℓ , i.e., $Y = BX$ a.e.

In the problems treated in the present chapter it is mathematically more convenient to study selection procedures in terms of γ^2_{yx} . When $q = 1$, which is probably the most common case in practice, selecting in terms of γ^2_{yx} is equivalent to selecting in terms of R^2_{yx} .

4.2. Selecting the best out of k populations with respect to the population vector coefficients of alienation

Consider k $(q + p_i)$-variate independent normal populations, $\begin{pmatrix} Y^i \\ X^i \end{pmatrix}$, with unknown population means and unknown population covariance matrices

$$\Sigma_i = \begin{pmatrix} \Sigma_{y_i} & \Sigma_{y_i x_i} \\ \Sigma_{x_i y_i} & \Sigma_{x_i} \end{pmatrix} \quad (1 \leq i \leq k) .$$

We assume $q \leq \min_j p_j$. Let the population squared vector coefficient of alienation between Y^i and X^i be

$$\gamma_i^2 = \frac{\det \Sigma_i}{\det \Sigma_{y_i} \det \Sigma_{x_i}} \quad (1 \leq i \leq k) ,$$

and let the ranked values of the γ_i^2 be $\gamma_{[1]}^2 \leq \cdots \leq \gamma_{[k]}^2$. It is assumed that the experimenter has no prior knowledge concerning the values of the γ_i^2 , or of the pairing of the $\gamma_{[j]}^2$ with the populations $\begin{pmatrix} Y^i \\ X^i \end{pmatrix}$ $(1 \leq i,j \leq k)$.

When $q = 1$, selecting in terms of the γ_i^2 is equivalent to selecting in terms of the population squared multiple correlation coefficients, as indicated in Section 4.1. These selection problems $(q = 1)$ have been considered by Gupta and Panchapakesan (1969) using the subset approach, and by Rizvi and Solomon (1973) using the indifference-zone approach. Both papers provide different treatments

than ours; in particular, our indifference-zone is distinct from Rizvi and Solomon's, being perhaps more natural.

Indifference-zone formulation

The experimenter's goal is to select the population associated with $\gamma^2_{[1]}$. He specifies $\{\theta^*, P^*\}$, $\theta^* > 1$, $1/k < P^* < 1$, before experimentation starts. If $PCS_R(\{\Sigma_i\})$ denotes the probability of a correct selection when decision rule R is used, we restrict consideration to decision procedures R which guarantee the probability requirement:

$$\inf_{\Omega} PCS_R(\{\Sigma_i\}) \geq P^*$$

where

$$\Omega = \{(\Sigma_1, \ldots, \Sigma_k) \mid \theta^* \gamma^2_{[1]} \leq \gamma^2_j \ , \ j \neq [1]\} \ .$$

Single-stage "natural" selection procedures will be used. We propose the following decision procedure:

A sample of N independent vector observations,

$$W^i_\alpha = \begin{pmatrix} Y^i_\alpha \\ X^i_\alpha \end{pmatrix} \quad (1 \leq i \leq k) \quad (1 \leq \alpha \leq N) \ ,$$

is taken from each population and one computes for $(1 \leq i \leq k)$,

$$S_i = \sum_{\alpha=1}^{N} (W^i_\alpha - \overline{W}_i)(W^i_\alpha - \overline{W}_i)^t/n \ , \ \overline{W}_i = \sum_{\alpha=1}^{N} W^i_\alpha/N \ ,$$

where $n = N - 1$, and the sample squared vector coefficient of alienation,

$$G_i^2 = \frac{\det S_i}{\det S_{y_i} \det S_{x_i}} .$$

<u>Rule R_{C1}</u>: Select the population associated with $G_{[1]}^2 = \min \{G_1^2, \ldots, G_k^2\}$, as the one corresponding to $\gamma_{[1]}^2$.

Our task is to determine the smallest sample size N which guarantees the probability requirement when R_{C1} is used.

<u>Subset formulation</u>

If the experimenter's goal is to select a subset of populations containing the one associated with $\gamma_{[1]}^2$, he specifies $\{P^*\}$, $1/k < P^* < 1$, prior to the start of experimentation. Then if $PCS_R(\{\Sigma_i\})$ has the same meaning as above, we restrict consideration to decision procedures R which guarantee the probability requirement:

$$\inf_{\Sigma_1, \ldots, \Sigma_k} PCS_R(\{\Sigma_i\}) \geq P^* .$$

We propose the following decision procedure:

<u>Rule R_{C2}</u>: Include the population associated with G_j^2 in the selected subset of populations if $G_j^2 \leq d^* G_{[1]}^2$, where $d^* > 1$ is a specified constant.

Our objective then is to determine the smallest sample size N which will guarantee the probability requirement when R_{C2} is employed.

It is clear that we may disregard the population means in what follows. It will be seen (Theorems 4.1 and 4.2) that we may assume, without loss of generality, that $\gamma_1^2 \leq \gamma_j^2$ $(j \neq 1)$.

Part of the following lemma is proved in Siotani, Chou and Geng (1971).

Lemma 4.1. The a.d. of

$$n^{1/2}(G_1^2 - \gamma_1^2, \ldots, G_k^2 - \gamma_k^2)$$

is multivariate normal with zero means, variances $4\gamma_i^4 \ell_i$, and zero correlations, where,

$$0 \leq \ell_i \equiv \operatorname{tr} \Sigma_{y_i}^{-1} \Sigma_{y_i x_i} \Sigma_{x_i}^{-1} \Sigma_{x_i y_i} \leq q \ .$$

Proof. Since the squared trace correlation between Y_i and X_i is $\rho_{y_i x_i}^2 = \frac{\ell_i}{q}$, it follows that $0 \leq \ell_i \leq q$. Using Lemma 3.1 of Chapter 3, with the notation introduced there, we have only to compute the asymptotic variances. The result follows noticing that

$$f_i(\Sigma_i) = \frac{\det \Sigma_i}{\det \Sigma_{y_i} \det \Sigma_{x_i}} ,$$

$$\phi_i(\Sigma_i) = \{\partial_{\alpha\beta}(\log \det \Sigma_i - \log \det \Sigma_{y_i} - \log \det \Sigma_{x_i})\}$$

$$= \Sigma_i^{-1} - \begin{pmatrix} \Sigma_{y_i}^{-1} & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{x_i}^{-1} \end{pmatrix} .$$

$$\operatorname{tr}(\phi_i(\Sigma_i)\Sigma_i)^2 = 2 \operatorname{tr} \Sigma_{y_i}^{-1} \Sigma_{y_i x_i} \Sigma_{x_i}^{-1} \Sigma_{x_i y_i} \doteq 2\ell_i$$

QED

Lemma 4.2. The a.d. of

$$(4.1) \qquad Y_j^1 = n^{1/2} \left\{ \frac{\log(G_1^2/G_j^2) - \log(\gamma_1^2/\gamma_j^2)}{2(\ell_1+\ell_j)^{1/2}} \right\} \quad (j \neq 1)$$

is standard multivariate normal with

$$\mathrm{corr}(Y_i^1, Y_j^1) = w_{ij} = \frac{\ell_1}{(\ell_1+\ell_i)^{1/2}(\ell_1+\ell_j)^{1/2}} \quad (i \neq j) .$$

Proof. This result follows using Lemma 4.1 and Lemma 2.9 of Chapter 2.

Theorem 4.1. If the experimenter uses Rule R_{C1}, an asymptotic $(N \to \infty)$ lower bound on the PCS_a is

$$(4.2) \qquad \inf_{\Omega} PCS_a \geq P\left(Y_j^1 \leq \frac{n^{1/2} \log \theta^*}{2(2q)^{1/2}} , \ j \neq 1\right)$$

where $\{Y_j^1 , j \neq 1\}$ is distributed as in (4.1), with $w_{ij} = 1/2$ $(i \neq j)$.

Proof. We shall only outline the proof, since it is very similar to the proof of Theorem 2.3.

In Ω , if $\gamma_1^2 < \gamma_j^2$ $(j \neq 1)$ we have

$$PCS_a = P(G_1^2 < G_j^2 , \ j \neq 1) = P\left(Y_j^1 < \frac{n^{1/2}\log(\gamma_j^2/\gamma_1^2)}{2(\ell_1+\ell_j)^{1/2}} , \ j \neq 1\right)$$

$$\geq P\left(Y_j^1 < \frac{n^{1/2}\log \theta^*}{2(\ell_1+\ell_j)^{1/2}} , \ j \neq 1\right)$$

$$= \Phi_{(w_{ij})}\left(\frac{c_{\theta^*}(n)}{(\ell_1+\ell_2)^{1/2}} , \ \ldots , \ \frac{c_{\theta^*}(n)}{(\ell_1+\ell_k)^{1/2}} \right)$$

where $c_{\theta^*}(n) = n^{1/2}(1/2)\log\theta^*$, and

$$\Phi_{(w_{ij})} = \Phi_{(w_{ij})}(c_2(n),\ldots,c_k(n))$$

$$= \int_{-\infty}^{c_2(n)} \cdots \int_{-\infty}^{c_k(n)} f_{(w_{ij})}(y_2,\ldots,y_k)\,dy_2\ldots dy_k ,$$

$$c_j(n) = c_{\theta^*}(n)(\ell_1+\ell_j)^{-1/2} \quad (j \neq 1) ,$$

and $f_{(w_{ij})}(y_2,\ldots,y_k)$ is the p.d.f. of the $\{Y_j^1 , j \neq 1\}$.

It is easy to check that

$$\frac{\partial\Phi_{(w_{ij})}}{\partial\ell_j} < 0 \quad (j \neq 1) .$$

Let $w^* = \ell_1/(\ell_1+q)$, $c(n) = c_{\theta^*}(n)/(\ell_1+q)^{1/2}$. Then it follows from the signs of the last derivatives that

$$\Phi_{(w_{ij})}(c_2(n),\ldots,c_k(n)) \geq \Phi_{(w^*)}(c(n),\ldots,c(n)) = \Phi_{(w^*)} .$$

Now,

$$\frac{\partial\Phi_{(w^*)}}{\partial\ell_1} = \frac{\partial\Phi_{(w^*)}}{\partial w^*}\bigg|_{\ell_1 \text{ fixed}} \cdot \frac{\partial w^*}{\partial\ell_1} + \frac{\partial\Phi_{(w^*)}}{\partial\ell_1}\bigg|_{w^* \text{ fixed}}$$

$$= \frac{\partial w^*}{\partial\ell_1}\,\frac{(k-1)(k-2)}{2}\int_{-\infty}^{c(n)}\cdots\int_{-\infty}^{c(n)} f_{(w^*)}(c(n),c(n),y_4,\ldots,y_k)\,dy_4\ldots dy_k$$

$$+ \frac{\partial c(n)}{\partial\ell_1}(k-1)\int_{-\infty}^{c(n)}\cdots\int_{-\infty}^{c(n)} f_{(w^*)}(c(n),y_3,\ldots,y_k)\,dy_3\ldots dy_k .$$

When $c_{\theta^*}(n) \to \infty$, we have

$$\frac{\partial\Phi(w^*)}{\partial\ell_1} < 0 .$$

Therefore, the infimum of PCS_a occurs when $\ell_i = q$ $(1 \le i \le k)$. QED

It is not easy to display an asymptotic least favorable configuration of the parameters. This is so because when $\ell_i = q$, $Y_i = BX_i$ a.e., and $\Sigma_{y_i} = B\Sigma_{x_i}B^t$, $\Sigma_{y_i x_i} = B\Sigma_{x_i}$, implying that $\gamma_i^2 = 0$. However, it is possible to use a limit argument to show that the least favorable configuration of these parameters occurs when $\gamma_i^2 \to 0$ and $\gamma_i^2/\gamma_1^2 \to \theta^*$.

<u>Theorem 4.2.</u> If Rule R_{C2} is used, an asymptotic $(N \to \infty)$ lower bound on the PCS_a is,

$$\inf_{\{\Sigma_i\}} PCS_a \ge P\left(Y_j^1 \le \frac{n^{1/2}\log d^*}{2(2q)^{1/2}} , j \ne 1\right) \tag{4.4}$$

where $\{Y_j^1 , j \ne 1\}$ is as in the previous theorem.

<u>Proof.</u> The result follows as in the proof of Theorem 4.1.

To obtain an asymptotic least favorable configuration of the parameters, we must take $\gamma_i^2 = \gamma_j^2$ $(i \ne j)$, and let $\gamma_i^2 \to 0$ for all i , so that $\ell_i \to q$, and then (4.4) follows.

<u>Lemma 4.3.</u> If $\tilde{S}$ denotes the size of the selected subset when R_{C2} if employed, we have

$$\text{(a)} \qquad E_a(\tilde{S}|\{\Sigma_i\}) = \sum_{i=1}^{k} P\left(Y_j^i \le \frac{n^{1/2}\log(d^*\gamma_j^2/\gamma_i^2)}{2(p_i+p_j)^{1/2}} , i \ne j\right)$$

where the $\{Y_j^i, i \neq j\}$ are distributed as in (4.1) with 1 replaced by i.

(b) $\sup E_a(\tilde{S}|\{\Sigma_i\}) = k$ which occurs when $Y_i = B_i X_i$ a.e., and

$$\Sigma_i = \begin{pmatrix} B_i \Sigma_{x_i} B_i^t & B_i \Sigma_{x_i} \\ \Sigma_{x_i} B_i^t & \Sigma_{x_i} \end{pmatrix}$$

in which case $G_i^2 = 0$ a.e. $(1 \leq i \leq k)$.

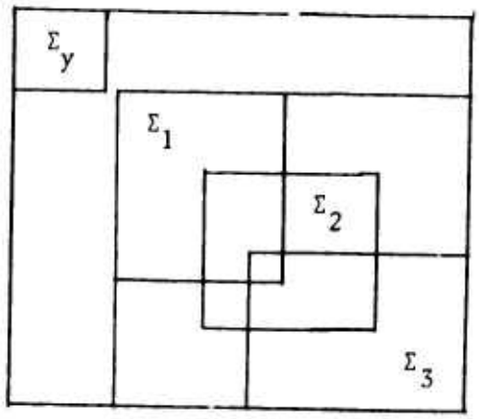
Proof. The result is a consequence of previous developments.

4.3. Selecting the best subclass of predictors (single population)

Consider a (q+p)-variate normal population, $\begin{pmatrix} Y \\ X \end{pmatrix}$ with unknown population mean vector and unknown population covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{pmatrix}.$$

Let X^j $(1 \leq j \leq k)$ be k subclasses of X of size p_j, no one of which is entirely contained in the other. Let Σ_j be the population covariance matrix of X^j $(1 \leq j \leq k)$. The following is a possible covariance matrix in the present setting:

$\Sigma =$ 

Denote the covariance matrix of $\begin{pmatrix} Y \\ X^j \end{pmatrix}$ by $\Sigma^j = \begin{pmatrix} \Sigma_y & \Sigma_{yj} \\ \Sigma_{jy} & \Sigma_j \end{pmatrix}$, and let the population conditional generalized variance of Y given X^j be

$$\lambda_j = \det \Sigma_{y \cdot j} = \frac{\det \Sigma^j}{\det \Sigma_j} \quad (1 \leq j \leq k) .$$

Let the ordered values of the λ_j be $\lambda_{[1]} \leq \cdots \leq \lambda_{[k]}$. We assume that the experimenter has no prior knowledge concerning the values of the λ_j, or of the pairing of the $\lambda_{[j]}$ with $\begin{pmatrix} Y \\ X^i \end{pmatrix}$ $(1 \leq i,j \leq k)$.

In the present context, selecting in terms of the conditional generalized variances, λ_i, is equivalent to selecting in terms of the squared vector coefficient of alienation between Y and X^i, since Y is a common factor to each pair $\begin{pmatrix} Y \\ X^i \end{pmatrix}$ $(1 \leq i \leq k)$. When $q = 1$ we are equivalently selecting in terms of the multiple

correlation coefficients between Y and X^i $(1 \leq i \leq k)$. Ramberg (1969) considered the problem $(q = 1)$ of selecting the subclass X^j associated with $\lambda_{[1]}$, for some special cases of Σ, and developed lower bounds on PCS_a, using an indifference-zone approach. Our bound (Theorem 4.3) is sharper than any of his, and we show that it is attained in some important cases.

A particular case of the theory we develop is the problem of selecting the "best" (corresponding to $\lambda_{[1]}$) subclass of X of size t, for which there are $\binom{p}{t}$ possible decisions. Arvensen (1971) devised a Bayesian procedure for a subset approach formulation of this problem, when $q = 1$. He used asymptotic distribution results of Siotani (1971), but his results are very cumbersome. Theorems 4.5 and 4.6 give a simple counterpart to his theory.

Indifference-zone formulation

The experimenter's goal is to select the subclass X^j $(1 \leq j \leq k)$ associated with $\lambda_{[1]}$. He specifies $\{\theta^*, P^*\}$, $\theta^* > 1$, $1/k < P^* < 1$, prior to experimentation. Then, if $PCS_{\mathcal{R}}(\Sigma)$ denotes the probability of a correct selection when decision procedure $\mathcal{R}$ is employed, we restrict consideration to procedures $\mathcal{R}$ which guarantee the probability requirement:

$$\inf_{\Omega} PCS_{\mathcal{R}}(\Sigma) \geq P^* ,$$

where,

$$\Omega = \{\Sigma | \theta^* \lambda_{[1]} \leq \lambda_j \ , \ j \neq [1]\} \ .$$

We propose the use of the following single-stage "natural" selection procedure for this <u>indifference-zone</u> goal:

A sample of N independent vector observations,

$$Z_\alpha = \begin{pmatrix} Y_\alpha \\ X_\alpha \end{pmatrix} \quad (1 \leq \alpha \leq N)$$

is taken. Let,

$$Z_\alpha^j = \begin{pmatrix} Y_\alpha \\ X_\alpha^j \end{pmatrix} \quad (1 \leq \alpha \leq N) \quad (1 \leq j \leq k) \quad \text{correspond to} \quad \begin{pmatrix} Y \\ X^j \end{pmatrix} .$$

For each $(1 \leq j \leq k)$ compute,

$$S^j = \sum_{\alpha=1}^{N} (Z_\alpha^j - \overline{Z}^j)(Z_\alpha^j - \overline{Z}^j)^t / n \equiv \begin{pmatrix} S_y & S_{yj} \\ S_{jy} & S_j \end{pmatrix} , \quad \overline{Z}_j = \sum_{\alpha=1}^{N} Z_\alpha^j / N ,$$

where $n = N - 1$, and the sample conditional generalized variances,

$$V_j \equiv \det S_{y \cdot j} = \frac{\det S^j}{\det S_j} .$$

<u>Rule R_{C3}</u>: Select the subclass X^j $(1 \leq j \leq k)$ associated with $V_{[1]} \equiv \min \{V_1, \ldots, V_k\}$, as the subclass corresponding to $\lambda_{[1]}$.

Our objective is to determine the smallest sample size N which will guarantee the probability requirement when R_{C3} is used.

Subset formulation

If the experimenter's goal is to select a subset of subclasses of X, X_j $(1 \leq j \leq k)$, which contains the subclass associated with $\lambda_{[1]}$, he specifies $\{P^*\}$, $1/k < P^* < 1$, prior to experimentation. Then, if $PCS_{\mathcal{R}}(\Sigma)$ is as defined above, we limit consideration to decision procedures $\mathcal{R}$ which guarantee the probability requirement:

$$\inf_{\Sigma} PCS_{\mathcal{R}}(\Sigma) \geq P^* .$$

We propose the following "natural" procedure for this subset goal:

Rule $\mathcal{R}_{C4}$: Include the subclass X^j $(1 \leq j \leq k)$ in the selected subset of subclasses if $V_j \leq d^* V_{[1]}$, where $d^* > 1$ is a specified constant.

Our objective is to determine the smallest N which will guarantee the probability requirement, when $\mathcal{R}_{C4}$ is used.

It is clear that the population means may be ignored in the following developments. It will follow from Theorems 4.3 and 4.5 below that we may assume, without loss of generality, that $\lambda_1 \leq \lambda_j$ $(j \neq 1)$.

Lemma 4.4. The a.d. of

$$n^{1/2}(V_1 - \lambda_1, \ldots, V_k - \lambda_k)$$

is multivariate normal, with zero means, variances $2q\lambda_j^2$, and

covariances $2\lambda_i\lambda_j\ell_{ij}$, where $\ell_{ij} \geq 0$.

Proof. We employ Lemma 3.1 of Chapter 3, with its special notation. In order to compute the variances, we note that,

$$f_j(\Sigma) = \det \Sigma^j / \det \Sigma_j ,$$

$$\phi_j(\Sigma) = \{\partial_{\alpha\beta}(\log \det \Sigma^j - \log \det \Sigma_j)\}$$

$$= (\Sigma^j)^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_j^{-1} \end{pmatrix} .$$

Hence,

$$\mathrm{tr}(\phi_j(\Sigma)\Sigma)^2 = \mathrm{tr}\left(I_{q+p_j} - \begin{pmatrix} 0 & 0 \\ \Sigma_j^{-1}\Sigma_{jy} & I_{p_j} \end{pmatrix}\right)^2 = q ,$$

and the variances equal $2q\lambda_j^2$ $(1 \leq j \leq k)$.

The covariances are computed similarly, since we define,

$$2\lambda_i\lambda_j \ \mathrm{tr}\ \phi_i(\Sigma)\Sigma\phi_j(\Sigma)\Sigma \equiv 2\lambda_i\lambda_i\ell_{ij} .$$

We have only to show that $\ell_{ij} \geq 0$. Since $\phi_j(\Sigma)$ can be easily shown to be symmetric nonnegative definite, it follows that

$\ell_{ij} = \mathrm{tr}\ \Sigma\phi_j(\Sigma)\Sigma\phi_i(\Sigma) \geq 0$. QED

Lemma 4.5. The a.d. of

$$(4.5) \qquad Y_j^1 = n^{1/2} \left\{ \frac{\log(V_1/V_j) - \log(\lambda_1/\lambda_j)}{2q^{1/2}(1-\ell_{1j})^{1/2}} \right\} \quad (j \neq 1)$$

is standard multivariate normal with

$$\operatorname{corr}(Y_i^1, Y_j^1) \equiv \gamma_{ij} = \frac{1-\ell_{1i}-\ell_{1j}+\ell_{ij}}{2(1-\ell_{1i})^{1/2}(1-\ell_{1j})^{1/2}} \quad (i \neq j) .$$

Proof. One uses Lemma 4.4 and Lemma 2.9 of Chapter 2.

Theorem 4.3. If the experimenter uses Rule R_{C3} , an asymptotic $(N \to \infty)$ lower bound on the PCS_a is

$$\inf_{\Omega} PCS_a \geq P(Y_j^1 \leq \frac{n^{1/2} \log \theta^*}{2q^{1/2}} , j \neq 1) , \tag{4.6}$$

where the $\{Y_j^1 , j \neq 1\}$ are as in (4.5) with $\gamma_{ij} = 1/2$ $(i \neq j)$.

Proof. This theorem is an immediate consequence of Lemma 4.5 and Theorem 2.3 of Chapter 2.

Lower bound (4.6) turns out to be a sharp bound for a very wide class of problems. Indeed, the only requirement is that each subclass X^j have a variate x_j of its own. More precisely, for $(1 \leq j \leq k)$, there exists x_j such that $x_j \in X^j$, but $x_j \notin X^i$ $(i \neq j)$. When this is the case, we will display an asymptotic $(N \to \infty)$ least favorable configuration of Σ . In order to do so, let y be any fixed component of Y and define,

$$\sigma_{yj} \equiv \operatorname{cov}(y, x_j) , \quad \sigma_{ij} \equiv \operatorname{cov}(x_i, x_j) \quad (1 \leq i,j \leq k) .$$

Theorem 4.4. An asymptotic $(N \to \infty)$ least favorable configuration of Σ , when each X^j has at least one variate of its own, is:

(i) $\sigma_{yy} = \sigma_{jj} = 1 \quad (1 \leq j \leq k)$

(ii) $\sigma_{y1} = \left(1 - \frac{\varepsilon}{\theta^* k}\right)^{1/2}$

(iii) $\sigma_{yj} = \left(1 - \frac{\varepsilon}{k}\right)^{1/2} \quad (j > 1)$

(iv) $\sigma_{1j} = \dfrac{1-\varepsilon/k-\varepsilon/\theta^* k}{(1-\varepsilon/k-\varepsilon/\theta^* k+\varepsilon^2/\theta^* k^2)^{1/2}}$

(v) $\sigma_{ij} = \dfrac{1-2\varepsilon/k}{1-\varepsilon/k} \quad (i,j > 1)$

(vi) all other diagonal elements of Σ_y equal to δ^{-1} ,

(vii) all other diagonal elements of Σ_x equal to 1 ,

(viii) all other elements of Σ equal to zero.

Finally, we take ε sufficiently small and let $\delta \to 0$.

<u>Proof.</u> In order to show that Σ so defined is positive semi-definite, three conditions must be satisfied:

$$\sigma_{ij} \geq -1/(k-2) \qquad (i,j > 1)$$

$$\frac{\sigma_{ij}-\sigma_{1j}^2}{1-\sigma_{1j}^2} \geq -1/(k-2) \qquad (i,j > 1)$$

$$\frac{\sigma_{ij}-\sigma_{yj}^2-(\sigma_{1j}-\sigma_{y1}\sigma_{yj})^2/(1-\sigma_{y1}^2)}{1-\sigma_{yj}^2-(\sigma_{1j}-\sigma_{y1}\sigma_{yj})/(1-\sigma_{y1}^2)} \geq -1/(k-2) \qquad (i,j > 1)$$

A tedious, but straightforward, computation shows that these conditions are satisfied when ε is sufficiently small. Next, we observe that,

$$\theta^*\lambda_1 = \theta^*\delta^{-(q-1)}(1-\sigma_{y1}^2) = \lambda_j = \delta^{-(q-1)}(1-\sigma_{yj}^2) \quad (j > 1) .$$

Finally, another tedious calculation shows that,

$$\ell_{ij} = (1-\sigma_{yi}^2)^{-1}(1-\sigma_{yj}^2)^{-1}\{(q-1)\delta+(1-\sigma_{yi}^2-\sigma_{yj}^2+\sigma_{yi}\sigma_{yj}\sigma_{ij})^2\} .$$

Since it may be checked that $(1-\sigma_{yi}^2-\sigma_{yj}^2+\sigma_{yi}\sigma_{yj}\sigma_{ij}) = 0$, as $\delta \to 0$, we have $\ell_{ij} \to 0$, for all i,j . QED

When $q = 1$, the limit argument $\delta \to 0$ is unnecessary.

<u>Theorem 4.5.</u> If Rule R_{C4} is used, an asymptotic $(N \to \infty)$ lower bound on the PCS_a is

$$\text{(4.7)} \qquad \inf PCS_a \geq P(Y_j^1 \leq \frac{n^{1/2}\log d^*}{2q^{1/2}} \ , \ j \neq 1) \ ,$$

where the $\{Y_j^1 \ , \ j \neq 1\}$ are as in Theorem 4.3.

<u>Proof.</u> The proof is similar to the proof of Theorem 4.3.

<u>Theorem 4.6.</u> Using the same notation as in Theorem 4.4, an asymptotic $(N \to \infty)$ least favorable configuration of Σ , when each X^j has at least one variate of its own, is:

(i) $\sigma_{yy} = \sigma_{jj} = 1 \quad (1 \leq j \leq k)$

(ii) $\sigma_{yj} = (1 - \varepsilon/k)^{1/2} \quad (1 \leq j \leq k)$

(iii) $\sigma_{ij} = \dfrac{1-2\varepsilon/k}{1-\varepsilon/k} \quad (i,j > 1)$

(iv) all other diagonal elements of Σ_y equal to δ^{-1} ,

(v) all other diagonal elements of Σ_x equal to 1 ,

(vi) all other elements of Σ equal to zero.

Finally, we take ε small and let $\delta \to 0$.

Proof. This proof is similar to the proof of Theorem 4.4. The conditions that Σ be positive semi-definite are:

$$\sigma_{ij} \geq -1/(k-1) \quad (i,j > 1)$$

$$\frac{\sigma_{ij}-\sigma_{yj}^2}{1-\sigma_{yj}^2} \geq -1/(k-1) \quad (i,j > 1) ,$$

which can be shown to be satisfied when ε is sufficiently small.

Moreover,

$$\lambda_j = \delta^{-(q-1)}(1-\sigma_{yj}^2) = \lambda_i \quad (i \neq j) .$$

Finally, for $(i \neq j)$,

$$\ell_{ij} = (1-\sigma_{yi}^2)^{-1}(1-\sigma_{yj}^2)^{-1}\{(q-1)\delta+(1-\sigma_{yi}^2-\sigma_{yj}^2+\sigma_{yi}\sigma_{yj}\sigma_{ij})^2\} \to 0$$

because $(1-\sigma_{yi}^2-\sigma_{yj}^2+\sigma_{yi}\sigma_{yj}\sigma_{ij}) = 0$ and $\delta \to 0$. QED

When $q = 1$, the limit argument $\delta \to 0$ is unnecessary.

Lemma 4.6. If $\tilde{S}$ denotes the size of the selected subset of subclasses when R_{C4} is used, we have

$$\text{(a)} \qquad E_a(\tilde{S}|\Sigma) = \sum_{i=1}^{k} P(Y_j^i \leq \frac{n^{1/2}\log(d^*\lambda_j/\lambda_i)}{2q^{1/2}(1-\ell_{ij})^{1/2}} , \ i \neq j)$$

where the $\{Y_j^i , i \neq j\}$ are distributed as in (4.5) with 1 replaced by i .

(b) $\sup_{\Sigma} E_a(S|\Sigma) = k$ when $Y = BX$ a.e., in which case,

$$S_{y \cdot j} = 0 \text{ a.e. } (1 \leq j \leq k) .$$

<u>Proof.</u> Consequence of previous developments.

BIBLIOGRAPHY

Anderson, T.W. (1958), An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York.

Arvensen, J.N. (1971), "A subset selection procedure for selecting the largest multiple correlation coefficient," Dept. Statist. Mimeo. Ser. No. 269, Purdue U., Lafayette, Indiana.

Bahadur, R.R. (1950), "On a problem in the theory of k populations," Ann. Math. Statist., 21, pp. 362-375.

Bahadur, R.R., and Goodman, L.A. (1952), "Impartial decision rules and sufficient statistics," Ann. Math. Statist., 23, pp. 553-562.

Bartlett, M.S., and Kendall, D.G. (1946), "The statistical analysis of variance-heterogeneity and the logarithm transformation," J. Roy. Statist. Soc. Suppl., 8, pp. 128-138.

Bechhofer, R.E. (1954), "A single-sample multiple decision procedure for ranking means of normal populations with known variances," Ann. Math. Statist., 25, pp. 16-39.

Bechhofer, R.E. (1968), "Single-stage procedures for ranking multiply-classified variances of normal populations," Technometrics, 10, pp. 693-714.

Bechhofer, R.E., Kiefer, J., and Sobel, M. (1968), Sequential Identification and Ranking Procedures, Statistical Research Monographs, Vol. III, The University of Chicago Press, Chicago.

Bechhofer, R.E., and Sobel, M. (1954), "A single-sample multiple-decision procedure for ranking variances of normal populations," Ann. Math. Statist., 25, pp. 273-289.

Dunnett, C.W. (1960), "On selecting the largest of k normal populations means," J. Roy. Statist. Soc. Ser. B., 22, pp. 1-40.

Eaton, M.L. (1967a), "Some optimum properties of ranking procedures," Ann. Math. Statist., 38, pp. 124-137.

Eaton, M.L. (1967b), "The generalized variance: testing and ranking problems," Ann. Math. Statist., 38, pp. 941-943.

Fabian, V. (1962), "On multiple decision methods for ranking population means," Ann. Math. Statist., 33, pp. 248-254.

Feller W. (1968), An Introduction to Probability Theory and Its Applications (3rd Edition), John Wiley and Sons, New York.

Gnanadesikan, M.R., and Gupta, S.S. (1970), "Selection procedures for multivariate normal distributions in terms of measures of dispersion," Technometrics, 12, pp. 103-117.

Gupta, S.S. (1956), "On a decision rule for a problem in ranking means," Inst. Stat. Mimeo. Ser. No. 150, Inst. Stat., University of N.C., Chapel Hill, N.C.

Gupta, S.S. (1963), "Probability integrals of multivariate normal and multivariate t ," Ann. Math. Statist., 34, pp. 792-828.

Gupta, S.S. (1965), "On some multiple decision (selection and ranking) rules," Technometrics, 7, pp. 225-245.

Gupta, S.S., and Panchapakesan, S. (1969), "Some selection and ranking procedures for multivariate normal populations," in Multivariate Analysis, Vol. 2, Academic Press, New York.

Gupta, S.S., and Panchapakesan, S. (1972), "On multiple decision procedures," Journal Math. Physical Sciences, 6, pp. 1-72.

Gupta, S.S., and Santner, T. J. (1972), "Selection of a restricted subset of normal populations containing the one with the largest mean," Dep. Statist. Mimeo. Ser. No. 299, Purdue U., Lafayette, Indiana.

Gupta, S.S., and Sobel, M. (1962), "On selecting a subset containing the population with the smallest variance," Biometrika, 49, pp. 495-507.

Hall, W.J. (1958), "Most economical multiple-decision rules," Ann. Math. Statist., 29, pp. 1079-1094.

Hall, W.J. (1959), "The most economical character of some Bechhofer and Sobel decision rules," Ann. Math. Statist., 30, pp. 964-969.

Hooper, J.W. (1959), "Simultaneous Equations and Canonical Correlation Theory," Econometrica, 27, pp. 245-256.

Hooper, J.W. (1962), "Partial Trace Correlations," Econometrica, 30, pp. 324-331.

Hotelling, H. (1936), "Relations between two sets of variates," Biometrika, 28, pp. 321-377.

Johnson, N.L., and Kotz, S. (1972), Distributions in Statistics: Continuous Multivariate Distributions, John Wiley and Sons, New York.

Lehmann, E.L. (1957), "A theory of some multiple decision problems, I and II," Ann. Math. Statist., 28, pp. 1-25 and pp. 547-572.

Lehmann, E. L. (1961), "Some model I problems of selection," Ann. Math. Statist., 32, pp. 990-1012.

Lehmann, E.L. (1966), "On a theorem of Bahadur and Goodman," Ann. Math. Statist., 37, pp. 1-6.

Mahamunulu, D.M. (1966), "Two properties of a subset selection procedure (Preliminary Report), Abstract, Ann. Math. Statist., 37, p. 1429.

Mahamunulu, D.M. (1967), "Some fixed-sample ranking and selection problems," Ann. Math. Statist., 38, pp. 1079-1091.

Milton, R.C. (1963), "Tables of the equally correlated multivariate normal probability integral," Tech. Rep. No. 27, Dep. of Stat., Univ. of Minn., Mpls., Minn.

National Bureau of Standards (1959), Tables of Bivariate Normal Distribution and Related Functions, Applied Math. Series 50, U.S. Government Printing Office, Washington.

Olkin, I., and Siotani, M. (1964), "Asymptotic distribution of functions of a correlation matrix," Tech. Rep. No. 6, Dep. Stat., Stanford U., Stanford, California.

Paulson, E. (1949), "A multiple decision procedure for certain problems in the analysis of variance," Ann. Math. Statist., 20, pp. 95-98.

Paulson, E. (1952a), "On the comparison of several experimental categories with a control," Ann. Math. Statist., 23, pp. 239-246.

Paulson, E. (1952b), "An optimum solution to the k-sample slippage problem for the normal distribution," Ann. Math. Statist., 23, pp. 610-616.

Paulson, E. (1964), "A sequential procedure for selecting the population with the largest mean from k normal populations," Ann. Math. Statist., 35, pp. 174-180.

Plackett, R. L. (1954), "A reduction formula for normal multivariate integrals," Biometrika, 41, pp. 351-360.

Ramberg, J.S. (1969), "A multiple decision approach to the selection of the best set of predictor variables," Tech. Rep. No. 79, Dep. Operations Research, Cornell U., Ithaca, N.Y.

Rao, C.R. (1968), Linear Statistical Inference and Its Applications, John Wiley and Sons, New York.

Rizvi, M.H., and Solomon, H. (1973), "Selection of largest multiple correlation coefficient: asymptotic case," Journal Am. Statist. Assoc., 68, pp. 184-188.

Siotani, M., and Hayakawa, T. (1964), "Asymptotic distributions of functions of Wishart matrix," Proc. Inst. Statist. Math., 12, pp. 191-198 (in Japanese with English abstract).

Siotani, M., Chou, C., and Geng, S. (1971), "Asymptotic joint distributions of vector correlation coefficients and of vector alienation coefficients," Tech Report No. 22, Dept. Statist. Comp. Science, Kansas State U., Manhattan Kansas.

Siotani, M. (1971), "Asymptotic joint distribution of $\binom{p}{t}$ multiple correlation coefficients between a certain variate and t variates among p other variates $(t < p)$," Tech. Report No. 16, Dept. Statist. Comp. Science, Kansas State U., Manhattan, Kansas.

Slepian, D. (1962), "The one-sided barrier problem for Gaussian noise," The Bell System Tech. Journal, 41, pp. 463-501.

Somerville, P.M. (1954)," Some problems of optimum sampling," Biometrika, 41, pp. 420-429.

Wald, A. (1950), Statistical Decision Functions, John Wiley and Sons, New York.