SUBSET SELECTION PROCEDURES: REVIEW AND ASSESSMENT*

by

Shanti S. Gupta and S. Panchapakesan
Purdue University Southern Illinois University

Technical Report #84-4

Department of Statistics
Purdue University

February 1984
(Revised September 1984)

*This research was supported by the Office of Naval Research Contracts N00014-75-C-0455 and N00014-84-C-0167 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

SUBSET SELECTION PROCEDURES: REVIEW AND ASSESSMENT*

Shanti S. Gupta and S. Panchapakesan
Purdue University Southern Illinois University

1. INTRODUCTION

It is now over three decades since statistical inference problems were first posed in the now familiar selection and ranking framework. More than 700 papers have been published over these years in journals and proceedings of international conferences. During the last fifteen years, five books and a categorized bibliography have been published. Starting with a handful of researchers in the 1950's, the area of selection and ranking procedures has gained the attention of numerous active researchers today.

Selection and ranking problems have generally been studied using either the indifference zone approach of Bechhofer (1954) or the so-called subset selection approach due mainly to Gupta (1956). A comprehensive survey of significant contributions using these two approaches covering a span of almost thirty years is given in Gupta and Panchapakesan (1979). The present paper is mainly concerned with the subset selection approach. Our aim is to provide a historical perspective, trace the major developments that took place in the subset selection theory over the years 1955-1984 divided into three periods, indicate recent trends and directions for future research, and discuss the impact of research in this area. In doing so, we will not be concerned with details of procedures but only with the nature and trend of the

*This research was supported by the Office of Naval Research Contracts N00014-75-C-0455 and N00014-84-C-0167 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

developments in each period. The periods themselves serve more as a general reference to several phases of growth of the theory rather than as precise partition of the entire period.

2. HISTORICAL PERSPECTIVE

In many practical situations, the experimenter is faced with the problem of comparing k (≥ 2) populations. These may, for example, represent different varieties of wheat in an agricultural experiment, or different competing coherent systems in engineering models, or different drugs for a certain ailment. In all these problems, each population is characterized by the value of a parameter θ. In the above-mentioned examples, this parameter θ may be the average yield of a variety of wheat, or the reliability function of a system, or an appropriate measure of the effectiveness of a drug.

The classical approach in the preceding situations has been to test the so-called homogeneity hypothesis H_0 that $\theta_1 = \ldots = \theta_k$, where $\theta_1, \ldots, \theta_k$ are the (unknown) values of the parameter θ for the k populations. If the populations are assumed to be normal with means $\theta_1, \ldots, \theta_k$ and a common unknown variance σ^2 (which is a nuisance parameter), we have the familiar one-way classification model and the test can be carried out using Fisher's analysis of variance technique. However, this usually does not serve the experimenter's real purpose which is not just to accept or reject the homogeneity hypothesis. The real goal often is to identify the best population (the variety with the largest average yield, the most reliable system and so on). As Bechhofer noted in his now classical 1954 paper, the deficiencies of the ANOVA 'do not lie in the design aspects of the procedure but rather in the types of decisions which are

made on the basis of the data'. Of course it was recognized (see Cochran and Cox, 1950, p. 5) that the hypothesis that there is no difference between different treatments is unrealistic and that the real problem is to obtain estimates of the sizes of the differences between the treatments. However, the method of estimating the sizes of differences was often used as an indirect way of attempting to reach the goal of finding the best treatment or treatments. The attempts to formulate the decision problem to answer this realistic goal set the stage for the development of the selection and ranking theory.

The two main approaches that have been used in formulating a selection and ranking problem are familiarly known as the indifference zone approach and the subset selection approach. Suppose there are k populations $\pi_1,\ldots,\pi_k$ where π_i is characterized by the distribution function F_{θ_i}, $i = 1,\ldots,k$, where θ_i is a real-valued parameter with a value in the set Ω. It is assumed that the θ_i are unknown. Let us denote the ordered θ_i by $\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}$ and the (unknown) population π_i associated with $\theta_{[i]}$ by $\pi_{(i)}$, $i = 1,\ldots,k$. The populations are ranked according to their θ-values. To be specific, $\pi_{(j)}$ is defined to be better than $\pi_{(i)}$ ($\pi_{(i)} \prec \pi_{(j)}$) if $i < j$ (that is, $\theta_{[i]} \leq \theta_{[j]}$). The experimenter is presumed to have no prior information regarding the true pairing between $(\theta_1,\ldots,\theta_k)$ and $(\theta_{[1]},\ldots,\theta_{[k]})$. The basic problem in the indifference zone approach is to select one of the k populations with a guarantee that the probability of selecting the best population, called the probability of a correct selection (PCS), is at least P^* ($1/k < P^* < 1$) whenever $\delta(\theta_{[k]},\theta_{[k-1]}) \geq \delta^*$; here $\delta(\theta_{[k]},\theta_{[k-1]})$ is an appropriate measure of the separation of the best population $\pi_{(k)}$ and the next best population $\pi_{(k-1)}$. The constants P^* and δ^* are specified by the experimenter in advance. The statistical problem is to define a selection

rule which really contains three parts: sampling rule, stopping rule for sampling and decision rule. If the rule is based on a single sample of fixed size n from each population, then the minimum value of n is determined so that the specified minimum PCS can be guaranteed. As we stated above, this guarantee is to be met when $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)$ belongs to a part of the parameter space Ω, namely, $\Omega_{\delta^*} = \{\underset{\sim}{\theta}: \delta(\theta_{[k]},\theta_{[k-1]}) \geq \delta^*\}$. The region Ω_{δ^*} is called the preference zone. It should be noted that no requirement is made of the PCS when $\underset{\sim}{\theta}$ belongs to the complement of Ω_{δ^*}. It is this fact that led to the original label of 'indifference zone' approach. There are several variations and generalizations of the basic goal discussed above. For details, reference can be made to Gupta and Panchapakesan (1979).

In the subset selection approach for selecting the best population the goal is to select a nonempty subset of the k populations so that the best population is included in the selected subset with a minimum guaranteed probability $P^*(\frac{1}{k} < P^* < 1)$. Here the size of the selected subset is not determined in advance but by the data themselves. Selection of any subset consistent with the goal (here selecting the best population) is called a correct selection (CS) and the probability of a correct selection using a rule R is denoted by $P(CS|R)$. The requirement that

$$P(CS|R) \geq P^* \tag{1}$$

is referred to as the basic probability requirement, the P*-requirement or the P*-condition.

Denoting the (random) selected subset by S, the requirement (1) can be written in the form

$$\Pr(\pi_{(k)} \in S) \geq P^* \tag{2}$$

which brings out its similarity to the probability statement associated with a confidence interval procedure. While P* corresponds to the confidence

coefficient, the size of S, denoted by $|S|$, corresponds to the 'length' of the confidence interval. Thus any subset selection rule for 'constructing' S meets the criterion of validity by satisfying (1) or (2) and $|S|$ serves as a measure of the sensitivity or performance of the rule. It should also be emphasized that in the subset selection framework there is no indifference zone specification; the validity criterion or the P*-condition must be satisfied, whatever the configuration of the unknown parameters. The configuration of the parameters which yields the infimum of the probability of a correct selection (PCS) is referred to as the least favorable configuration (LFC).

Besides being a goal in itself, selecting a subset containing the best can also serve as a first-stage screening in a two-stage procedure designed to choose one population as the best; see, for example, Alam (1970), and Tamhane and Bechhofer (1977).

To point out some other differences between the indifference zone and the subset selection approaches, consider the problem of selecting the population associated with the largest mean from k normal populations with unknown means $\theta_1,\ldots,\theta_k$ and a common variance σ^2. When σ^2 is known, Bechhofer (1954) proposed a single stage procedure based on samples of size n each from the k populations. When σ^2 is not known, a two-stage procedure is necessary to guarantee the probability requirement using the indifference zone approach. On the other hand, one can solve the problem by single stage procedures for both cases in the subset selection approach. Also the subset approach can be used when the sample size $n \geq 2$ has already been chosen without regard to the type of analysis to be used for the data.

Besides the problem of selecting the best of k given populations, another problem that has been investigated from the early period is that of comparing k

experimental treatments (populations) with a standard or a control treatment. The goal is to select a subset of the experimental treatments that contains all treatments that are better than the standard or the control.

3. EARLY DEVELOPMENTS (1955-1965)

Early investigations of subset selection rules predictably centered around well-known parametric families of distributions, namely, normal, binomial and gamma. Gupta (1956) considered a procedure for selecting the population with the largest mean from k normal populations with means $\mu_1,\ldots,\mu_k$ and a common variance σ^2. He considered the case of known as well as unknown σ^2. Based on samples of size n from these populations, his rule in the case of known σ^2 is

$$R_1: \text{ Select } \pi_i \text{ if and only if } \bar{X}_i \geq \max_{1\leq j\leq k} \bar{X}_j - \frac{d\sigma}{\sqrt{n}}, \tag{3}$$

where $\bar{X}_i$ is the mean of the sample from π_i, $i = 1,\ldots,n$, and $d > 0$ is the smallest constant such that the probability requirement (1) is satisfied. The smallest constant d satisfying the requirement is given by

$$\inf_{\Omega} P(CS|R) = P^* \tag{4}$$

where Ω denotes the parametric space. When σ^2 is not known, the rule R_2 of Gupta (1956) is of the same form as R_1 except that σ is replaced by s, where s^2 is the usual pooled unbiased estimator of σ^2. Of course, the constant d will have a different value now.

For selecting the population with the largest scale parameter from k gamma populations with (unknown) scale parameters $\theta_1,\ldots,\theta_k$ and a common known shape parameter ν, Gupta (1963) proposed the rule

R_3: Select π_i if and only if $\bar{X}_i \geq c \max_{1 \leq j \leq k} \bar{X}_j$ (5)

where $\bar{X}_i$ is the mean of a random sample of size n from π_i, $i = 1,\ldots,n$, and $c \in (0,1)$ is to be determined to satisfy the P*-requirement.

The rules such as R_1, R_2, and R_3 are all referred to as Gupta's maximum type rules. Of course, these have their counterparts for the problem of selecting the population with the smallest parameter of interest. These maximum type rules have been investigated extensively in the literature; their optimal properties have also been studied.

As opposed to the maximum type rules, average type rules were proposed by Seal (1955, 1957, 1958a). In the case of selecting the normal population with the largest mean when the common variance σ^2 is unknown, the average-type rule is

R_4: Select π_i if and only if

$$\bar{X}_i \geq \sum_{r=1}^{k-1} c_r \bar{X}_{[r]}^{(i)} - st, \tag{6}$$

where $\bar{X}_{[1]}^{(i)} \leq \bar{X}_{[2]}^{(i)} \leq \ldots \leq \bar{X}_{[k-1]}^{(i)}$ denote the ordered sample means after deleting $\bar{X}_i (i = 1,\ldots,k)$, s^2 is the usual pooled unbiased estimator of σ^2, $c_1,\ldots,c_{k-1}$ are nonnegative constants subject to the constraint $\sum_{i=1}^{k-1} c_i = 1$, and $t = t(k,P^*,c_1,\ldots,c_{k-1})$ is chosen to satisfy the P*-requirement. However, as we will discuss later, the maximum type rules are found to be approximately Bayes optimal under reasonable loss functions. The additional simplicity in determining the constants associated with these rules makes them more appealing and useful.

The initial investigations of the rules for normal means, normal variances and gamma scale parameters were concerned with derivations of the properties of the rules such as monotonicity and of results relating to the supremum of the

expected size of the selected subset for these specific distributions.

The first paper in the direction of a unified treatment was by Gupta (1965) who treated subset selection in terms of location and scale parameters; a similar exposition for the indifference zone approach was given by Barr and Rizvi (1966a). It was assumed that the selection statistics used in the rule have distributions differing in a location or a scale parameter. Let T_i be the statistic associated with the sample from π_i $(i = 1,\ldots,k)$ with distribution $F(x-\theta_i)$ in the location case and $F(x/\theta_i)$ in the scale case; here θ_i's are the parameters to be ranked. The rules investigated by Gupta (1965) are R_5 (location case) and R_6 (scale case) given below.

R_5: Select π_i if and only if

$$T_i \geq \max_{1\leq j\leq k} T_j - d \tag{7}$$

and

R_6: Select π_i if and only if

$$T_i \geq c \max_{1\leq j\leq k} T_j \tag{8}$$

where $d > 0$ and $c \in (0,1)$ are to be determined so that the P*-requirement is satisfied.

Gupta (1965) showed that the infimum of the PCS in either case is attained when the parameters are equal and this infimum is independent of their common value. He also derived the following important properties of these procedures.

(1) The procedures are <u>monotone</u>, i.e., for $\theta_i > \theta_j$, the probability of including π_i in the selected subset is at least as large as that of including π_j.

(2) The probability of selecting the best population in the selected subset of size $|S|$ (not known in advance) is maximum among all possible subsets of size $|S|$.

(3) If the distribution $F(x,\theta)$ possesses a density $f(x,\theta)$ that has a monotone likelihood ratio in x, then the $E(|S|)$ is maximized over all parametric configurations when the θ_i are equal and this maximum is kP*.

For selecting the binomial population with the largest success probability, Gupta and Sobel (1960) proposed a location type rule. Let X_i be the number of successes in n trials associated with π_i, i = 1,...,k. Their rule is

R_7: Select π_i if and only if

$$X_i \geq \max_{1\leq j\leq k} X_j - d \tag{9}$$

where d is the smallest nonnegative integer for which the P*-requirement is met.

An interesting feature of Procedure R_7 is that the infimum of the PCS occurs when all the parameters are equal but it is not independent of their common value, say, p. For k = 2, Gupta and Sobel (1960) showed that the infimum of PCS over p occurs when $p = \frac{1}{2}$. When k > 2, the common value p for which this infimum takes place is not known. However, it is known that this common value $p \to \frac{1}{2}$ as $n \to \infty$. This difficulty regarding the infimum of the PCS led to the investigations of conditional selection rules which will be discussed in the next section.

The investigations of these early period were mainly under the assumption that the sample sizes are equal and that the nuisance parameters (such as σ_i^2 for the normal means problem) are equal.

Besides the problem of selecting the best of k given populations, procedures were proposed and investigated also for the problem of selecting a subset containing all the populations that are better than a control, and that of partitioning a set of populations with respect to a control. The early contributors are Bhattacharya (1956), Gupta and Sobel (1958), and Seal (1958b).

4. YEARS OF MAIN GROWTH (1965-1975)

The decade 1965-1975 witnessed a very significant growth of the ranking and selection theory, in general, and the subset selection theory, in particular. This period also marks the advent of the 'second generation' of researchers coming mostly out of Cornell University, University of Minnesota and Purdue University. The research during this period encompassed many facets of the subset selection theory. The main developments during this period can be broadly categorized into (i) unified results for the existing theory, (ii) generalizations and modifications in the formulation of the problem and the goal, (iii) decision-theoretic formulations, Bayes and empirical Bayes procedures, (iv) selection procedures for multivariate normal and multinomial populations, (v) development of conditional procedures, (vi) nonparametric procedures, (vii) selection from restricted families and (viii) sequential procedures. As one can see, many of the developments that took place in the theory had their beginnings in this period. We will discuss these briefly. For more details on these results, the reader is referred to Gupta and Panchapakesan (1979).

4.1 Unified Theory

In Section 3, we referred to Gupta (1965) who presented unified results for location and scale parameter families. These results were further generalized by Gupta (1966). This was followed by a more comprehensive unified theory by Gupta and Panchapakesan (1972). Let $\pi_1,\ldots,\pi_k$ have absolutely continuous distributions $F_{\theta_1},\ldots,F_{\theta_k}$, respectively, where the θ_i belong to an open interval Θ of the real line. It is assumed that $\{F_\theta\}$, $\theta \in \Theta$, is a stochastically increasing family in θ. Let $h \equiv h_{c,d}$, $c \in [1,\infty)$, $d \in [0,\infty)$ be a class of real-valued functions defined on the real line satisfying the following conditions: For every x belonging to the support of F_θ,

(i) $h_{c,d}(x) \geq x$, (ii) $h_{1,0}(x) = x$, (iii) $h_{c,d}(x)$ is continuous in c and d, and (iv) $\lim_{d\to\infty} h_{c,d}(x) = \infty$, c fixed, and/or $\lim_{c\to\infty} h_{c,d}(x) = \infty$, d fixed, $x \neq 0$.

Using the above class of functions h, Gupta and Panchapakesan (1972) considered the following class of procedures whose typical member is denoted by R_h.

R_h: Select the population π_i if and only if

$$h(x_i) \geq \max_{1\leq j\leq k} x_j, \tag{10}$$

where x_i is an observation from π_i, i = 1,...,k. The PCS is minimized when $\theta_1 = \ldots = \theta_k = \theta$. In general, the value of the PCS depends on θ. Under certain regularity conditions (see Gupta and Panchapakesan, 1979, p. 206) Gupta and Panchapakesan (1972) obtained a sufficient condition for the PCS to be monotonically increasing (or decreasing) in θ. When θ is a location or a scale parameter, the PCS is independent of θ. Gupta and Panchapakesan also obtained a sufficient condition for the supremum of the expected subset size to occur when the parameters are equal. This latter sufficient condition implies the one for the monotonicity of the PCS in θ. Besides the cases of location and scale parameters earlier discussed by Gupta (1965), the general results have been applied to the case where the density $f_\theta(x)$ is a convex mixture of the form $\sum_{j=0}^{\infty} w(\theta,j)g_j(x)$. Here $g_j(x)$, j = 0,1,..., is a sequence of density functions and the $w(\theta,j)$ are nonnegative weights such that $\sum_{j=0}^{\infty} w(\theta,j) = 1$. The results for the convex mixture directly apply to the procedures for selection from multivariate normal populations by Gupta and Panchapakesan (1969) in terms of the multiple correlation coefficient of one component with respect to the others, and by Gupta and Studden (1970) in terms of the Mahalanobis distance function. It should also be noted that

the class of functions h includes the usual choices made earlier, namely, $h(x) = cx$, $c \geq 1$, and $h(x) = x+d$, $d \geq 0$. The class also includes $h(x) = cx+d$, $c \geq 1$, $d \geq 0$, which was used by some authors later.

4.2 Generalizations and Modifications of Basic Formulation

Deverman and Gupta (1969) considered a generalization of the basic subset selection goal. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ be the ordered parameters of k populations. The populations associated with t largest θ_i's are the t best populations. Any subset of a fixed size s is called an s-subset. The goal is to select a subcollection from the $\binom{k}{s}$ s-subsets with a minimum guaranteed probability P* so that the chosen subcollection contains at least one s-subset having at least c of the t best populations. Obviously, for a meaningful problem, the integers c, s, t, and k must be such that $k \geq 2$ and $\max(1, s+t+1-k) \leq c \leq \min(s,t)$. Also, $P^* \geq \sum_{i=c}^{\min(s,t)} \binom{t}{i}\binom{k-i}{s-i}/\binom{k}{s}$. When $s = t = c = 1$, we get the basic problem of selecting a subset to contain the best.

In the basic formulation we select a nonempty subset of the k given populations. When the parameters θ_i are all very close to one another, we are likely to select all the populations. So it is meaningful to put a restriction that the size of the selected subset will not exceed m ($1 < m < k$). Even otherwise, one may want to select a nonempty subset of a random size subject to a maximum of m. Such a formulation is called a restricted subset size formulation. The general theory was developed by Santner (1973, 1975) and the normal means selection problem was investigated by Gupta and Santner (1973). An important feature of this formulation is that an indifference zone is introduced. The minimum guaranteed PCS is required when the parametric vector $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ belongs to the preference zone. The minimum sample size

n and the constant associated with the selection rule are to be determined. The general theory of Santner (1975) formally reduces to give the results of Bechhofer (1954) for m = 1 and, those of Gupta (1956, 1965) for m = k if the indifference zone is allowed to vanish.

To illustrate the restricted subset selection problem,consider k normal populations with unknown means $\mu_1,\ldots,\mu_k$ and a common known variance σ^2. We want to select a subset of size not exceeding m ($1 < m < k$) such that the best population (the one associated with $\mu_{[k]}$) is selected with a probability at least equal to P* whenever $\mu_{[k]} - \mu_{[k-1]} \geq \delta$ where $\delta > 0$ is specified in advance. The Gupta-Santner rule (1973) is

R_8: Select π_i if and only if

$$\bar{X}_i \geq \max\{\bar{X}_{[k-m+1]},\ \bar{X}_{[k]} - \frac{d\sigma}{\sqrt{n}}\} \tag{11}$$

where $\bar{X}_1,\ldots,\bar{X}_k$ and $\bar{X}_{[1]} \leq \ldots \leq \bar{X}_{[k]}$ are the unordered and the ordered sample means based on samples of size n. For a specified value of δ, d will depend on k, P*, and n.

Another modification is to relax the goal of selecting <u>the</u> best population. If $\theta_1,\ldots,\theta_k$ are the parameters to be ranked, one may be content with selecting populations that are nearly as good as the best (the one associated with $\theta_{[k]}$). Lehmann (1963a)used this idea, though not for a subset selection goal. Priority in introducing this concept goes to Fabian (1962) who defined a Δ-correct ranking for the problem of Bechhofer (1954). Let us consider the case of location parameters. Lehmann (1963a)defined a good population as any population π_i for which $\theta_i \geq \theta_{[k]} - \Delta$, $\Delta > 0$. Desu (1970) defined superior and inferior populations by $\theta_i \geq \theta_{[k]} - \Delta_1$ and $\theta_i \leq \theta_{[k]} - \Delta_2$, respectively, where $0 < \Delta_1 < \Delta_2$. His goal is to select a nonempty subset of the k given populations that excludes all inferior populations with a

minimum guaranteed probability P*. The performance of a procedure satisfying the P*-requirement can be evaluated, for example, by the expected number of superior populations included in the selected subset. Carroll, Gupta and Huang (1975) considered eliminating inferior populations with respect to the t best, i.e., those π_i's for which $\theta_i \leq \theta_{[k-t+1]} - \Delta$, $\Delta > 0$. They called these populations strictly non-t-best. These definitions are modified in an obvious way to handle scale parameters. Panchapakesan and Santner (1977) introduced a generalization by defining a good population relative to the t-th best as one for which $\theta_i \geq p(\theta_{[k-t+1]})$ where p is a function possessing certain general properties. They considered two goals: (i) selecting a nonempty subset containing only good ones, and (ii) selecting a subset whose size does not exceed m ($1 < m < k$) and which will include at least one good population. Their treatment complements the unified results of Gupta and Panchapakesan (1972) and Santner (1975).

4.3 Decision-theoretic formulation; Bayes and empirical Bayes Procedures

During the period of main growth, early contributions to the decision-theoretic formulation were also made, and some Bayes and empirical Bayes procedures were derived. These early contributions may be considered modest compared to the growth of the literature on the classical procedures during this period. However, they gave the impetus to the developments that would follow in the subsequent periods.

To describe the decision-theoretic setup, let π_i ($i = 1,\ldots,k$) be characterized by the probability space $(\mathcal{X}, \mathcal{A}, P_i)$, where P_i belongs to some family $\mathcal{P}$ of probability measures. Let us assume that the family $\mathcal{P}$ is stochastically ordered; in other words, there is a stochastic ordering between any pair (P_i, P_j) from $\mathcal{P}$. The stochastically largest among

$\pi_1,\ldots,\pi_k$ is the best population. In the case of more than one contender, we assume that one of them is tagged as the best. We observe $\underset{\sim}{X} = (X_1,\ldots,X_k)$ where X_i is an observation from π_i, $i = 1,\ldots,k$. The space of the observation $\underset{\sim}{X}$ is $\mathcal{X}^k = \{\underset{\sim}{x} = (x_1,\ldots,x_k): x_i \in \mathcal{X},\ i = 1,\ldots,k\}$. The decision space $\mathcal{D}$ consists of the 2^k subsets d of the set $\{1,2,\ldots,k\}$; in other words, $\mathcal{D} = \{d \mid d \subseteq \{1,\ldots,k\}\}$. Thus a decision d corresponds to the selection of a subset (possibly the empty set) of the k given populations. Any decision $d \in \mathcal{D}$ is a correct selection if $j \in d$ where π_j is the best population. A selection procedure is a measurable function δ defined on $\mathcal{X}^k \times \mathcal{D}$ such that for each $\underset{\sim}{x} \in \mathcal{X}^k$, we have $\delta(\underset{\sim}{x},d) \geq 0$ for any $d \in \mathcal{D}$ and $\sum_{d\in\mathcal{D}} \delta(\underset{\sim}{x},d) = 1$. Here $\delta(\underset{\sim}{x},d)$ is the probability that the subset d is selected when $\underset{\sim}{x}$ is observed. The individual selection probability $p_i(\underset{\sim}{x})$ for the population π_i is given by $p_i(\underset{\sim}{x}) = \sum_{d\ni i} \delta(\underset{\sim}{x},d)$, the summation being over all subsets that contain i. While, in general, the individual selection probabilities do not uniquely determine the selection procedure $\delta(\underset{\sim}{x},d)$, they do so when the $p_i(\underset{\sim}{x})$ take on only values 0 and 1 (see Gupta and Panchapakesan, 1979, p. 212).

Studden (1967) studied optimum selection rules assuming that $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)$ is a permutation of a k-vector of known elements. He assumed a loss function $L(\underset{\sim}{\theta},d) = \sum_{i\in d} L_i(\underset{\sim}{\theta}) + L(1-I)$, where $L_i(\underset{\sim}{\theta})$ is the loss whenever π_i is selected and I = 1 or 0 according as a correct selection is or is not made; here, L is a positive constant. This loss function is also assumed to be permutation-invariant. Studden (1967) obtained the best (in the sense of minimizing the risk) invariant selection rule.

Nagel (1970) defined a concept of <u>just</u> selection rules. His definition [see also Gupta and Nagel (1971)] involves a partial ordering defined on $\mathcal{X}$ and is applicable when the dimension of $\mathcal{X}$ is one or more. In the one-dimensional case, a selection rule R, defined by its individual selection probabilities

$p_i(\underset{\sim}{x})$, $i = 1,\ldots,k$, is called just if and only if the following holds:

$$x_i \leq y_i,\ x_j \geq y_j \text{ for all } j \neq i \Rightarrow p_i(\underset{\sim}{y}) \geq p_i(\underset{\sim}{x}).$$

Let Ω denote the space of the parameter vector $\underset{\sim}{\theta}$ and Ω_0 denote the part of Ω in which all the parameters are equal. Nagel (1970) showed that, under stochastic ordering on the parameter space, for any just rule R

$$\inf_{\Omega} P(CS|R) = \inf_{\Omega_0} P(CS|R). \tag{12}$$

The result in (12) is a reasonable property to be expected of a rule. Nagel also showed that a permutation-invariant just rule is monotone.

Deely and Gupta (1968) obtained Bayes procedures considering linear loss functions of the type

$$L(S,\theta) = \sum_{j\in S} a_j(\theta_{[k]}-\theta_j) \tag{13}$$

where S denotes the set of indices of the selected populations. Deely (1965) investigated empirical Bayes procedures and derived these procedures in several special cases.

4.4 Selection Procedures for Multivariate Normal and Multinomial Distributions

Several problems were investigated relating to the best component of the mean vector of a single multivariate normal population and the best of several multivariate normal populations. For ranking multivariate normal populations several criteria were considered, such as the Mahalanobis distance function (Alam and Rizvi, 1966; Gupta, 1966; Gupta and Studden, 1970), generalized variance (Gnanadesikan and Gupta, 1970), and multiple correlation coefficient between a particular component and the remaining ones (Gupta and Panchapakesan, 1969). However, in some of these problems the exact infimum of the PCS was not derived in general. For selecting the best component of a single

multivariate normal population, Gnanadesikan (1966) considered a location type procedure based on sample component means. Except in the case of bivariate normal, only a lower bound of the PCS is used to obtain a conservative value of the constant to be used in the procedure even in the case of known correlation matrix Σ. The difficulty is due to the fact that the association between the ranked components and the known correlations is unknown. If we assume that the components have the same variance and are equally correlated with correlation $\rho > 0$, then the exact solution is available (Gupta, Nagel and Panchapakesan, 1973). For selecting the best of several p-variate normal distributions, $N(\underset{\sim}{\mu}_i, \Sigma_i)$, $i = 1,\ldots,k$, in terms of the Mahalanobis distance function, Alam and Rizvi (1966), Gupta (1966), and Gupta and Studden (1970) proposed procedures when the covariance matrices Σ_i are known as well as when they are unknown. The special case of common Σ ($\Sigma_i = \Sigma$ for all i) was not solved. This was later solved in an approximate sense by Chattopadhyay (1981). A few other measures were considered (Frischtak, 1973; Gnanadesikan, 1966) for ranking multivariate normal populations but the results in these cases are very limited in scope or are asymptotic in nature. For selecting the populations better than a standard, Krishnaiah and Rizvi (1966) considered, as criteria, linear combinations of the elements of mean vectors and distance functions whereas Krishnaiah (1967) considered linear combinations of the elements of the covariance matrices.

For selecting the most (or the least) probable cell in a multinomial distribution, Gupta and Nagel (1967) proposed a single stage procedure. Let $X_1,\ldots,X_k$ denote the cell counts based on n independent observations from a k-cell multinomial distribution with unknown cell probabilities $p_1,\ldots,p_k$. Gupta and Nagel (1967) proposed and investigated the following rules R_9 and R_{10} for selecting the cell with the largest and the smallest p_i, respectively.

R_9: Select π_i if and only if

$$X_i \geq \max_{1 \leq j \leq k} X_j - D \tag{14}$$

where $D = D(k,n,P^*)$ is the smallest nonnegative integer for which the P^*-condition is satisfied.

R_{10}: Select π_i if and only if

$$X_i \leq \min_{1 \leq j \leq k} X_j + C \tag{15}$$

where $C = C(k,n,P^*)$ is the smallest nonnegative integer for which the P^*-condition is satisfied.

The first interesting point to emerge about R_9 and R_{10} is that, unlike in earlier problems such as normal means, normal variances, etc., the analysis in the minimum case does not exactly parallel that in the maximum case. Also, for $k > 2$, the LFC was not completely determined. Gupta and Nagel (1967) showed that the LFC (in terms of the ordered p_i) is of the type $(0,\ldots,0,s,p,\ldots,p)$, $s \leq p$, in the case of R_9 and is of the type $(p,\ldots,p,q)$, $p \leq q$, in the case of R_{10}. An alternative to R_9 is the inverse sampling selection rule of Panchapakesan (1971, 1973) for which the infimum of the PCS occurs when all the cell probabilities are equal.

Multinomial selection rules are also important in the sense that they provide distribution-free procedures. Let $\pi_1,\ldots,\pi_k$ have continuous distributions F_{θ_i}, $i = 1,\ldots,k$, which belong to a stochastically increasing family in θ. Let p_i denote the probability that in a set of k observations, one from each distribution, the observation from π_i is the largest, $i = 1,\ldots,k$. Selecting the stochastically largest population is equivalent to selecting the population associated with the largest p_i. By taking observations, vector at a time,

and noting which population yielded the largest observation, we can convert the problem, in an obvious manner, to that of selecting the most probable multinomial cell.

4.5 Conditional Selection Procedures

In Section 3, we noted that, for the Gupta-Sobel rule [R_7 defined by (9)], the infimum of the PCS occurs when all the success probabilities associated with the k binomial populations are equal to p, but this common value p at which the infimum takes place is not known when $k > 2$. Thus, there was no result available earlier that gave a reasonable conservative value of the associated constant for any given n. Similar difficulties arise also with procedures for Poisson populations. There are also a few other interesting points about the usual procedures in this case. Let us briefly mention them here.

Let $X_1,\ldots,X_k$ denote the numbers of occurrences from k Poisson populations with parameters $\lambda_1,\ldots,\lambda_k$, respectively. Suppose we want to select the population with the largest λ_i. Here the usual location and scale type procedures cannot be found to satify the P*-condition for all permissible values of P*. Gupta and Huang (1975a) proposed a modified procedure given below.

R_{11}: Select π_i if and only if

$$X_i+1 \geq c \max_{1\leq j\leq k} X_j \tag{16}$$

where $0 < c = c(k,P^*) < 1$ is to be chosen subject to the P*-requirement. The motivation behind this procedure comes from a result of Chapman (1952) which says that there is no unbiased estimator of λ_1/λ_2 but $X_1/(X_2+1)$ is "almost unbiased." Gupta and Huang (1975a) have shown that the infimum of the PCS occurs when $\lambda_1 =\ldots= \lambda_k = \lambda$; however, the common value λ at which the infimum occurs is not established.

Since the common value of the parameters at which the infimum of the PCS occurs is not known for these rules for the binomial and Poisson populations, the natural question is: Can we find conservative values for the constants defining the procedures? An answer in the affirmative follows from the use of conditional selection rules which form a part of the important contributions of the period under review.

Gupta and Nagel (1971) first proposed conditional subset selection rules in the case of binomial, Poisson, and negative binomial populations. Since their rules are randomized just rules, they satisfy (12). For selecting the binomial population with the largest success probability, their rule specified by the individual selection probabilities is as follows.

R_{12}: Select the population characterized by success probability θ_i with probability

$$p_i(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } x_i > c_t \\ \rho & \text{if } x_i = c_t, \ i = 1,\ldots,k, \\ 0 & \text{if } x_i < c_t \end{cases} \tag{17}$$

with $t = x_1 +\ldots+ x_k$ and the constants ρ and c_t are determined to satisfy

$$E(p_k(\underset{\sim}{X}) | T = t) = P^* \tag{18}$$

where $T = X_1 +\ldots+ X_k$. The important fact to note about R_{12} and similar Gupta-Nagel randomized procedures for Poisson and negative binomial populations is that the infimum of the PCS occurs when the parameters under consideration of the k populations are equal and the constant associated with the rule (depending on the value of the statistic T on which the conditioning is done) is independent of the common value of the parameters.

For the binomial selection problem Gupta, Huang and Huang (1976) proposed a nonrandomized conditional rule

R_{13}: Select π_i if and only if

$$X_i \geq \max_{1\leq j\leq k} X_j - D(t), \tag{19}$$

given $T \equiv \sum_{i=1}^{k} X_i = t$, where $D(t) > 0$ is to be chosen to satisfy the P*-condition. Exact result for the infimum of the PCS is obtained for only $k = 2$; in this case, the infimum is attained when $p_1 = p_2 = p$ and is independent of the common value p. For $k > 2$, Gupta, Huang and Huang (1976) obtained a conservative value for $D(t)$. They also obtained a conservative value for the constant d of the unconditional rule R_7 in Section 3. It should be noted that, in using the conditioning argument to obtain a conservative value of d, one can always guarantee the P*-condition. The values of the constant d tabulated by Gupta and Sobel (1960) for $k \geq 3$ are based on normal approximation and thus may lead to a drop of the PCS below P*.

Conditional rules for Poisson populations were given by Gupta and Huang (1975a). These are similar to R_{11} given by (16) with c(t) in place of c, given $T \equiv \sum_{i=1}^{k} X_i = t$. It is well known that, if $X_1,\ldots,X_k$ are independent Poisson variables with parameters $\lambda_1,\ldots,\lambda_k$, respectively, then the conditional joint distribution of $X_1,\ldots,X_k$ given $X_1 +\ldots+ X_k = N$ is multinomial with cell-probabilities $p_i = \lambda_i/\Sigma\lambda_i$. The conditional selection rule for Poisson was exploited by Gupta and Huang (1975a) to provide a selection rule for selecting the most probable multinomial cell which selects the cell π_i if and only if $X_i+1 \geq c \max_{1\leq j\leq k} X_j$, where $c = c(k,N,P^*) \in (0,1)$. A conservative value of c can be obtained from their results for the conditional selection rule for Poisson populations.

Although conditioning on $T = \Sigma X_i$ is used in the preceding procedures, the ultimate purpose is to guarantee a specified minimum unconditional probability of a correct selection. This is somewhat different in spirit from the approach of Kiefer (1977a) who has discussed selection and ranking problems

(using indifference zone as well as subset selection approach) to illustrate applications of his general theory of conditional confidence and estimated confidence in multidecision problems. For details, refer to two excellent papers by Kiefer (1977a,b).

4.6 Nonparametric Procedures

The first nonparametric subset selection procedure was studied by Rizvi and Sobel (1967) for the problem of selecting the population having the largest α-quantile ($0 < \alpha < 1$) from k populations having continuous distributions. Assume that the size n of the sample from each population is sufficiently large so that $1 \leq (n+1)\alpha \leq n$ and define a positive integer r by the inequalities $r \leq (n+1)\alpha < r+1$. This implies that $1 \leq r \leq n$. Let $Y_{j,i}$ denote the jth order statistic from π_i, $j = 1,\ldots,n$; $i = 1,\ldots,k$. The procedure proposed by Rizvi and Sobel (1967) is interesting in the sense that it differs from the usual maximum type. Their rule is

R_{14}: Select π_i if and only if

$$Y_{r,i} \geq \max_{1\leq j\leq k} Y_{r-c,j} \tag{20}$$

where c is the smallest integer with $1 \leq c \leq r-1$ for which the P*-condition is satisfied. However, a c-value satisfying the P*-condition exists only if a permissible value of P* does not exceed a value P_1 depending on n, α, and k. The procedure is monotone and the expected subset size is maximized when all the distributions are identical. Nonparametric procedures for the related problem of comparison with a control were studied by Rizvi, Sobel and Woodworth (1968).

Gupta and McDonald (1970) assumed that the distributions F_i, $i = 1,\ldots,k$, belong to a location or a scale parameter family. Let R_{ij}, $j = 1,\ldots,n_i$, denote the ranks of n_i independent observations taken from F_i, $i = 1,\ldots,k$, in the pooled sample of $N = n_1 +\ldots+ n_k$ observations. Let $Z(1) \leq Z(2) \leq\ldots\leq Z(N)$ denote an ordered sample of size N from any continuous distribution G such that $-\infty < a(r) = E(Z(r)|G) < \infty$, $r = 1,\ldots,N$. Define $H_i = n_i^{-1} \sum_{j=1}^{n_i} a(R_{ij})$, $i = 1,\ldots,k$. For selecting the population associated with the largest parameter, Gupta and McDonald (1970) proposed three procedures based on H_i's. Two of these are the usual maximum type procedures, one for the location case and the other for the scale case. The best that can be said about the LFC for these procedures is that it occurs when $\theta_{[k-1]} = \theta_{[k]}$. In general, the LFC for these procedures is not an equi-parameter one unless $k = 2$. It was inadvertently claimed to be so by some authors earlier. The same difficulty arises in the indifference zone formulation. Formal counterexamples were given by Rizvi and Woodworth (1970). Gupta and McDonald (1970) gave bounds for the infimum of the PCS. The third procedure of Gupta and McDonald (1970) is

R_{15}: Select π_i if and only if

$$H_i \geq d. \tag{21}$$

For this rule, the infimum of the PCS is attained when $\theta_1 =\ldots= \theta_k$; however, this rule may select an empty subset unless P* is sufficiently large. Some other related papers include McDonald (1972, 1974), Blumenthal and Patterson (1969), and Puri and Puri (1968, 1969).

If we have distributions from a location parameter family, we can use procedures based on one-sample Hodges-Lehmann estimators. For these procedures, the LFC can be determined. Ghosh (1973) studied such procedures with goals involving selection of a fixed number of populations. Gupta and D.-Y. Huang (1974) studied such a procedure for the goal of eliminating populations which are strictly inferior to the t best.

A review of the procedures described above and a few other related results are given by Gupta and McDonald (1982).

4.7 Selection from Restricted Families

A restricted family of probability distributions is defined by a partial order relation with respect to a known distribution. Such families provide characterizations of life-length distributions and thus are very important in reliability studies. Selection rules for such restricted families were first considered by Barlow and Gupta (1969). In order to make our discussion of the selection procedures for these families adequately self-contained, we will define the partial orderings that have been used. For more details and related references, the reader is referred to Gupta and Panchapakesan (1979, Chapter 16).

We define the following partial order ($\underset{\sim}{\prec}$) relations for two distributions F and G assumed to be absolutely continuous.

Definitions 4.1. (1) F is said to be convex with respect to G ($F \underset{c}{\prec} G$) if and only if $G^{-1}F(x)$ is convex on the support of F.

(2) F is said to be star-shaped with respect to G ($F \underset{*}{\prec} G$) if and only if $F(0) = G(0) = 0$, $G^{-1}F(x)/x$ is increasing in $x \geq 0$ on the support of F.

(3) F is said to be r-ordered with respect to G ($F \underset{r}{\prec} G$) if and only if $F(0) = G(0) = \frac{1}{2}$ and $G^{-1}F(x)/x$ is increasing (decreasing) in x positive (negative).

(4) F is said to be tail-ordered with respect to G ($F \underset{t}{\prec} G$) if and only if $F(0) = G(0) = \frac{1}{2}$ and $G^{-1}F(x)-x$ is increasing on the support of F.

It is known that convex ordering implies star-ordering. Further, when $G(x) = 1-e^{-x} (x \geq 0)$, $F \underset{c}{\prec} G$ is equivalent to saying that F has an increasing failure rate (IFR) and $F \underset{*}{\prec} G$ is equivalent to saying that F has an increasing failure rate on the average (IFRA). Of course, if F is IFR, then it is also IFRA.

Let $\pi_1,\ldots,\pi_k$ have the associated absolutely continuous distributions $F_1,\ldots, F_k$, respectively. The best population is defined in terms of a characteristic such as the mean or quantile of a given order. Let $F_{[k]}$ denote the distribution function of the best population. We assume that $F_{[k]}$ is stochastically larger than the rest and that $F_i \precsim G$, $i = 1,\ldots,k$.

Under the above setup, Barlow and Gupta (1969) proposed a procedure for selecting the population having the largest α-quantile ($0 < \alpha < 1$) when all the F_i are star-shaped with respect to a known G. Let T_{ji} denote the jth order statistic based on n independent observations from π_i, $i = 1,\ldots,k$, where $j \leq (n+1)\alpha < j+1$. The Barlow-Gupta procedure (1969) is

R_{16}: Select π_i if and only if

$$T_{j,i} \geq c \max_{1\leq r\leq k} T_{j,r} \tag{22}$$

where $c = c(k, P^*, n, j)$ is the largest constant in (0,1) for which the P*-condition is satisfied. Tables for the constant c and the constant for the analogous procedure for selecting the population with the smallest α-quantile are given by Barlow, Gupta and Panchapakesan (1969) for the special case of exponential G,i.e. for the IFRA family of distributions. Another interesting special case of G is the half-normal obtained by folding $N(0,\sigma^2)$ at the origin, where σ is assumed to be known. The class of distributions which are star-shaped with respect to this folded normal is a subclass of IFRA distributions. Selection in terms of quantiles for this restricted family was considered by Gupta and Panchapakesan (1975).

Barlow and Gupta (1969) considered also the selection of the population with the largest median from a set of distributions that have lighter tails than a specified G. The definition of F_i having a lighter tail than G used by them implies that F_i centered at its median is tail-ordered with respect to G. The procedure of Barlow and Gupta (1969) applies equally to the

problem of selecting the population with the largest median from a set of populations which, centered at their respective medians, are tail-ordered with respect to G. This is discussed by Gupta and Panchapakesan (1974), who also provide the values of the constant when G is the logistic distribution, that is, $G(x) = [1+e^{-x}]^{-1}$.

Some important unified results were obtained by Gupta and Panchapakesan (1974). They defined a general partial order relation called $\mathcal{H}$-ordering through a class of functions $\mathcal{H} = \{h\}$ and discussed a related selection problem. The $\mathcal{H}$-ordering includes the star- and tail-orderings as special cases. The selection rule is of the type R_h defined in (10) using a member h of $\mathcal{H}$. In Section 4.1, the general results of Gupta (1966) were mentioned. His third special case, besides the location and scale parameter cases, is really the case of a restricted family using $\mathcal{H}$-ordering, even though the description there is not in terms of the partial ordering.

4.8 Sequential Procedures

Barron (1968) and Barron and Gupta (1972) investigated a noneliminating type sequential procedure for selecting the population with the largest mean from k normal populations with unknown means $\theta_1,\ldots,\theta_k$ and common known variance σ^2. However, it was assumed that the successive differences of the ordered θ_i are known. The sampling for their procedure is done by taking one observation from each population at each stage. At any stage, each population that has not been so far classified as accepted or rejected, is subject to one of three possible decisions: accept, reject, or postpone classification. Sampling continues until all the populations are classified either as accepted or as rejected. All populations that are accepted constitute the selected subset. It should be noted that until all populations are classified, the sampling is made from all populations, previously classified or not.

Swanepoel and Geertsema (1973) considered a sequential procedure for selecting the normal population with the largest mean from k populations, $N(\theta_i,\sigma_i^2)$, where all the parameters are unknown. They defined a selection sequence using the idea of a confidence sequence introduced by Robbins (1970). For each $n \geq 1$, let B_n denote a subset of the k populations defined by n observations from each population. Any sequence $\{B_n\}$ is a <u>selection sequence</u> if

$$\Pr(\pi_{(k)} \in B_n \text{ for all } n \geq 1) \geq P^* \tag{23}$$

for all $\theta_1,\ldots,\theta_k$. Let $|S(n)|$ denote the size of the subset B_n and let r denote the number of populations tied for the best population. Then, for the selection sequence defined by Swanepoel and Geertsema, $|S(n)| \to r$ a.s. (almost surely) as $n \to \infty$, and $B_n = \{\pi_{(k-r+1)},\ldots,\pi_k\}$ a.s. for large n. The size of the selected subset can be restricted to a maximum of m ($1 \leq m < k$) by defining a stopping variable N as the first integer $n \geq 1$ such that $|S(n)| \leq m$. If $r \leq m$, then $N < \infty$ a.s. and the subset B_N (which contains at most m populations) includes the best population with a minimum probability P*. However, if $r \geq m+1$, then $N = \infty$ with positive probability.

Gupta and Huang (1975b) discussed three sequential procedures of which two are parametric and the third is nonparametric. The nonparametric and one of the parametric procedures are of the nonelimination type. The goal of their parametric procedures is to select what they called mildly t best populations. Suppose that $\theta_1,\ldots,\theta_k$ are unknown location parameters of k given populations. Then π_i is called <u>mildly t best</u> if $\theta_i \geq \theta_{[k-t+1]} - \Delta$, where $\Delta > 0$ is specified. For $t = 1$, π_i has been called a superior population by Desu (1970) and a good population by Lehmann (1963). Gupta and Huang (1975b) have discussed their procedures in a general setup and obtained special results for selecting from normal populations in terms of their means or variances. Their nonparametric procedure is for selecting the population with the largest location parameter when the k populations have

absolutely continuous distributions $F(x-\theta_i)$, $i = 1,\ldots,k$. It is assumed that $F(\cdot)$ is symmetric about the origin and that the densities have monotone likelihood ratio property and are differentiable almost everywhere. Carroll (1974) has discussed some asymptotically nonparametric sequential selection procedures.

4.9 Other Developments

In the early investigations, detailed results were obtained only for procedures which used samples of a common size from the populations under consideration. Also, in the case of selection in terms of the means from k normal populations, the early investigations assumed equal variances. When the variances are not equal (that is, under heteroscedasticity), the only trivial case is when they are all known and the sample sizes are the same. To handle various other situations that arise, several procedures were proposed and investigated.

Let $\pi_1,\ldots,\pi_k$ be k normal populations with unknown means $\theta_1,\ldots,\theta_k$ and (known or unknown) variances $\sigma_1^2,\ldots,\sigma_k^2$. Let $\bar{X}_i$ and s_i^2 denote the mean and the variance (divisor n_i-1) of a random sample of size n_i from π_i, $i = 1,\ldots,k$. Let $s^2 = \sum_{i=1}^{k}(n_i-1)s_i^2/(N-k)$, where $N = \sum_{i=1}^{k} n_i$.

Let us first consider the case of known variances. In describing various procedures, we have used the same letter d to denote the constant in each case. This constant d is the smallest positive constant for which the P*-condition is satisfied. Also, if σ rather than σ_i appears, it it assumed that $\sigma_1 = \ldots = \sigma_k = \sigma$. Gupta and W.-T. Huang (1974) proposed the rule

R_{17}: Select π_i if and only if $\bar{X}_i \geq \max_{1\leq j\leq k} \bar{X}_j - \frac{d\sigma}{\sqrt{n_i}}$.

Later Gupta and D. Y. Huang (1976a) proposed the rule

R_{18}: Select π_i if and only if $\bar{X}_i \geq \max_{1\leq j\leq k} (\bar{X}_i - d\sigma\sqrt{\frac{1}{n_i}+\frac{1}{n_j}})$.

For the case of unequal variances, Gupta and Wong (1976) proposed the rule

R_{19}: Select π_i if and only if $\bar{X}_i \geq \max_{1\leq j\leq k} (\bar{X}_j - d\sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}})$.

Chen, Dudewicz and Lee (1976) proposed a rule assuming σ to be unknown. In the case of known σ, their rule would be

R_{20}: Select π_i if and only if $\bar{X}_i \geq \max_{1\leq j\leq k} - d\sigma\sqrt{\frac{1}{n_i} + \frac{1}{a}}$

where the constant a is nonnegative and is usually chosen between $n_{[1]}$ and $n_{[k]}$, both inclusive.

For all the above procedures (R_{17} through R_{20}), the respective authors have obtained lower bounds for the infimum of the PCS. For k = 2, Gupta and D.-Y. Huang (1976a) have shown R_{18} to be more efficient than R_{17} in terms of the supremum of the expected subset size. Berger and Gupta (1980) considered minimax subset selection rules using the criterion $M = \max_{1\leq j\leq k-1} P(\pi_{(j)}$ is selected). They have shown that R_{18}, and R_{19} (when $\sigma_1 = \ldots = \sigma_k = \sigma$) are minimax with respect to M in the class of nonrandomized, just, and translation invariant rules which satisfy the P*-condition. The rules R_{17} and R_{20} are not minimax, in general.

Next, let us consider the case of unknown variances. The counterparts of the rules R_{17}, R_{18} and R_{20} were proposed by the respective authors where the new rules R'_{17}, R'_{18} and R'_{20} were of similar forms with s replacing σ. Of course, the constant d will have a different value in each case. Gupta and Wong (1982) proposed a rule R'_{19} which selects π_i if and only if

$$\bar{X}_i \geq \max_{1\leq j\leq k} \bar{X}_j - c \max_{1\leq j\leq k} s_i.$$

As in the case of known σ_i's, these authors have given only lower bounds for the infimum of the PCS. Some comparisons of R'_{17}, R'_{18} and R'_{20} are given by Chen, Dudewicz and Lee (1976).

When the variances are unknown and unequal, and the sample sizes are unequal, Dudewicz and Dalal (1975) proposed a two-stage procedure for selecting the population with the largest mean under the indifference zone

formulation. Let n_0 be the first-stage sample size for each population, and $n_i - n_0$ the second-stage sample size from π_i, $i = 1,\ldots,k$. Their procedure is based on the statistics $\tilde{\tilde{X}}_i$, where $\tilde{\tilde{X}}_i$ is a weighted average based on all the n_i observations from π_i, $i = 1,\ldots,k$. The weights are chosen subject to certain conditions. They also proposed the following subset selection rule

$$R_{21}: \text{ Select } \pi_i \text{ if and only if } \tilde{\tilde{X}}_i \geq \max_{1\leq j\leq k} \tilde{\tilde{X}}_j - d$$

where $d \geq 0$. For this procedure, the P*-condition is satisfied irrespective of the choice of the positive constant d. So one has to impose some additional restriction in order to have a meaningful choice of d. One possibility is to introduce a restriction on the expected subset size in some configuration of the means. It is worth noting that a two-stage procedure is not necessary for satisfying the P*-condition in the subset selection approach, whereas it is necessary for the indifference zone approach.

Among other problems studied during this period is the selection from k normal populations, $N(\mu_i,\sigma_i^2)$, $i = 1,\ldots,k$, when the parameters to be ranked are $\theta_i = |\mu_i|/\sigma_i$, $i = 1,\ldots,k$. The parameters θ_i are the signal-to-noise ratios if the populations represent different communication systems. For this problem, Rizvi (1971) studied a location-type procedure (of the type R_1 in (1)) based on the absolute sample means. His main results deal with the case of $\sigma_1^2 = \ldots = \sigma_k^2 = \sigma^2$ (known or unknown) and samples of equal sizes.

5. YEARS OF FURTHER STRIDES (1975-1980)

Of the many important contributions made during 1975-1980, the foremost concerned the development of the decision-theoretic approach to subset selection. Besides Bayes procedures, several minimax and Γ-minimax rules were derived. The first paper on locally optimal rules also appeared during this period. Several contributions were also made to classical procedures for specific families of distributions.

5.1 Bayes Procedures

In Section 4.3, we discussed the early developments of Bayes procedures using linear loss functions. The first papers to appear with nonlinear loss functions are Bickel and Yahav (1977), Chernoff and Yahav (1977), Goel and Rubin (1977), and Gupta and Hsu (1978). They use different linear combinations of four components of loss, namely, (i) the simple loss due to an incorrect selection, which takes value 0 or 1 according as a correct selection is or is not made, (ii) $|S|$, the size of the selected subset, (iii) $\theta_{[k]} - \max_{j\in S} \theta_j$, which expresses a measure of loss in using in the future the populations that are selected, and (iii) $\theta_{[k]} - \sum_{j\in S} \theta_j/|S|$, which is an 'average' loss in using in the future the populations that are selected. The components used in the linear combination are: (i) and (iv) by Bickel and Yahav (1977), (iii) and (iv) by Chernoff and Yahav (1977), (ii) and (iii) by Goel and Rubin (1977), and (i) and (ii) by Gupta and Hsu (1978).

Goel and Rubin (1977) assumed that the k distributions had densities and belonged to a family with montone likelihood ratio property. The vector $\underset{\sim}{\theta}$ of the k parameters was assumed to have an exchangeable prior distribution. They derived the Bayes rule under this setup and obtained further simplifications in the case of the prior distribution of $\underset{\sim}{\theta}$ being a mixture of i.i.d. random variables. They also obtained an 'approximate' Bayes rule R, which selects larger subsets than the Bayes rule but is the Bayes rule for k = 2. This approximate Bayes rule, under a further assumption regarding the form of the posterior distribution of θ_i, reduces to the classical procedure of Gupta (1956) except that the constant that is involved depends only on the cost per population. Goel and Rubin (1977) also discussed the special case of normal populations, $N(\theta_i,\sigma^2)$, $i = 1,\ldots,k$, where σ^2 is known and $\underset{\sim}{\theta}$ has an exchangeable multivariate normal prior for all k.

Chernoff and Yahav (1977) considered selecting the population with the largest mean from k normal populations with means $\theta_1,\ldots,\theta_k$ and a common known variance σ^2, where $\underset{\sim}{\theta}$ has an exchangeable normal prior. They compared their Bayes procedure with the (random size) subset selection procedure of Gupta (1956) and the fixed size subset selection procedure of Desu and Sobel (1968). Their results were based on Monte Carlo studies of the integrated risks with respective to different exchangeable normal priors.

Bickel and Yahav (1977) also considered k normal populations with means $\theta_1,\ldots,\theta_k$ and a common known variance σ^2. They investigated the optimal solution when k goes to infinity under the assumption that the "empirical distribution " of the means θ_i, i = 1,...,k, converge in a suitable sense to a smooth limiting probability distribution. Their asymptotic solution is: Select the populations that generated the last $100\lambda_0$ percent of the order statistics, where λ_0 depends on the limiting distribution of the θ_i and on the penalty associated with a wrong decision.

Gupta and Hsu (1978) studied the comparative performances of the maximum type procedure of Gupta (1956) and the average type procedure of Seal (1955) with their Bayes procedure in the case of normal means with exchangeable normal priors. Their Monte Carlo results indicate that the maximum type procedures do almost as well as the Bayes procedures. Similar results have been found under different loss functions by Chernoff and Yahav (1977), and Hsu (1977). These studies are useful because an easy-to-implement procedure whose performance is close to that of Bayes procedure is most welcome from a practical point of view; Bayes procedures typically require numerical integrations to implement them and are difficult to compute explicitly.

In other developments, Gupta and Hsu (1977) using the same loss function as in their 1978 paper established the monotonicity of Bayes subset selection

procedures, under certain generalized monotone likelihood ratio property assumption, for a restricted class of priors. Miescke (1979) assumed certain additive loss functions and showed that, in the normal case with symmetric priors, the Gupta procedure is the limit of Bayes rules as the sample size tends to infinity.

5.2 Minimax and Γ-minimax Rules

For the class of subset selection rules for which the PCS is at least P*, Berger (1979) investigated minimaxity taking the loss as measured by the subset size. Under certain mild conditions, he showed that the minimax value cannot be less than kP*. Applying this to location and scale problems, he showed that, under the monotone likelihood ratio assumption, the rules of Gupta (1965) are minimax. He also obtained some necessary conditions for minimaxity. One of these conditions is related to (12) which is an important property of just selection rules. It should be pointed out that if a rule is minimax with respect to the subset size $|S|$, then it is minimax also with respect to $|S'|$, where S' consists of all the non-best populations that are selected.

Berger and Gupta (1980) obtained minimax rules in the class of nonrandomized, just, and invariant rules when the risk is measured by the maximum probability of including a non-best population. These rules are of the form proposed and studied by Gupta (1965) in location and scale parameter problems. Berger and Gupta (1980) used their results to examine the minimaxity of several rules for the normal means problem when the variances are known but not necessarily equal and the sample sizes are unequal. We have referred to these results in Section 4.9.

Bjørnstad (1980) compared three minimax procedures for the normal means problem where the common known variance $\sigma^2 = 1$. Let $\theta_1,\ldots,\theta_k$ denote the means. The three procedures are the maximum type procedure of Gupta

(1956), the average type procedure of Seal (1955) and a procedure of Studden (1967). The performances of these three were compared by using the expected number of bad populations (that is, those for which $\theta_i < \theta_{[k]}-\Delta$, $\Delta > 0$ given) as the criterion, while controlling the PCS when there is only one good population. The numerical comparisons made for several slippage configurations showed that the average type procedure is inferior to the other two. While these two other rules seem comparable, the maximum type rule performs better when Δ is small.

The use of partial or incomplete prior information in statistical inference has led to the so-called Γ-minimax criterion, a term initially used by Blum and Rosenblatt (1967). The basic idea of the criterion, however, is due to Robbins (1951). Here the prior distribution is assumed to be a member of the subset Γ of the class of all priors. The Γ-minimax rule is obtained by minimizing the maximum expected risk over Γ. When Γ consists of a single prior, one has the Bayes rule for that prior. On the other hand, if Γ consists of all priors, then we have the usual minimax rule.

Gupta and Huang (1977) derived a Γ-minimax procedure for selecting the best population. Let $\pi_1,\ldots,\pi_k$ be k populations with densities f_{θ_i}, $i = 1,\ldots,k$, respectively. Define $\tau_i = \max_{\substack{1\le j\le k \\ j\neq i}} \tau_{ij}$, where τ_{ij} is a measure of separation between π_i and π_j. The population π_i such that $\tau_i = \min_{1\le j\le k} \tau_j$ is called the best population. Let $\Omega_i = \{\underset{\sim}{\theta} | \tau_i \le \tau_0\}$, $i = 1,\ldots,k$, where τ_0 is a given constant. The parameter space Ω is partitioned into $\Omega_0 \cup \Omega_1 \cup \ldots \cup \Omega_k$, where Ω_0 can possibly be the empty set. Here τ_0 is appropriately chosen so that Ω_i corresponds to configurations where π_i is the best, $i = 1,\ldots,k$. When $\underset{\sim}{\theta} \in \Omega_0$, selection of any one of the populations is considered a correct

selection. For deriving their Γ-minimax rule, Gupta and Huang (1977) assumed that $\Gamma = \{\rho(\underset{\sim}{\theta}) \mid \int_{\Omega_i} d\rho(\underset{\sim}{\theta}) = q_i,\ i = 1,\ldots,k\}$, where $\rho(\underset{\sim}{\theta})$ is a prior distribution and $q_1,\ldots,q_k$ are nonnegative and $\sum_{i=1}^{k} q_i \leq 1$.

Gupta and Kim (1980) considered minimax and Γ-minimax rules for partitioning k populations $\pi_1,\ldots,\pi_k$ in comparison with a standard or a control π_0. Let π_i have density $f_i(x-\theta_i)$, where f_i is symmetric about the origin and is strongly unimodal (that is, f_i is log-concave on the real line). Any population π_i is superior, equivalent, or inferior to π_0 according as $\theta_i-\theta_0 \geq \Delta$, or $-\Delta < \theta_i-\theta_0 < \Delta$, or $\theta_i-\theta_0 \leq -\Delta$, where $\Delta > 0$ is given. Gupta and Kim (1980) under appropriate losses for misclassifications of the populations derived Γ-minimax and minimax rules for both the cases when θ_0 is known or unknown. When θ_0 is unknown, attention was restricted to the class of rules for which the decision about π_i depends only on the observations from π_i and π_0, i = 1,...,k.

5.3 Construction of Optimal Selection Procedures and an Essentially Complete Class

Gupta and Huang (1980a) obtained a selection procedure under certain optimality conditions. Though their results are obtained in a general setup, we will describe it in terms of the normal means problem for simplicity. Let $\pi_1,\ldots,\pi_k$ be normal populations with unknown means $\theta_1,\ldots,\theta_k$ and a common variance $\sigma^2 = 1$. The population associated with the largest θ_i is the best population. A selection procedure is specified by the individual selection probabilities for the populations. Gupta and Huang (1980a) derived an optimal rule in the class of rules for which the PCS is at least $\gamma(0 < \gamma < 1)$ by minimizing the supremum of the expected subset size. In the general setup, the result requires a generalized version of the monotone likelihood ratio

for the multidimensional case.

Gupta and Huang (1980b) considered the class $\mathcal{C}$ of rules for which the size of the selected subset is controlled when the distributions are identical. The goal is to obtain a rule in this class which maximizes the infimum of the PCS over the parameter space Ω. Under certain assumptions, Gupta and Huang (1980b) obtained an essentially complete class of rules for this problem. In this regard, a rule δ_1 is defined to be as good as δ_2 if $\inf_{\Omega} P(CS|\delta_1) \geq \inf_{\Omega} P(CS|\delta_2)$ where both δ_1 and δ_2 belong to the class $\mathcal{C}$. The essentially complete class obtained by Gupta and Huang is the class of"monotone"selection procedures. Suppose $x_1,\ldots,x_k$ are observations from k populations with densities $f(x-\theta_i)$, $i = 1,\ldots,k$. Let $\underset{\sim}{y}_i = (y_{i1},\ldots,y_{ik})$, where $y_{ij} = x_i - x_j$, $j = 1,\ldots,k$; $j \neq i$. Let δ_i denote the individual selection probability for π_i, $i = 1,\ldots,k$. Then the selection rule $\delta = (\delta_1,\ldots,\delta_k)$ is monotone if, for any i and j, $\delta_i(\underset{\sim}{y}_i)$ is nondecreasing in y_{ij}, when all other components of $\underset{\sim}{y}_i$ are kept fixed.

5.4 Locally Optimal Subset Selection Rules

Gupta, Huang and Nagel (1979) were the first to investigate locally optimal subset selection rules. They were interested in obtaining such rules based on ranks while still assuming that the distributions associated with the populations were known. This is appealing especially in situations in which the order of the observations is more readily available than the actual measurements themselves due, perhaps, to excessive cost or other physical constraints.

Let $f(x,\theta_i)$ be the density associated with π_i, $\theta_i \in \Theta$ $i = 1,\ldots,k$, where Θ is an interval containing the origin. Let $\Delta = (\Delta_1,\ldots,\Delta_N)$ denote the rank configuration of the pooled sample of $N = kn$ observations obtained by taking a sample of size n from each population. Here $\Delta_i = j$ means that

the ith smallest observation in the pooled sample came from π_j. Let $\Omega_0 = \{\underset{\sim}{\theta}: \theta_1 = \ldots = \theta_k\}$. The goal is to derive a permutation-invariant rule δ based on the rank configuration Δ such that

$$\inf_{\underset{\sim}{\theta}\in\Omega_0} P_{\underset{\sim}{\theta}}(CS|\delta,\Delta) = P^* \tag{24}$$

subject to the condition:

$$\text{maximize } P_{\underset{\sim}{\theta}}(CS|\delta,\Delta) \text{ for all } \underset{\sim}{\theta} \in A_0 \tag{25}$$

where A_0 denotes a neighborhood of any $\underset{\sim}{\theta}_0 \in \Omega_0$. The motivation in seeking locally optimal rules satisfying (24) is due to the fact that, for many classical subset selection rules in the literature, the LFC is an equi-parameter point. Since the distribution of the ranks does not depend on the underlying distributions when $\underset{\sim}{\theta} \in \Omega_0$, the condition (24) implies that the PCS is constant for $\underset{\sim}{\theta} \in \Omega_0$. So A_0 can be taken as a neighborhood of $\underset{\sim}{\theta}_0 = (0,\ldots,0)$. Gupta, Huang and Nagel (1979) derived a locally optimal (in the sense of (25)) rule under certain regularity conditions on $f(x,\theta)$. If $f(x,\theta) = e^{-(x-\theta)}/[1+e^{-(x-\theta)}]^2$, the logistic density, their rule becomes: Select π_i if and only if $R_i \geq c$, where R_i is the rank-sum statistic for the sample from π_i, $i = 1,\ldots,k$. This is the procedure R_{15} defined by (21) of Gupta and McDonald (1970).

Some other locally optimal subset selection rules with different optimality criteria have been recently obtained by Huang and Panchapakesan (1982b) and Huang, Panchapakesan and Tseng (1984). These will be discussed in the next section.

5.5 Modified Goal for Subset Selection, and a Complete Ranking Formulation

In Section 4.2, we discussed the restricted subset selection formulation of Santner (1975) whose aim was to restrict the size of the selected subset by an upper bound $m \leq k-1$. Huang and Panchapakesan (1976) studied a modified formulation in which besides controlling the PCS, an upper bound $\beta \in (0, k-1]$ is imposed on the supremum of the expected subset size whenever the

parametric vector $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)$ belongs to a preference zone Ω_{δ^*} defined appropriately for a specified δ^*. The treatment of the problem by Huang and Panchapakesan (1976) is in a general setup that includes location and scale parameter cases. As in the restricted subset size formulation of Santner (1975), one has to determine a constant associated with the rule as well as the smallest sample size needed to meet the requirements. Specific application to selection in terms of treatment effects in a two-way layout is also discussed by Huang and Panchapakesan (1976).

In another paper, Huang and Panchapakesan (1978) have considered a subset selection formulation of the complete ranking problem. Let $\theta_1,\ldots,\theta_k$ be the unknown parameters of k populations. The interest is in ranking the k populations according to their θ-values. Any permutation of the set of integers $\{1,2,\ldots,k\}$ corresponds to a ranking of the populations. Huang and Panchapakesan (1978) considered the problem of selecting a nonempty subset (of a random size) of all the k! possible permutations such that the correct ranking is included in the selected subset of permutations with a minimum probability $P^* \in (1/k!, 1)$. They have discussed location and scale parameter cases. If $X_1,\ldots,X_k$ are the observations from $\pi_1,\ldots,\pi_k$ with densities $f(x-\theta_i)$, $i = 1,\ldots,k$, the procedure of Huang and Panchapakesan (1978) is

R_{22}: Include the ranking $(i_1,i_2,\ldots,i_k)$ in the selected subset if and only if

$$\max_{2\le s\le k} \max_{1\le r\le s} (X_{i_r} - X_{i_s}) \le d, \tag{26}$$

where d is the smallest positive constant for which the P*-condition is satisfied. The infimum of the PCS is attained when $\theta_1 = \ldots = \theta_k$.

5.6 Entropy-based Selection

Selection in terms of an information measure was first considered by Gupta and Huang (1976b) and Gupta and Wong (1977a). The former paper was concerned with binomial populations and the latter with multinomial populations. The significance of these papers is due not only to the importance of information-theoretic measures in practice but also to the illustration of using concepts of majorization and Schur functions for obtaining probability inequalities in selection problems.

Let $\pi_1,\ldots,\pi_k$ be k multinomial populations each with m cells and let the unknown cell-probabilities of π_i be $p_{i1},\ldots,p_{im}$; $i = 1,\ldots,k$. The Shannon entropy function associated with π_i is $H_i \equiv H(p_{i1},\ldots,p_{ik}) = -\sum_{j=1}^{m} p_j \log p_j$. This function is a measure of the uncertainty with regard to the nature of the outcomes from π_i, $i = 1,\ldots, k$. The populations are to be ranked in terms of the entropy function. For $m = 2$, the problem of selecting the population with the largest H_i reduces to that of selecting the binomial population associated with the largest $\psi(\theta_i) = -\theta_i \log \theta_i - (1-\theta_i)\log(1-\theta_i)$, where θ_i is the success probability. Gupta and Huang (1976b) studied a selection rule in this case. Their rule is

R_{23}: Select π_i if and only if

$$\psi(\frac{X_i}{n}) \geq \max_{1\leq j\leq k} \psi(\frac{X_j}{n}) - d, \tag{27}$$

where X_i is the number of successes in n trials associated with π_i, and $d = d(k,n,P^*)$ is the smallest nonnegative constant such that $0 < d \leq \psi([n/2]/n)$. Here $[n/2]$ denotes the largest integer $\leq n/2$. For this rule, the infimum of the PCS takes place when $\theta_1 = \ldots = \theta_k = \theta$.

However, as in the case of R_7 of Section 3, the common value of θ for which the infimum is attained is not known. Gupta and Huang (1976b) have obtained a conservative value for d as was done by Gupta, Huang and Huang (1976) in the case of R_7.

To discuss the general case of $m \geq 2$, we need the following definitions. Let $\underset{\sim}{a} = (a_1,\ldots,a_m)$ and $A_r = \sum_{i=r}^{m} a_{[i]}$, where $a_{[1]} \leq \ldots \leq a_{[m]}$ denote the ordered components.

Definitions 4.2. A vector $\underset{\sim}{a} = (a_1,\ldots,a_m)$ is said to majorize another vector $\underset{\sim}{b} = (b_1,\ldots,b_m)$ of same dimension (written $\underset{\sim}{a} \succ \underset{\sim}{b}$ or $\underset{\sim}{b} \prec \underset{\sim}{a}$) if $A_r \geq B_r$ for $r = 2,\ldots,m$, and $A_1 = B_1$. A function f is said to be Schur-convex (Schur-concave) if $f(\underset{\sim}{x}) \geq f(\underset{\sim}{x}')$ $(f(\underset{\sim}{x}) \leq f(\underset{\sim}{x}'))$ whenever $\underset{\sim}{x} \succ \underset{\sim}{x}'$.

In our selection problem, we assume that there is a population whose associated vector of cell-probabilities is majorized by the associated vector of any other population. Such a population will have the largest H_i because the entropy function is Schur-concave. Gupta and Wong (1977a) proposed the rule

R_{24}: Select π_i if and only if

$$\varphi\left(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}\right) \geq \max_{1\leq j\leq k} \varphi\left(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}\right) - d, \tag{28}$$

where $X_{i1},\ldots,X_{im}$ are the cell-counts based on n independent observations from π_i $(i = 1,\ldots,k)$, φ is a Schur-concave function, and $d = d(k,m,n,P^*)$ is the smallest positive constant for which the P*-condition is satisfied. Gupta and Wong (1977a) obtained a conservative value of d using the idea of conditioning as in the paper of Gupta and Huang (1976b).

5.7 Other Developments

Here we discuss several developments dealing with various aspects of subset selection procedures. These relate to selection procedures for Poisson processes, selection from restricted families, selection procedures based on medians, robust nonparametric procedures, selecting a good subset of the predictor variables, and subset selection used for screening in a two-stage procedure for selecting one population as the best.

Selection procedures for Poisson processes. Let $\pi_1,\ldots,\pi_k$ be k Poisson processes with unknown mean rates $\mu_1,\ldots,\mu_k$, respectively. Gupta and Wong (1977b) investigated four different procedures for selecting the process with the largest μ_i. Two of these procedures are based on the number of occurrences $X_i(t_0)$, $i = 1,\ldots,k$, during time t_0 from these processes and are essentially Rule R_{11} defined by (16) and its conditional analogue, both discussed in Section 4.5. A third procedure is

R_{25}: Observe the processes continuously until $\max_{1\le i\le k} X_i(t) = N$.

Select π_i if and only if

$$X_i(t) \ge N-c \tag{29}$$

where N is a positive integer specified in advance, and $c = c(k,P^*,N)$ is the smallest nonnegative integer for which the P*-condition is satisfied. The infimum of the PCS for the rule R_{25} takes place when $\mu_1 = \ldots = \mu_k = \mu$ and is independent of the common value μ. The constant c is the smallest integer ($0 \le c \le N$) which satisfies

$$\int_0^\infty [1-G_N(t)]^{k-1} dG_{N-c}(t) \ge P^*, \tag{30}$$

where $G_r(t)$ is the cdf of the gamma distribution with unit scale parameter and shape parameter r.

The fourth procedure of Gupta and Wong (1977b) is based on observing the processes at successive intervals of time, t = 1,2,..., until the first time t_0 when $X_i(t_0) \geq N$ for some i. The selection procedure is based on the waiting times for N and N-c occurrences, where c is the constant of the rule R_{25}. The details of this procedure are omitted here. The infimum of the PCS for this rule is the same as in the case of R_{25}, namely, the left-hand side of (30).

Selection from restricted families. Let $\pi_1,\ldots,\pi_k$ be k given populations with distributions $F_1,\ldots,F_k$, respectively. It is assumed that there is one among them which is stochastically larger than any of the rest. This population, denoted by $F_{[k]}$, is defined to be the best. It is also assumed that $F_{[k]} \underset{c}{\prec} G$, where G is a known distribution. All distributions in the context are assumed to be absolutely continuous with the positive real line as the support. Let $X^{(i)}_{j,n}$ ($Y_{j,n}$) denote the jth order statistic in a random sample of size n from $F_i(G)$. Define

$$T = \sum_{j=1}^{r} a_j Y_{j,n}, \quad T_i = \sum_{j=1}^{r} a_j X^{(i)}_{j,n}, \quad i = 1,\ldots,k, \tag{31}$$

where r ($1 \leq r \leq n$) is a fixed integer,

$$\left.\begin{aligned} a_j &= gG^{-1}\left(\tfrac{j-1}{n}\right) - gG^{-1}\left(\tfrac{j}{n}\right), \quad j = 1,\ldots,r-1, \\ a_r &= gG^{-1}\left(\tfrac{r-1}{n}\right), \end{aligned}\right\} \tag{32}$$

and g is the density associated with G.

If $G(y) = 1-e^{-y}$, $y \geq 0$, then $a_j = 1/n$ for $j = 1,\ldots,r-1$, and $a_r = (n-r+1)/n$. Thus $nT_i = X^{(i)}_{1,n} + \ldots + X^{(i)}_{r-1,n} + (n-r+1)X^{(i)}_{r,n}$, which is the well-known total life statistic until the rth failure from F_i.

Now, for selecting a subset containing $F_{[k]}$, Gupta and Lu (1979) proposed the rule

R_{26}: Select π_i if and only if

$$T_i \geq c \max_{1\leq j\leq k} T_j, \tag{33}$$

where c is the largest number in (0,1) for which the P*-condition is satisfied. They have shown that, if $a_j \geq 0$ for $j = 1,\ldots,r$, $g(0) \leq 1$ and $a_r \geq c$, then

$$\inf_{\Omega} P(CS|R_{26}) = \int_0^\infty G_T^{k-1}(y/c)dG_T(y), \tag{34}$$

where G_T is the cdf of T, and Ω is the space of all the k-tuples $(F_1,\ldots,F_k)$ such that there is one among them which is stochastically larger than the others and is convex with respect to G. Thus, the constant c is chosen to be the largest number in (0,1) such that $gG^{-1}(\frac{r-1}{n}) \geq c$ and the right-hand side of (34) is equal to P*. For the special case of $G(y) = 1-e^{-y}$, $y \geq 0$, the condition $a_j \geq c$ becomes $c \leq (n-r+1)/n$. This special case is a slight generalization of the results of Patel (1976).

Hooper and Santner (1979) considered selection of good populations in terms of α-quantiles for star- and tail-ordered distributions using the restricted subset size approach discussed in Section 4.2. Let π_i have the distribution F_i and let $F_{[i]}$ denote the distribution having the ith smallest α-quantile. Denoting the α-quantile of any distribution F by $x_\alpha(F)$, π_i is called a good population if $x_\alpha(F_i) > c^* x_\alpha(F_{[k-t+1]})$, $0 < c^* < 1$, in the case of star-ordered families, and if $x_\alpha(F_i) > x_\alpha(F_{[k-t+1]}) - d^*$, $d^* > 0$, in the case of tail-ordered families. The goal of Hooper and Santner (1979) is to select a subset of size not exceeding m ($1 \leq m < k-1$) that contains at least one good population.

Selection procedures based on medians. Gupta and Leong (1979), Gupta and Singh (1980), and Lorenzen and McDonald (1981) investigated subset selection procedures based on sample medians for selecting the population with the largest location parameter θ_i from k given populations belonging to several specific location parameter families. Let T_i be the sample median based on n independent observations π_i, i = 1,...,k. Then the procedure studied by all these authors is

$$R_{27}: \text{ Select } \pi_i \text{ if and only if } T_i \geq \max_{1\leq j\leq k} T_j - d, \tag{35}$$

where d is the smallest nonnegative number for which the P*-condition is satisfied.

Gupta and Leong (1979) considered the case of double exponential populations, namely, $f(x-\theta_i) = \frac{1}{2}\exp[-|x-\theta_i|]$, i = 1,...,k. Gupta and Singh (1980) considered normal populations with means $\theta_1,\ldots,\theta_k$, and a common known variance. They also studied performance characteristics of R_{27} in the double exponential case. Lorenzen and McDonald (1981) discussed the case of logistic distributions with means $\theta_1,\ldots,\theta_k$ and a common known variance. Gupta and Singh (1980) made a numerical study of the efficiency of R_{27} compared to the Gupta procedure based on sample means. Their study indicates that the procedure based on sample medians, though ordinarily less efficient than the procedure based on sample means, does better when the normal populations are contaminated. Lorenzen and McDonald (1981) compared R_{27} with the procedure based on means, and the rank-sum procedure (in the case of k = 2) of Gupta and McDonald (1970). The general nature of their findings are that the median procedure does better than the means procedure when there is contamination and it does better than the rank procedure when the θ_i are believed to be roughly in an equally spaced configuration.

Robust nonparametric procedures. In Section 4.6, we discussed the difficulty in establishing the LFC for maximum type procedures based on ranks. Hsu (1980) proposed a rule based on pairwise (rather than joint) ranking of the samples for which the PCS is minimized when the distributions are identical. Let $\pi_1,\ldots,\pi_k$ have distributions $F_{\theta_1},\ldots,F_{\theta_k}$, respectively. Let $X_{i1},\ldots,X_{in}$ denote the observations from π_i, $i = 1,\ldots,k$. For $1 \le i,j \le k$, $i \neq j$, define $R_j^{(i)}$ to be rank-sum of the sample from π_i in the combined sample from π_i and π_j; also, let $D^{ji}_{(1)} <\ldots< D^{ji}_{(n^2)}$ denote the n^2 ordered differences $X_{i\alpha}-X_{j\beta}$, $\alpha, \beta = 1,\ldots,n$. Then the rule of Hsu (1980) is

R_{28}: Select π_i if and only if

$$\text{either } T_i \ge \max_{1\le j\le k} T_j \quad \text{or} \quad \max_{j\neq i} R_j^{(i)} < r, \tag{36}$$

where $T_i = \sum_{j=1}^{k} D^{ji}_{med}/k$ and

$$D^{ji}_{med} = \begin{cases} D^{ji}_{(p+1)} & \text{if } n^2 = 2p+1 \\ (D^{ji}_{(p)}+D^{ji}_{(p+1)})/2 & \text{if } n^2 = 2p. \end{cases}$$

Here $D^{ii}_{med} = 0$. The constant $r = r(n,P^*)$ is the smallest integer such that

$$P_0\{\max_{j\neq 1} R_j^{(1)} \ge r\} \le 1-P^*, \tag{37}$$

where P_0 denotes the probability evaluated when the distributions are identical. It should be pointed out that D^{ji}_{med} is the usual Hodges-Lehmann estimator of $\theta_i-\theta_j$; the averaged estimator T_i of θ_i was introduced by Lehmann (1963b) to make the estimators compatible.

In another paper, Hsu (1981a) proposed a class of nonparametric subset selection procedures based on two-sample linear rank statistics for the

case of unequal sample sizes.

Selection of variables in linear regression. Applications of regression analysis in practice for prediction purposes often involve a large number of independent variables. In such situations, a subset of these predictor variables may be sufficient for "adequate" prediction. In this sense, there arises a problem of selecting a "good" subset of these variables. For nice reviews of several criteria and techniques that have been used in practice, reference may be made to Hocking (1976) and Thompson (1978a,b). The ad hoc procedures described in the papers of Hocking and Thompson, however, were not designed to select a good subset with any control on the probability of selecting the important variables. The first papers to formulate this problem in the framework of Gupta's subset selection were McCabe and Arvesen (1974), and Arvesen and McCabe (1975).

Consider the linear model

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}, \tag{38}$$

where X is an Nxp known matrix of rank $p \leq N$, $\underset{\sim}{\beta}$ is a pxl parameter vector, and $\underset{\sim}{\epsilon} \sim N(0, \sigma^2 I_N)$, I_N being an identity matrix. The model (38) which includes p independent variables is considered to be the "true" model. Consider all $k = \binom{p}{t}$ reduced models that are obtained by retaining t out of the p variables. The comparisons of these reduced models are made under the true model assumptions. Let $\sigma_1^2, \ldots, \sigma_k^2$ denote the error variances of these reduced models. The best subset of size t of the p variables is defined to be the set for which the error variance of the corresponding reduced model is $\sigma_{[1]}^2$. We will call this model the best reduced model of size t. Arvesen and McCabe (1975) proposed a rule to select a nonempty subset of all reduced models of size t so that the best reduced model

is included with a minimum guaranteed probability P*. They proposed a Gupta type scale procedure based on the error sums of squares associated with these models. However, these sums of squares are not independent and an exact evaluation of the infimum of the PCS is difficult. Arvesen and McCabe showed that the PCS is asymptotically ($N \to \infty$) minimized when $\beta_1 = \ldots = \beta_p = 0$. Still the constant associated with the procedure is not simple to compute. An algorithm for determining the constant using Monte Carlo methods is given by McCabe and Arvesen (1974).

Subset selection for screening in two-stage procedures. Suppose we have k normal populations with unknown means $\theta_1,\ldots,\theta_k$ and a common variance σ^2. The population associated with the largest θ_i is the best population. For selecting a single population as the best under the indifference zone formulation of Bechhofer (1954), a two-stage procedure is necessary if σ^2 is unknown. The main purpose of the first stage is to obtain an estimate of σ^2 so that the total sample size necessary to satisfy the P*-condition can be determined. If σ^2 is known, then the single stage procedure of Bechhofer (1954) applies. However, in this latter situation, one can use a two-stage procedure where the first stage is meant to effectively screen out inferior populations. Obviously, this is done by using a subset selection procedure designed to retain superior populations. Early investigations of Cohen (1959) and Alam (1970) were mostly confined to the special case of k = 2. The 1977 paper of Tamhane and Bechhofer for $k \geq 2$ renewed the interest in such procedures. Their procedure is described below.

R_{29}: Take a first-stage sample of n_1 observations from each population. Retain the populations π_i for which $\bar{X}_i \geq \max_{1 \leq j \leq k} \bar{X}_j - h\sigma$, where the $\bar{X}_i$ are the sample means based on n_1 observations and $h > 0$ is to be specified prior to the experimentation.

Let S denote the set of populations thus retained. If S consists of only one population, stop sampling and select this population. If S consists of more than one population, take an additional sample of size n_2 from each population in S. Select the population associated with the largest $\bar{Y}_i$, where the $\bar{Y}_i$ are the means based on $n_1 + n_2$ observations for the populations in S.

It should be noted that the fixed-sample procedure of Bechhofer (1954) is obtained as a special case of R_{29} by letting h = 0 or ∞. For Rule R_{29}, PCS is guaranteed to be at least P* whenever $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)$ belongs to $\Omega_{\delta^*} = \{\underset{\sim}{\theta}: \theta_{[k]} - \theta_{[k-1]} \geq \delta^*\}$; however, for given (k, δ^*, P*), there are an infinite number of combinations of (n_1, n_2, h) which will give the above guarantee for PCS. Tamhane and Bechhofer (1977) used a minimax criterion, namely, minimize $\sup_{\Omega}$ E(T) subject to $\inf_{\Omega_{\delta^*}}$ PCS being equal to P*, where $T = kn_1 + |S|n_2$, the total sample size required. However, the LFC is shown to be $\theta_{[1]} = \ldots = \theta_{[k-1]} = \theta_{[k]} - \delta^*$ only in the case of k = 2. For k > 2, Tamhane and Bechhofer (1977) obtained conservative solution by taking the infimum over Ω_{δ^*} of a lower bound of the PCS; in a subsequent paper, Tamhane and Bechhofer (1979) provided an improved yet conservative solution by using a sharper lower bound for the PCS. Their numerical study shows that the procedure R_{29} is very effective as a screening procedure, especially as k increases.

We initially pointed out that, when σ^2 is unknown, the first stage of a two-stage procedure is utilized for estimating σ^2 and determining the total sample size. If one wants to further use the idea of screening, it can be done by a three-stage procedure where the first stage is used to determine the additional sample sizes necessary in the subsequent stages, the second stage is used to eliminate inferior populations by a subset rule, and the third stage (if necessary) to make the final decision. Such procedures have been studied by Tamhane (1976) and Hochberg and Marcus (1981).

6. RECENT DEVELOPMENTS (1980-1984)

Developments that have taken place in the last three or four years constitute continuation of research on several topics that we discussed in the preceding sections as well as some new aspects and directions.

6.1 Decision-Theoretic Developments

In this subsection, we will discuss minimax, Γ-minimax, Bayes and empirical Bayes procedures for three different goals, namely, (i) selecting the best populations, (ii) selecting good populations, and (iii) selecting good populations in comparison to a control. However, results relating to two-stage and sequential procedures will be discussed separately along with some results for classical two-stage procedures. Also, some locally optimal selection procedures will be reviewed elsewhere in this section.

Selecting the best populations. Let $\pi_1,\ldots,\pi_k$ have densities $f(x,\theta_i)$, where θ_i belongs to an interval of the real line, $i = 1,\ldots,k$. It is assumed that $f(x,\theta)$ has a montone likelihood ratio in x. Bjørnstad (1981a) considered the goal of selecting the t best populations, namely, those associated with t largest θ_i's. Here the decision space consists of all subsets of $\{1,\ldots,k\}$ of size $\geq t$. Bjørnstad considered nonnegative, semi-additive loss functions of the form $L(\underset{\sim}{\theta},d) = \alpha(|d|)\sum_{i\in d} L_i(\underset{\sim}{\theta})$, where d denotes the subset selected and $|d|$ its size. Here $\alpha(|d|) \geq 0$ and $L_i(\underset{\sim}{\theta}) \geq 0$. Bjørnstad (1981a) has shown that, under certain conditions, the procedure that selects the t populations corresponding to the t largest X_i's (X_i is an observation from π_i) minimizes the risk uniformly in $\underset{\sim}{\theta}$ in the class of permutation-invariant procedures. He has also shown that a class of likelihood-ratio type procedures are admissible for monotone loss functions.

In Section 5.1, we discussed asymptotically optimal rules of Bickel and Yahav (1977) for selecting the best population. In a recent paper, Bickel and Yahav (1982) showed that the same rules also minimize the asymptotic risk for a wide class of smooth "monotone" loss functions within the class of procedures with PCS bounded below by a specified P*. They also showed that Gupta's maximum type rule with P* as the minimum PCS is asymptotically optimal within the same class of procedures and for the same class of loss functions for essentially any prior for which the empirical distribution function of the means tends to a fixed distribution function with prior probability 1, and whose essential supremum is finite.

For selecting the best population in the context of a randomized complete block design, Huang and Tseng (1983) have obtained Γ-minimax procedures.

Selecting good populations. Let $\pi_1,\ldots,\pi_k$ be normal populations with unknown means $\theta_1,\ldots,\theta_k$ and a common known variance σ^2. A population π_i is said to be good if $\theta_i \geq \theta_{[k]}-\Delta$, $\Delta > 0$ given, and bad otherwise. Gupta and Kim (1981) considered the loss function

$$L(\underset{\sim}{\theta},d) = \sum_{i\in d} L_B(\theta_i-\theta_{[k]}+\Delta) + \sum_{i\notin d} L_G(\theta_i-\theta_{[k]}+\Delta),$$

where d is the selected subset of $\{1,\ldots,k\}$, L_B is nonincreasing, L_G is nondecreasing, $L_B(y) = 0$ for $y \geq 0$, and $L_G(y) = 0$ for $y < 0$. Assuming that $\underset{\sim}{\theta}$ has an exchangeable normal prior, Gupta and Kim (1981) have shown that Gupta's maximum type rules are extended Bayes. For a simple loss function, they have made Monte Carlo comparisons of the maximum type and the average type rules with the Bayes rule. As in the studies of Chernoff and Yahav (1977), and Gupta and Hsu (1978), the study of Gupta and Kim (1981) indicates that the maximum type rules do, indeed, perform well.

Bjørnstad (1981b) developed a theory for a class of procedures, called Schur procedures, and applied it to certain minimax problems. Let $\pi_1,\ldots,\pi_k$ have densities $f(x-\theta_i)$, $i = 1,\ldots,k$. We assume that $f(x-\theta)$ has monotone likelihood ratio in x. Given the observation $\underset{\sim}{x} = (x_1,\ldots,x_k)$, a selection rule is defined by $\delta(\underset{\sim}{x}) = (\delta_1(\underset{\sim}{x}),\ldots,\delta_k(\underset{\sim}{x}))$, where $\delta_i(\underset{\sim}{x})$ denotes the individual selection probability for π_i, $i = 1,\ldots,k$. We consider the class $\mathcal{C}$ of just, permutation-invariant and translation-invariant procedures. Now, $\delta \in \mathcal{C}$ if and only if there exists a permutation-symmetric function δ': $\mathbb{R}^{k-1} \to \mathbb{R}$, which is nonincreasing in each component such that for every $i, \delta_i(\underset{\sim}{x}) = \delta'(x_1-x_i,\ldots,x_{i-1}-x_i,x_{i+1}-x_i,\ldots,x_k-x_i)$. A subset selection procedure $\delta = (\delta_1,\ldots,\delta_k)$ is said to be a Schur-procedure if $\delta \in \mathcal{C}$ and δ' is a Schur-concave function. Bjørnstad (1981b) has obtained several properties of Schur procedures. Subject to controlling the minimum expected number of good populations selected or the probability that the best population is in the selected subset, he has obtained minimax procedures using the criterion of minimizing the expected number of bad populations (or a similar criterion).

Selecting good populations compared to a control. Let $\pi_1,..,\pi_k$ be the populations that are compared with the control π_0. Let π_i be characterized by θ_i, $i = 0,1,\ldots,k$. Gupta and Hsiao (1981) considered the case of normal populations with unknown means $\theta_0,\theta_1,\ldots,\theta_k$ and known variances. They called a population π_i good if $|\theta_i-\theta_0| \leq \Delta$ and bad if $|\theta_i-\theta_0| \geq \Delta+\epsilon$, where $\Delta > 0$ and $\epsilon > 0$ are specified constants. They used a simple additive loss function which is made up of loss L_1 for every good population that is not selected, and loss L_2 for every bad population that is included. They considered both cases of known and unknown θ_0 and obtained minimax, Γ-minimax and Bayes rules. Their Bayes rule was derived under a normal prior for $\underset{\sim}{\theta}$.

In another paper, Gupta and Hsiao (1983) considered uniform distributions on $(0,\theta_i)$, $i = 0,1,\ldots,k$. They defined a population π_i to be good if $\theta_i > \theta_0$ and bad otherwise. They considered the loss function $L(\underset{\sim}{\theta},s) = L_1 \sum_{i\in A}(\theta_i-\theta_0) + L_2 \sum_{i\in B}(\theta_0-\theta_i)$, where L_1 and L_2 are known positive constants, s denotes the selected subset, A denotes the set of good populations that are not selected, and B denotes the set of bad populations that are selected. Gupta and Hsiao (1983) derived empirical Bayes rules in both cases of known and unknown θ_0.

Gupta and Leu (1983a) also considered selection from uniform distributions on $(0,\theta_i)$, $i = 0,1,\ldots,k$. But they defined π_i to be good if $|\theta_i-\theta_0| < \Delta$ and bad otherwise. They derived Bayes and empirical Bayes procedures (for both θ_0 known and unknown cases) using a loss function $L(\underset{\sim}{\theta},s)$ given by

$$
\begin{aligned}
L(\underset{\sim}{\theta},s) = & \sum_{i\in s} c_1(\theta_0-\Delta-\theta_i)I_{\{\theta_i\le\theta_0-\Delta\}}(\theta_i) + \\
& \sum_{i\in s} c_2(\theta_i-\theta_0-\Delta)I_{\{\theta_0+\Delta\le\theta_i\}}(\theta_i) + \\
& \sum_{i\notin s} c_3(\theta_i-\theta_0+\Delta)I_{\{\theta_0-\Delta<\theta_i\le\theta_0\}}(\theta_i) + \\
& \sum_{i\notin s} c_4(\theta_0+\Delta-\theta_i)I_{\{\theta_0<\theta_i<\theta_0+\Delta\}}(\theta_i),
\end{aligned}
$$

where s is the selected subset, c_i's are positive constants, and I is the indicator function. Subset selection from uniform distributions, using the classical approach, was studied by Barr and Rizvi (1966b) with an extension of their procedures to a large class of non-regular distributions.

6.2 Isotonic Procedures

In this section as well as in the previous sections, we have discussed the problem of comparing several populations with a control and the contributions of several authors in this respect. It has been assumed by these authors that there is no information about the ordering of the unknown parameters θ_i of these populations. In some situations, we may have partial prior information

in the form of a simple or partial order relationship among the unknown parameters. This is typical, for example, in experiments involving different dose levels of a drug where the treatment effects will have a known ordering. Let $\pi_1,\ldots,\pi_k$ be the k populations that are compared with the control population π_0. Let π_i have distribution F_{θ_i}, i = 0,1,...,k. It is assumed that θ_i, i = 1,...,k, are unknown, but it is known that $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$. A population π_i (i = 1,...,k) is defined to be good if $\theta_i \geq \theta_0$ and bad otherwise. The goal is to select all good populations. We would expect any reasonable procedure R to have the property: If R selects π_i then it selects all populations π_j for j > i. This is the isotonic behavior of the procedure R. Naturally, one would propose rules based on isotonic estimators of $\theta_1,\ldots,\theta_k$. Such procedures have been recently investigated by Gupta and Yang (1984) in the case of normal means (common variance σ^2 may be known or unknown), by Gupta and Huang (1982) in the case of binomial populations with success probabilities θ_i, and by Gupta and Leu (1983b) in the case of two-parameter exponential populations with location parameters (guarantee times) θ_i and a common (known or unknown) scale parameter . All these papers treat both cases of known and unknown θ_0.

6.3 Locally Optimal Subset Selection Rules

In Section 5.4, we discussed a locally optimal subset selection rule based on ranks given by Gupta, Huang and Nagel (1979). Their local optimality criterion involved maximizing the PCS in a neighborhood of an equi-parameter point. Locally optimal rules involving different optimality criteria have been recently investigated by Huang and Panchapakesan (1982b) and Huang, Panchapakesan and Tseng (1984).

Huang and Panchapakesan (1982b) considered two goals, namely, selecting the best from k populations $\pi_1,\ldots,\pi_k$, and selecting from $\pi_1,\ldots,\pi_k$ those populations, if any, that are better than π_0 which is the control population. Let π_i have density $f(x,\theta_i)$, $i = 0,1,\ldots,k$, satisfying certain regularity conditions. For the first goal, the best population is the one associated with the largest among $\theta_1,\ldots,\theta_k$. For the second goal, π_i is defined to be better than π_0 if $\theta_i > \theta_0$ and inferior to π_0 otherwise. As in Gupta, Huang and Nagel (1979), it is assumed that the functional form of $f(x,\theta)$ is known. For selecting the best population, Huang and Panchapakesan (1982b) derived a permutation-invariant rule δ such that $\inf_{\underset{\sim}{\theta} \in \Omega_0} P_{\underset{\sim}{\theta}}(CS|\delta) = P^*$ where $\Omega_0 = \{\underset{\sim}{\theta}: \theta_1 = \ldots = \theta_k\}$. Their rule is locally optimal in the sense that it is strongly monotone in a neighborhood of any point $\underset{\sim}{\theta}_0$ in Ω_0. For selecting populations better than a standard, it is assumed that θ_0 is unknown but $\theta_0 \leq \theta_0^*$, a known quantity. Huang and Panchapakesan considered the class of rules δ for which

$$P_{\underset{\sim}{\theta}}(\pi_i \text{ is selected} | \underset{\sim}{\theta} \in \Omega_0) \leq \gamma, \quad i = 1,\ldots,k,$$

and obtained a locally optimal rule in this class which is optimal in the sense that it maximizes

$$\sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} P_{\underset{\sim}{\theta}}(\pi_i \text{ is selected} | \theta_j = \theta_0^* < \theta_i,\ j \neq i)\Big|_{\theta_i = \theta_0^*}.$$

The above criterion of local optimality is related to the ability of a rule in choosing a population which is 'distinctly better' than the control while all others are not.

Huang, Panchapakesan and Tseng (1984) obtained a locally optimal rule for selecting populations better than a control. Their rule is not based on ranks but on statistics T_{i0} suitably defined to indicate the difference

between π_i and π_0, i = 1,...,k. Also, their procedure does not require equal sample sizes. They considered the class of rules for which the probability of selecting π_i when $\theta_1 = \ldots = \theta_k = \theta_0$ is fixed at the level γ_i, i = 1,...,k. The local optimality criterion used by them amounts to maximizing the efficiency (in a certain sense) of the rule in picking out the superior populations in the direction of each π_i being superior while all others are identical to the control. Huang, Panchapakesan and Tseng have applied their general results to the following special cases: (a) normal means comparison - common known variance, (b) normal means comparison - common unknown variance, (c) gamma scale parameters comparison - known (unequal) shape parameters, and (d) regression slopes.

6.4 Two-Stage and Sequential Rules

In Section 5.7, we descirbed a two-stage procedure (R_{29}) of Tamhane and Bechhofer (1977) for normal populations where the first stage involves a subset selection rule employed to eliminate inferior populations. Such rules have been called elimination type rules. We also noted the difficulty in establishing the LFC when $k \geq 3$. Consider the problem and the goal of Tamhane and Bechhofer (1977). They used Gupta's maximum type rule for screening based on the first stage sample. Let us call this procedure S_1. Gupta and Miescke (1982) considered S_1 and two other screening procedures S_2 and S_3. S_2 retains populations that yield the t largest $\bar{X}_i$, $2 \leq t \leq k-1$, t fixed, and S_3 retains π_i if and only if $\bar{X}_i \geq c_i$, i = 1,...,k. Let d_1 and d_2 denote two decision rules at the second stage, both choosing the population that yielded the largest sample mean, d_1 using only the second stage samples and d_2 using combined samples of both stages. The Tamhane - Bechhofer rule corresponds to (S_1,d_2). Gupta and Miescke (1982) proved that for (S_α,d_β),

$\alpha = 2,3$, $\beta = 1,2$, the LFC in Ω_{δ^*} is the slippage configuration $\underset{\sim}{\theta}_{\delta^*}$ given by $\theta_{[1]} = \ldots = \theta_{[k-1]} = \theta_{[k]} - \delta^*$. For (S_1, d_1), they have obtained a lower bound for the PCS which is minimized over Ω_{δ^*} at $\underset{\sim}{\theta}_{\delta^*}$, a result similar to that of Tamhane and Bechhofer (1979).

Gupta and Kim (1984) proposed a two-stage elimination type procedure for the normal means problem when the common variance σ^2 is unknown. It should be noted that for this problem, Tamhane (1976) and Hochberg and Marcus (1981) proposed three-stage procedures as pointed out in Section 5.7. Gupta and Kim gave a new design criterion and obtained a sharp lower bound on the PCS.

For the normal means problem (common known variance), Gupta and Miescke (1984a) studied two-stage procedures with screening at the first stage using a Bayesian approach. The sampling is done as in the case of the procedure R_{29}. Under the assumption of a loss function which includes the cost of sampling, they derived a Bayes procedure with respect to i.i.d. normal priors.

In another paper, Gupta and Miescke (1984b) studied permutation-invariant sequential procedures for selection from $\pi_1, \ldots, \pi_k$ belonging to an exponential family, with associated parameters $\theta_1, \ldots, \theta_k$, respectively. For the class of procedures considered, at stage m (m = 1,2,...) observations are drawn from all eligible populations at that stage. Then the procedure either stops and makes a <u>final</u> subset selection from the eligible populations, or it selects a subset from the eligible populations and proceeds to stage m+1. Under a general loss structure (favoring large parameters), Gupta and Miescke (1984b) have shown that at all stages the finally selected subsets have to be associated with the largest sufficient statistics from the eligible populations. In fact, these natural final decisions have been proved to be uniformly optimum in terms of the associated risk.

For a survey of these and other multi-stage selection procedures, reference may be made to Miescke (1982).

6.5 Other Developments

In Section 5.7, we discussed the problem of selecting a set of good predictor variables. In the formulation of Arvesen and McCabe (1975) only reduced models involving t (fixed) out of p independent variables were considered. Huang and Panchapakesan (1982a) formulated the problem as one of eliminating all inferior models (compared with the full model which is called the true model) using the criterion of expected residual sum of squares to define inferior models. Hsu and Huang (1982) investigated a sequential procedure for selecting good regression models. Recently, Gupta, Huang and Chang (1984) have discussed selection of an optimal subset of predictor variables using the criterion of expected residual mean squares to define an inferior model. Their treatment of the problem differs from that of earlier papers in the sense that they use simultaneous tests for a family of hypotheses.

In Section 3, we defined a rule R_4 which is really a class of rules based on contrasts. Let $\mathcal{C}$ denote this class. The procedure R_4 can select an empty subset unless P* is suitably (depending on k, $c_1,\ldots,c_{k-1}$) large. Let $\mathcal{C}_+$ denote this subclass of rules that select a nonempty subset. Deely and Gupta (1968) showed that, for the normal means problem (common known variance), Rule R_1 of Gupta (1956) is optimal (in the sense of minimizing the expected subset size) in the class $\mathcal{C}_+$ when the means are in a slippage configuration $(\theta,\ldots,\theta,\theta+\delta)$, $\delta > 0$, provided $\sqrt{n}\,\delta$ is greater than a constant depending on k and P* only. This result (which essentially amounts to considering k = 2, if we consider all configurations) was extended by Gupta and Miescke (1981) to k = 3 normal populations. They proved the following result: Let

$P^* \in (0,1)$ ($P^* \in [2/3,1)$) and n be fixed. Then R_1 is optimal within $\mathcal{C}$ ($\mathcal{C}_+$) at every configuration $\underset{\sim}{\theta} = (\theta_1,\theta_2,\theta_3)$ such that $\theta_{[3]}-\theta_{[1]}$ is 'sufficiently' large.

Gupta and W. T. Huang (1981) presented a survey of results on mixtures of distributions and considered selection from $\pi_1,\ldots,\pi_k$ where π_i has the density $f_i(x) = \sum_{j=1}^{m} \alpha_{ij} g_j(x)$, where $g_j(x)$, $j = 1,\ldots,m$, are known densities and $\alpha_{i1},\ldots,\alpha_{im}$ are unknown constants in (0,1) such that $\sum_{j=1}^{m} \alpha_{ij} = 1$, $i = 1,\ldots,k$. They considered several procedures for selecting the population associated with the largest λ_i where $\lambda_i = \lambda(\alpha_{i1},\ldots,\alpha_{im})$, $i = 1,\ldots,k$. For $m = 2,3$, they studied small samples procedures with $\lambda(\alpha_{i1},\ldots,\alpha_{im}) = \alpha_{i1}$. For any real-valued function λ on $(0,1)^m$, they considered procedures based on large samples when $m \geq 2$.

Bjørnstad (1983) considered a class of estimators called expansion estimators to be used in defining a subset selection procedure. These estimators of the population parameters are obtained by employing preliminary multiple comparisons procedures, and they tend to expand the differences between them, compared to the usual estimators. Bjørnstad (1983) has studied a class of maximum type procedure based on these expansion estimators.

Amer and Dudewicz (1980) studied the problem of selecting from several populations a subset containing (i) those which are better than a dual proportion standard, and (ii) those which are better than an upper proportion standard and none which are worse than a lower proportion standard.

Dudewicz and Chen (1983) investigated subset selection procedures where the correct selection corresponds to including all best populations in the selected subset. They considered the location and scale parameter cases, with specific applications to normal and exponential populations. Their procedures are related to those of Desu (1970), Carroll, Gupta and Huang (1975), and

Panchapakesan and Santner (1977), all discussed in Section 4.2. The goal of selecting all best populations is a limiting case of selecting good populations.

Mishra and Dudewicz (1983) have studied subset selection procedures for simultaneous selection of extreme populations, that is, populations associated with the largest and the smallest parametric values. Their results for normal means include both cases of known and unknown variances (not necessarily equal).

Bröstrom (1981) investigated a technique, called sequential rejection, for selection procedures. This technique is applicable to "all or nothing" type goals, such as selecting a subset containing all good populations, or selecting a subset containing no bad populations. Chotai (1980a,b) has studied several procedures based on likelihood ratio.

In related developments, Hsu (1981b) obtained parametric and nonparametric simultaneous confidence intervals for all distances from the best under the location model. In the parametric case, he has shown that these intervals represent a substantial strengthening of the probability statements associated with the procedures of Bechhofer (1954) and Gupta (1956, 1965). Jeyaratnam and Panchapakesan (1984) have discussed the problem of estimation after the subset selection using the Gupta rule (1956) for $k = 2$ normal populations with common known variance.

7. IMPACT OF DEVELOPMENTS AND FUTURE - AN ASSESSMENT

In the preceding sections, we have attempted to provide an introduction to the beginnings of the ranking and selection theory and to trace the developments in the subset selection theory. The literature on the subject of ranking and selection on the whole has grown enormously over these years, thanks to vigorous pursuit of many research workers. The

research workers in this field are no longer confined to a few schools or a few geographical parts of the world. It serves well to take stock of past in order to visualize the potential for future. Our general assessment that follows is not confined to subset selection alone but encompasses the ranking and selection field as a whole.

We earlier referred to the indifference zone and subset selection approaches for the selection and ranking problems. The inadequacies of the tests of homogeneity and the multiple comparisons techniques to answer the concerns of the experimenter regarding the best population or subset of best populations had been recognized by the late 1940's. The slippage problems of Mosteller (1948) and Paulson (1952) were early efforts in the direction of more meaningful formulations. The early papers dealing with choosing the best population were Bahadur (1950) and Bahadur and Robbins (1950). The stage was set for the development of the field of selection and ranking when Bechhofer published his now classical 1954 paper.

Some of the early applications of selection and ranking procedures were in the areas of animal science and agriculture. A few papers worth noting in this respect are Becker (1961, 1963), and Putter and Soller (1964, 1965). Several papers have appeared dealing with applications to tournaments, traffic fatality data, system performance evaluation, accounting, reliability models, education and psychology. A list of papers with such specific applications is given in the recent categorized bibliography of Dudewicz and Koo (1982, pp. 88-92). A bibliography of applications is also given in Gibbons, Olkin and Sobel (1977). For some papers advocating selection and ranking in practice, see Chew (1977a,b).

Although tables needed to implement the procedures were available in many papers starting with Bechhofer (1954), the application of the theory by the user had been rather slow. Part of this problem, until recently,

was due to unavailability of books that presented these new techniques and results in a way easily accessible to users at various levels. The monograph by Bechhofer, Kiefer and Sobel (1968) was the first book to deal exclusively with the subject. Though written for the theoretician and the practitioner, the book represents significant contributions to sequential procedures by the authors and thereby, perhaps, is accessible only to sophisticated users. Also, the period 1965 - 1975 marked the main growth of the field and as such it would have been rather premature for the appearance of a methods-oriented book for the practitioner or a comprehensive book on various developments. So it was not until the late 1970's that the next two books came out fulfilling these objectives. The book by Gibbons, Olkin and Sobel (1977) brings the basic methodology (mostly using the indifference zone approach) to practitioners and others. The text of Gupta and Panchapakesan (1979) provides a comprehensive survey of various developments in all aspects of the theory, with a special chapter on Guide to Tables. These are followed by books authored by Buringer, Martin and Schriever (1980) and Gupta and D.-Y. Huang (1981). Besides these, the text of Dudewicz (1976) devotes a chapter to selection and ranking, and the book by Kleijnen (1975) discusses uses of several selection procedures in the context of simulation. In addition to these books, several expository articles have appeared in journals from time to time; these are either overviews of the subject or surveys of developments dealing with certain aspects of the theory. A special issue of Communications in Statistics - Theory and Methods (Volume A6, Number 11) was devoted to selection and ranking procedures.

The books and special issues mentioned above have certainly contributed to further developments in the theory. Equally important have been forums for exchange and dissemination of ideas provided by special meetings and workshops. In this respect, special mention should be made of the special course on selection and ranking under the auspices of The American Statistical Association during its annual meeting in 1979 at Washington, D.C., and a similar special

course organized in the same year at the U. S. Naval Postgraduate School, Monterey, California. The lecturers in these two shourt courses were: Bechhofer, Gibbons, Gupta, and Olkin. The course organized by the ASA is available on video tape from the Continuing Education Section of the association. Also to be noted is the special advanced workshop and Conference on Inference Procedures Associated with Statistical Ranking and Selection organized by Dudewicz, and held in Honolulu, Hawaii in July 1982. The proceedings of this conference are to be published by the American Sciences Press under the title: The Frontiers of Modern Statistical Inference Procedures, as Volume 10 in their Series in Mathematical and Management Sciences.

In this connection, mention should be made of the proceedings of three symposia held at Purdue University in 1970, 1976 and 1981. These are published by Academic Press under the title: Statistical Decision Theory and Related Topics. Each of these three volumes include quite a few papers dealing with selection and ranking. The activities described above have been helpful in bringing the developments in the field to the attention of research workers in industry, government and academia.

Although for several standard situations, tables are available in the original papers and in the book by Gibbons, Olkin and Sobel (1977), it is desirable to develop computer packages for applications. One such package, commercially available, is RANKSEL produced by Edwards. It is an interactive package for the design and implementation of (a) selection of the best normal population when the common variance is known under the indifference zone formulation (Bechhofer, 1954), (b) selection of a subset of normal populations when the common variance is known or unknown (Gupta, 1956, 1965), (c) selection of all normal populations better than a control when the common variance is known or unknown (Gupta and Sobel, 1958). For further details on this package, see Edwards (1984a,b).

Another computer package, named RS-MCB, has been recently developed by Hsu and Gupta for (simultaneous) implementation of ranking and selection (RS) and multiple comparisons with the best (MCB) in terms of means for normal populations having a common unknown variance. At any specified confidence level, the RS part of the package (a) implements the Gupta subset selection procedure, and (b) decides whether or not to choose the population that yielded the largest sample mean as the best. The MCB part prints and plots constrained two-sided simultaneous confidence intervals for each mean minus the best of the remaining ones (Hsu, 1981b and 1984). This package can handle unequal sample sizes. For details of this package, reference can be made to Gupta and Hsu (1984).

Considering the fact that many of the activities that we have described in the preceding paragraphs took place within the last five years, it is perhaps too early to be pessimistic about the absence of dramatic change in the attitudes of the users. The major hurdle, if we may call it so, in adopting the selection and ranking formulation lies, on the part of many applied statisticians, in giving up the rush to testing of a 'null hypothesis'.

Finally, as our review would indicate, there are several situations in which more satisfactory solutions are needed. Some of the areas where not much has been done are multivariate problems, reliability models, and selection of the best predictor variables. Little attention has been paid to problems that arise with regard to model selection, time series data and signals in communications.

ACKNOWLEDGEMENT

This research was supported by the Office of Naval Research Contracts N00014-75-C-0455 and N00014-84-C-0167 at Purdue University. The authors wish to thank the Editor, Dr. M. Haseeb Rizvi,and the referees for valuable comments and suggestions; these have helped us in improving the presentation.

REFERENCES

Alam, K. (1970). A two-sample procedure for selecting the population with the largest mean from k normal populations. Annals of Institute of Statistical Mathematics, 22, 127-136.

Alam, K. and Rizvi, M. H. (1966). Selection from multivariate populations. Annals of Institute of Statistical Mathematics, 18, 307-318.

Amer, P. D. and Dudewicz, E. J. (1980). Two new goals for selection based on proportions. Communications in Statistics-Theory and Methods, A9, 1461-1472.

Arvesen, J. N. and McCabe, G.P., Jr. (1975). Subset selection problems of variances with applications to regression analysis. Journal of American Statistical Association, 70, 166-170.

Bahadur, R. R. (1950). On a problem in the theory of k populations. Annals of Mathematical Statistics, 21, 362-375.

Bahadur, R. R. and Robbins, H. (1950). The problem of the greater mean. Annals of Mathematical Statistics, 21, 469-487. Correction: 22 (1951), 301.

Barlow, R. E. and Gupta, S. S. (1969). Selection procedures for restricted families of distributions. Annals of Mathematical Statistics, 40, 905-917.

Barlow, R. E., Gupta, S. S. and Panchapakesan, S. (1969). On the distribution of the maximum and minimum of ratios of order statistics. Annals of Mathematical Statistics, 40, 918-934.

Barr, D. R. and Rizvi, M. H. (1966a). An introduction to selection and ranking procedures. Journal of American Statistical Association, 61, 640-646.

Barr, D. R. and Rizvi, M. H. (1966b). Ranking and selection problems of uniform distributions. Trabajos Estadistica, 17, 15-31.

Barron, A. M. (1968). A Class of Sequential Multiple Decision Procedures. Ph.D. Thesis (Also Mimeograph Series No. 169), Department of Statistics, Purdue University, West Lafayette, Indiana.

Barron, A. M. and Gupta, S. S. (1972). A class of non-eliminating sequential multiple decision procedures. Operations Research-Verfahren (Henn, Künzi and Schubert, eds.), Verlag Anton Hein, Meisenheim am Glan, Germany, 11-37.

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Annals of Mathematical Statistics, 25, 16-39.

Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968). Sequential Identification and Ranking Procedures (with special reference to Koopman-Darmois populations). The University of Chicago Press, Chicago and London.

Becker, W. A. (1961). Comparing entries in random sample tests. Poultry Science, 40, 1507-1514.

Becker, W. A. (1963). Ranking all-or-none traits in random sample tests. Poultry Science, 41, 1437-1438.

Berger, R. L. (1979). Minimax subset selection for loss measured by subset size. Annals of Statistics, 7, 1333-1338.

Berger, R. L. and Gupta, S. S. (1980). Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. Scandinavian Journal of Statistics, 7, 21-26.

Bhattacharya, P. K. (1956). Comparison of the means of k normal populations. Calcutta Statistical Association Bulletin, 7, 1-20.

Bickel, P. J. and Yahav, J. A. (1977). On selecting a subset of good populations. Statistical Decision Theory and Related Topics-II (S. S. Gupta and D. S. Moore, eds.), Academic Press, New York, 37-55.

Bickel, P. J. and Yahav, J. A. (1982). Asymptotic theory of selection procedures and optimality of Gupta's rules. Statistics and Probability: Essays in Honor of C. R. Rao (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.), North-Holland, Amsterdam, 109-124.

Bjørnstad, J. F. (1980). Comparison of three minimax subset selection procedures. Technometrics, 22, 617-620.

Bjørnstad, J. F. (1981a). A decision-theoretic approach to subset selection. Communications in Statistics - Theory and Methods, A10, 2411-2433.

Bjørnstad, J. F. (1981b). A class of Schur procedures and minimax theory for subset selection. Annals of Statistics, 9, 777-791.

Bjørnstad, J. F. (1983). On the use of expansion-estimators in subset selection. A Festschrift for Erich L. Lehmann (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.), Wadsworth, Belmont, California, 62-78.

Blum, J. R. and Rosenblatt, J. (1967). On partial information in statistical inference. Annals of Mathematical Statistics, 38, 1671-1678.

Blumenthal, S. and Patterson, D. W. (1969). Rank order procedures for selecting a subset containing the population with the smallest scale parameter. Sankhyā Series A, 31, 37-42.

Broström, G. (1981). On sequentially rejective subset selection procedures. Communications in Statistics - Theory and Methods, A10, 203-221.

Buringer, H., Martin, H. and Schriever, K.-H. (1980). Nonparametric Sequential Selection Procedures. Birkhauser, Boston, Massachusetts.

Carroll, R. J. (1974). Some Contributions to the Theory of Parametric and Nonparametric Sequential Ranking and Selection. Ph.D. Thesis (Also Mimeograph Series No. 359), Department of Statistics, Purdue University, West Lafayette, Indiana.

Carroll, R. J., Gupta, S. S. and Huang, D. Y. (1975). On selection procedures for the t best populations and some related problems. Communications in Statistics, 4, 987-1008.

Chapman, D. G. (1952). On tests and estimates for the ratio of Poisson means. Annals of Institute of Statistical Mathematics, 4, 45-49.

Chattopadhyay, A. K. (1981). Selecting the normal population with largest (smallest) value of Mahalanobis distance form the origin. Communications in Statistics - Theory and Methods, A10, 31-37.

Chen, H. J., Dudewicz, E. J. and Lee, Y. J. (1976). Subset selection procedures for normal means under unequal sample sizes. Sankhyā Series B, 38, 249-255.

Chernoff, H. and Yahav, J. (1977). A subset selection problem employing a new criterion. Statistical Decision Theory and Related Topics-II (S. S. Gupta and D. S. Moore, eds.), Academic Press, New York, 93-119.

Chew, V. (1977a). Comparisons Among Treatment Means in an Analysis of Variance. Report ARS/H/6, Agricultural Research Service, U. S. Department of Agriculture, Washington, DC.

Chew, V. (1977b). Statistical hypothesis testing: An academic exercise in futility. Proceedings of Florida State Horticultural Society, 90, 214-215.

Chotai, J. (1980a). Subset selection based on likelihood from uniform and related populations. Communications in Statistics - Theory and Methods, A9, 1147-1164.

Chotai, J. (1980b). Selection and Ranking Procedures Based on Likelihood Ratios. Ph.D. Thesis, Department of Mathematical Statistics, University of Umea, Umea, Sweden.

Cochran, W. C. and Cox, M. (1950). Experimental Designs. John Wiley, New York.

Cohen, D. S. (1959). A Two-Sample Decision Procedure for Ranking Means of Normal Populations with a Common Known Variance. M.S. Thesis, Department of Operations Research, Cornell University, Ithaca, New York.

Deely, J. J. (1965). Multiple Decision Procedures from an Empirical Bayes Approach. Ph.D. Thesis (Also Mimeograph Series No. 45), Department of Statistics, Purdue University, West Lafayette, Indiana.

Deely, J. J. and Gupta, S. S. (1968). On the properties of subset selection procedures. Sankhyā Series A, 30, 37-50.

Desu, M. M. (1970). A selection problem. Annals of Mathematical Statistics, 41, 1596-1603.

Desu, M. M. and Sobel, M. (1968). A fixed subset-size approach to a selection problem. Biometrika, 55, 401-410. Corrections and amendments: 63 (1976), 685.

Deverman, J. N. and Gupta, S. S. (1969). On a selection procedure concerning the t best populations. Technical Report, Sandia Laboratories, Livermore, California. Also Abstract: Annals of Mathematical Statistics 40 (1969), 1870.

Dudewicz, E. J. (1976). Introduction to Statistics and Probability. American Sciences Press, Columbus, Ohio.

Dudewicz, E. J. and Chen, H. J. (1983). A new subset selection theory for all best populations and its applications. Technical Report No. 158, Department of Statistics and Computer Science, The University of Georgia, Athens, Georgia. To appear in the Proceedings of the First Saudi Symposium on Statistics and Its Applications held in May 1983.

Dudewicz, E. J. and Dalal, S. R. (1975). Allocation of observations in ranking and selection with unequal variances. Sankhyā Series B, 37, 28-78.

Dudewicz, E. J. and Koo, J. O. (1982). The Complete Categorized Guide to Statistical Selection and Ranking Procedures. Series in Mathematical and Management Sciences, Vol. 6, American Sciences Press, Columbus, Ohio.

Edwards, H. P. (1984a). RANKSEL: A ranking and selection package. The American Statistician, 38, 158-159.

Edwards, H. P. (1984b). RANKSEL - An interactive computer package of ranking and selection procedures. The Frontiers of Modern Statistical Inference Procedures (E. J. Dudewicz, ed.), Volume 10, Series in Mathematical and Management Sciences, American Sciences Press, Columbus, Ohio (to appear).

Fabian, V. (1962). On multiple decision methods for ranking population means. Annals of Mathematical Statistics, 33, 248-254.

Frischtak, R. M. (1973). Statistical Multiple-Decision Procedures for Some Multivariate Selection Problems. Ph.D. Thesis (Also Technical Report No. 187), Department of Operations Research, Cornell University, Ithaca, New York.

Ghosh, M. (1973). Nonparametric selection procedure for symmetric location parameter populations. Annals of Statistics, 1, 773-779.

Gibbons, J. D., Olkin, I. and Sobel, M. (1977). Selecting and Ordering Populations: A New Statistical Methodology. John Wiley, New York.

Gnanadesikan, M. (1966). Some Selection and Ranking Procedures for Multivariate Normal Populations. Ph.D. Thesis, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gnanadesikan, M. and Gupta, S. S. (1970). A selection procedure for multivariate normal distributions in terms of the generalized variances. Technometrics, 12, 103-117.

Goel, P. K. and Rubin, H. (1977). On selecting a subset containing the best population. Annals of Statistics, 5, 969-983.

Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Mimeograph Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Gupta, S. S. (1963). On a selection and ranking procedure for gamma populations. Annals of Institute of Statistical Mathematics, 14, 199-216.

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

Gupta, S. S. (1966). On some selection and ranking procedures for multivariate normal populations using distance functions. Multivariate Analysis (P. R. Krishnaiah, ed.), Academic Press, New York, 457-475.

Gupta, S. S. and Hsiao, P. (1981). On Γ-minimax, minimax, and Bayes procedures for selecting populations close to a control. Sankhyā Series B, 43, 291-318.

Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. Journal of Statistical Planning and Inference, 8, 87-101.

Gupta, S. S. and Hsu, J. C. (1977). On the monotonicity of Bayes subset selection procedures. Proceedings of the 41st Session of the International Statistical Institute, Vol. 47, Book 4, 208-211.

Gupta, S. S. and Hsu, J. C. (1978). On the performance of some subset selection procedures. Communications in Statistics - Simulation and Computation, B7, 561-591.

Gupta, S. S. and Hsu, J. C. (1984). A computer package for ranking, selection, and multiple comparisons with the best. Technical Report No. 294, Department of Statistics, The Ohio State University, Columbus, Ohio.

Gupta, S. S. and Huang, D. Y. (1974). Nonparametric subset selection procedures for the t best populations. Bulletin of the Institute of Mathematics, Academia Sinica, 2, 377-386.

Gupta, S. S. and Huang, D.-Y. (1975a). On subset selection procedures for Poisson populations and some applications to the multinomial selection problems. Applied Statistics (R. P. Gupta, ed.), North-Holland, Amsterdam, 97-109.

Gupta, S. S. and Huang, D.-Y. (1975b). On some parametric and nonparametric sequential subset selection procedures. Statistical Inference and Related Topics, Vol. 2 (M. L. Puri, ed.), Academic Press, New York, 101-128.

Gupta, S. S. and Huang, D.-Y. (1976a). Subset selection procedures for the means and variances of normal populations: Unequal sample sizes case. Sankhyā Series B, 38, 112-128.

Gupta, S. S. and Huang, D.-Y. (1976b). On subset selection procedures for the entropy function associated with the binomial populations. Sankhyā Series A, 38, 153-173.

Gupta, S. S. and Huang, D.-Y. (1977). On some Γ-minimax selection and multiple comparison procedures. Statistical Decision Theory and Related Topics-II (S. S. Gupta and D. S. Moore, eds.), Academic Press, New York, 139-155.

Gupta, S. S. and Huang, D.-Y. (1980a). A note on optimal subset selection procedures. Annals of Statistics, 8, 1164-1167.

Gupta, S. S. and Huang, D.-Y. (1980b). An essentially complete class of multiple decision procedures. Journal of Statistical Planning and Inference, 4, 115-121.

Gupta, S. S. and Huang, D.-Y. (1981). Multiple Decision Theory: Recent Developments. Lecture Notes in Statistics Vol. 6, Springer-Verlag, New York.

Gupta, S. S., Huang, D.-Y. and Chang, C.-L. (1984). Selection procedures for optimal subsets of regression variables. Design of Experiments: Ranking and Selection (T. J. Santner and A. C. Tamhane, eds.), Marcel Dekker, New York, 67-75.

Gupta, S. S., Huang, D.-Y. and Huang, W.-T. (1976). On ranking and selection procedures and tests of homogeneity for binomial populations. Essays in Probability and Statistics (S. Ikeda, T. Hayakawa, H. Hudimoto, M. Okamoto, M. Siotani and S. Yamamoto, eds.), Shinko Tsusho Co. Ltd., Tokyo, Japan, Chapter 33, 501-533.

Gupta, S. S., Huang, D.-Y. and Nagel, K. (1979). Locally optimal subset selection procedures based on ranks. Optimizing Methods in Statistics (J. S. Rustagi, ed.), Academic Press, New York, 251-260.

Gupta, S. S. and Huang, W.-T. (1974). A note on selecting a subset of normal populations with unequal sample sizes. Sankhyā Series A, 36, 389-396.

Gupta, S. S. and Huang, W.-T. (1981). On mixtures of distributions: A survey and some new results on ranking and selection. Sankhyā Series B, 43, 245-290.

Gupta, S. S. and Huang, W.-T. (1982). On isotonic selection rules for binomial populations better than a standard. Technical Report 82-22, Department of Statistics, Purdue University. To appear in the Proceedings of the First Saudi Symposium on Statistics and Its Applications, held in May 1983.

Gupta, S. S. and Kim, W.-C. (1980). Γ-minimax and minimax decision rules for comparison of treatments with a control. Recent Developments in Statistical Inference and Data Analysis (K. Matusita, ed.), North-Holland, 55-71.

Gupta, S. S. and Kim, W.-C. (1981). On the problem of selecting good populations. Communications in Statistics - Theory and Methods, A10, 1043-1077.

Gupta, S. S. and Kim, W.-C. (1984). A two-stage elimination type procedure for selecting the largest of several normal means with a common unknown variance. Design of Experiments: Ranking and Selection (T. J. Santner and A. C. Tamhane, eds.), Marcel Dekker, New York, 77-93.

Gupta, S. S. and Leong, Y.-K. (1979). Some results on subset selection procedures for double exponential populations. Decision Information (C. P. Tsokos and R. M. Thrall, eds.), Academic Press, New York, 277-305.

Gupta, S. S. and Leu, L.-Y. (1983a). On Bayes and empirical Bayes rule for selecting good populations. Technical Report No. 83-37, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. and Leu, L.-Y. (1983b). Isotonic procedures for selecting populations better than a standard: two-parameter exponential distributions. Technical Report No. 83-46, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. and Lu, M.-W. (1979). Subset selection procedures for restricted families of probability distributions. Annals of Institute of Statistical Mathematics, 31, 235-252.

Gupta, S. S. and McDonald, G. C. (1970). On some classes of selection procedures based on ranks. Nonparametric Techniques in Statistical Inference (M. L. Puri, ed.), Cambridge University Press, London, 491-514.

Gupta, S. S. and McDonald, G. C. (1982). Nonparametric procedures in multiple decisions (ranking and selection procedures). Colloquia Mathematica Societatis János Bolyai, 32: Nonparametric Statistical Inference, Vol. I (B. V. Gnedenko, M. L. Puri and I. Vincze, eds.), North-Holland, Amsterdam, 361-389.

Gupta, S. S. and Miescke, K. J. (1981). Optimality of subset selection procedures for ranking means of three normal populations. Sankhyā Series B, 43, 1-17.

Gupta, S. S. and Miescke, K. J. (1982). On the least favorable configurations in certain two-stage selection procedures. Statistics and Probability: Essays in Honor of C. R. Rao (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.), North-Holland, 295-305.

Gupta, S. S. and Miescke, K. J. (1984a). On two-stage Bayes selection procedures. Sankhyā Series B, 46,

Gupta, S. S. and Miescke, K. J. (1984b). Sequential selection procedures - A decision-theoretic approach. Annals of Statistics, 12, 336-350.

Gupta, S. S. and Nagel, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. Sankhyā Series B, 29, 1-34.

Gupta, S. S. and Nagel, K. (1971). On some contributions to multiple decision theory. Statistical Decision Theory and Related Topics (S. S. Gupta and J. Yackel, eds.), Academic Press, New York, 79-102.

Gupta, S. S., Nagel, K. and Panchapakesan, S. (1973). On the order statistics from equally correlated normal random variables. Biometrika, 60, 403-413.

Gupta, S. S. and Panchapakesan, S. (1969). Some selection and ranking procedures for multivariate normal populations. Multivariate Analysis-II (P. R. Krishnaiah, ed.), Academic Press, New York, 475-505.

Gupta, S. S. and Panchapakesan, S. (1972). On a class of subset selection procedures. Annals of Mathematical Statistics, 43, 814-822.

Gupta, S. S. and Panchapakesan, S. (1974). Inference for restricted families: (a) multiple decision procedures; (b) order statistics inequalities. Reliability and Biometry (F. Proschan and R. J. Serfling, eds.), SIAM, Philadelphia, 503-596.

Gupta, S. S. and Panchapakesan, S. (1975). On a quantile selection procedure and associated distribution of ratios of order statistics from a restricted family of probability distributions. Reliability and Fault Tree Analysis (R. E. Barlow, J. B. Fussell and N. D. Singpurwalla, eds.), SIAM, Philadelphia, 557-576.

Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations. John Wiley, New York.

Gupta, S. S. and Santner, T. J. (1973). On selection and ranking procedures - a restricted subset selection rule. Proceedings of the 39th Session of the International Statistical Institute, Vol. 45, Book I, 478-486.

Gupta, S. S. and Singh, A. K. (1980). On rules based on sample medians for selection of the largest location parameter. Communications in Statistics - Theory and Methods, A9, 1277-1298.

Gupta, S. S. and Sobel, M. (1958). On selecting a subset which contains all populations better than a standard. Annals of Mathematical Statistics, 29, 235-244.

Gupta, S. S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. Contributions to Probability and Statistics (I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann, eds.), Stanford University Press, Stanford, California, Chapter 20, 224-248.

Gupta, S. S. and Studden, W. J. (1970). On a ranking and selection procedure for multivariate populations. Essays in Probability and Statistics (R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao and K. J. C. Smith, eds.), University of North Carolina Press, Chapel Hill, Chapter 16, 327-338.

Gupta, S. S. and Wong, W.-Y. (1982). Subset selection procedures for the means of normal populations with unequal variances: Unequal sample sizes case. Selecta Statistica Canadiana, Vol. 6, 109-149.

Gupta, S. S. and Wong, W.-Y. (1977a). Subset selection procedures for finite schemes in information theory. Colloquia Mathematica Societatis János Bolyai, 16: Topics in Information Theory (I. Csiszár and P. Elias, eds.), 279-291.

Gupta, S. S. and Wong, W.-Y. (1977b). On subset selection procedures for Poisson processes and some applications to the binomial and multinomial problems. Operations Research Verfahren (R. Henn et al., eds.), Verlog Anton Hain, Meisenheim am Glan, Germany, 49-70.

Gupta, S. S. and Yang, H.-M. (1984). Isotonic procedures for selecting populations better than a control under ordering prior. Statistics: Applications and New Directions, Proceedings of the Indian Statistical Institute Golden Jubilee International Conference.

Hochberg, Y. and Marcus, R. (1981). Three stage elimination type procedures for selecting the best normal population when variances are unknown. Communications in Statistics - Theory and Methods, A10, 597-612.

Hocking, R. R. (1976). The analysis and selection of variables in regression analysis. Technometrics, 9, 531-540.

Hooper, J. H. and Santner, T. J. (1979). Design of experiments for selection from ordered families of distributions. Annals of Statistics, 7, 615-643.

Hsu, J. C. (1977). On Some Decision-Theoretic Contributions to the Problem of Subset Selection. Ph.D. Thesis (Also Mimeograph Series No. 491), Department of Statistics, Purdue University, West Lafayette, Indiana.

Hsu, J. C. (1980). Robust and nonparametric subset selection procedures. Communications in Statistics - Theory and Methods, A9, 1439-1459.

Hsu, J. C. (1981a). A class of nonparametric subset selection procedures. Sankhyā Series B, 43, 235-244.

Hsu, J. C. (1981b). Simultaneous confidence intervals for all distances from the "best". Annals of Statistics, 9, 1026-1034.

Hsu, J. C. (1984). Constrained simultaneous confidence intervals for multiple comparisons with the best. Annals of Statistics, 12,

Hsu, T.-A. and Huang, D.-Y. (1982). Some sequential selection procedures for good regression models. Communications in Statistics - Theory and Methods, A11, 411-421.

Huang, D.-Y. and Panchapakesan, S. (1976). A modified subset selection formulation with special reference to one-way and two-way layout experiments. Communications in Statistics - Theory and Methods, A5, 621-633.

Huang, D.-Y. and Panchapakesan, S. (1978). A subset selection formulation of the complete ranking problem. Journal of Chinese Statistical Association, 16, 5801-5810.

Huang, D.-Y. and Panchapakesan, S. (1982a). On eliminating inferior regression models. Communications in Statistics - Theory and Methods, A11, 751-759.

Huang, D.-Y. and Panchapakesan, S. (1982b). Some locally optimal subset selection rules based on ranks. Statistical Decision Theory and Related Topics - III, Vol. 2 (S. S. Gupta and J. O. Berger, eds.), Academic Press, New York, 1-14.

Huang, D.-Y., Panchapakesan, S. and Tseng, S.-T. (1984). Some locally optimal subset selection rules for comparison with a control. Journal of Statistical Planning and Inference, 9, 63-72.

Huang, D.-Y. and Tseng, S.-T. (1983). Γ-optimal decision procedures for selecting the best population in randomized complete block design. Communications in Statistics - Theory and Methods, 12, 341-354.

Jeyaratnam, S. and Panchapakesan, S. (1984). An estimation problem relating to subset selection from normal populations. Design of Experiments: Ranking and Selection (T. J. Santner and A. C. Tamhane, eds.), Marcel Dekker, New York, 287-302.

Kiefer, J. (1977a). Conditional confidence and estimated confidence in multidecision problems (with applications to selection and ranking). Multivariate Analysis-IV (P. R. Krishnaiah, ed.), North-Holland Publishing Company, Amsterdam, 143-158.

Kiefer, J. (1977b). Conditional confidence statements and confidence estimators. Journal of American Statistical Association, 72, 789-808.

Kleijnen, J. P. C. (1976). Statistical Techniques in Simulation, Vol. 2. Marcel Dekker, New York.

Krishnaiah, P. R. (1967). Selection procedures based on covariance matrices of multivariate normal populations. Blanch Anniversary Volume, Aerospace Research Laboratories, U. S. Air Force, Dayton, Ohio, 147-160.

Krishnaiah, P. R. and Rizvi, M. H. (1966). Some procedures for selection of multivariate normal populations better than a control. Multivariate Analysis (P. R. Krishnaiah, ed.), Academic Press, New York, pp. 477-490.

Lehmann, E. L. (1963a). A class of selection procedures based on ranks. Mathematische Annalen, 150, 268-275.

Lehmann, E. L. (1963b). Robust estimation in analysis of variance. Annals of Mathematical Statistics, 34, 957-966.

Lorenzen, T. J. and McDonald, G. C. (1981). Selecting logistic populations using the sample medians. Communications in Statistics - Theory and Methods, A10, 101-124.

McCabe, G. P., Jr. and Arvesen, J. N. (1974). Subset selection procedure for regression variables. Journal of Statistical Computation and Simulation, 3, 137-146.

McDonald, G. C. (1972). Some multiple comparison selection procedures based on ranks. Sankhyā Series A, 53-64.

McDonald, G. C. (1974). Characteristics of three selection rules based on ranks in a small sample exponential case. Sankhyā Series B, 36, 261-266.

Miescke, K. J. (1979). Bayesian subset selection for additive and linear loss functions. Communications in Statistics - Theory and Methods, A8, 1205-1226.

Miescke, K. J. (1982). Recent results on multi-stage selection procedures. Technical Report No. 82-25, Department of Statistics, Purdue University, West Lafayette, Indiana.

Mishra, S. N. and Dudewicz, E. J. (1983). Simultaneous selection of extreme populations: a subset selection approach. Technical Report, Department of Mathematics, Syracuse University, Syracuse, New York.

Mosteller, F. (1948). A k-sample slippage test for an extreme population. Annals of Mathematical Statistics, 19, 58-65.

Nagel, K. (1970). On Subset Selection Rules with Certain Optimality Properties. Ph.D. Thesis (Also Mimeograph Series No. 222), Department of Statistics, Purdue University, West Lafayette, Indiana.

Panchapakesan, S. (1971). On a subset selection procedure for the most probable event in a multinomial distribution. Statistical Decision Theory and Related Topics (S. S. Gupta and J. Yackel, eds.), Academic Press, New York, 275-298.

Panchapakesan, S. (1973). On a subset selection procedure for the best multinomial cell and related problems. Abstract, Bulletin of Institute of Mathematical Statistics, 2, 112-113.

Panchapakesan, S. and Santner, T. J. (1977). Subset selection procedures for Δ_p-superior populations. Communications in Statistics - Theory and Methods, A6, 1081-1090.

Patel, J. K. (1976). Ranking and selection of IFR populations based on means. Journal of American Statistical Association, 71, 143-146.

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

Puri, M. L. and Puri, P. S. (1969). Multiple decision procedures based on ranks for certain problems in analysis of variance. Annals of Mathematical Statistics, 40, 619-632.

Puri, P. S. and Puri, M. L. (1968). Selection procedures based on ranks: scale parameter case. Sankhyā Series A, 30, 291-302.

Putter, J. and Soller, M. (1964). On the probability that the best chicken stock will come out best in a single random sample tests. Poultry Science, 43, 1425-1427.

Putter, J. and Soller, M. (1965). Probability of correct selection of sires having highest transmitting ability. Journal of Dairy Science, 48, 747-748.

Rizvi, M. H., (1971). Some selection problems involving folded normal distribution. Technometrics, 13, 355-369.

Rizvi, M. H. and Sobel, M. (1967). Nonparametric procedures for selecting a subset containing the population with the largest α-quantile. Annals of Mathematical Statistics, 38, 1788-1803.

Rizvi, M. H., Sobel, M. and Woodworth, G. G. (1968). A nonparametric ranking procedure for comparison with a control. Annals of Mathematical Statistics, 39, 2075-2093.

Rizvi, M. H. and Woodworth, G. G. (1970). On selection procedures based on ranks: counterexamples concerning the least favorable configurations. Annals of Mathematical Statistics, 41, 1942-1951.

Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (J. Neyman, ed.), University of California Press, Berkeley and Los Angeles, 131-148.

Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. Annals of Mathematical Statistics, 41, 1397-1409.

Santner, T. J. (1973). A Restricted Subset Selection Approach to Ranking and Selection Problems. Ph.D. Thesis (Also Mimeograph Series No. 318), Department of Statistics, Purdue University, West Lafayette, Indiana.

Santner, T. J. (1975). A restricted subset selection approach to ranking and selection problems. Annals of Statistics, 3, 334-349.

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

Seal, K. C. (1957). An optimum decision rule for ranking means of normal populations. Calcutta Statistical Association Bulletin, 7, 131-150.

Seal, K. C. (1958a). On ranking parameters of scale in Type III populations. Journal of American Statistical Association, 53, 164-175.

Seal, K. C. (1958b). Selection of a subset superior, inferior or equivalent to a control population. Calcutta Statistical Association Bulletin, 8, 20-30.

Studden, W. J. (1967). On selecting a subset of k populations containing the best. Annals of Mathematical Statistics, 38, 1072-1078.

Swanepoel, J. W. and Geertsema, J. C. (1973). Selection sequences for selecting the best k normal populations. South African Statistical Journal, 7, 11-21.

Tamhane, A. C. (1976). A three-stage elimination type procedure for selecting the largest normal mean (common unknown variance) case. Sankhya Series B, 38, 339-349.

Tamhane, A. C. and Bechhofer, R. E. (1977). A two-stage minimax procedure with screening for selecting the largest normal mean. Communications in Statistics - Theory and Methods, A6, 1003-1033.

Tamhane, A. C. and Bechhofer, R. E. (1979). A two-stage minimax procedure with screening for selecting the largest normal mean (II): An improved PCS lower bound and associated tables. Communications in Statistics - Theory and Methods, A8, 337-358.

Thompson, M. L. (1978a). Selection of variables in multiple regression: Part I. A review and evaluation. International Statistical Review, 46, 1-19.

Thompson, M. L. (1978b). Selection of variables in multiple regression: Part II. Chosen procedures, computations and examples. International Statistical Review, 46, 129-146.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #84-4	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): SUBSET SELECTION PROCEDURES: REVIEW AND ASSESSMENT	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER: Technical Report #84-4
7. AUTHOR(s): Shanti S. Gupta and S. Panchapakesan	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0455 and N00014-84-C-0167
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Purdue University, Department of Statistics, West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, DC	12. REPORT DATE: (Revised September 1984)
	13. NUMBER OF PAGES: 75
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Selection and ranking, subset selection, historical perspectives, early developments, years of main growth, years of further strides, recent developments, impact on users, references.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is well over three decades since statistical inference problems were first posed in the now familiar selection and ranking framework. This paper deals with the subset selection theory mainly developed by Gupta and his associates. Our aim is to provide a historical perspective (Section 2), and trace the major developments that took place in the subset selection theory over the years 1950-1980 which is divided into three periods representing early developments (Section 3), years of main growth (Section 4) and years of further strides (Section 5). We also discuss recent developments (Section 6) and briefly assess the impact of these developments on the users (Section 7).

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)