UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

Mathematical Sciences Directorate
Air Force Office of Scientific Research
Washington 25, D. C.

AFOSR Report No.

933

A MINIMUM AVERAGE RISK SOLUTION FOR THE PROBLEM OF CHOOSING THE LARGEST MEAN

By

Richard Park Bland

April, 1961

Contract No. AF 49(638)-929

The problem of choosing the largest of n means is considered as a multiple decision problem which is generated from n component two-decision problems. With additive losses Bayes rules for the component problems yield Bayes rules for the multiple decision problem. Some properties of these Bayes rules are found. Also a conservative-near-Bayes rule is presented with tabled values for any number of means.

Institute of Statistics
Mimeograph Series No. 280

A C K N O W L E D G E M E N T S

This short note is to acknowledge encouragement, guidance and other forms of help from various sources. To attempt any listing except the most fragmentary would be foolhardy,so the following is by no means complete. Of primary importance is Dr. David Duncan for suggesting the problem and having patience with my feeble attempts at a solution. Dr. Wm. J. Hall made many valuable suggestions and criticisms. The National Science Foundation, Office of Naval Research and Air Force Office of Scientific Research were kind enough to grant financial aid at different stages of the preparation for this work. The unpleasant task of typing the manuscript fell upon Mrs. Patricia Ziolkowski and I am grateful for her pleasant execution of this chore. To those who know Martha Jordan I need not explain how helpful she has been in the interpretation of and guidance through administrative regulations; to those who do not know her, words cannot convey the debt.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

It has been generally recognized for many years that the classical treatment of the problem of comparison of means by a test of the hypothesis that all the means are equal does not provide the answers which the experimenter usually desires. As a result much work has been done on a class of problems which are felt to be more applicable. For example, in some situations an experimenter may conduct an experiment to choose from a group of populations the one with the largest mean. Usually, in the interests of not missing the population with the largest mean, he would be willing to choose a group of one or more populations such that any one of them could have the largest mean. This he might term the superior group of populations as distinct from the group of populations not chosen which could be called the inferior group. He would reason that due to experimental error it is quite possible for any of the populations in the superior group to be the population with the largest mean so he would treat these populations equally and either do further experimenting to choose between them or decide on the basis of some other criteria which one population to choose. Thus in this case a statistical problem would be to find a fixed sample size decision procedure for dividing a group of n populations into a superior group, which can contain from one to n of the populations, and an inferior group containing the remaining populations.

It is this decision problem that we study in this paper. We will refer to problems of this type as problems of the largest mean. In other situations an experimenter may conduct an experiment not to determine just the population with the largest mean but to determine the complete ranking of all the population means. We will refer to problems of this type as problems of ranking means.

With the development by Wald [27]* of a general theory of statistical decisions attempts have been made to find decision procedures for problems of the above types in the framework of this theory but it has usually been necessary to use a simplified approach in order to obtain a solution. Mosteller [14] considered a 'slippage' problem where it is desired to find which one, if any, of k populations has a larger value of a location parameter than the other populations. The decision procedure proposed was not claimed to be optimum but only easy and rapid to apply. Paulson [18] considered the same multiple decision problem as considered in this paper, that of dividing a set of populations into a "superior" and an "inferior" group. The decision procedure proposed was only claimed to be quite reasonable on an intuitive basis. In [19] Paulson considered the k-decision problem of finding the 'best' of k populations when k-1 experimental populations are compared with a control population. A decision procedure was obtained which

---

*The numbers in square brackets refer to the bibliography.

had a probability $\geq 1 - \alpha$ of selecting the control population when it is the superior population. He also determined the sample size needed so that the probability of selecting a superior experimental population as best is $\geq 1 - \beta$. In [20] Paulson gave a decision procedure for the slippage problem of Mosteller that had optimum properties. Bahadur [1] considered the same decision problem as this paper and showed for a particular type of loss function what the best decision procedure was in the class of impartial decision procedures. Bechhofer [2] considered the determination of sample size for a single sample experiment so as to make decisions about various rankings of means so that under specified conditions the proportion of correct statements associated with the decision procedure is equal or greater than a preassigned value. Dunnett [8] considered the same problem for the case where the means have prior distributions. In [3] Bechhofer, Dunnett and Sobel gave a two sample decision procedure for ranking means similar in method to Bechhofer's earlier single sample decision procedure. Bechhofer and Sobel [4] considered a sequential decision procedure for the problem of ranking means. Hall [11] formulated the same problems as considered by Bechhofer, Dunnett and Sobel [2],[3], [4] but, instead of placing a guarantee on the probability of making the correct choice, placed a guarantee on the probability of making a distinctly erroneous choice. Seal [24], working with Bose, gave a general class of decision procedures for the same decision problem as this paper and showed that the members of this class possess some desirable properties. Gupta [10], another student of Bose, considered one decision procedure from the class considered by Seal and showed it has some other desirable properties. Decision procedures for the problem of ranking means have also been proposed by Fisher [9], Newman [17], Duncan [5], [6] and Tukey [25], [26], while Scheffé [23] proposed a decision procedure for making comparisons between all contrasts of the means.

In all of the above the approach has either been to find the properties of some proposed decision procedure or to find a decision procedure which has some specified good properties. None of them have given a complete treatment in terms of Wald's theory. However in [7] Duncan formulated the problem of ranking means as a pairwise-multiple comparisons problem in terms of Wald's theory and, for a particular type of loss function and prior distribution of the means, found a Bayes (or minimum average risk) decision procedure. The present paper considers the problem of choosing the population with the largest mean from a group of n normal populations in terms of Wald's theory as a multiple decision problem and, for a particular type of loss function and prior distribution of the means, presents a Bayes decision procedure. A key development is that of generating the multiple decision problem involved as the restricted product (Lehmann [12], [13]) of n simple two-decision problems. In [7] Duncan used a similar technique in dealing with the problem of ranking the means. Also both of these techniques are similar to the 'union-intersection' principle of test construction proposed by Roy [22].

In Chapter II the multiple decision problem to be studied together with the loss functions and prior distributions of the means are given. Then it is shown that Bayes decision procedures for some two-decision 'component' problems generate Bayes decision procedures for the original multiple decision problem. The Bayes decision procedures for these 'component' problems are given only implicitly with the final solution depending on the evaluation of certain integrals.

In Chapter III numerical values are given for this decision procedure for the particular case where there are three normal populations with a common and known variance. Tables are provided to facilitate the use of this decision procedure. For the case where there are two populations the pairwise-multiple comparisons problem as formulated by Duncan [7] and the problem of choosing the largest mean as formulated here are exactly the same so Duncan's results provide the Bayes decision procedure for our problem in this beginning special case. The table of Duncan's results is included. Also it is noted that in contrast to previous approaches to the problem of choosing the largest mean the present approach leads to 'significant differences' that decrease with the number of populations included in the problem.

In Chapter IV we find, by the union-intersection method of test construction proposed by Roy [22], a class of decision procedures for our original multiple decision problem. These decision procedures are not Bayes but are 'conservatively-near' Bayes (in a sense given in Chapter IV). Some interesting relations between these decision procedures and the Bayes decision procedure are given. In particular it is shown that the decreasing-with-n nature of the Bayes decision procedure noted in Chapter III holds in general. Also the ratio of the risk of one particularly simple decision procedure to the risk of the Bayes decision procedure is presented for the case of three normal populations with a common and known variance.

Chapter V presents a study of the variation of the Bayes decision procedure with changes in the prior distribution of the means. This is done for the case of two populations so it is applicable not only to the problem considered in this paper but also to the problem of ranking means as formulated by Duncan [7].

CHAPTER II

THE PROBLEM OF CHOOSING THE LARGEST MEAN

2.1 Statement of the problem.

2.1a The observed data and decision system.

Let $\pi_1, \pi_2, \ldots, \pi_n$ be n normal populations with means $\mu_1, \mu_2, \ldots, \mu_n$ and a common variance σ_o^2, where $\underline{\mu} = (\mu_1, \ldots, \mu_n)$ is unknown and σ_o^2 is either known (case 1) or unknown (case 2). Suppose there are m random observations x_{ij} $(j = 1, \ldots, m)$ from each of the n normal populations π_i $(i = 1, \ldots, n)$. (Let S denote the set of the n μ's, $S = \{\mu_1, \ldots, \mu_n\}$, and $\overline{\{\mu_{j_1}, \mu_{j_2}, \ldots, \mu_{j_k}\}}$ denote the complementary set

$$S - \{\mu_{j_1}, \mu_{j_2}, \ldots, \mu_{j_k}\} .)$$

With the information obtained from the nm random observations we consider the problem of finding a Bayes decision procedure for choosing one of the $p = 2^n - 1$ decisions:

$$d_i : \underline{\mu} \in \omega_i , \quad (i = 1, 2, \ldots, p),$$

the $p = \Sigma\, p_j$ subsets ω_i being defined as follows: the first $p_1 = \binom{n}{1} = n$ subsets are those in which one specified mean exceeds the remaining n-1 means,

$$\omega_1 = \left\{\underline{\mu};\ \mu_1 > \max \overline{\{\mu_1\}}\right\}, \ldots, \omega_p = \left\{\underline{\mu};\ \mu_n > \max \overline{\{\mu_n\}}\right\},$$

the next $p_2 = \binom{n}{2} = n!/2!(n-2)!$ subsets are those in which the larger of two specified means exceeds the remaining n-2 means,

$$\omega_{p_1+1} = \left\{ \underline{\mu};\ \max\{\mu_1,\mu_2\} > \max \overline{\{\mu_1,\mu_2\}} \right\} \quad ,\ldots$$

$$\omega_{p_1+n-1} = \left\{ \underline{\mu};\ \max\{\mu_1,\mu_n\} > \max \overline{\{\mu_1,\mu_n\}} \right\} \quad ,\ldots$$

$$\omega_{p_1+p_2} = \left\{ \underline{\mu};\ \max\{\mu_{n-1},\mu_n\} > \max \overline{\{\mu_{n-1},\mu_n\}} \right\} \quad ,$$

the next $p_3 = \binom{n}{3} = n!/3!(n-3)!$ subsets are those in which the largest of three specified means exceeds the remaining n-3 means,

$$\omega_{p_1+p_2+1} = \left\{ \underline{\mu};\ \max\{\mu_1,\mu_2,\mu_3\} > \max \overline{\{\mu_1,\mu_2,\mu_3\}} \right\} \quad ,\ldots$$

$$\omega_{p_1+p_2+n-2} = \left\{ \underline{\mu};\ \max\{\mu_1,\mu_2,\mu_n\} > \max \overline{\{\mu_1,\mu_2,\mu_n\}} \right\} \ ,$$

$$\omega_{p_1+p_2+n-1} = \left\{ \underline{\mu};\ \max\{\mu_1,\mu_3,\mu_4\} > \max \overline{\{\mu_1,\mu_3,\mu_4\}} \right\} \ ,\ldots$$

$$\omega_{p_1+p_2+p_3} = \left\{ \underline{\mu};\ \max\{\mu_{n-2},\mu_{n-1},\mu_n\} > \max \overline{\{\mu_{n-2},\mu_{n-1},\mu_n\}} \right\} \quad ,$$

. . .

the next to the last $p_{n-1} = \binom{n}{n-1} = n$ subsets are those in which the largest of (n-1) specified means exceeds the remaining mean,

$$\omega_{p_1+p_2+\ldots+p_{n-2}+1} = \left\{ \underline{\mu};\ \max \overline{\{\mu_n\}} > \mu_n \right\} \quad ,\ldots$$

$$\omega_{p_1+p_2+\ldots+p_{n-2}+p_{n-1}} = \left\{ \mu;\ \max \overline{\{\mu_1\}} > \mu_1 \right\} \quad ,$$

the last $p_n = \binom{n}{n} = 1$ subset is the set in which any one of the n means may be the largest of all the means,

$$\omega_p = \Omega \quad \text{(the entire parameter space of } \underline{\mu}\text{)}.$$

These p decisions can be viewed as decisions classifying the n populations into a superior group, which is to contain the population with the largest mean, and an inferior group. (The terms superior and inferior are as used by Paulson [18]). From this point of view the p decisions may be described as:

n decisions specifying exactly one population in the superior group,

$\binom{n}{2} = n!/2!(n-2)!$ decisions specifying exactly two populations in the superior group,

. . .

$\binom{n}{n-1} = n$ decisions specifying exactly n-1 populations in the superior group,

and one decision specifying all n populations in the superior group.

This p-decision system is the same decision system considered previously by Paulson [18], very briefly by Duncan [5, p. 18], Bahadur [1], Seal [24], Gupta [10] and others.

For n = 3, the case considered in more detail in the next chapter, there are the following p = 7 decisions:

$$d_1 : \underline{\mu} \in \omega_1 = \left\{ \underline{\mu};\ \mu_1 > \max\ (\mu_2, \mu_3) \right\} ,$$

that is, π_1 is the superior group,

$$d_2 : \underline{\mu} \in \omega_2 = \left\{ \underline{\mu};\ \mu_2 > \max\ (\mu_1, \mu_3) \right\} ,$$

that is, π_2 is the superior group,

$$d_3 : \underline{\mu} \in \omega_3 = \left\{ \underline{\mu};\ \mu_3 > \max\ (\mu_1, \mu_2) \right\} ,$$

that is, π_3 is the superior group,

$d_4 : \underline{\mu} \in \omega_4 = \{ \underline{\mu}; \max (\mu_1, \mu_2) > \mu_3 \}$,

that is, π_1 and π_2 are the superior group ,

$d_5 : \underline{\mu} \in \omega_5 = \{ \underline{\mu}; \max (\mu_1, \mu_3) > \mu_2 \}$,

that is, π_1 and π_3 are the superior group ,

$d_6 : \underline{\mu} \in \omega_6 = \{ \underline{\mu}; \max (\mu_2, \mu_3) > \mu_1 \}$,

that is, π_2 and π_3 are the superior group ,

$d_7 : \underline{\mu} \in \omega_7 = \Omega$,

that is, π_1, π_2 and π_3 are the superior group.

2.1b <u>Loss functions</u>.

For any one of the p decisions the loss function is defined in the following manner:

let $\mu_{max} = \max (\mu_1, \ldots, \mu_n)$,

let d be the decision that

$\pi_{j_1}, \pi_{j_2}, \ldots, \pi_{j_k}$ are the superior group ,

and $\pi_{j_{k+1}}, \pi_{j_{k+2}}, \ldots, \pi_{j_n}$ are the inferior group,

(where $(j_1, \ldots, j_n)$ is a permutation of $(1, \ldots, n)$),

that is, d is the decision that

$$\underline{\mu} \in \left\{ \underline{\mu}; \max (\mu_{j_1}, \ldots, \mu_{j_k}) > \max(\mu_{j_{k+1}}, \ldots, \mu_{j_n}) \right\},$$

then the loss incurred by d at any point $\underline{\mu}$ is

$$L(d; \underline{\mu}) = \sum_{i=1}^{k} (k_o/\sigma)(\mu_{max} - \mu_{j_i})$$

$$+ \sum_{i=1}^{n-k} (k_1/2\sigma) \left[(\mu_{j_{k+i}} - \max \overline{\{ \mu_{j_{k+i}} \}}) + | \mu_{j_{k+i}} - \max \overline{\{ \mu_{j_{k+i}} \}} | \right].$$

where $k_1 > k_o > 0$ are loss proportionality factors, and $\sigma = \sigma_o/\sqrt{m}$.

That is, the total loss for a given decision is the sum of the losses made with respect to each individual population. The contribution of π_i to the loss is as follows: if π_i is included in the superior group a zero loss is made if $\mu_i = \mu_{max}$ and a positive loss proportional to $(\mu_{max} - \mu_i)$ is made if $\mu_i < \mu_{max}$; if π_i is included in the inferior group a zero loss is made if $\mu_i < \mu_{max}$ and a positive loss proportional to $(\mu_i - \max\{\mu_i\})$ is made if $\mu_i = \mu_{max}$. To illustrate with a simple example: for $n = 3$ suppose decision d_2 is made when $\underline{\mu} = (5, 7, 10)$ where decision d_2 is to state $\underline{\mu} \in \omega_2 = \{\underline{\mu};\ \mu_2 > \max(\mu_1, \mu_3)\}$, that is, π_2 is the superior group. Then the loss is:

$$\begin{aligned} L(d_2;\ 5,7,10) &= (k_o/\sigma)(10 - 7) \\ &+ (k_1/2\sigma)[(5 - 10) + |5 - 10|] \\ &+ (k_1/2\sigma)[(10 - 7) + |10 - 7|] \\ &= 3(k_o/\sigma) + 0 + 3(k_1/\sigma) . \end{aligned}$$

This includes a loss of $3(k_o/\sigma)$ for incorrectly retaining π_2 in the superior group, a zero loss for correctly placing π_1 in the inferior group and a loss of $3(k_1/\sigma)$ for incorrectly placing π_3 in the inferior group.

As will be seen only the ratio $k_1/k_o = k$ is needed to be known and not k_o and k_1 . This ratio will be called the loss ratio or error-seriousness ratio (terms first used by Duncan [7] for a similar ratio). (As seen later we must require that k be not less than three for the case n = 3.)

2.1c Prior distributions.

As the final assumption needed for a Bayes formulation of this p-decision problem the μ's are given independent identical normal prior distributions, each with an unknown mean c' and variance $\gamma^2\sigma^2$ where γ^2 is known. $\sigma^2 = \sigma_0^2/m$ is taken as fixed and hence has no prior distribution.

While it is not difficult to formulate this problem as above, in the framework of the general theory of statistical decisions as developed by Wald, a direct approach to find a Bayes solution is very difficult. Others in attempting to surmount this difficulty have used simplified approaches to find decision procedures for the p-decision system that have some desirable properties. However it will be shown in the following section that with an indirect approach we will be able to find a Bayes decision procedure without the need of any further simplifications.

2.2 The p-decision problem as the restricted product of n two-decision problems.

2.2a Generation of the p decision subsets.

The subsets of the p-decision problem as formulated in the previous section can be generated as the restricted product (a term introduced by Lehmann [12], [13] to be described below) of the subsets of n two-decision component problems.

For this purpose, let us consider a two-decision problem (which will be called the i^{th} component two-decision problem) where the pair of subsets are:

$$\omega_{i0} = \Omega \text{ , the entire parameter space of } \underline{\mu},$$

and

$$\omega_{i1} = \left\{ \underline{\mu};\ \mu_i < \mu_{max} \right\} \text{ , that is, in } \omega_{i1} \text{ some } \mu_j \text{ is larger than } \mu_i \text{ .}$$

The corresponding decisions for this component problem may be denoted by:

$$d_{i0} : \underline{\mu} \in \omega_{i0} \text{ ,}$$

and $d_{i1} : \underline{\mu} \in \omega_{i1}$. Viewed as decisions to classify the populations into a superior group and an inferior group, where initially all the populations are considered as being in the superior group, decision d_{i0} is the decision to leave π_i in the superior group while decision d_{i1} is the decision to place π_i in the inferior group.

For $i = 1,2,\ldots,n$ there are n such two-decision problems. Now if these problems are considered simultaneously, they generate a multiple decision problem where the different decisions are statements about $\underline{\mu}$ being in the intersection of the n component subsets. Thus if $(j_1,\ldots,j_n)$ is a permutation of $(1,\ldots,n)$ and if in the $j_1^{th},\ldots,j_k^{th}$ component problems decisions $d_{j_1 0},\ldots,d_{j_k 0}$ are made and in the $j_{k+1}^{th},\ldots,j_n^{th}$ component problems decisions $d_{j_{k+1} 1},\ldots,d_{j_n 1}$ are made then in the multiple decision problem the decision that corresponds to these component decisions being made simultaneously is

$$d_k : \underline{\mu} \in \omega_k = \bigcap_{i=1}^{k} \omega_{j_i 0} \bigcap_{i=k+1}^{n} \omega_{j_i 1} ,$$

that is,

$$d_k : \underline{\mu} \in \omega_k = \Omega \bigcap_{i=k+1}^{n} \omega_{j_i 1} \text{ (since } \omega_{i0} = \Omega \text{ for all i).}$$

Viewed as a decision to classify the populations into a superior and an inferior group, this decision classifies $\pi_{j_1},\ldots,\pi_{j_k}$ in the superior group and $\pi_{j_{k+1}},\ldots,\pi_{j_n}$ in the inferior group.

Not all of the products of the component subsets form decisions in the p-decision problem. If d_{i1} is the decision made in the i^{th} component problem (i=1,...,n) then considered simultaneously this is a decision to assert

$$\underline{\mu} \in \bigcap_{i=1}^{n} \omega_{i1} .$$

But since $\bigcap \omega_{i1} = 0$ this decision represents an impossibility and is not desired in the product problem. (This decision says for each and every mean that there is another mean which is larger.) Therefore in general a product decision problem is restricted to decisions about nonempty subsets. Lehmann [12], [13] refers to such a generation of a multiple decision problem from a set of two-decision problems as a restricted product decision problem. More specifically the subset system generated by the nonempty intersections of the component subsets is termed the restricted product of the component subset systems. In the case under consideration only one ($\bigcap_{i=1}^{n} \omega_{i1}$) of the component subset intersections is empty. The rest all form elements in the p-decision system.

To illustrate the foregoing, for n=3 there are three component two-decision problems with the following decision pairs:

for the first component problem

$$d_{10} : \underline{\mu} \in \omega_{10} = \Omega ,$$

that is, leave π_1 in the superior group,

$$d_{11} : \underline{\mu} \in \omega_{11} = \left\{ \underline{\mu};\ \mu_1 < \max\ (\mu_2,\mu_3) \right\} ,$$

that is, place π_1 in the inferior group,

for the second component problem

$$d_{20} : \underline{\mu} \in \omega_{20} = \Omega ,$$

that is, leave π_2 in the superior group,

$$d_{21} : \underline{\mu} \in \omega_{21} = \left\{ \underline{\mu};\ \mu_2 < \max\ (\mu_1,\mu_3) \right\} ,$$

that is, place π_2 in the inferior group,

for the third component problem

$$d_{30} : \underline{\mu} \in \omega_{30} = \Omega ,$$

that is, leave π_3 in the superior group,

$$d_{31} : \underline{\mu} \in \omega_{31} = \left\{ \underline{\mu};\ \mu_3 < \max\ (\mu_1,\mu_2) \right\} ,$$

that is, leave π_3 in the inferior group.

The p = 7 decisions in the restricted product decision problem are generated as follows:

$$d_1 = (d_{10}, d_{21}, d_{31}): \underline{\mu} \in \omega_1 = \omega_{10} \cap \omega_{21} \cap \omega_{31},$$

$$d_2 = (d_{11}, d_{20}, d_{31}): \underline{\mu} \in \omega_2 = \omega_{11} \cap \omega_{20} \cap \omega_{31},$$

$$d_3 = (d_{11}, d_{21}, d_{30}): \underline{\mu} \in \omega_3 = \omega_{11} \cap \omega_{21} \cap \omega_{30},$$

$$d_4 = (d_{10}, d_{20}, d_{31}): \underline{\mu} \in \omega_4 = \omega_{10} \cap \omega_{20} \cap \omega_{31},$$

$$d_5 = (d_{10}, d_{21}, d_{30}): \underline{\mu} \in \omega_5 = \omega_{10} \cap \omega_{21} \cap \omega_{30},$$

$$d_6 = (d_{11}, d_{20}, d_{30}): \underline{\mu} \in \omega_6 = \omega_{11} \cap \omega_{20} \cap \omega_{30},$$

$$d_7 = (d_{10}, d_{20}, d_{30}): \underline{\mu} \in \omega_7 = \omega_{10} \cap \omega_{20} \cap \omega_{30},$$

The remaining product decision

$$d_8 = (d_{11}, d_{21}, d_{31}): \underline{\mu} \in \omega_8 = \omega_{11} \cap \omega_{21} \cap \omega_{31} = 0$$

states that $\underline{\mu}$ is in an empty set hence the product decision problem is restricted so as not to include this decision.

With this method of generating the decision system of the original p-decision problem as the restricted product of n two-decision problems and, with the loss functions for these component problems as given below, a considerable reduction is made in the difficulty of the subsequent derivation.

2.2b Generation of the p-decision problem loss functions.

For the i^{th} of the two-decision component problems let loss functions be defined as follows:

$$L(d_{i0}; \underline{\mu}) = 0, \qquad \underline{\mu} \in \Omega - \omega_{i1},$$

$$L(d_{i0}; \underline{\mu}) = (k_o/\sigma)(\mu_{max} - \mu_i), \qquad \underline{\mu} \in \omega_{i1},$$

$$L(d_{i1}; \underline{\mu}) = 0, \qquad \underline{\mu} \in \omega_{i1},$$

$$L(d_{i1}; \underline{\mu}) = (k_1/\sigma)[\mu_i - \max(\mu_1, \ldots, \mu_{i-1}, \mu_{i+1}, \ldots, \mu_n)], \quad \underline{\mu} \in \Omega - \omega_{i1}.$$

Thus for the i^{th} component two-decision problem, a zero loss is made if decision d_{i0} is made when μ_i is the largest mean or if decision d_{i1} is made when μ_i is not the largest mean. A positive loss proportional to the difference between the largest mean and μ_i is made if decision d_{i0} is made when μ_i is not the largest mean and a positive loss proportional to the difference between μ_i and the next largest mean is made if decision d_{i1} is made when μ_i is the largest mean. That is, a zero loss is made if π_i is put in the correct group while a positive loss is made if π_i is left in the superior group when $\mu_i < \mu_{max}$ or if π_i is placed in the inferior group when $\mu_i = \mu_{max}$.

With these loss functions for the component problems it is seen that the loss functions for each of the decisions of the restricted product decision problem is the sum of the loss functions for the decisions of the component two-decision problems that generate that particular decision. That is, if d_i is a decision of the restricted product decision problem and $d_{1i_1},\ldots,d_{ni_n}$ are the n decisions from the n two-decision problems (one from each problem, $i_k = 0$ or 1) such that

$$d_i \text{ is } (d_{1i_1},\ldots,d_{ni_n})$$

than

$$L(d_i;\underline{\mu}) = \sum_{k=1}^{n} L(d_{ki_k};\underline{\mu}), \qquad (i_k = 0 \text{ or } 1).$$

This will be referred to as additive loss.

To illustrate let us refer to the example used previously for the case $n = 3$, with decision d_2:

$\underline{\mu} \in \omega_2 = \left\{\underline{\mu};\ \mu_2 > \max(\mu_1, \mu_3)\right\}$ being made when $\underline{\mu} = (5, 7, 10)$. (The decision d_2 is to leave π_2 in the superior group and place π_1 and π_3 in the inferior group.) As seen before, the loss for this particular example is:

$$L(d_2;\ 5, 7, 10) = 3(k_o/\sigma) + 3(k_1/\sigma).$$

The decision d_2 in the $p = 7$ decision problem results from making the component decisions d_{11}, d_{20}, d_{31} simultaneously. The losses for these component decisions are:

$$L(d_{11};\ \underline{\mu}) = 0$$
$$L(d_{20};\ \underline{\mu}) = (k_o/\sigma)(10 - 7) = 3(k_o/\sigma)$$
$$L(d_{31};\ \underline{\mu}) = (k_1/\sigma)(10 - 7) = 3(k_1/\sigma),$$

which add to the given loss for d_2. For the first component a zero loss is made when π_1 is placed in the inferior group since μ_1 is not the largest mean; for the second component a positive loss of $3(k_o/\sigma)$ is made by leaving π_2 in the superior group since μ_2 is 3 less than the largest mean and for the third component a positive loss of $3(k_1/\sigma)$ is made by not leaving π_3 in the superior group since μ_3 is the largest mean and is 3 larger than the next smaller mean.

2.2c <u>Generation of p-decision procedures as the intersection of n two-decision component procedures</u>.

If for each component problem we have a decision procedure such that considered simultaneously these decision procedures do not lead to any inconsistencies, that is, they do not make the statement that $\underline{\mu}$ is in an empty set, then they form a compatible set of decision procedures.

As proved by Lehmann [12], decision procedures for the restricted product decision problem obtained as the intersection of a family of compatible decision procedures for the component decision problems are in one-to-one correspondence with the generating families of compatible decision procedures for the component decision problems. That is, a set of compatible decision procedures for the component decision problems generates a decision procedure for the restricted product decision problem. And since the decisions of the restricted product problem are the same as the decisions of the original p-decision problem, it follows that a set of compatible decision procedures for the component decision problems generates a decision procedure for the original p-decision problem.

To insure the compatibility of the n two-decision procedures the condition

$$\Pr\left\{d_{p+1} = (d_{11}, d_{21}, \ldots, d_{n1})\right\} = 0$$

must be met. That is, each of the two-decision procedures must be chosen in such a way that when considered simultaneously the probability of placing all the populations in the inferior group is zero. Following Lehmann [12] this will be referred to as the condition of compatibility for the component decision procedures.

Not only does a family of compatible decision procedures for the component two-decision problems generate a decision procedure for the restricted product problem but, with the same prior distribution for the μ's in the component problems as assumed in the original p-decision problem and with additive loss, a family of compatible Bayes decision procedures for the component problems generates a Bayes decision procedure for the restricted product decision problem. This most useful and important result is a simple consequence of the fact that with additive loss the risk is additive and, since expectation is a linear operator, the average risk is additive. Hence if each component of the sum of average risks is minimized then the sum is minimized. Therefore minimum-average-risk decision procedures, that is, Bayes decision procedures for the component two-decision problems generate a minimum-average-risk decision procedure for the restricted product decision problem. (A similar result was noted and used in another problem by Duncan [7] and the general result for any restricted product decision problem was proved by Lehmann [12].)

Thus to obtain a Bayes decision procedure for the p-decision problem we need only to obtain compatible Bayes decision procedures for the n two-decision component problems and to apply these decision procedures simultaneously. It is to be noted that there is complete symmetry among all of the n two-decision problems (as desired, since we are assuming there is no prior reason to distinguish between the means of the various populations). Therefore a Bayes decision procedure for any one of the two-decision problems will yield a Bayes decision procedure for any other, hence for all the others.

Thus the original problem of finding a Bayes decision procedure for a p-decision problem is reduced to that of finding a Bayes decision procedure for one two-decision problem, it being understood that the condition of compatibility must also be satisfied.

2.3 Invariance of decision procedure.

In addition to the requirement that the decision procedure be Bayes with respect to the specified loss functions and prior distributions we require that it be invariant with respect to changes in scale and changes in location.

2.4 Summary of assumptions.

All the preceding assumptions are listed below with comments as to how they could be relaxed.

1. Fixed, equal size (=m) random samples from n independent normal populations with a common variance, σ_0^2.
2. Additive loss, that is, the loss function for one of the product decisions is the sum of the loss functions for its generating decisions.
3. Loss functions for the individual decisions in the two-decision problems are proportional to $(\mu_{max} - \mu_i)$ if d_{i0} is made and proportional to

$$[\mu_i - \max(\mu_1, \ldots, \mu_{i-1}, \mu_{i+1}, \ldots, \mu_n)]$$

if d_{i1} is made. (Proportionality constant is zero if the correct decision is made.)

4. The population means have independent identical normal prior distributions, each with mean c' (unknown) and variance $\gamma^2\sigma^2$ (γ^2 known), ($\sigma^2 = \sigma_o^2/m$).

5. Case 1, σ_o^2 known.

   Case 2, σ_o^2 unknown.

Assumption 1 can be relaxed somewhat as follows. From the random samples the set of sufficient statistics $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ and s_o^2, estimate of σ_o^2 (if σ_o^2 unknown), are obtained. Thus if we let $\underline{z} = (z_1, \ldots, z_{n-1})$ be the vector of n-1 orthogonal comparisons in the Helmert transformation

$$\begin{pmatrix} z_o \\ \underline{z}' \end{pmatrix} = \begin{pmatrix} \underline{k} \\ K \end{pmatrix} \underline{\bar{x}}'$$

of the $\bar{x}$'s, and if $\underline{z}'$ and s_o^2 are a set of sufficient statistics for $\underline{\zeta}' = K\,\underline{\mu}'$ and σ_o^2 then it is necessary to assume only that $\underline{z}'$ has a multivariate normal distribution with mean $\underline{\zeta}' = K\,\underline{\mu}'$ and variance $I\sigma^2$. Also instead of having equal size samples from normal populations with a common variance, it is necessary only that the sample means have a common variance.

With this relaxation of assumption 1 it is necessary to assume in assumption 4 only that the differences of the μ's

$$(\mu_i - \mu_j,\ i \neq j,\ i = 1,2,\ldots,n\ j = 1,2,\ldots,n)$$

have independent identical normal distributions each with mean 0 and variance $2\gamma^2\sigma^2$.

What follows is the same under either the original or the relaxed assumptions. However, for ease of exposition in treating the more essential features of the problem the original simpler assumptions will be used.

2.5 Bayes decision procedure.

As shown in the preceding sections the problem of finding a Bayes decision procedure for the original p-decision problem is reduced to that of finding a Bayes decision procedure for one two-decision problem. For a well-behaved two-decision problem, such as this one, the method of finding the Bayes decision procedure is fairly straight forward. A derivation of this method is next outlined.

Let $\bar{x}_i = \Sigma\, x_{ij}/m$ be the mean of the sample from π_i,

$$(i = 1,2,\dots,n)$$

and s_o^2 = the estimate of σ_o^2. Then $(\bar{x}_1,\dots,\bar{x}_n,\ s_o^2)$ is a sufficient statistic for the problem. From the requirement of invariance under changes in scale and invariance under changes in location attention may be restricted to decision procedures depending on the random observations through the statistic $\underline{y} = \underline{z}/s$ where $\underline{z} = (z_1,\dots,z_{n-1})$ is the vector of n-1 orthogonal comparisons in the Helmert transformation

$$\begin{pmatrix} z_o \\ \underline{z}' \end{pmatrix} = \begin{pmatrix} \underline{k} \\ K \end{pmatrix} \underline{\bar{x}}'$$

of the $\bar{x}$'s and $s = s_o/\sqrt{m}$. (In the Helmert transformation,

$\underline{k} = (1/\sqrt{n} \quad 1/\sqrt{n} \quad . \quad . \quad . \quad 1/\sqrt{n})$

and

$$K = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & \cdots & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & \cdots & 0 \\ \vdots & & & & \\ 1/\sqrt{n(n-1)} & 1/\sqrt{n(n-1)} & 1/\sqrt{n(n-1)} & \cdots & -(n-1)/\sqrt{n(n-1)} \end{bmatrix} .)$$

For the case where σ_o^2 is known $\sigma = \sigma_o/\sqrt{m}$ is used instead of s in the above, that is, the statistic $\underline{y}^* = \underline{z}/\sigma$ is used. (In the following the symbol $\underline{y}$ will be used, with the understanding that it would be replaced by $\underline{y}^*$ for σ_o^2 known.) Also let $\underline{\theta} = \underline{\zeta}/\sigma$ where $\underline{\zeta}$ is the vector of n-1 orthogonal comparisons in the Helmert transformation of the μ's.

Let $f(\underline{y};\ \underline{\theta})$ be the probability density function of $\underline{y}$ given $\underline{\theta}$,

$\xi(\underline{\theta})$ the prior probability density function of $\underline{\theta}$,

$L(d_i;\underline{\theta})$ the loss due to decision d_i when $\underline{\theta}$ is the parameter point,

Y sample space of $\underline{y}$ (or of $\underline{y}^*$ if σ_o^2 known)

and Ω_θ parameter space of $\underline{\theta}$,

(Ω_θ is the image of Ω under the transformation made from the μ's to the θ's).

Then for a two-decision problem,

$$\text{risk} = L(d_0;\ \underline{\theta})\ P(d_0;\ \underline{\theta}) + L(d_1;\ \underline{\theta})\ P(d_1;\ \underline{\theta})$$

where $P(d_i;\ \underline{\theta})$ is the probability of making the decision d_i $(i = 0,\ 1)$ when $\underline{\theta}$ is the parameter point.

Introduce the test function $\alpha(\underline{y})$ where,

$$\alpha(\underline{y}) = \begin{cases} 0, \text{ for } d_0 \\ 1, \text{ for } d_1 \ . \end{cases}$$

Then $P(d_0;\ \underline{\theta}) = E_{\underline{y}}(1 - \alpha;\ \underline{\theta}) = \int_Y (1 - \alpha)\ f(\underline{y};\ \underline{\theta})\ d\underline{y}$

and $P(d_1;\ \underline{\theta}) = E_{\underline{y}}(\alpha;\ \underline{\theta}) = \int_Y \alpha\ f(\underline{y};\ \underline{\theta})\ d\underline{y}$. Therefore,

$$\begin{aligned} \text{risk} &= L(d_0;\ \underline{\theta})\ E_{\underline{y}}(1 - \alpha;\ \underline{\theta}) + L(d_1;\ \underline{\theta})\ E_{\underline{y}}(\alpha;\ \underline{\theta}) \\ &= \int_Y (1 - \alpha)\ L(d_0;\ \underline{\theta})\ f(\underline{y};\ \underline{\theta})\ d\underline{y} \\ &\quad + \int_Y \alpha\ L(d_1;\ \underline{\theta})\ f(\underline{y};\ \underline{\theta})\ d\underline{y}\ . \end{aligned}$$

$$\begin{aligned} \text{Average risk} &= \int_{\Omega_\theta} \text{risk}\ \xi(\underline{\theta})\ d\underline{\theta} \\ &= \int_{\Omega_\theta} \left[\int_Y ((1 - \alpha)\ L(d_0;\underline{\theta}) + \alpha\ L(d_1;\underline{\theta}))\ f(\underline{y};\underline{\theta})d\underline{y} \right] \xi(\underline{\theta})d\underline{\theta}. \end{aligned}$$

Then if the order of integration can be changed (as can be done in the cases herein considered) we obtain:

average risk=

$$\int_Y \left[\int_{\Omega_\Theta} ((1-\alpha)\, L(d_0;\underline{\theta}) + \alpha\, L(d_1;\underline{\theta}))\, f(\underline{y};\underline{\theta})\, \xi(\underline{\theta}) d\theta \right] d\underline{y}$$

$$= \int_Y \left[\int_{\Omega_\Theta} L(d_0;\underline{\theta})\, f(\underline{y};\underline{\theta})\, \xi(\underline{\theta})\, d\underline{\theta} \right] d\underline{y}$$

$$+ \int_Y \alpha \left[\int_{\Omega_\Theta} (L(d_1;\underline{\theta}) - L(d_0;\underline{\theta}))\, f(\underline{y};\underline{\theta})\, \xi(\underline{\theta})\, d\underline{\theta} \right] d\underline{y}\ .$$

Let

$$g(\underline{y}) = \int_{\Omega_\Theta} (L(d_1;\ \underline{\theta}) - L(d_0;\ \underline{\theta}))\, f(\underline{y};\ \underline{\theta})\, \xi(\underline{\theta})\, d\underline{\theta}\ ,$$

then the average risk is minimized by the test:

$$\alpha = 1 \text{ for } g(\underline{y}) < 0\ ,$$
$$= 0 \text{ for } g(\underline{y}) > 0\ ,$$
$$\text{arbitrary for } g(\underline{y}) = 0\ .$$

To remove the arbitrariness let $\alpha = 0$ where $g(\underline{y}) = 0$. That is, made decision d_0 if $g(\underline{y}) \geq 0$ and make decision d_1 if $g(\underline{y}) < 0$.

Thus for, say, the i^{th} component problem,

$$d_{i0} : \underline{\mu} \in \omega_{i0}\ ,\ (\underline{\theta} \in \omega_{i0}\ (\underline{\theta})\ )$$

$$d_{i1} : \underline{\mu} \in \omega_{i1}\ ,\ (\underline{\theta} \in \omega_{i1}\ (\underline{\theta})\ ),$$

we have

$$g_i(\underline{y}) = \int_{\Omega_\Theta} \left| L(d_{i1};\ \underline{\theta}) - L(d_{i0};\ \underline{\theta}) \right| f(\underline{y};\ \underline{\theta})\, \xi(\underline{\theta})\, d\underline{\theta}$$

$$= \int_{\Omega^*_\Theta} \left| L(d_{i1};\ \underline{\theta}) - L(d_{i0};\ \underline{\theta}) \right| f(\underline{y};\ \underline{\theta})\, \xi(\underline{\theta})\, d\underline{\theta}$$

where $\Omega_\Theta^* = \Omega_\Theta - \{$points where two or more of the Θ's are equal$\}$ (this does not affect the results since the ξ-measure of the omitted points is zero). Therefore

$$g_i(\underline{y}) = -(k_0/\sigma) \sum_{\substack{j=1 \\ j\neq i}}^{n} \int_A [\mu_j(\underline{\Theta}) - \mu_i(\underline{\Theta})]\, f(\underline{y};\, \underline{\Theta})\, \xi(\underline{\Theta})\, d\underline{\Theta}$$

$$+(k_1/\sigma) \sum_{\substack{j=1 \\ j\neq i}}^{n} \int_B [\mu_i(\underline{\Theta}) - \mu_j(\underline{\Theta})]\, f(\underline{y};\, \underline{\Theta})\, \xi(\underline{\Theta})\, d\underline{\Theta}$$

(where $\mu_i(\underline{\Theta})$ and $\mu_j(\underline{\Theta})$ are functions of $\underline{\Theta}$ given by the inverse of the Helmert transformation from the μ's to the Θ's,

$$A = \Omega_\Theta^* \cap \{\underline{\Theta};\ \mu_j > \max\ \{\overline{\mu_j}\}\}$$ is the set of points in Ω_Θ^* where $\mu_j = \mu_{max}$ and

$$B = \Omega_\Theta^* \cap \{\underline{\Theta};\ \mu_i > \mu_j > \max\ \{\overline{\mu_i,\ \mu_j}\}\}$$ is the set of points in Ω_Θ^* where $\mu_i = \mu_{max}$ and $\mu_j = \max(\mu_1, \ldots, \mu_{i-1}, \mu_{i+1}, \ldots, \mu_n)$).

With the function g_i the Bayes decision procedure for the i^{th} component problem is given by the following: make decision d_{i0} if $g_i(\underline{y}) \geq 0$ and make decision d_{i1} if $g_i(\underline{y}) < 0$. The Bayes decision procedure for the other component problems follows from this one, as noted previously, due to the uniform treatment of the μ's. Finally, if the condition for compatibility of these component decision procedures is satisfied then the Bayes decision procedure for the p-decision problem is the product of these n two-decision procedures.

To carry out the necessary integration to permit practical use of the decision procedure presents some difficulties. Due to the non-elementary nature of these integrals this integration must be done by numerical methods, however, for the case n=3 and σ_o^2 known, the tables [15] and [16] have been used to evaluate the integrals, and the solution for this case is given in Chapter III. Cases where $n > 3$ or $n = 3$ and σ_o^2 unknown must wait on some numerical integration work for the exact procedures. However in Chapter IV a decision procedure is given that is 'conservative' and 'near' to the Bayes decision procedure. For this decision procedure results are presented that can be applied to practical situations for all n and σ_o^2 known and unknown.

CHAPTER III

BAYES DECISION PROCEDURE FOR THE CASE $n = 3$, σ_o^2 KNOWN

As in the last chapter

let

$$g_i(\underline{y}) = \int_{\Omega_\theta^*} [L(d_{i1}; \underline{\theta}) - L(d_{i0}; \underline{\theta})] f(\underline{y}, \underline{\theta}) \xi(\underline{\theta}) d\underline{\theta}$$

where $\underline{y} = \underline{y}* = \underline{z}/\sigma$ since σ_o^2 is known.

As shown in the last chapter $g_i(\underline{y})$ determines the Bayes decision procedure for any of the three (=n) two-decision problems and, due to the symmetry of these component problems, it is necessary to find the Bayes decision procedure for only one. Let us then consider the third component problem with the two decisions

$$d_{30} : \underline{\mu} \in \omega_{30} \quad (\underline{\theta} \in \omega_{30}(\underline{\theta}))$$

$$d_{31} : \underline{\mu} \in \omega_{31} \quad (\underline{\theta} \in \omega_{31}(\underline{\theta})) .$$

For this component problem the loss functions as previously defined are:

$$L(d_{30}; \underline{\mu}) = \begin{cases} 0 & , \mu_3 = \mu_{max} \\ (k_o/\sigma)(\mu_1-\mu_3), & \mu_1 \geq \mu_2 \text{ and } \mu_1 > \mu_3 \\ (k_o/\sigma)(\mu_2-\mu_3), & \mu_2 \geq \mu_1 \text{ and } \mu_2 > \mu_3 \end{cases}$$

$$L(d_{31};\ \underline{\mu}) = \begin{cases} 0 & , \mu_1 > \mu_3 \text{ or } \mu_2 > \mu_3 \\ (k_1/\sigma)(\mu_3-\mu_1), & \mu_3 \geq \mu_1 \geq \mu_2 \\ (k_1/\sigma)(\mu_3-\mu_2), & \mu_3 \geq \mu_2 \geq \mu_1 \ . \end{cases}$$

In terms of $\underline{\theta} = (\theta_1, \theta_2)$, where $\theta_1 = (\mu_1 - \mu_2)/\sqrt{2}\sigma$ and $\theta_2 = (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{6}\sigma$, these loss functions are:

$$L(d_{30};\ \underline{\theta}) = \begin{cases} 0 & , \theta_2 < \theta_1/\sqrt{3} \text{ and } \theta_2 < -\theta_1/\sqrt{3} \\ k_o(\theta_1 + \sqrt{3}\theta_2)/\sqrt{2}, & \theta_1 \geq 0 \text{ and } \theta_2 > -\theta_1/\sqrt{3} \\ k_o(\sqrt{3}\theta_2 - \theta_1)/\sqrt{2}, & \theta_1 \leq 0 \text{ and } \theta_2 > \theta_1/\sqrt{3} \end{cases}$$

$$L(d_{31};\ \underline{\theta}) = \begin{cases} 0 & , \theta_2 > \theta_1/\sqrt{3} \text{ or } \theta_2 > -\theta_1/\sqrt{3} \\ -k_1(\theta_1 + \sqrt{3}\theta_2)/\sqrt{2}, & \theta_1 \geq 0 \text{ and } \theta_2 \leq -\theta_1/\sqrt{3} \\ -k_1(\sqrt{3}\theta_2 - \theta_1)/\sqrt{2}, & \theta_1 \leq 0 \text{ and } \theta_2 \leq \theta_1/\sqrt{3} \ . \end{cases}$$

Thus

$$g_3(\underline{y}) = -\int_0^\infty d\theta_1 \int_{-\theta_1/\sqrt{3}}^\infty [k_o(\theta_1 + \sqrt{3}\theta_2)/\sqrt{2}]\ f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})\ d\theta_2$$

$$-\int_{-\infty}^0 d\theta_1 \int_{\theta_1/\sqrt{3}}^\infty [k_o(\sqrt{3}\theta_2 - \theta_1)/\sqrt{2}]\ f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})\ d\theta_2$$

$$-\int_{-\infty}^0 d\theta_1 \int_{-\infty}^{\theta_1/\sqrt{3}} [k_1(\sqrt{3}\theta_2 - \theta_1)/\sqrt{2}]\ f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})\ d\theta_2$$

$$-\int_0^\infty d\theta_1 \int_{-\infty}^{-\theta_1/\sqrt{3}} [k_1(\theta_1 + \sqrt{3}\theta_2)/\sqrt{2}]\ f(\underline{y};\underline{\theta})\ \xi(\underline{\theta})\ d\theta_2$$

where $f(\underline{y};\ \underline{\theta}) = (2\pi)^{-1} \exp\left\{-1/2[(y_1-\theta_1)^2+(y_2-\theta_2)^2]\right\}$

and $\xi(\underline{\theta}) = (2\pi\gamma^2)^{-1} \exp\left\{-(2\gamma^2)^{-1}(\theta_1^2+\theta_2^2)\right\}$.

Since it is desired to determine the points (y_1, y_2) where $g_3(y_1, y_2) \leq 0$ positive proportionality terms that will cancel out and not affect the final result can be discarded in this determination.

By completing the squares of θ_1 and θ_2 in the exponent of $f(\underline{y};\ \theta)\ \xi(\underline{\theta})$ we have (where $\beta^2 = (\gamma^2+1)/\gamma^2$):

$$f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta}) = (2\pi\gamma)^{-2} \exp\left\{(-\beta^2/2)[(\theta_1-y_1/\beta^2)^2+(\theta_2-y_2/\beta^2)^2\right\}$$

$$\exp\left\{-[2(1+\gamma^2)]^{-1}(y_1^2+y_2^2)\right\} .$$

Hence omitting the positive factor

$$(k_o/\sqrt{2})(2\pi\gamma\beta^2)^{-1} \exp\left\{-[2(1+\gamma^2)]^{-1}[y_1^2+y_2^2]\right\}$$

which does not affect the sign of g_3, we have:

$$g_3 \propto -I_1-I_2-kI_3-kI_4$$

where $k = k_1/k_o$,

$$I_1 = \int_0^{\infty} d\theta_1 \int_{-\theta_1/\sqrt{3}}^{\infty} (\theta_1+\sqrt{3}\theta_2) \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2,\ \beta^{-1})\ d\theta_2 ,$$

$$I_2 = \int_{-\infty}^{0} d\theta_1 \int_{\theta_1/\sqrt{3}}^{\infty} (\sqrt{3}\theta_2-\theta_1) \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2,\ \beta^{-1})\ d\theta_2 ,$$

$$I_3 = \int_{-\infty}^{0} d\theta_1 \int_{-\infty}^{\theta_1/\sqrt{3}} (\sqrt{3}\theta_2 - \theta_1) \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2, \beta^{-1})\, d\theta_2,$$

$$I_4 = \int_{0}^{\infty} d\theta_1 \int_{-\infty}^{-\theta_1/\sqrt{3}} (\theta_1 + \sqrt{3}\theta_2) \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2, \beta^{-1}) d\theta_2,$$

and $\phi(\theta_i;\ y_i/\beta^2, \beta^{-1}) = (2\pi)^{-1/2}\beta \exp\left\{-(\beta^2/2)(\theta_i - y_i/\beta^2)^2\right\}$ $(i=1,2)$.

Consider $I_1 + kI_4$.

$$I_1 + kI_4 = k\int_{0}^{\infty} d\theta_1 \int_{-\infty}^{\infty} (\theta_1 + \sqrt{3}\theta_2) \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2, \beta^{-1}) d\theta_2$$

$$-(k-1)\int_{0}^{\infty} d\theta_1 \int_{-\theta_1/\sqrt{3}}^{\infty} (\theta_1 + \sqrt{3}\theta_2) \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2, \beta^{-1}) d\theta_2$$

$$= kI_5 - (k-1)I_6, \text{ defining } I_5 \text{ and } I_6 .$$

Subtract and add the factor

$$(y_1/\beta^2 + \sqrt{3}y_2/\beta^2) \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2, \beta^{-1})$$

to the integrands in I_5 and I_6 to obtain:

$$I_5 = \int_{0}^{\infty} (\theta_1 - y_1/\beta^2) d\theta_1 \int_{-\infty}^{\infty} \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2, \beta^{-1}) d\theta_2$$

$$+\sqrt{3}\int_{0}^{\infty} d\theta_1 \int_{-\infty}^{\infty} (\theta_2 - y_2/\beta^2) \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2, \beta^{-1}) d\theta_2$$

$$+[(y_1 + \sqrt{3}y_2)/\beta^2] \int_{0}^{\infty} d\theta_1 \int_{-\infty}^{\infty} \prod_{i=1}^{2} \phi(\theta_i;\ y_i/\beta^2, \beta^{-1}) d\theta_2$$

I_6 = similar expression except lower limit of integration for θ_2 is $-\theta_1/\sqrt{3}$.

In I_5 and I_6, let $v = \beta(\theta_1 - y_1/\beta^2)$ and $w = \beta(\theta_2 - y_2/\beta^2)$ then

$$I_5 = \beta^{-1}\int_{-y_1/\beta}^{\infty} v\,\phi(v)\,dv\int_{-\infty}^{\infty}\phi(w)\,dw + \sqrt{3}\beta^{-1}\int_{-y_1/\beta}^{\infty}\phi(v)dv\int_{-\infty}^{\infty}w\phi(w)\,dw$$

$$+ (y_1+\sqrt{3}y_2)\beta^{-2}\int_{-y_1/\beta}^{\infty}\phi(v)dv\int_{-\infty}^{\infty}\phi(w)\,dw,$$

where $\phi(v) = \phi(v;0,1)$ and $\phi(w) = \phi(w;0,1)$,

I_6 = similar expression except lower limit of integration for w in each term is $(-v/\sqrt{3})-(y_1+\sqrt{3}y_2)/(\sqrt{3}\beta)$

$$= \beta^{-1}I_7 + \sqrt{3}\beta^{-1}I_8 + (y_1+\sqrt{3}y_2)\beta^{-2}I_9,$$ defining I_7, I_8 and I_9.

But $\int_{-\infty}^{\infty}\phi(w)dw = 1$, and $\int_{-\infty}^{\infty}w\phi(w)dw = 0$. Therefore

$$I_5 = \beta^{-1}\int_{-y_1/\beta}^{\infty}v\phi(v)dv + (y_1+\sqrt{3}y_2)\beta^{-2}\int_{-y_1/\beta}^{\infty}\phi(v)dv$$

$$= \beta^{-1}\phi(-y_1/\beta) + (y_1+\sqrt{3}y_2)\beta^{-2}[1 - \Phi(-y_1/\beta)]$$

$$= \beta^{-1}\phi(y_1/\beta) + (y_1+\sqrt{3}y_2)\,\beta^{-2}\,\Phi(y_1/\beta),$$

since $\phi(-a) = \phi(a) = (2\pi)^{-1/2}\exp(-2^{-1}a^2)$,

$\int_a^{\infty} v\,\phi(v)\,dv = \phi(a)$ and $1 - \Phi(-a) = \Phi(a)$, where $\Phi(a) = \int_{-\infty}^{a}\phi(x)dx$.

$$I_7 = \int_{-y_1/\beta}^{\infty} v\phi(v)dv \int_{(-v/\sqrt{3})-(y_1+\sqrt{3}y_2)/(\sqrt{3}\beta)}^{\infty} \phi\,(w)dw \quad .$$

Integrate by parts by letting

$$dr = v\ \phi(v)dv$$

and

$$s = \int_{(-v/\sqrt{3})-(y_1+\sqrt{3}y_2)/(\sqrt{3}\beta);}^{\infty} \phi(w)dw = 1 - \Phi\left\{(-v/\sqrt{3}) - (y_1+\sqrt{3}y_2)/(\sqrt{3}\beta)\right\} \quad ;$$

then $$I_7 = -\ \phi(v)\ [1 - \Phi\left\{(-v/\sqrt{3}) - (y_1+\sqrt{3}y_2)/(\sqrt{3}\beta)\right\}] \Big|_{v=-y_1/\beta}^{v=\infty}$$

$$+(\sqrt{3})^{-1}\int_{-y_1/\beta}^{\infty} \phi(v)\ \phi\left\{(-v/\sqrt{3})-(y_1+\sqrt{3}y_2)/(\sqrt{3}\beta)\right\} dv.$$

But $\phi(v)\ \phi\left\{(-v/\sqrt{3})-(y_1+\sqrt{3}y_2)/(\sqrt{3}\beta\ )\right\}$

$$= (2\pi)^{-1} \exp\left\{\ -(1/2)[v^2+ \left\{(-v/\sqrt{3})-(y_1+\sqrt{3}\ y_2)/\sqrt{3}\beta\right\}^2]\right\}$$

$$= (2\pi)^{-1} \exp\left\{\ -(1/2)[4/3\ \left\{v+(y_1+\sqrt{3}y_2)/(4\beta)\right\}^2 + \left\{(y_1+\sqrt{3}y_2)/(2\beta)\right\}^2]\right\}.$$

Therefore

$$I_7 = \phi(y_1/\beta)[1 - \Phi\left\{y_1/(\sqrt{3}\beta)-(y_1+\sqrt{3}y_2)/(\sqrt{3}\beta)\right\}]$$

$$+ (\sqrt{3}\ )^{-1}\phi\left\{(y_1+\sqrt{3}y_2)/2\beta\right\}\ (\sqrt{3}/2)$$

$$\int_{-y_1/\beta}^{\infty} \phi(v;\ (-y_1-\sqrt{3}y_2)/4\beta\ ,\ \sqrt{3}/2)dv.$$

Let $u = (2/\sqrt{3})\left\{v+(y_1+\sqrt{3}y_2)/4\beta\right\}$ and the integral in this expression becomes

$$\int_{(y_2-\sqrt{3}y_1)/2\beta}^{\infty} \phi(u)\, du$$

so $$I_7 = \phi(y_1/\beta)\, \Phi\, (y_2/\beta)$$

$$+ (1/2)\ \phi\left\{(y_1+\sqrt{3}y_2)/2\beta\right\} \Phi\left\{(\sqrt{3}y_1 - y_2)/2\beta\right\} .$$

$$I_8 = \int_{-y_1/\beta}^{\infty} \phi(v)\, dv \int_{(-v/\sqrt{3})-(y_1+\sqrt{3}y_2)/\sqrt{3}\beta}^{\infty} w\, \phi(w)\, dw$$

$$= \int_{-y_1/\beta}^{\infty} \phi(v)\ \phi\left\{(-v/\sqrt{3})-(y_1+\sqrt{3}\ y_2)/(\sqrt{3}\beta)\right\} dv,$$

but this is $\sqrt{3}$ times the second integral in I_7. Therefore

$$I_8 = (\sqrt{3}/2)\ \phi\left\{(y_1+\sqrt{3}y_2)/2\beta\right\}\ \Phi\left\{(\sqrt{3}y_1 - y_2)/2\beta\right\} .$$

$$I_9 = \int_{-y_1/\beta}^{\infty} \phi(v)dv \int_{(-v/\sqrt{3})-(y_1+\sqrt{3}y_2)/\sqrt{3}\beta}^{\infty} \phi(w)dw \quad .$$

Let $v = v$ and $u = (v+\sqrt{3}w)/2$;

then

$$I_9 = \int_{-y_1/\beta}^{\infty}\left(\int_{-(y_1+\sqrt{3}y_2)/2\beta}^{\infty} (2\pi)^{-1}(\sqrt{3}/2)^{-1} \exp\left\{-(2/3)(v^2-vu+u^2)\right\} du\right) dv.$$

But this is just the integral of a bivariate normal distribution and can be found using the tables [16]. Using the notation of these tables

$$I_9 = L(-y_1/\beta,\ -(y_1+\sqrt{3}y_2)/2\beta,\ 1/2)\ .$$

Collecting and combining the preceding results, we have

$$\begin{aligned}I_1+kI_4 &= k\beta^{-1}\,\phi(y_1/\beta) + k\beta^{-2}(y_1+\sqrt{3}y_2)\,\Phi(y_1/\beta)\\ &\quad -(k-1)\beta^{-1}\,\phi(y_1/\beta)\,\Phi\,(y_2/\beta)\\ &\quad -2(k-1)\beta^{-1}\phi\{(y_1+\sqrt{3}y_2)/2\beta\}\,\Phi\,\{(\sqrt{3}y_1-y_2)/2\beta\}\\ &\quad -(k-1)\beta^{-2}(y_1+\sqrt{3}y_2)\,L(-y_1/\beta,\ -(y_1+\sqrt{3}y_2)/2\beta, 1/2)\ .\end{aligned}$$

For I_2+kI_3 we have

$$\begin{aligned}I_2+kI_3 &= k\int_{-\infty}^{0} d\theta_1 \int_{-\infty}^{\infty}(\sqrt{3}\theta_2-\theta_1)\prod_{i=1}^{2}\phi(\theta_i;\ y_i/\beta^2,\beta^{-1})d\theta_2\\ &\quad -(k-1)\int_{-\infty}^{0} d\theta_1 \int_{\theta_1/\sqrt{3}}^{\infty}(\sqrt{3}\theta_2-\theta_1)\prod_{i=1}^{2}\phi(\theta_i;\ y_i/\beta^2,\beta^{-1})d\theta_2.\end{aligned}$$

Let $\theta_1' = -\theta_1$ and $\theta_2' = \theta_2$, then

$$\begin{aligned}I_2+kI_3 &= k\int_{\infty}^{0}(-d\theta_1')\int_{-\infty}^{\infty}(\theta_1'+\sqrt{3}\theta_2')\phi(\theta_1';-y_1/\beta^2,\beta^{-1})\phi(\theta_2';y_2/\beta^2,\beta^{-1})d\theta_2'\\ &\quad -(k-1)\int_{\infty}^{0}(-d\theta_1')\int_{-\theta_1'/\sqrt{3}}^{\infty}(\theta_1'+\sqrt{3}\theta_2')\ \phi(\theta_1';-y_1/\beta^2,\beta^{-1})\phi(\theta_2';y_2/\beta^2,\beta^{-1})d\theta_2'.\end{aligned}$$

Then interchanging the limits of integration for θ_1' and dropping the primes,

$$I_2+kI_3 = k\int_0^{\infty}d\Theta_1\int_{-\infty}^{\infty}(\Theta_1+\sqrt{3}\Theta_2)\ \phi(\Theta_1;-y_1/\beta^2,\beta^{-1})\phi(\Theta_2;y_2/\beta^2,\beta^{-1})d\Theta_2$$

$$-(k-1)\int_0^{\infty}d\Theta_1\int_{-\Theta_1/\sqrt{3}}^{\infty}(\Theta_1+\sqrt{3}\Theta_2)\phi(\Theta_1;-y_1/\beta^2,\beta^{-1})\phi(\Theta_2;y_2/\beta^2,\beta^{-1})d\Theta_2,$$

but this is the same as I_1+kI_4 except y_1/β^2 in I_1+kI_4 has been replaced by $-y_1/\beta^2$. Therefore using the results for I_1+kI_4,

$$I_2+kI_3 = k\beta^{-1}\ \phi(y_1/\beta) + k\beta^{-2}(\sqrt{3}y_2-y_1)[1-\Phi(y_1/\beta)]$$

$$-(k-1)\beta^{-1}\ \phi(y_1/\beta)\ \Phi(y_2/\beta)$$

$$-2(k-1)\beta^{-1}\ \phi\left\{(y_1-\sqrt{3}y_2)/2\beta\right\}\Phi\left\{-(\sqrt{3}y_1+y_2)/2\beta\right\}$$

$$+(k-1)\beta^{-2}(y_1-\sqrt{3}y_2)\ L(y_1/\beta,\ (y_1-\sqrt{3}y_2)/2\beta,\ 1/2).$$

But $g\ (y_1,\ y_2) \propto -(I_1+kI_4) - (I_2+kI_3)$,

hence by combining terms, letting $r = y_1/\beta$ and $s = y_2/\beta$ and omitting the positive factor $(k-1)/\beta$, we obtain

$$\begin{aligned} g_3(y_1,\ y_2) \propto\ & -2k(k-1)^{-1}\ \phi(r) + k(k-1)^{-1}\ (r-\sqrt{3}s) \\ & - 2k(k-1)^{-1}\ r\Phi(r) + 2\phi(r)\ \Phi\ (s) \\ & + 2\phi\left\{(r+\sqrt{3}s)/2\right\}\ \Phi\ \left\{(\sqrt{3}r-s)/2\right\} \\ & + 2\phi\ \left\{(r-\sqrt{3}s)/2\right\}\ \Phi\ \left\{-(\sqrt{3}r+s)/2\right\} \\ & + (r+\sqrt{3}s)\ L(-r,\ -(r+\sqrt{3}s)/2,\ 1/2) \\ & - (r-\sqrt{3}s)\ L(r,\ (r-\sqrt{3}s)/2,\ 1/2). \end{aligned}$$

Let $h(r, s)$ be the expression on the right. Thus finally we have

$$g_3(y_1, y_2) \propto h(r, s)$$

where positive proportionality terms have been omitted and $r = y_1/\beta$, $s = y_2/\beta$. Therefore

$$g_3(y_1, y_2) \geq 0$$

if and only if

$$h(r, s) \geq 0 .$$

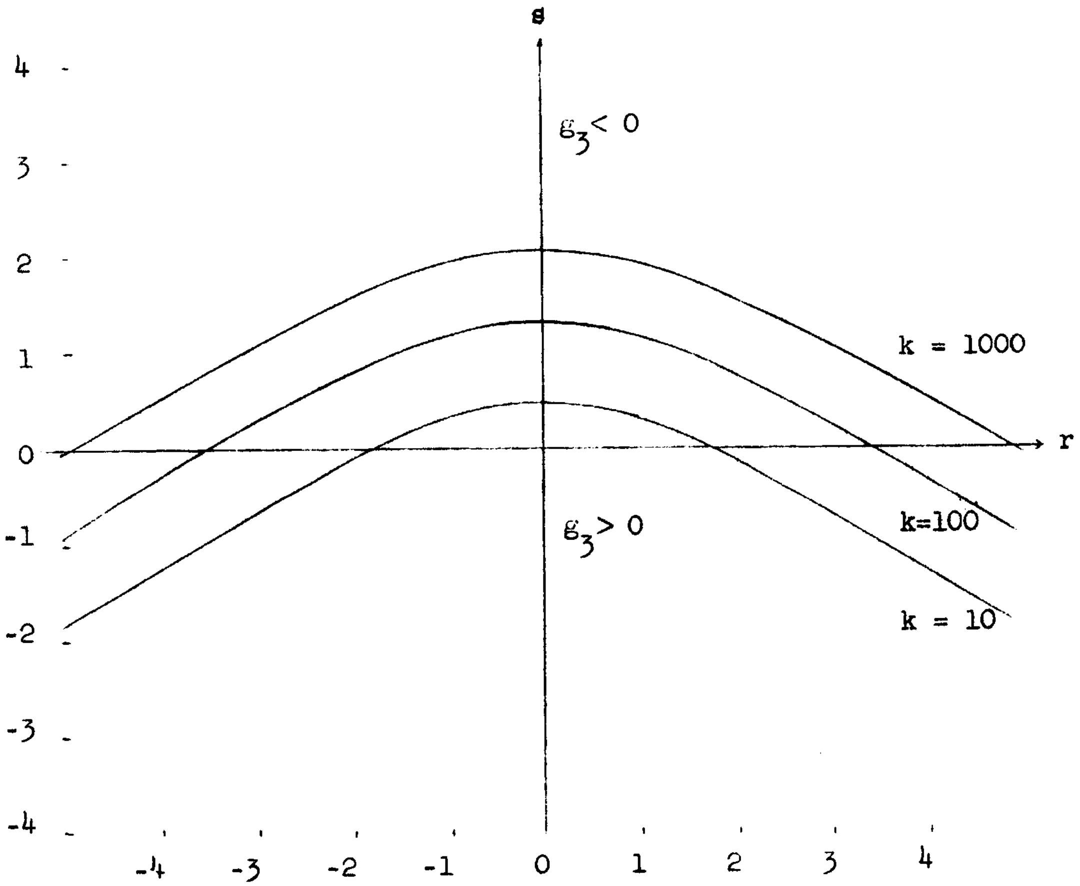
Since h is continuous in r and s, it follows that the set of points where $h = 0$ is the boundary between the set of points where $h < 0$ and the set where $h > 0$, that is, this boundary divides the (r, s) space into two regions, the region where $h < 0$ (that is, where decision d_{i1} is made) and the region where $h \geq 0$ (that is, where decision d_{i0} is made). As it turns out, each of these regions is a connected set so the critical region (the retion where decision d_{i1} is made) is determined by these boundary points together with the determination of the sign of h at one other point. This boundary was determined by an iterative procedure with the aid of the tables [15] and [16], and the solution was found for k=10, 100, 1000. These boundaries in the (r, s) space are given in Figure 1,(page 39). The regions in the (r, s) space were then mapped onto the (y_1, y_2) space for $\gamma^2 = 0.5$, 1, 3, and ∞. (Since $\beta^2 = (\gamma^2 + 1)/\gamma^2$, $\beta = 1$ when $\gamma^2 = \infty$ so for other values of γ^2 the regions in the (y_1, y_2) space can be easily obtained from the regions for the case $\gamma^2 = \infty$. Thus Figure 1 gives the boundaries for k=10, 100, 1000 and $\gamma^2 = \infty$ in the (y_1, y_2) space.)

Figure 1

Bayes decision procedure critical regions

($n = 3$, σ_0^2 known)

(third component problem)



Viewed as above, the decision procedure is: If the sample means are such that $(y_1 = (\bar{x}_1 - \bar{x}_2)/\sqrt{2}\sigma,\ y_2 = (\bar{x}_1 + \bar{x}_2 - 2\bar{x}_3)/\sqrt{6}\sigma)$ is a point in the region where $g_3 < 0$, then make decision d_{31}; that is, place π_3 in the inferior group. Otherwise make decision d_{30}; that is, leave π_3 in the superior group. However, it is easy to see that an equivalent way of viewing this decision procedure is: make decision d_{31} if $\bar{x}_3/\sqrt{2}\sigma < \max(\bar{x}_1/\sqrt{2}\sigma, \bar{x}_2/\sqrt{2}\sigma) - t$, where t is a number determined by k, γ^2 and $\Delta = |(\bar{x}_1/\sqrt{2}\sigma) - (\bar{x}_2/\sqrt{2}\sigma)|$; otherwise make decision d_{30}. (In (y_1, y_2) space t is the distance from the boundary where $g_3 = 0$, to the line $\bar{x}_3 = \bar{x}_1$ if $\bar{x}_1 \geq \bar{x}_2$, or to the line $\bar{x}_3 = \bar{x}_2$ if $\bar{x}_2 \geq \bar{x}_1$.) In the following this formulation of the test procedure will be used since it is more useful for practical purposes and is easier to generalize to cases where $n > 3$.

Table 1 (page 45) presents the values of t for this decision procedure for k=10, 100, 1000, $\gamma^2 = 0.5, 1, 3, \infty$ and various values of $\Delta = |(\bar{x}_1/\sqrt{2}\sigma) - (\bar{x}_2/\sqrt{2}\sigma)|$.

Thus we have the decision procedure for the third component problem. The decision procedure for the first component problem is obtained from that for the third by interchanging $\bar{x}_1$ and $\bar{x}_3$ in the decision procedure, that is, for the first component problem make decision d_{11} if $\bar{x}_1/\sqrt{2}\sigma < \max(\bar{x}_2/\sqrt{2}\sigma, \bar{x}_3/\sqrt{2}\sigma) - t$, otherwise make decision d_{10}, where t is now a function of k, γ^2 and $\Delta = |(\bar{x}_2/\sqrt{2}\sigma) - (\bar{x}_3/\sqrt{2}\sigma)|$. Similarly for the second component problem, make decision d_{21} if $\bar{x}_2/\sqrt{2}\sigma < \max(\bar{x}_1/\sqrt{2}\sigma, \bar{x}_3/\sqrt{2}\sigma) - t$, otherwise make decision d_{20}, where t is a function of k, γ^2 and $\Delta = |(\bar{x}_1/\sqrt{2}\sigma) - (\bar{x}_3/\sqrt{2}\sigma)|$. It is clear that these component decision procedures are compatible if and only if t is not less than zero and it is easy to determine that $t \geq 0$ for $k \geq 3$ while t is not always non-negative for $k < 3$. Thus to insure the compatibility of the component decision procedures $k = k_1/k_0$ must be ≥ 3. Finally for $k \geq 3$ these component two-decision procedures considered simultaneously give the following decision procedure for the seven-decision problem:

if $\bar{x}_i/\sqrt{2}\sigma < \max(\bar{x}_1/\sqrt{2}\sigma, \bar{x}_2/\sqrt{2}\sigma, \bar{x}_3/\sqrt{2}\sigma) - t$

then place π_i in the inferior group,

otherwise leave π_i in the superior group $(i = 1,2,3)$

where t is determined by k, γ^2 and

$$\Delta = \max(\{\bar{x}_1/\sqrt{2}\sigma, \bar{x}_2/\sqrt{2}\sigma, \bar{x}_3/\sqrt{2}\sigma\} - \{\bar{x}_i/\sqrt{2}\sigma\})$$
$$-\min(\{\bar{x}_1/\sqrt{2}\sigma, \bar{x}_2/\sqrt{2}\sigma, \bar{x}_3/\sqrt{2}\sigma\} - \{\bar{x}_i/\sqrt{2}\sigma\}).$$

Another way of stating this decision procedure is the following:

let $\bar{x}_{(1)} \geq \bar{x}_{(2)} \geq \bar{x}_{(3)}$ be the ranked sample means and $\pi_{(1)}, \pi_{(2)}, \pi_{(3)}$ be the corresponding populations, then the superior group is

(1) $\pi_{(1)}$

if $(\bar{x}_{(2)}/\sqrt{2}\sigma) < (\bar{x}_{(1)}/\sqrt{2}\sigma) - t_1$,

(2) $\pi_{(1)}$ and $\pi_{(2)}$

if $(\bar{x}_{(2)}/\sqrt{2}\sigma) \geq (\bar{x}_{(1)}/\sqrt{2}\sigma) - t_1$

and $(\bar{x}_{(3)}/\sqrt{2}\sigma) < (\bar{x}_{(1)}/\sqrt{2}\sigma) - t_2$,

and (3) $\pi_{(1)}$, $\pi_{(2)}$ and $\pi_{(3)}$

if $(\bar{x}_{(3)}/\sqrt{2}\sigma) \geq (\bar{x}_{(1)}/\sqrt{2}\sigma) - t_2$,

where t_1 is a function of k, γ^2 and $(\bar{x}_{(1)}/\sqrt{2}\sigma) - (\bar{x}_{(3)}/\sqrt{2}\sigma)$, and t_2 is a function of k, γ^2 and $(\bar{x}_{(1)}/\sqrt{2}\sigma) - (\bar{x}_{(2)}/\sqrt{2}\sigma)$. Figure 2 (page 43) illustrates the seven regions in (r, s) space corresponding to the seven decisions. For sample means such that (r, s) is a point in region a_i make decision d_i $(i = 1,\ldots,7)$.

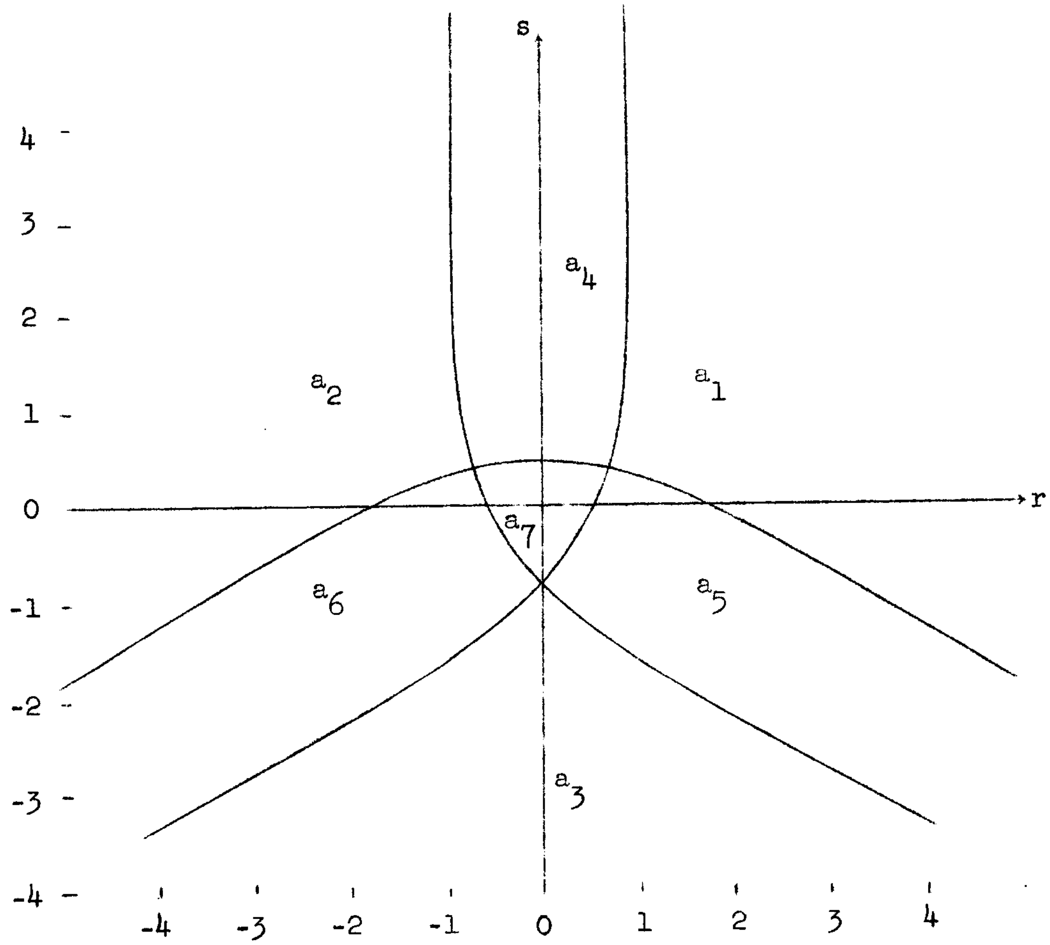
Table 2 (page 49) presents the values of t for the decision procedure for the case n = 2 and various values of k and γ^2 and σ_o^2 known and unknown (ν is the degrees of freedom for the estimate s_o^2). These t values were obtained by Duncan [7] for the pairwise-multiple comparisons problem as he treated it and are the appropriate values to use in the p-decision problem treated here for n = 2. For n = 2 t is a function of k, γ^2 and ν but does not vary with any function of the sample means (as in the case n = 3).

A comparison of t values for n = 3 with the ones for equivalent conditions for the case n = 2 shows that $t(n = 3) < t(n = 2)$. This is of considerable interest in that it runs counter to a general tendency in previous approaches (to the problem of choosing the largest mean) for such critical values to increase with n.

Figure 2

Sample space regions for Bayes decision procedure

($n = 3$, σ_o^2 known)



This increasing-with-n nature of the previous decision procedures results from the method of reducing the type-2-like error probability subject to the fixing of the type-1-like error probability in the component problems. A more balanced emphasis placed on the two types of errors in the Bayes approach results in the decreasing-with-n nature of the critical values in the present decision procedure. In Chapter IV it will be shown that this is a general property of the Bayes decision procedure for all n.

Table 1

Bayes decision procedure t values, $n = 3$, σ_0^2 known

$(\gamma^2 = \infty)$

Δ	k = 10	100	1000
0	.423	1.159	1.804
.1	.470	1.210	1.856
.2	.517	1.260	1.905
.3	.558	1.303	1.952
.4	.598	1.345	1.993
.5	.632	1.383	2.030
.6	.666	1.418	2.066
.7	.696	1.451	2.100
.8	.722	1.481	2.134
.9	.746	1.508	2.164
1.0	.769	1.535	2.194
1.2	.806	1.577	2.244
1.4	.835	1.614	2.288
1.6	.857	1.646	2.325
1.8	.872	1.669	2.355
2.0	.882	1.685	2.379
2.2	.889	1.697	2.397
2.4	.895	1.705	2.409
2.6	.898	1.711	2.418
2.8	.900	1.715	2.425
3.0	.901	1.718	2.430
3.5	.901	1.719	2.434
5.0	.901	1.721	2.436

Table 1 (cont'd)

Bayes decision procedure t values, n = 3, σ_0^2 known

($\gamma^2 = 3$)

Δ	k = 10	100	1000
0	.423	1.159	1.804
.1	.464	1.203	1.850
.2	.503	1.245	1.892
.3	.539	1.283	1.933
.4	.575	1.321	1.971
.5	.608	1.357	2.004
.6	.638	1.389	2.036
.7	.667	1.421	2.068
.8	.695	1.450	2.099
.9	.717	1.475	2.127
1.0	.738	1.498	2.154
1.2	.777	1.543	2.203
1.4	.807	1.580	2.247
1.6	.833	1.612	2.286
1.8	.853	1.640	2.318
2.0	.867	1.663	2.345
2.2	.877	1.679	2.368
2.4	.885	1.690	2.388
2.6	.891	1.699	2.400
2.8	.895	1.706	2.410
3.0	.898	1.711	2.418
3.5	.901	1.718	2.430
5.0	.901	1.720	2.436

Table 1 (cont'd)

Bayes decision procedure t values, n = 3, σ_0^2 known
($\gamma^2 = 1.$)

Δ	k = 10	100	1000
0	.423	1.159	1.804
.1	.456	1.195	1.842
.2	.489	1.230	1.878
.3	.521	1.263	1.912
.4	.552	1.295	1.945
.5	.579	1.324	1.974
.6	.606	1.353	2.000
.7	.630	1.380	2.026
.8	.654	1.405	2.052
.9	.677	1.430	2.078
1.0	.697	1.453	2.103
1.2	.734	1.493	2.149
1.4	.766	1.532	2.190
1.6	.794	1.563	2.228
1.8	.817	1.591	2.261
2.0	.837	1.625	2.292
2.2	.853	1.640	2.317
2.4	.865	1.658	2.340
2.6	.873	1.673	2.360
2.8	.880	1.684	2.377
3.0	.886	1.692	2.391
3.5	.896	1.707	2.411
5.0	.901	1.720	2.435

Table 1 (cont'd)

Bayes decision procedure t values, n = 3, σ_o^2 known

(γ^2 = .5)

Δ	k = 10	100	1000
0	.423	1.159	1.804
.1	.450	1.188	1.834
.2	.477	1.217	1.864
.3	.503	1.245	1.892
.4	.529	1.272	1.920
.5	.553	1.297	1.947
.6	.575	1.321	1.971
.7	.597	1.345	1.994
.8	.618	1.368	2.015
.9	.638	1.389	2.036
1.0	.658	1.410	2.057
1.2	.695	1.450	2.099
1.4	.724	1.483	2.137
1.6	.751	1.514	2.172
1.8	.777	1.543	2.203
2.0	.798	1.568	2.233
2.2	.817	1.590	2.260
2.4	.833	1.612	2.286
2.6	.847	1.632	2.308
2.8	.858	1.648	2.327
3.0	.867	1.663	2.345
3.5	.882	1.687	2.381
5.0	.900	1.717	2.427

Table 2*

Bayes decision procedure t values, n = 2

γ^2	ν	Log k: 0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5
∞	1	0.0	.375	.807	1.353	2.102	3.160	4.685	6.854
	2	0.0	.413	.860	1.379	2.012	2.814	3.851	5.208
	4	0.0	.434	.884	1.367	1.900	2.502	3.197	4.010
	6	0.0	.443	.891	1.356	1.848	2.374	2.948	3.580
	14	0.0	.451	.898	1.340	1.779	2.217	2.654	3.099
	∞	0.0	.457	.902	1.326	1.721	2.091	2.436	2.759
3	1	0.0	.444	1.053	2.503	∞	∞	∞	∞
	2	0.0	.484	1.060	1.926	4.077	∞	∞	∞
	4	0.0	.506	1.056	1.718	2.623	4.178	9.595	∞
	6	0.0	.515	1.053	1.653	2.370	3.308	4.732	7.706
	14	0.0	.522	1.047	1.582	2.136	2.724	3.360	4.074
	∞	0.0	.528	1.041	1.531	1.987	2.414	2.813	3.186

*Reprinted from [7].

Table 2 (cont'd)

Bayes decision procedure t values, n = 2

γ^2	ν	Log k 0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5
1	1	0.0	.572	1.930	∞	∞	∞	∞	∞
	2	0.0	.610	1.532	8.741	∞	∞	∞	∞
	4	0.0	.629	1.395	2.648	8.592	∞	∞	∞
	6	0.0	.637	1.353	2.303	3.980	13.625	∞	∞
	14	0.0	.642	1.308	2.030	2.859	3.891	5.326	7.818
	∞	0.0	.646	1.275	1.875	2.433	2.957	3.445	3.902
.5	1	0.0	.767	∞	∞	∞	∞	∞	∞
	2	0.0	.785	2.292	∞	∞	∞	∞	∞
	4	0.0	.791	1.963	9.243	∞	∞	∞	∞
	6	0.0	.794	1.800	3.777	∞	∞	∞	∞
	14	0.0	.792	1.653	2.693	4.162	6.670	∞	∞
	∞	0.0	.792	1.562	2.296	2.980	3.622	4.219	4.779

CHAPTER IV

SOME DERIVED DECISION PROCEDURES AND THEIR PROPERTIES

4.1 Progressive method for generating a class of decision procedures for the p-decision problem.

The following notation will be used:

L: p-decision problem of choosing the largest mean, as presented in Chapter II,

L_i: the i^{th} component two-decision problem used to generate L,

D_i: a decision procedure for L_i in which the region for making d_{i1} is denoted by A_i. (Using the language of hypothesis testing A_i will be called the critical region.)

D: a decision procedure for L,

$D_i(B)$: a Bayes decision procedure for L_i,

D(B): a Bayes decision procedure for L.

As shown in Chapter II, a compatible set of Bayes decision procedures, $D_i(B)$ $(i = 1,2,\ldots,n)$, for the component two-decision problems L_i $(i = 1,\ldots,n)$ generates a Bayes decision procedure, D(B), for the p-decision problem L. Moreover, any compatible set of decision procedures, D_i $(i = 1,\ldots,n)$, for the component two-decision problems, L_i $(i = 1,\ldots,n)$, generates a decision procedure, D, for the p-decision problem, L. Also, due to the symmetrical way in which the means and loss functions have been treated in these component problems, it follows that a decision procedure for any component problem, with only changes in notation, will yield a decision procedure for each one of the other component problems. That is, from a decision procedure for one component two-decision problem we have decision procedures for all n of the component two-decision problems and, if these two-decision procedures are compatible, they generate a decision procedure for the p-decision problem. Therefore, to obtain a class of decision procedures for the p-decision problem with a symmetrical treatment of the means it will be sufficient to obtain a class of decision procedures for a component two-decision problem.

Consider the i^{th} component two-decision problem, L_i, with the two decisions:

$$d_{i0}: \underline{\mu} \in \omega_{i0} = \Omega$$

$$d_{i1}: \underline{\mu} \in \omega_{i1} = \{\underline{\mu};\ \mu_i < \mu_{max}\} .$$

Remove one population other than π_i, say π_j $(j \neq i)$, and consider the decision problem L_i^j, which is a two-decision problem where the mean of π_i is compared with the means of all the other populations except π_j. That is, for this two-decision problem the two decisions are

$$d_{i0}^j: \underline{\mu} \in \omega_{i0}^j = \Omega$$

$$d_{i1}^j: \underline{\mu} \in \omega_{i1}^j = \{\underline{\mu};\ \mu_i < \max(\mu_1,\ldots,\mu_{j-1},\mu_{j+1},\ldots\mu_n)\} .$$

Let A_i^j be the critical region for a decision procedure D_i^j for this two-decision problem. From D_i^j decision procedures D_i^k $(k = 1,\ldots,n;\ k \neq i)$ are easily obtainable due to the symmetrical treatment of the μ's. Then for the two-decision problem D_i we can generate the following decision procedure:

make decision d_{i1} if d_{i1}^j is made for any j,

and make decision d_{i0} if d_{i0}^j is made for all j,

$(j = 1,\ldots,i-1,\ i+1,\ldots,n)$,

that is, the region A_i where decision d_{i1} is made is the union

$$A_i = \bigcup_{\substack{j=1 \\ j\neq i}}^{n} A_i^j$$

and the region $Y-A_i$ where decision d_{i0} is made is the intersection

$$Y-A_i = \bigcap_{\substack{j=1 \\ j\neq i}}^{n} (Y-A_i^j).$$

(This is the union-intersection method of test construction as proposed by Roy [22].)

Thus a one-versus-(n-2) procedure (that is, a decision procedure for a two-decision problem in which the mean of one population is compared with the means of (n-2) other populations) can be used to generate a one-versus-(n-1) procedure. Carrying this one step further, a one-versus-(n-3) procedure can be used to generate a one-versus-(n-2) procedure. Combining these results, a one-versus-(n-3) procedure can be used to generate a one-versus-(n-1) procedure.

Hence, it is easy to see that a one-versus-(n-m-1) procedure ($1 \leq m \leq n-2$) can be used to generate a one-versus-(n-1) procedure (which is a decision procedure for a component two-decision problem).

In the above terminology, the Bayes decision procedure obtained in Chapter III for a component problem with three populations is a one-versus-two procedure. Also the Bayes decision procedure obtained by Duncan [7] is a one-versus-one procedure. Hence for each of these known decision procedures a decision procedure for the p-decision problem can be obtained as above. Also, as the necessary integration is done and the Bayes decision procedures are found for cases in which there are more than three populations, each can be used to generate a decision procedure for the general p-decision problem.

In section 4.2 some important and interesting relations between the decision procedures generated as above from Bayes decision procedures for subcomponent problems will be given. These relations will in turn show the significant-difference-decreasing-with-n nature of the Bayes decision procedures. In section 4.3 special attention will be given to the decision procedure for the p-decision problem obtained as above from the Bayes one-versus-one decision procedure.

4.2 Some properties of the generated decision procedures.

The following notation will be used in the remainder of this chapter:

D_i : a decision procedure for the component two-decision problem L_i,

$D_i(B)$: a Bayes decision procedure for the same problem,

$D_i^{j_1,\ldots,j_m}$: a one-versus-(n-m-1) procedure for the two-decision problem where μ_i is compared with the means of all the other populations except $\pi_{j_1},\ldots,\pi_{j_m}$ ($j_1,\ldots,j_m$ are some m of the integers $1,\ldots,i-1, i+1,\ldots,n$)

$D_i^{j_1,\ldots,j_m}(B)$: a Bayes decision procedure for the same problem,

$D_i(B_{n-m})$: a decision procedure for the component problem L_i obtained by the union-intersection method of test construction from the Bayes decision procedure $D_i^{j_1,\ldots,j_m}(B)$ $(1 \leq m \leq n-2)$,

D : a decision procedure for the p-decision problem L,

D(B): a Bayes decision procedure for the p-decision problem L.

$D(B_{n-m})$: the decision procedure for the p-decision problem L generated by the decision procedure $D_i(B_{n-m})$ for the component two-decision problem L_i $(1 \leq m \leq n-2)$.

The following definitions are made:

(1) The decision procedure D_i (for the component two-decision problem L_i) is said to be <u>conservative</u> with respect to a second

decision procedure D_i' (for the same problem) if the critical region A_i for D_i is contained in the critical region A_i' for D_i'. (That is, $A_i \subset A_i'$ and $A_i \neq A_i$.)

(2) The decision procedure D (for the p-decision problem L) is said to be <u>conservative</u> with respect to a second decision procedure D' (for the same problem), where D and D' are the decision procedures for L generated by the decision procedures D_i and D_i', if D_i is conservative with respect to D_i' for $i = 1,\ldots,n$.

If it is recalled that for a point in the critical region the decision to place π_i in the inferior group is to be made while for points outside the critical region the decision to leave π_i in the superior group is to be made then if D_i is conservative with respect to D_i' it is seen that D_i leaves π_i in the superior group more often than D_i'. It follows that if D is conservative with respect to D' then for a given set of observations D makes a decision which never leaves fewer populations in the superior group than D'(moreover, the superior group obtained with D contains the superior group obtained with D'). It follows from definition 1 that if D_i is conservative with respect to D_i' and D_i' is conservative with respect to D_i'' than D_i is conservative with respect to D_i'' (since if $A_i \subset A_i'$ and $A_i' \subset A_i''$ then $A_i \subset A_i''$). From this and definition 2 it follows that if D is conservative with respect to D' and D' is conservative with respect to D'' then D is conservative with respect to D''.

The important result that $D_i(B_{n-1})$ is conservative with respect to $D_i(B)$ is given by the following theorem.

Theorem

Let $D_i(B)$ be a Bayes decision procedure for the i^{th} component two-decision problem with the two-decisions

$$d_{i0}: \underline{\mu} \in \omega_{i0} = \Omega$$

$$d_{i1}: \underline{\mu} \in \omega_{i1} = \{\underline{\mu};\ \mu_i < \mu_{max}\}$$

and loss functions and prior distributions as given in Chapter II. Let $D_i(B_{n-1})$ be a decision procedure for the same decision system obtained by the union-intersection method of test construction from the Bayes decision procedure for the two-decision problem with the two decisions

$$d_{i0}^j: \underline{\mu} \in \omega_{i0}^j = \Omega$$

$$d_{i1}^j: \underline{\mu} \in \omega_{i1}^j = \{\underline{\mu};\ \mu_i < \max \overline{\{\mu_j\}}\}$$

and loss functions and prior distributions as given in Chapter II. Then the decision procedure $D_i(B_{n-1})$ is conservative with respect to the decision procedure $D_i(B)$.

Before proving the theorem we first illustrate it with an example. For n = 3 and σ_o^2 known $D_3(B)$ is as given in Chapter III either by Figure 1 or Table 1. The critical region A_3 (in (y_1, y_2) space) for $D_3(B)$ is given in Figure 3 for a particular k and γ^2 (which need not concern us here). $D_3(B)$ is a Bayes decision procedure for the problem where μ_3 is compared with μ_1 and μ_2 . For a Bayes decision procedure for the problem where μ_3 is compared with μ_1 we have the decision procedure $D_2(B)$ from Table 2. Figure 4 gives the critical region for this decision procedure with the same k and γ^2 as used for $D_3(B)$ in Figure 3. Similarly Figure 5 gives the critical region for $D_2(B)$ for the problem where μ_3 is compared with μ_2. The critical region of $D_3(B_{3-1})$ is the union of the critical regions in Figures 4 and 5 and is given in Figure 6 together with the critical region of $D_3(B)$. From this figure it appears that the critical region of $D_3(B_{3-1})$ is contained in the critical region of $D_3(B)$. The theorem states that this is indeed the case and holds for any n.

We now return to the proof of the theorem.

Proof:

The result of the theorem will follow if it can be shown that the critical region, $A_i(B_{n-1})$, for the decision procedure $D_i(B_{n-1})$ is contained in the critical region, $A_i(B)$, for the decision procedure $D_i(B)$. With no loss of generality let i = 1. Since

$$A_1(B_{n-1}) = \bigcup_{j=2}^{n} A_1^j(B),$$

$$A_1(B_{n-1}) \subset A_1(B)$$

if

$$A_1^j(B) \subset A_1(B) \qquad \text{for } j = 2,\dots,n.$$

The critical regions

$A_3(B)$, $A_3^2(B)$, $A_3^1(B)$, $A_3(B_2)$

Figure 3

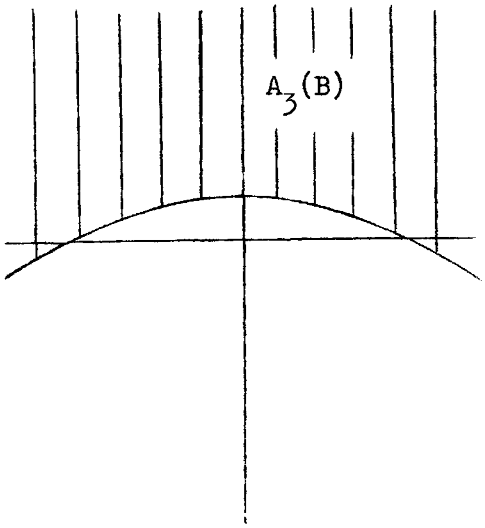


Figure 4

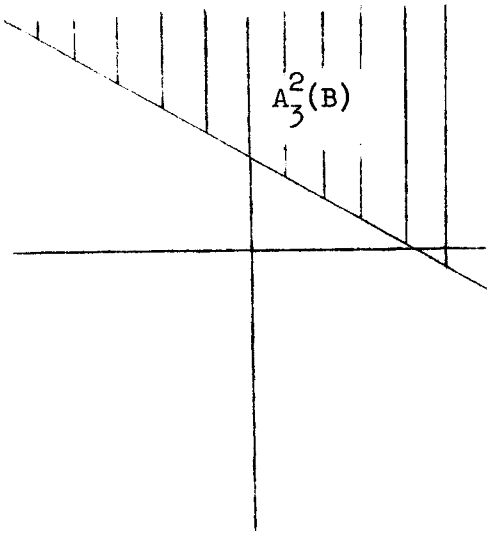


Figure 5

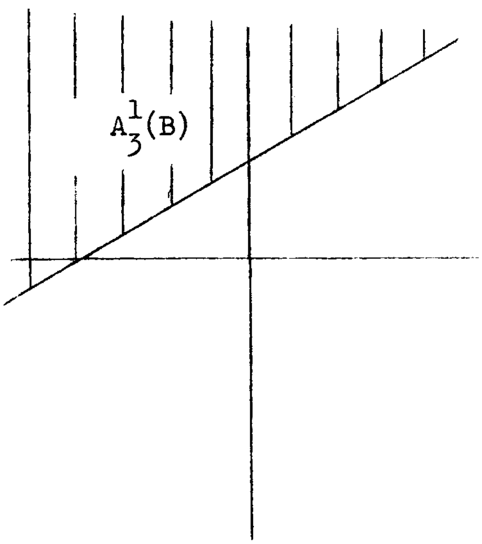
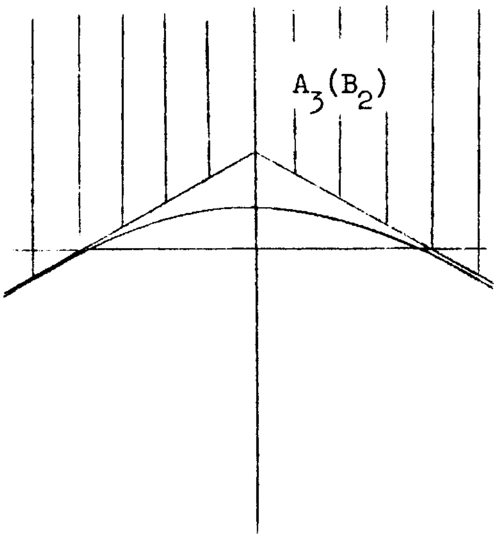


Figure 6



But due to the symmetry of our treatment of the μ's and the loss functions it is easy to see that if

$$A_1^j(B) \subset A_1(B) \quad \text{for any } j$$

then

$$A_1^j(B) \subset A_1(B) \quad \text{for all } j \ (j = 2,\dots,n).$$

Therefore it is sufficient to prove that

$$A_1^n(B) \subset A_1(B) \ .$$

But $\underline{y} \in A_1(B)$ if and only if $g_1(\underline{y}) < 0$ and $\underline{y} \in A_1^n(B)$ if and only if $g_1(y_1,\dots,y_{n-1}) < 0$ (where the function g_1 is as given in section 2.5).

Therefore

$$A_1^n(B) \subset A_1(B)$$

if

$$g_1(y_1, \dots , y_n) < g_1(y_1, \dots , y_{n-1}) \ .$$

Since if $g_1(y_1, \dots , y_n) < g_1(y_1, \dots , y_{n-1})$

then $g_1(y_1, \dots , y_{n-1}) < 0$ implies $g_1(y_1, \dots , y_n) < 0$

so $\underline{y} \in A_1^n(B)$ implies $\underline{y} \in A_1(B)$,

that is,

$$A_1^n(B) \subset A_1(B) \ .$$

From section 2.5

$$g_1(y_1,\ldots,y_n) = -(k_o/\sigma)\sum_{j=2}^{n}\int_{\Omega_\theta^* \cap \{\underline{\theta};\mu_j > \max \overline{\{\mu_j\}}\}} (\mu_j(\underline{\theta}) - \mu_1(\underline{\theta}))f(\underline{y};\underline{\theta})\xi(\underline{\theta})d\underline{\theta}$$

$$+(k_1/\sigma)\sum_{j=2}^{n}\int_{\Omega_\theta^* \cap \{\underline{\theta};\mu_1>\mu_j>\max \overline{\{\mu_1,\mu_j\}}\}} (\mu_1(\underline{\theta}) - \mu_j(\underline{\theta}))f(\underline{y};\underline{\theta})\xi(\underline{\theta})d\underline{\theta}$$

and

$$g_1(y_1,\ldots,y_{n-1}) = -(k_o/\sigma)\sum_{j=2}^{n-1}\int_{\Omega_\theta^* \cap \{\underline{\theta};\mu_j>\max \overline{\{\mu_j,\mu_n\}}\}} (\mu_j(\underline{\theta})-\mu_1(\underline{\theta}))\ f(\underline{y};\underline{\theta})\xi(\underline{\theta})d\underline{\theta}$$

$$+(k_1/\sigma)\sum_{j=2}^{n-1}\int_{\Omega_\theta^* \cap \{\underline{\theta};\ \mu_1>\mu_j>\max \overline{\{\mu_1,\mu_j,\mu_n\}}\}} (\mu_1(\underline{\theta})-\mu_j(\underline{\theta}))\ f(\underline{y};\underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}$$

In the expression for $g_1(y_1,\ldots,y_n)$ separate the (n-1) terms of the first sum $(-(k_o/\sigma)\Sigma\ldots)$ into the sum of the first (n-2) terms and the term with $j = n$. Next divide the region of integration for this last term (where $j = n$) into the (n-1) regions,

$$\Omega_\theta^* \cap \{\underline{\theta};\ \mu_n > \mu_i > \max \overline{\{\mu_i,\ \mu_n\}}\}\ ,\ (i = 1,\ldots,n-1),$$

and then group the (n-1) integrals obtained from this subdivision of the region of integration into two parts, the first with $i = 1$ and the second consisting of the remainder. Separate the (n-1) terms of the second sum $((k_1/\sigma)\Sigma\ldots)$ into the sum of the first (n-2) terms and the term with $j = n$, then divide the region of integration for this last term (where $j = n$) into the (n-2) regions,

$$\Omega_\theta^* \cap \{\underline{\theta};\ \mu_1 > \mu_n > \mu_i > \max \overline{\{\mu_1,\ \mu_i,\ \mu_n\}}\}\ ,\ (i = 2,\ldots,n-1),$$

to obtain

$$g_1(y_1,\ldots,y_n) = -(k_o/\sigma)\sum_{j=2}^{n-1}\int_{\Omega_\theta^*\cap\{\underline{\theta};\mu_j>\max\overline{\{\mu_j\}}\}}(\mu_j(\underline{\theta})-\mu_1(\underline{\theta}))f(\underline{y},\underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}$$

$$-(k_o/\sigma)\int_{\Omega_\theta^*\cap\{\underline{\theta};\mu_n>\mu_1>\max\overline{\{\mu_1,\mu_n\}}\}}(\mu_n(\underline{\theta})-\mu_1(\underline{\theta}))\ f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}$$

$$-(k_o/\sigma)\sum_{i=2}^{n-1}\int_{\Omega_\theta^*\cap\{\underline{\theta};\mu_n>\mu_i>\max\overline{\{\mu_i,\mu_n\}}\}}(\mu_n(\underline{\theta})-\mu_1(\underline{\theta}))f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}$$

$$+(k_1/\sigma)\sum_{j=2}^{n-1}\int_{\Omega_\theta^*\cap\{\underline{\theta};\mu_1>\mu_j>\max\overline{\{\mu_1,\mu_j\}}\}}(\mu_1(\underline{\theta})-\mu_j(\underline{\theta}))f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}$$

$$+(k_1/\sigma)\sum_{i=2}^{n}\int_{\Omega_\theta^*\cap\{\underline{\theta};\mu_1>\mu_n>\mu_i>\max\overline{\{\mu_1,\mu_i,\mu_n\}}\}}(\mu_1(\underline{\theta})-\mu_n(\underline{\theta}))f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}\ .$$

In the expression for $g_1(y_1,\ldots,y_{n-1})$ divide the region of integration for the (n-2) terms in the first sum$((-k_o/\sigma)\Sigma\ldots)$ into the two regions

$$\Omega_\theta^*\cap\{\underline{\theta};\ \mu_j>\max\overline{\{\mu_j\}}\}$$

and

$$\Omega_\theta^*\cap\{\underline{\theta};\ \mu_n>\mu_j>\max\overline{\{\mu_j,\ \mu_n\}}\}\ ;$$

divide the region of integration for the (n-1) terms in the second sum $((k_1/\sigma)\Sigma\ldots)$ into the three regions

$$\Omega_\theta^*\cap\{\underline{\theta};\ \mu_1>\mu_j>\max\overline{\{\mu_1,\ \mu_j\}}\}\ ,$$

$$\Omega_\Theta^* \cap \left\{\underline{\Theta};\ \mu_1 > \mu_n > \mu_j > \max \overline{\{\mu_1,\ \mu_j,\ \mu_n\}}\right\}$$

and
$$\Omega_\Theta^* \cap \left\{\underline{\Theta};\ \mu_n > \mu_1 > \mu_j > \max \overline{\{\mu_1,\ \mu_j,\ \mu_n\}}\right\};$$

to obtain

$$\begin{aligned}
g_1(y_1,\ldots,y_{n-1}) = &-(k_o/\sigma) \sum_{j=2}^{n-1} \int\limits_{\Omega_\Theta^* \cap \left\{\underline{\Theta};\mu_j > \max \overline{\{\mu_j\}}\right\}} (\mu_j(\underline{\Theta})-\mu_1(\underline{\Theta}))f(\underline{y};\ \underline{\Theta})\xi(\underline{\Theta})d\underline{\Theta} \\
&-(k_o/\sigma) \sum_{j=2}^{n-1} \int\limits_{\Omega_\Theta^* \cap \left\{\underline{\Theta};\mu_n > \mu_j > \max \overline{\{\mu_j,\mu_n\}}\right\}} (\mu_j(\underline{\Theta})-\mu_1(\underline{\Theta}))\ f(\underline{y};\ \underline{\Theta})\xi(\underline{\Theta})d\underline{\Theta} \\
&+(k_1/\sigma) \sum_{j=2}^{n-1} \int\limits_{\Omega_\Theta^* \cap \left\{\underline{\Theta};\mu_1 > \mu_j > \max \overline{\{\mu_1,\mu_j\}}\right\}} (\mu_1(\underline{\Theta})-\mu_j(\underline{\Theta}))f(\underline{y};\ \underline{\Theta})\xi(\underline{\Theta})d\underline{\Theta} \\
&+(k_1/\sigma) \sum_{j=2}^{n-1} \int\limits_{\Omega_\Theta^* \cap \left\{\underline{\Theta};\mu_1 > \mu_n > \mu_j > \max \overline{\{\mu_1,\mu_j,\mu_n\}}\right\}} (\mu_1(\underline{\Theta})-\mu_j(\underline{\Theta}))f(\underline{y};\ \underline{\Theta})\ \xi(\underline{\Theta})d\underline{\Theta} \\
&+(k_1/\sigma) \sum_{j=2}^{n-1} \int\limits_{\Omega_\Theta^* \cap \left\{\underline{\Theta};\mu_n > \mu_1 > \mu_j > \max \overline{\{\mu_1,\mu_j,\mu_n\}}\right\}} (\mu_1(\underline{\Theta})-\mu_j(\underline{\Theta}))f(\underline{y};\ \underline{\Theta})\ \xi(\underline{\Theta})d\underline{\Theta}\ .
\end{aligned}$$

Then $g_1(y_1,\ldots,y_{n-1}) - g_1(y_1,\ldots,y_n) =$

$$(k_o/\sigma) \int_{\Omega_\Theta^* \cap \{\underline{\theta};\mu_n > \mu_1 > \max \overline{\{\mu_1,\mu_n\}}\}} (\mu_n(\underline{\theta})-\mu_1(\underline{\theta}))f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}$$

$$+(k_o/\sigma) \sum_{j=2}^{n-1} \int_{\Omega_\Theta^* \cap \{\underline{\theta};\ \mu_n > \mu_j > \max \overline{\{\mu_j,\mu_n\}}\}} [(\mu_n(\underline{\theta})-\mu_1(\underline{\theta})) - (\mu_j(\underline{\theta})-\mu_1(\underline{\theta}))] f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}$$

$$+(k_1/\sigma) \sum_{j=2}^{n-1} \int_{\Omega_\Theta^* \cap \{\underline{\theta};\mu_1 > \mu_n > \mu_j > \max \overline{\{\mu_1,\mu_j,\mu_n\}}\}} [(\mu_1(\underline{\theta})-\mu_j(\underline{\theta})) - (\mu_1(\underline{\theta})-\mu_n(\underline{\theta}))] f(\underline{y};\ \underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}$$

$$+(k_1/\sigma) \sum_{j=2}^{n-1} \int_{\Omega_\Theta^* \cap \{\underline{\theta};\mu_n > \mu_1 > \mu_j > \max \overline{\{\mu_1,\mu_j,\mu_n\}}\}} (\mu_1(\underline{\theta})-\mu_j(\underline{\theta}))f(y;\ \underline{\theta})\ \xi(\underline{\theta})d\underline{\theta}\ .$$

In this difference the first term is positive since the integrand is positive over the region of integration. The next (n-2) terms are each positive for the same reason. Similarly for the next (n-2) terms. Finally, the last (n-2) terms are each positive for the same reason. Therefore we have

$$g_1(y_1,\ \ldots\ ,\ y_{n-1}) - g_1(y_1,\ \ldots\ ,y_n) > 0\ ,$$

that is,

$$g_1(y_1,\ \ldots\ ,\ y_n) < g_1(y_1,\ \ldots\ ,\ y_{n-1})$$

which implies

$$A_1^n(B) \subset A_1(B)$$

which implies that $D_1(B_{n-1})$ is conservative with respect to $D_1(B)$, and since there was no loss of generality in considering the first component problem we have that $D_i(B_{n-1})$ is conservative with respect to $D_i(B)$.

(It should be noted that the above argument is not limited to cases with the loss functions and probability density functions $f(\underline{y};\ \underline{\theta})$ and $\xi(\underline{\theta})$ of Chapter II but applies to other much more general situations. Instead of the loss being proportional to the difference between two means it could be any non-decreasing function of the difference. Also the probability density functions $f(\underline{y};\ \underline{\theta})$ and $\xi(\underline{\theta})$ can be any probability density functions with only the restriction that the Bayes decision procedure is determined by the functions $g(\underline{y})$. Since Ω_θ^* does not contain the parameter points where any of the μ's are equal, these points must have ξ-measure zero. (However the above argument can be modified to overcome this difficulty.))

In terms of the t values used in the decision procedures the theorem states that for a given set of sample means the t value used with the decision procedure $D_i(B_{n-1})$, say $t\,[D_i(B_{n-1})]$, is greater than the t value used with the decision procedure $D_i(B)$, say $t\,[D_i(B)]$, that is $t[D_i(B_{n-1})] > t[D_i(B)]$.

From the theorem and the transitive property of conservativeness the following corollary results.

Corollary 1

The decision procedure $D_i(B_{n-m})$ is conservative with respect to the decision procedure $D_i(B)$. In particular, $D_i(B_2)$ is conservative with respect to $D_i(B)$, $(1 \leq m \leq n-2)$.

In terms of the t values used in these decision procedures this corollary states that

$$t[\bar{D}_i(B)] < t[\bar{D}_i(B_{n-1})] < \ldots < t[\bar{D}_i(B_2)].$$

There are two interpretations of $t[\bar{D}_i(B_{n-m})]$. The first is that $t[\bar{D}_i(B_{n-m})]$ is the t value for a decision procedure for the two-decision problem L_i where μ_i is compared with (n-1) other means. (This decision procedure is obtained by the union-intersection method of test construction from the Bayes decision procedure where μ_i is compared with (n-m-1) other means.) The second interpretation is that $t[\bar{D}_i(B_{n-m})]$ is the t value used in the Bayes decision procedure for a two-decision component problem L_i when there are only (n-m) populations in the problem. With this second interpretation we have the following corollary.

Corollary 2

When the number of populations included in the problem is increased the t value for the Bayes decision procedure decreases. This is the generalization of the result noted in Chapter III with two and three populations. As noted there this decreasing-with-n behavior of the Bayes decision procedure t values is of considerable interest since it is opposite to the behavior of decision procedures proposed for the problem of choosing the largest mean by previous workers using fixed-type-1-like-error-probability approaches.

4.3 The decision procedures $D_i(B_2)$ and $D(B_2)$.

In this section we will obtain the decision procedure $D_i(B_2)$ for the two-decision problem L_i obtained from the Bayes decision procedure for the two-decision problem where the mean of π_i is compared with the mean of one other population (say π_j where $j \neq i$). For $D_i(B_2)$ $(i = 1,\ldots,n)$ the condition of compatibility is determined and then the decision procedure $D(B_2)$ for the p-decision problem L is obtained.

For the two-decision problem where the mean of π_i is compared with the mean of π_j the two decisions are

$$d_{i0}(j): \underline{\mu} \in \omega_{i0}(j) = \Omega$$

$$d_{i1}(j): \underline{\mu} \in \omega_{i1}(j) = \{\underline{\mu};\ \mu_i < \mu_j\} \quad .$$

A Bayes decision procedure for this problem with assumptions which imply the assumptions listed in section 2.4 was found by Duncan [7]. Thus the decision procedure found by Duncan is the Bayes decision procedure for n = 2 for a component problem that generates the p-decision problem L. This decision procedure can be written as

$$\text{make decision } d_{i0}(j) \text{ if } (\bar{x}_i/\sqrt{2}s) \geq (\bar{x}_j/\sqrt{2}s)-t$$

$$\text{and make decision } d_{i1}(j) \text{ if } (\bar{x}_i/\sqrt{2}s) < (\bar{x}_j/\sqrt{2}s)-t,$$

where t is a function of k, ν and γ^2. Therefore, using the union-intersection method of test construction, the decision procedure for the two-decision problem where the mean of π_i is compared with the means of two other populations, say π_{j_1} and π_{j_2} is

$$\text{make decision } d_{i0}(j_1,j_2) \text{ if } (\bar{x}_i/\sqrt{2}s) \geq (\bar{x}_{j_1}/\sqrt{2}s)-t$$

$$\text{and if } (\bar{x}_i/\sqrt{2}s) \geq (\bar{x}_{j_2}/\sqrt{2}s)-t$$

$$\text{and make decision } d_{i1}(j_1,j_2) \text{ if } (\bar{x}_i/\sqrt{2}s) < (\bar{x}_{j_1}/\sqrt{2}s)-t$$

$$\text{or if } (\bar{x}_i/\sqrt{2}s) < (\bar{x}_{j_2}/\sqrt{2}s)-t$$

that is,

$$\text{make decision } d_{i0}(j_1,j_2) \text{ if } (\bar{x}_i/\sqrt{2}s) \geq \max_{k=1,2} (\bar{x}_{j_k}/\sqrt{2}s)-t$$

$$\text{and make decision } d_{i1}(j_1,j_2) \text{ if } (\bar{x}_i/\sqrt{2}s) < \max_{k=1,2} (\bar{x}_{j_k}/\sqrt{2}s)-t.$$

Hence it is easy to see that the decision procedure $D_i(B_2)$ for the component two-decition problem L_i, generated from the above decision procedure as outlined in section 4.1, is

$$\text{make decision } d_{i0} \text{ if } (\bar{x}_i/\sqrt{2}s) \geq \max_{j=1,\ldots,n} (\bar{x}_j/\sqrt{2}s)-t$$

$$\text{and make decision } d_{i1} \text{ if } (\bar{x}_i/\sqrt{2}s) < \max_{j=1,\ldots,n} (\bar{x}_j/\sqrt{2}s)-t.$$

These component decision procedures ($i = 1,\ldots,n$) are compatible if t is not less than zero which is true for $k = k_1/k_o \geq 1$. Thus for $k \geq 1$ the decision procedure $D(B_2)$ for the p-decision problem generated by these n compatible two-decision procedures is:

let $\bar{x}_{(1)} \geq \bar{x}_{(2)} \geq \cdots \geq \bar{x}_{(n)}$ be the ranked sample means,

$\pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(n)}$ be the corresponding populations and let j (an integer between 1 and n) be such that

$$(\bar{x}_{(j)}/\sqrt{2}s) \geq (\bar{x}_{(1)}/\sqrt{2}s)-t$$

$$\text{and } (\bar{x}_{(j+1)}/\sqrt{2}s) < (\bar{x}_{(1)}/\sqrt{2}s)-t;$$

then the superior group is

$$\pi_{(1)},\ldots,\pi_{(j)}$$

and the inferior group is

$$\pi_{(j+1)},\ldots,\pi_{(n)}$$

(where t is a function of k, ν and γ^2 but independent of the sample means and is given in Table 2).

For reasons explained below we will refer to this decision procedure as a conservative-near-Bayes decision procedure.

A comparison of this conservative-near-Bayes decision procedure with the Bayes decision procedure for the case n = 3 as found in Chapter 3 shows that it is somewhat easier to apply since here t does not vary with any function of the sample means as it does in the Bayes decision procedure. It can also be seen from a comparison of the t values from Tables 1 and 2 that for the case n = 3 if the ranked sample means, $\bar{x}_{(1)} \geq \bar{x}_{(2)} \geq \bar{x}_{(3)}$, are such that $(\bar{x}_{(1)} - \bar{x}_{(3)})/\sqrt{2}\sigma$ is large then the Bayes t value (Table 1) and the conservative-near-Bayes decision procedure t value (Table 2) are approximately the same. This is quite reasonable to expect since if $(x_{(1)} - x_{(3)})/\sqrt{2}\sigma$ is large then the problem reduces essentially to a problem with only the two means $\bar{x}_{(1)}$ and $\bar{x}_{(2)}$.

For this reason, among the class of decision procedures for the p-decision problem which have t values independent of the sample means, the decision procedure $D(B_2)$ is the 'closest' conservative approximation to $D(B)$ for all n. (Where 'closest' is used in the sense that $t[D(B_2)]-t[D(B)]$ is smaller than for any other conservative decision procedure in the class of decision procedures with t independent of the sample means.) This gives justification to the use of the term conservative-near-Bayes decision procedure for $D(B_2)$.

4.4 Comparison of risk functions of $D(B)$ and $D(B_2)$.

Let $R_1(\underline{\Theta})$ be the risk function for the conservative-near-Bayes decision procedure $D(B_2)$ and $R_2(\underline{\Theta})$ be the risk function for the Bayes decision procedure $D(B)$; then it is of interest to find the ratio R_1/R_2 as another measure of the closeness of $D(B_2)$ to $D(B)$. This ratio has been evaluated for various values of $\underline{\Theta}$ in the case $n = 3$ and σ_0^2 known with $k = 10$ and $\gamma^2 = 1$.

As pointed out in section 2.2c, since the loss functions for the restricted product problem are the sum of the loss functions for the component problems and since expectation is a linear operator, the expected loss, or risk, for the restricted product problem is the sum of the risks for the component problems. That is,

$$R_j(\underline{\Theta}) = \sum_{i=1}^{3} R_{ij}(\underline{\Theta}) \qquad (j = 1, 2)$$

where $R_{ij}(\underline{\Theta})$ is the risk function for the i^{th} component decision procedure which generates the decision procedure $D(B_2)$ (conservative-near-Bayes decision procedure) for $j = 1$ and generates the decision procedure $D(B)$ (Bayes decision procedure) for $j = 2$. For the i^{th} component problem

$$R_{ij}(\underline{\theta}) = \sum_{k=0}^{1} L(d_{ik}; \underline{\theta})\, P_j(d_{ik}; \underline{\theta});$$

where $L(d_{ik}; \underline{\theta})$ is the loss incurred by making decision d_{ik} when $\underline{\theta}$ is the parameter point and is as defined in section 2.2c,

$$P_j(d_{i0}; \underline{\theta}) = \int_{Y-A_{ij}} f(\underline{y}; \underline{\theta})d\underline{y}\ ,$$

and

$$P_j(d_{i1}; \underline{\theta}) = \int_{A_{ij}} f(\underline{y}; \underline{\theta})d\underline{y},$$

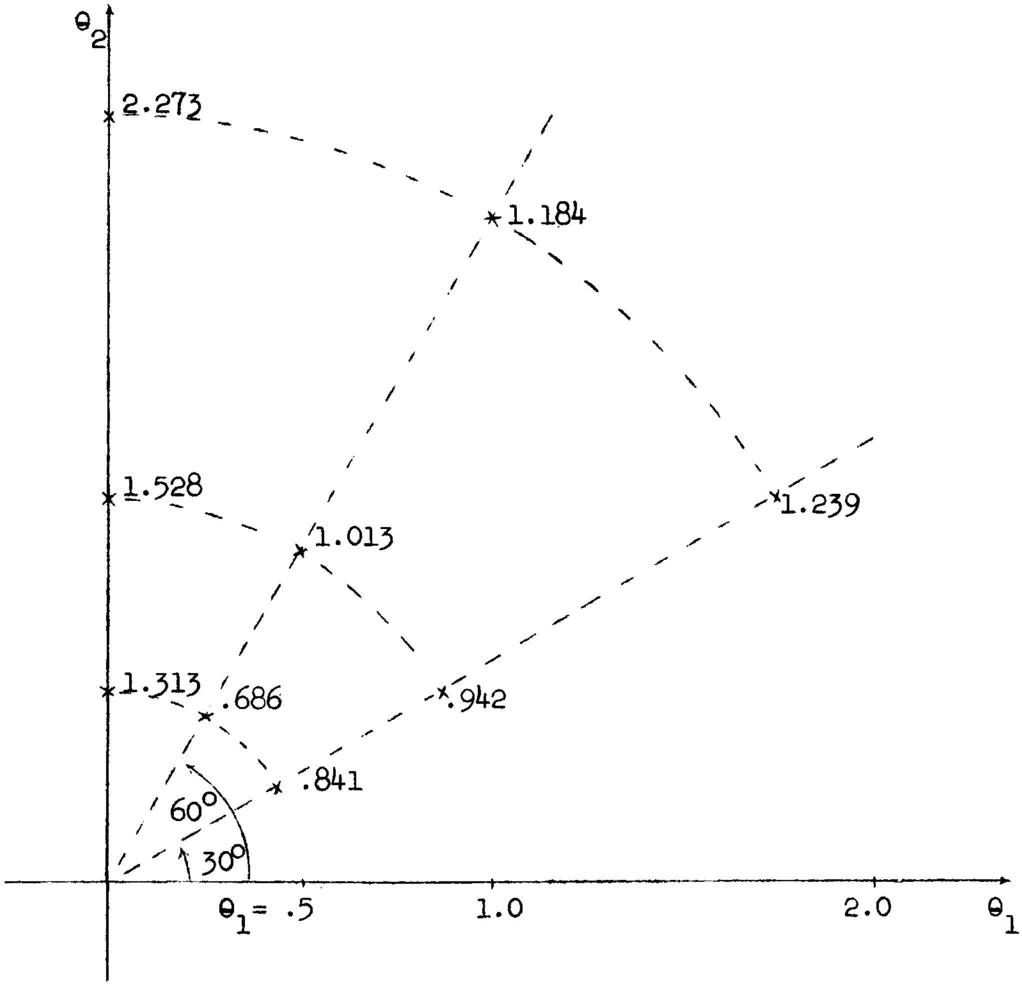
where A_{ij} is the critical region for the i^{th} component problem using the decision procedure $D_i(B_2)$ for $j = 1$ and the decision procedure $D_i(B)$ for $j = 2$. ($P_j(d_{ik}; \underline{\theta})$ is the probability of making the decision d_{ik} for the two different decision procedures, $D_i(B_2)$ for $j = 1$ and $D_i(B)$ for $j = 2$.)

Using the tables [16] and the method described therein for determining bivariate normal probabilities over polygons $P_1(d_{ik}; \underline{\theta})$ ($k = 0, 1$) can be evaluated exactly since the boundaries of A_{i1} ($i = 1, 2, 3$) are straight lines. To evaluate $P_2(d_{ik}; \underline{\theta})$ ($k = 0, 1$) the boundaries of A_{i2} ($i = 1, 2, 3$) can be approximated by a series of straight line segments and the above tables and method used. It is easy to show that $R_j(\underline{\theta}) = R_j(\underline{\theta}^{T_1}) = R_j(\underline{\theta}^{T_2})$ where $\underline{\theta}^{T_1}$ and $\underline{\theta}^{T_2}$ are the images of $\underline{\theta}$ under 120° and 240° rotations of the (θ_1, θ_2) axes. Thus $R_j(\underline{\theta})$ needs only to be evaluated over a 120° sector of the $\underline{\theta}$ plane to determine it everywhere. Also it can be shown that for the proper choice of this 120° sector there is a symmetry under a reflection in the bisector of the sector. Hence $R_j(\underline{\theta})$ needs only to be evaluated over a 60° sector.

This evaluation has been done for various points in the 60° sector $\theta_2 \geq \theta_1/\sqrt{3} \geq 0$ and the resulting values of $R_1(\underline{\theta})/R_2(\underline{\theta})$ are as given in Figure 7. From the 60° sector in the figure a reflection in the line $\theta_2 = \theta_1/\sqrt{3}$ will give values for the 60° sector $\theta_1 \geq 0$, $-\theta_1/\sqrt{3} \leq \theta_2 \leq \theta_1/\sqrt{3}$ and this sector together with the sector in the figure under rotations of 120° and 240° will give values for the remainder of the θ-plane.

Figure 7

Ratio of risk for conservative-near-Bayes decision procedure to risk for Bayes decision procedure ($n = 3$, σ_o^2 known)



CHAPTER V

THE EFFECT OF DEPARTURES FROM NORMALITY IN THE PRIOR DISTRIBUTION

5.1 Introduction.

In this chapter results are given of a study made of the effect on the Bayes decision procedure due to departures from normality in the prior distribution of the μ's for the case $n = 2$ and σ_o^2 known. Since $n = 2$ these results apply not only to the Bayes decision procedure for the p-decision problem for $p = 3$ $(= 2^2 - 1)$ but also to the pairwise multiple comparison problem as treated by Duncan [7] for all n and to what we have called the conservative-near-Bayes decision procedure $(D(B_2))$ for all n.

5.2 Choice of a family of prior distributions.

Let

$$\eta = E(\mu_i)$$

$$\mathrm{var}(\mu_i) = E(\mu_i - \eta)^2$$

$$\alpha_3 = E(\mu-\eta)^3/[\mathrm{var}(\mu)]^{3/2}, \quad \text{(skewness coefficient)},$$

and

$$\alpha_4 = E(\mu-\eta)^4/[\mathrm{var}(\mu)]^2, \quad \text{(kurtosis coefficient)}.$$

For the study of this chapter it was desired to choose a family of prior distributions of the μ's which would give wide ranges of values of α_3 and α_4 (since they are often used as measures of departure from normality). Also it was desired that this prior distribution should not differ much from the previous normal prior distribution if α_3 and α_4 did not differ much from α_3 and α_4 for a normal distribution (where $\alpha_3 = 0$, $\alpha_4 = 3$). With such a prior distribution the study is essentially reduced to a study of the variation of the Bayes decision procedure with the variation of α_3 and α_4 of the new prior distribution. Since the moments, and especially only the first four moments, do not always completely determine a distribution the above reduction in the study means that it is not complete since it will not be made with all possible prior distributions. However the study is at least a first approximation to a complete one. From a practical point of view with such a prior distribution an experimenter who is unwilling to assume that the prior distribution is normal but who is willing to assume that the prior distribution is determined close enough for his purposes by the first four moments could use the Bayes decision procedure with this new prior distribution.

A relatively simple distribution with the desired properties is the following one obtained as a mixture of two normal distributions,

$$(5.2.1) \qquad \psi(\mu_i) = p(2\pi b_1^2\gamma^2\sigma^2)^{-1/2}\exp\left\{-(2b_1^2\gamma^2\sigma^2)^{-1}[\mu_i-\gamma\sigma(c+e)]^2\right\}$$

$$+(1-p)(2\pi b_2^2\gamma^2\sigma^2)^{-1/2}\exp\left\{-(2b_2^2\gamma^2\sigma^2)^{-1}[\mu_i-\gamma\sigma(c-e)]^2\right\}$$

where $0 \leq p \leq 1$, $b_1 > 0$, $b_2 > 0$ $(i = 1, 2)$.

However this distribution is not completely satisfactory for our purposes since the first four moments do not completely specify all the parameters. To overcome this the condition

$$pb_1^2 = (1-p)b_2^2$$

was imposed. (This condition was suggested by having $\alpha_3 = 0$ imply symmetry of the distribution.) Therefore the prior distribution of the population means used is given by equation 5.2.1 with the additional restriction that

$$pb_1^2 = (1-p)b_2^2 \ .$$

(For this distribution, $\alpha_3 = 0$ implies that the distribution is symmetrical about the mean.)

With the proper choice of parameters in this distribution the variance can be set equal to the value used previously and a complete range of values of α_3 and α_4 obtained. Also when $\alpha_3 = 0$, $\alpha_4 = 3$ the above distribution reduces to a normal distribution as desired. The choice of this distribution is in keeping with some previous work of Karl Pearson's [21] in which a mixture of two normal distributions was used to fit some non-normal distributions. Finally with this choice of prior distribution the necessary numerical integration can be carried out with the aid of existing tables [15].

5.3 Parameter values used.

For this study α_3 and α_4 values were used as indicated in Table 3.

Table 3

α_3, α_4 values used

$\alpha_4 \backslash \alpha_3$	0	1	2	3
2	X	X*		
3	X	X		
5			X*	
6	X	X	X	
10				X*
11	X	X	X	X
18	X	X	X	X

('X' indicates α_3, α_4 combination used.)

Those (α_3, α_4) combinations with an asterisk are for the limiting form of the prior distribution where $b_1 = b_2 = 0$. This limiting form is

$$\begin{aligned} \psi(\mu_i) &= \quad p \quad , \mu_i = \gamma\sigma(c+e) \\ &= (1-p) \ , \mu_i = \gamma\sigma(c-e) \\ &= \quad 0 \quad , \text{elsewhere} \qquad (i = 1, 2), \end{aligned}$$

that is, the distribution where μ_i takes the value $\gamma\sigma(c+e)$ with probability p, the value $\gamma\sigma(c-e)$ with probability (1-p), and other values with probability zero.

It is easy to see for the distribution in equation 5.2.1 that

$$\mathrm{var}(\mu) = [pb_1^2+(1-p)b_2^2+4p(1-p)e^2]\gamma^2\sigma^2$$

$$E(\mu-\eta)^3 = [6p(1-p)e(b_1^2-b_2^2)+8p(1-p)(1-2p)e^3]\gamma^3\sigma^3$$

and $$E(\mu-\eta)^4 = \{3[pb_1^4+(1-p)b_2^4]+24p(1-p)e^2[(1-p)b_1^2+pb_2^2] +16p(1-p)[1-3p+3p^2]e^4\}\gamma^4\sigma^4 .$$

Thus imposing the condition

$$pb_1^2 = (1-p)b_2^2$$

and setting $\mathrm{var}(\mu) = \gamma^2\sigma^2$

as before it readily follows that

$$\alpha_3 = (1-2p)[3e-4p(1-p)e^3]$$

and $$\alpha_4 = 3[4p(1-p)]^{-1}+ 6(2p-1)^2e^2- 4p(1-p)(12p^2- 12p+5)e^4 .$$

For the limiting ($b_1= b_2= 0$) distribution

$$\mathrm{var}(\mu) = 4p(1-p)e^2\gamma^2\sigma^2$$

$$E(\mu-\eta)^3 = (1-2p)8p(1-p)(e\gamma\sigma)^3$$

and $$E(\mu-\eta)^4 = 16p(1-p)[1-3p(1-p)](e\gamma\sigma)^4$$

thus $$\alpha_3 = (1-2p)[p(1-p)]^{-1/2}$$

and $$\alpha_4 = [p(1-p)]^{-1}- 3 = \alpha_3^2 + 1 .$$

From the above formulas the parameter values corresponding to the α_3, α_4 values given in Table 3 were found and are as given in Table 4 (page 79). (For our purposes c need not be specified.) The distributions for these parameter values (with $c = 0$) are sketched in Figure 8 (page 80).

Table 4

Parameter values used

in non-normal prior distributions

(α_3, α_4)	e	p	b_1	b_2
$(b_1 > 0,\ b_2 > 0)$				
(0, 2)	.6180	.5	.7862	.7862
(0, 3)	0	.5	1.	1.
(0, 6)	0	.1464	1.8478	.7654
(0,11)	0	.0736	2.6064	.7347
(0,18)	0	.0436	3.3878	.7230
(1, 3)	.9899	.2727	.6387	.3912
(1, 6)	.5270	.1666	1.5932	.7123
(1,11)	.4019	.0787	2.4607	.7192
(1,18)	.3695	.0453	3.2823	.7151
(2, 6)	1.3217	.1452	.6758	.2785
(2,11)	.9112	.0955	1.9324	.6279
(2,18)	.7728	.0513	2.9360	.6825
(3,11)	1.7323	.0837	.6904	.2086
(3,18)	1.3155	.0615	2.2093	.5656
$(b_1 = b_2 = 0)$				
(1, 2)	1.1180	.2764	0	0
(2, 5)	1.4142	.1464	0	0
(3,10)	1.8028	.0840	0	0

Figure 8

Non-normal prior distributions

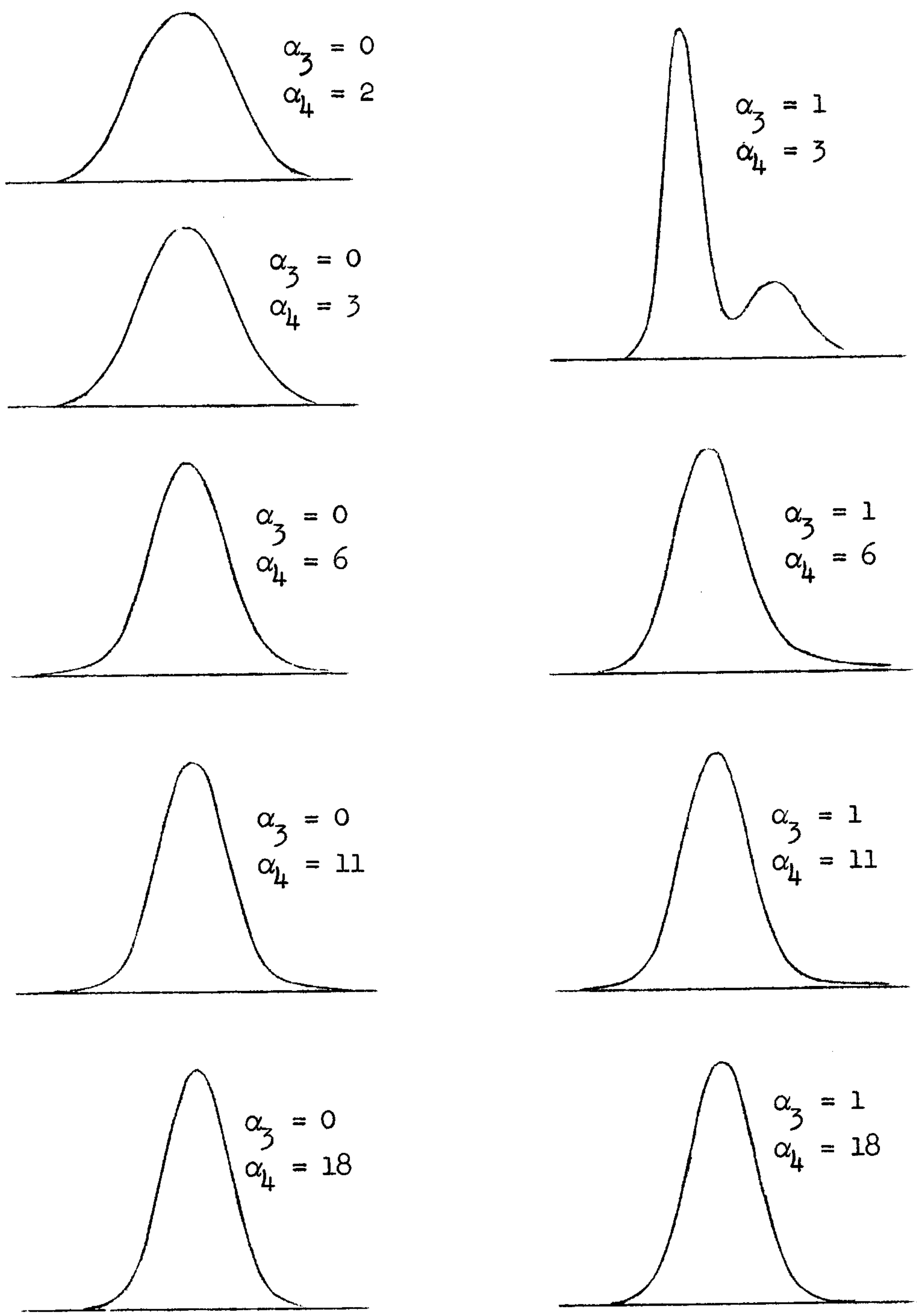


Figure 8 (cont'd)

Non-normal prior distributions

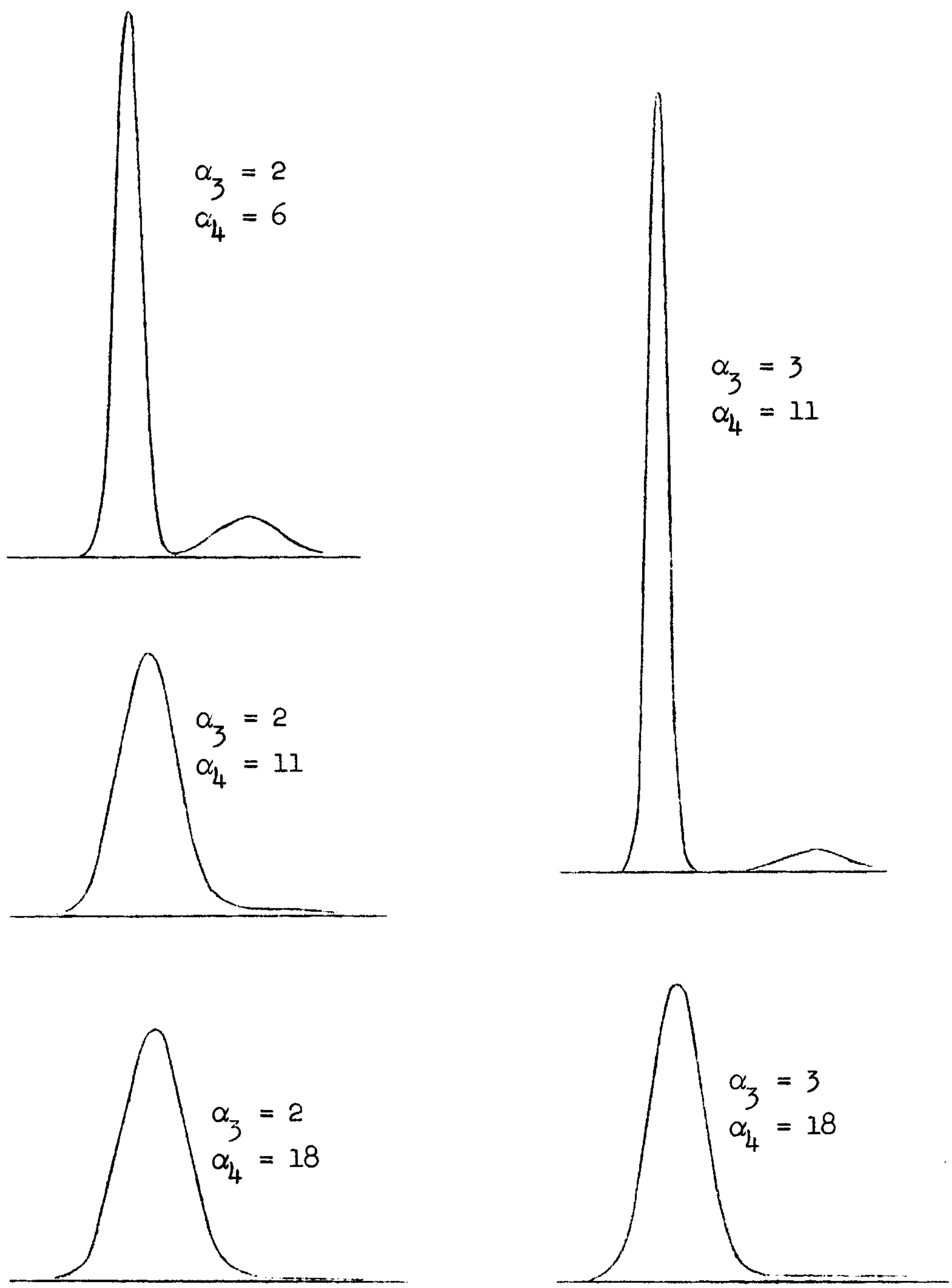
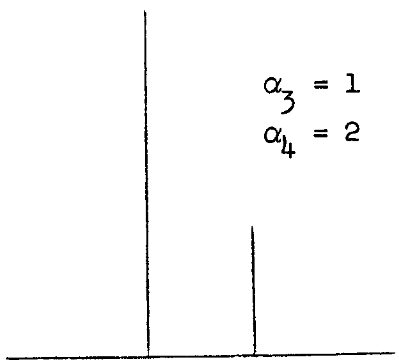
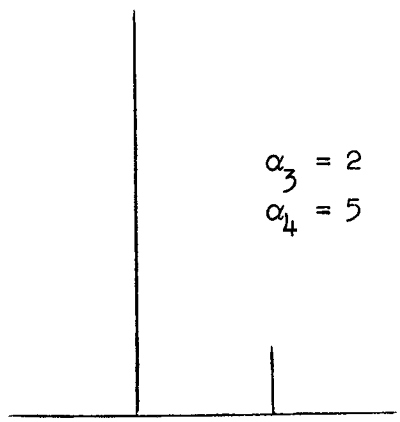
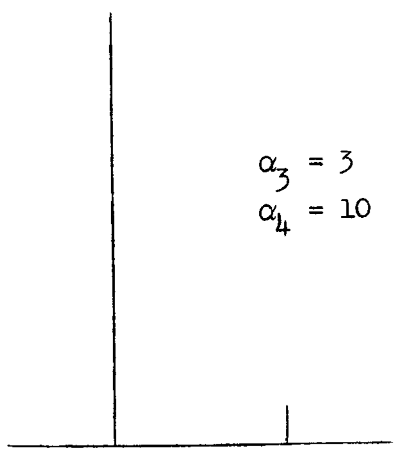


Figure 8 (cont'd)

Non-normal prior distributions







From this it can be seen that quite large deviations from the normal distribution have been obtained with the family of distributions employed.

5.4 Bayes decision procedure.

As described in section 2.5 the Bayes decision procedure for, say, the first (of the n = 2) component problems is:

make decision d_{10} if $g_1 \geq 0$

and make decision d_{11} if $g_1 < 0$,

where

$$g_1(\underline{y}) = -(k_0/\sigma)\int_{\mu_2>\mu_1} [\mu_2(\underline{\theta})-\mu_1(\underline{\theta})] f(\underline{y};\ \underline{\theta})\xi(\underline{\theta})d\underline{\theta}$$

$$+(k_1/\sigma)\int_{\mu_1>\mu_2} [\mu_1(\underline{\theta})-\mu_2(\underline{\theta})] f(\underline{y};\ \underline{\theta})\xi(\underline{\theta})d\underline{\theta} \quad .$$

Also

$$\theta_0 = (\mu_1+\mu_2)/\sqrt{2}\sigma \qquad y_0 = (\bar{x}_1+\bar{x}_2)/\sqrt{2}\sigma$$

$$\underline{\theta} = \theta_1 = (\mu_1-\mu_2)/\sqrt{2}\sigma \qquad \underline{y} = y_1 = (\bar{x}_1-\bar{x}_2)/\sqrt{2}\sigma \ ,$$

$$f(\underline{y};\ \underline{\theta}) = f(y_1;\ \theta_1) = (2\pi)^{-1/2}\exp\left\{-1/2(y_1-\theta_1)^2\right\} ,$$

and

$$\xi(\underline{\theta}) = \xi((\mu_1-\mu_2)/\sqrt{2}\sigma).$$

Where $\xi(\underline{\theta})$ is the prior distribution of the standarized difference of the two population means, while $\psi(\mu)$ is the prior distribution of one population mean.

(1) <u>$b_1 > 0,\ b_2 > 0$ case.</u>

$\xi(\theta)$ can be found from $\psi(\mu_i)$ with the use of characteristic functions.

$$E[\exp(it\theta)] = E[\exp(it\mu_1/\sqrt{2}\sigma)]\ E[\exp(-it\mu_2/\sqrt{2}\sigma)]$$

$$= \{p\ \exp[i(c+e)\gamma t/\sqrt{2}]\ \exp(-b_1^2\gamma^2t^2/4) + (1-p)\exp[i(c-e)\gamma t/\sqrt{2}]\ \exp(-b_2^2\gamma^2t^2/4)\}$$

$$\{p\ \exp[-i(c+e)\gamma t/\sqrt{2}]\ \exp(-b_1^2\gamma^2t^2/4) + (1-p)\exp[-i(c-e)\gamma t/\sqrt{2}]\ \exp(-b_2^2\gamma^2t^2/4)\}$$

$$= p^2\ \exp(-b_1^2\gamma^2t^2/2)+(1-p)\exp(-b_2^2\gamma^2t^2/2)$$
$$+p(1-p)\exp[i\ \sqrt{2}e\gamma t]\ \exp[-(b_1^2+b_2^2)\gamma^2t^2/4]$$
$$+p(1-p)\exp[-i\ \sqrt{2}e\gamma t]\ \exp[-(b_1^2+b_2^2)\gamma^2t^2/4].$$

From which it follows that

$$\xi(\theta) = p^2(2\pi)^{-1/2}(b_1\gamma)^{-1}\ \exp[-(2b_1^2\gamma^2)^{-1}\ \theta^2]$$
$$+(1-p)^2(2\pi)^{-1/2}(b_2\gamma)^{-1}\ \exp[-(2b_2^2\gamma^2)^{-1}\theta^2]$$
$$+p(1-p)(2\pi)^{-1/2}\sqrt{2}(b_1^2+b_2^2)^{-1/2}\gamma^{-1}\exp[-(b_1^2+b_2^2)^{-1}\gamma^{-2}(\theta-e\gamma)^2]$$
$$+p(1-p)(2\pi)^{-1/2}\sqrt{2}(b_1^2+b_2^2)^{-1/2}\gamma^{-1}\exp[-(b_1^2+b_2^2)^{-1}\gamma^{-2}(\theta+e\gamma)^2].$$

Therefore

$$f(y;\ \theta)\ \xi(\theta) =$$
$$\sum_{i=1}^{4} (2\pi)^{-1/2}\exp[-(y-\theta)^2/2]a_{1i}(2\pi)^{-1/2}(a_{2i})^{-1}\exp[-(\theta-a_{3i})^2/2a_{2i}^2]$$

where

$$a_{11} = p^2, \qquad a_{12} = (1-p)^2,$$
$$a_{13} = a_{14} = p(1-p),$$
$$a_{21} = b_1\gamma, \qquad a_{22} = b_2\gamma,$$
$$a_{23} = a_{24} = \sqrt{b_1^2+b_2^2}\,\gamma/\sqrt{2},$$
$$a_{31} = a_{32} = 0,$$
$$a_{33} = \sqrt{2}e\gamma, \qquad a_{34} = -\sqrt{2}e\gamma .$$

By completing the square of Θ in the exponent of $f(y;\Theta)\xi(\Theta)$ we have

$f(y;\Theta)\xi(\Theta) =$

$$\sum_{i=1}^{4} a_{1i}(2\pi)^{-1/2}c_i^{-1}\exp[-(y-a_{3i})^2/2c_i^2](2\pi)^{-1/2}s_i^{-1}\exp[-(\Theta-m_i)^2/2s_i^2]$$

where

$$c_i = (1+a_{2i}^2)^{-1/2}$$
$$m_i = (a_{21}^2 y + a_{3i})/(1+a_{2i}^2)$$
$$s_i = a_{2i}/\sqrt{(1+a_{2i}^2)} \qquad (i = 1, 2, 3, 4).$$

Let $\phi(x;a,b) = (2\pi)^{-\frac{1}{2}}b^{-1}\exp[-(x-a)^2/2b^2]$

then

$$g_1(y) = (k_0/\sigma)\int_{-\infty}^{0}(\sqrt{2}\sigma\Theta)\,f(y;\Theta)\xi(\Theta)d\Theta + (k_1/\sigma)\int_{0}^{\infty}(\sqrt{2}\sigma\Theta)f(y;\Theta)\xi(\Theta)d\Theta$$

$$= \sqrt{2}k_0\sum_{i=1}^{4}a_{1i}\,\phi(y;a_{3i},c_i)\int_{-\infty}^{0}\Theta\,\phi(\Theta;m_i,s_i)d\Theta$$

$$+ \sqrt{2}k_1\sum_{i=1}^{4}a_{1i}\,\phi(y;a_{3i},c_i)\int_{0}^{\infty}\Theta\,\phi(\Theta;m_i,s_i)d\Theta .$$

Let $z_i = (\theta - m_i)/s_i$

then

$$g_1(y) = \sqrt{2}k_o \sum_{i=1}^{4} a_{1i}\, \phi(y;\, a_{3i}, c_i) \int_{-\infty}^{-m_i/s_i} (s_i z_i + m_i)\, \phi(z_i) dz_i$$

$$+ \sqrt{2}k_1 \sum_{i=1}^{4} a_{1i}\, \phi(y;\, a_{3i}, c_i) \int_{-m_i/s_i}^{\infty} (s_i z_i + m_i)\, \phi(z_i) dz_i$$

$$= \sqrt{2}k_o \sum_{i=1}^{4} a_{1i}\, \phi(y;\, a_{3i}, c_i)\; s_i \int_{-\infty}^{-m_i/s_i} z_i\, \phi(z_i) dz_i$$

$$+ \sqrt{2}k_o \sum_{i=1}^{4} a_{1i}\, \phi(y;\, a_{3i}, c_i)\; m_i \int_{-\infty}^{-m_i/s_i} \phi(z_i) dz_i$$

$$+ \sqrt{2}k_1 \sum_{i=1}^{4} a_{1i}\, \phi(y;\, a_{3i}, c_i)\; s_i \int_{-m_i/s_i}^{\infty} z_i\, \phi(z_i) dz_i$$

$$+ \sqrt{2}k_1 \sum_{i=1}^{4} a_{1i}\, \phi(y;\, a_{3i}, c_i)\; m_i \int_{-m_i/s_i}^{\infty} \phi(z_i) dz_i \;.$$

But
$$\int_{-\infty}^{-a} z\, \phi(z) dz = -\phi(a),$$

$$\int_{-\infty}^{-a} \phi(z)\, dz = 1 - \Phi(a)\;,$$

$$\int_{-a}^{\infty} z\, \phi(z) dz = \phi(a)$$

and
$$\int_{-a}^{\infty} \phi(z) dz = \Phi(a)$$

where $\phi(a) = (2\pi)^{-1/2} \exp(-a^2/2)$

and $\Phi(a) = \int_{-\infty}^{a} (2\pi)^{-1/2} \exp(-z^2/2)dz.$

It readily follows from the above that

$$g_1(y) = \sqrt{2}\ (k_1-k_o) \sum_{i=1}^{4} a_{1i}\ \phi(y;\ a_{3i},\ c_i)s_i\ \phi(m_i/s_i)$$

$$= \sqrt{2}\ (k_1-k_o) \sum_{i=1}^{4} a_{1i}\ \phi(y;\ a_{3i},\ c_i)m_i\ \Phi(m_i/s_i)$$

$$+ \sqrt{2}\ k_o \sum_{i=1}^{4} a_{1i}\ \phi(y;\ a_{3i},\ c_i)m_i\ .$$

Using the tables [15] and an iterative technique y was found such that $g_1(y) = 0$. From this we have the Bayes decision procedure:

make decision d_{10} if $(\bar{x}_1/\sqrt{2}\sigma) \geq (\bar{x}_2/\sqrt{2}\sigma) - t$

and make decision d_{11} if $(\bar{x}_1/\sqrt{2}\sigma) < (\bar{x}_2/\sqrt{2}\sigma) - t$,

where t is the value of y such that $g_1(t) = 0$ and $g_1(y) > 0$ for $y > t$.

(2) $\underline{b_1 = b_2 = 0 \text{ case.}}$

For this case

$$\begin{aligned} \psi(\mu) &= p\ , & \mu &= \gamma c(c+e)\ , \\ &= 1-p\ , & \mu &= \gamma c(c-e)\ , \\ &= 0\ , & & \text{otherwise.} \end{aligned}$$

and $f(y;\,\theta)\,\xi(\theta) = p(1-p)\,\phi(y;\,-\sqrt{2}e\gamma,1)$, $\theta = -\sqrt{2}e\gamma$,

$= [p^2+(1-p)^2]\,\phi(y)$, $\theta = 0$,

$= p(1-p)\,\phi(y;\,\sqrt{2}e\gamma,1)$, $\theta = \sqrt{2}e\gamma$,

$= 0$, otherwise.

Thus

$$g_1(y) = -\sqrt{2}\,k_o\int_{-\infty}^{0}(-\theta)f(y;\,\theta)\xi(\theta)d\theta$$

$$+\sqrt{2}k_1\int_{0}^{\infty}\theta\,f(y;\,\theta)\,\xi(\theta)d\theta$$

$$= -\sqrt{2}k_o(\sqrt{2}e\gamma)p(1-p)\,\phi(y;\,-\sqrt{2}e\gamma,1)$$

$$+\sqrt{2}k_1(\sqrt{2}e\gamma)p(1-p)\,\phi(y;\,\sqrt{2}e\gamma,1)\ .$$

Thus $g_1(y) \geq 0$

if $\quad -\phi(y;\,-\sqrt{2}e\gamma,1)+k\,\phi(y;\,\sqrt{2}e\gamma,1) \geq 0$

that is, if $\quad \exp\left\{-[(y+2e\gamma)^2+(y-2e\gamma)^2]/2\right\} \leq k$

where $k = k_1/k_o$. Taking natural logarithms, $g_1(y) \geq 0$

if $\quad -2\sqrt{2}ye\gamma \leq \ln k$,

that is, if $\quad -y \leq (\ln k)/2\sqrt{2}e\gamma$.

But $\quad \mathrm{var}(\mu) = 4p(1-p)e^2\gamma^2\sigma^2$ which is to be $\gamma^2\sigma^2$.

Therefore $\quad e^2 = [4p(1-p)]^{-1}$

and $\quad (\ln k)/2\sqrt{2}e\gamma = \sqrt{p(1-p)}\,(\ln k)/\sqrt{2}\gamma$.

Therefore $g_1(y) \geq 0$

if $\quad y \geq -\sqrt{p(1-p)}\,(\ln k)/\sqrt{2}\gamma$.

Let $t = \sqrt{p(1-p)}\,(\ln k)/\sqrt{2}\gamma$

and we have the Bayes decision procedure,

$$\text{make decision } d_{10} \text{ if } (\bar{x}_1/\sqrt{2}\sigma) \geq (\bar{x}_2/\sqrt{2}\sigma) - t$$

$$\text{and make decision } d_{11} \text{ if } (\bar{x}_1/\sqrt{2}\sigma) < (\bar{x}_2/\sqrt{2}\sigma) - t.$$

The t values for the Bayes decision procedure with the non-normal prior distribution are given in Table 5 for the general ($b_1 > 0$, $b_2 > 0$) case and the limiting ($b_1 = b_2 = 0$) case. For $k = 10$, $\gamma^2 = 1$ the t values are given for all the (α_3, α_4) values of Table 3 while for other k and γ^2 values the t values are given for a selected number of these (α_3, α_4) values.

In Table 5 the t value for $\alpha_3 = 0$, $\alpha_4 = 3$ is for the Bayes decision procedure with normal prior distribution. Comparing the values of t for other α_3, α_4 values with this value of t it is seen that t does vary a great deal with changes in α_3 and α_4. However from the illustrations (Figure 8) showing the shapes of the prior distributions it is seen that the bigger changes in t are induced by quite large departures from normality. The t values for the limiting $b_1 = b_2 = 0$ case are typical examples of this.

In view of this it would seem reasonable to use the Bayes decision procedure with normal prior distribution if large departures from normality are not suspected in the prior distribution, while if the prior distribution is far removed from normality the Bayes decision procedure with the more general prior distribution should be used.

Table 5

Bayes decision procedure t values

non-normal prior distribution, n = 2, σ_o^2 known

(α_3, α_4)	k = 10, $\gamma^2 = .5$	k = 10, $\gamma^2 = 1$	k = 10, $\gamma^2 = 3$	k = 100, $\gamma^2 = 1$	k = 1000, $\gamma^2 = 1$
$(b_1 > 0, b_2 > 0)$					
(0, 2)		1.266			
(0, 3)	1.562	1.275	1.041	2.433	3.445
(0, 6)		1.344			
(0,11)	1.681	1.414	1.123	2.584	3.459
(0,18)		1.464			
(1, 3)	1.498	1.253	1.133	2.402	3.444
(1, 6)		1.355			
(1,11)		1.421			
(1,18)		1.470			
(2, 6)		1.346			
(2,11)	1.689	1.466	1.179	2.597	3.429
(2,18)		1.495			
(3,11)		1.542			
(3,18)		1.593			
$(b_1 = b_2 = 0)$					
(1, 2)	1.030	.728	.420	1.456	2.184
(2, 5)	.814	.576	.332	1.152	1.728
(3,10)	.634	.452	.261	.904	1.356

BIBLIOGRAPHY

[1] Bahadur, Raghu Raj, "On a problem in the theory of k populations," Annals of Mathematical Statistics, Vol. 21 (1950), pp. 362-375.

[2] Bechhofer, Robert E., "A single-sample multiple decision procedure for ranking means of normal populations with known variances," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 16-39.

[3] Bechhofer, R. E., Dunnett, C. W. and Sobel, M., "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance," Biometrika, Vol. 41 (1954), pp. 170-176.

[4] Bechhofer, R. E. and Sobel, M. "A sequential multiple decision procedure for selecting the population with the largest mean from k normal populations with a common unknown variance (Preliminary Report)," abstract, Annals of Mathematical Statistics, Vol. 27 (1956), pp. 218-219.

[5] Duncan, David B., "A significance test for differences between ranked treatments in an analysis of variance," The Virginia Journal of Science, Vol. 2 (new series) (1951), pp. 171-189.

[6] Duncan, D. B., "Multiple range and multiple F tests," Biometrics, Vol. 11 (1955), pp. 1-42.

[7] Duncan, D. B., "Bayes rules for a common multiple comparisons problem and related Student-t problems," Institute of Statistics, University of North Carolina, Mimeograph Series, No. 267, (1960).

[8] Dunnett, Charles W., "On selecting the largest of k normal population means," Journal of the Royal Statistical Society, Series B, Vol. 22 (1960), pp. 1-30.

[9] Fisher, R. A., The Design of Experiments, (sixth edition) Oliver and Boyd Ltd., Edinburgh (1951).

[10] Gupta, Shanti S., "On a decision rule for a problem in ranking means," Ph.D. Thesis, University of North Carolina, Chapel Hill (1956).

[11] Hall, Wm. Jackson, "Multiple-decision procedures which control all kinds of errors," Invited address at 1959 annual meetings of Institute of Mathematical Statistics, Washington, D. C..

[12] Lehmann, E. L., "A theory of some multiple decision problems, I," Annals of Mathematical Statistics, Vol. 28 (1957), pp. 1-25.

[13] Lehmann, E. L., "A theory of some multiple decision problems, II," Annals of Mathematical Statistics, Vol. 28 (1957), pp. 547-572.

[14] Mosteller, Frederick, "A k-sample slippage test for an extreme population," Annals of Mathematical Statistics, Vol. 19 (1948), pp. 58-65.

[15] National Bureau of Standards, Tables of Normal Probability Functions, Applied Mathematics Series 23, U.S. Government Printing Office, Washington, D.C., (1953).

[16] National Bureau of Standards, Tables of the Bivariate Normal Distribution Function and Related Functions, Applied Mathematics Series 50, U. S. Government Printing Office, Washington, D. C., (1959).

[17] Newman, D., "The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of the standard deviation," Biometrika, Vol. 31 (1939), pp.20-30.

[18] Paulson, Edward, "A multiple decision procedure for certain problems in the analysis of variance," Annals of Mathematical Statistics, Vol. 20 (1949), pp. 95-98.

[19] Paulson, E., "On the comparison of several experimental categories with a control," Annals of Mathematical Statistics, Vol. 23 (1952), pp. 239-246.

[20] Paulson, E., "An optimum solution to the k-sample slippage problem for the normal distribution," Annals of Mathematical Statistics, Vol. 23 (1952), pp. 610-616.

[21] Pearson, Karl, "Contributions to the mathematical theory of evolution," Philosophical Transactions of the Royal Society of London, Series A, Vol. 185 (1894), pp. 71-110 (reprinted in: Karl Pearson's Early Statistical Papers, Cambridge University Press, London, (1948).

[22] Roy, S. N., "On a heuristic method of test construction and its use in multivariate analysis," Annals of Mathematical Statistics, Vol. 24 (1953), pp. 220-238.

[23] Scheffé, Henry, "A method for judging all contrasts in the analysis of variance," Biometrika, Vol. 40 (1953), pp. 87-104.

[24] Seal, K. C., "On a class of decision procedures for ranking means of normal populations," Annals of Mathematical Statistics, Vol. 26 (1955), pp. 387-398.

[25] Tukey, John W., "Quick and dirty methods in statistics, Part II, simple analysis for standard designs," Proceedings Fifth Annual Convention, American Society for Quality Control, (1951), pp. 189-197.

[26] Tukey, J. W., "The problem of multiple comparisons," unpublished dittoed notes, 396 pp., Princeton University (1953).

[27] Wald, Abraham, Statistical Decision Functions, John Wiley and Sons, Inc., New York, (1950).