AN INTRODUCTION TO RANKING AND SELECTION

by

Jean D. Gibbons, Ingram Olkin, and Milton Sobel

TECHNICAL REPORT NO. 136

OCTOBER 1978

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT

MPS 75-09450

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
An Introduction to Ranking and Selection

by

Jean D. Gibbons*, Ingram Olkin**, and Milton Sobel***

1. Introduction

Suppose we have $k$ populations indexed by a parameter $\theta$ as

$$f(x_1; \theta_1), f(x_2; \theta_2), \ldots, f(x_k; \theta_k)$$

so that $\theta_j$ is the parameter of the $j$th population. We assume at the outset that the populations do not all have the same $\theta$ values, and that they can be ordered in some meaningful way with respect to these $\theta$ values, like from worst to best in some well-defined sense of best (say from smallest to largest). A random sample of size $n$ is taken from each population. A statistical selection procedure uses this sample data to make a selection of populations in such a way that we can assert with some specified level of confidence that the populations selected are the ones with the best $\theta$ values. A statistical ranking procedure uses this sample data to make an ordering of populations in such a way that we can assert with some specified level of confidence that the ordering made is correct. More generally, ranking and selection procedures are statistical techniques for comparing the parameters of some $k$ populations under the assumption that these parameters are not all the same.

Ranking and selection procedures are particularly appropriate for answering questions like the following. Which one of $k$ different drugs produces the best response? Which subgroup of the $k$ drugs produces a better response than a placebo? Which two of $k$ types of advertising media

*The University of Alabama, University, AL 35486
**Stanford University, Stanford, CA 94305
***University of California, Santa Barbara, CA 93106
reach the highest proportion of potential buyers of a particular product? Which one of k different learning techniques produces the best comprehension? How do a group of k candidates for a certain political office rank in popularity at a specific point in time? Which one of k different types of ski bindings has the lowest accident rate?

In each of these situations the k populations are the k alternatives or types of items under consideration and each item has a parameter that measures how "good" it is. These parameters, $\theta_1, \theta_2, \ldots, \theta_k$, represent the same type of description, attribute or response for each population.

To illustrate the situation more specifically, we use an application reported in the literature. Becker (1961) considered the problem of selecting the best one out of $k = 10$ different poultry stocks, where best is defined as the stock with the largest hen-house egg production. The experimental procedure was to draw a sample of n chickens from each stock, place them in floor pens such that the environment is identical for each chick, and measure the egg production for each chick after 500 days. The data on egg production for the jth stock are assumed to be normally distributed with mean $\mu_j$ and known variance $\sigma^2 = 5225.0$ for each $j = 1, 2, \ldots, 10$. Here the k populations are the $k = 10$ poultry stocks, the $\theta$ values are the means $\mu_1, \mu_2, \ldots, \mu_{10}$, the distributions are all the same except for these means, and the problem is to select the population with the largest $\mu$ value.

The classical and conventional statistical approach to such a problem is a test of homogeneity, where we test the null hypothesis $H_0$ that the $k$ parameters $\theta_1, \theta_2, \ldots, \theta_k$ are all equal, that is,

$$H_0: \theta_1 = \theta_2 = \ldots = \theta_k.$$

The analysis of variance test is of this type where the $\theta$ values are
interpreted as the means of the $k$ populations. If a test of homogeneity is the primary and final goal of an investigation or experiment (which is rarely the case), then alternative methods of statistical analysis are not needed. However, if the question of interest is one of those mentioned earlier, say to identify the one best population, then a test of homogeneity is inadequate for, or at least not pertinent to, the problem at hand, because it cannot answer the question that we really want answered. The homogeneity test can only tell us whether or not the populations are equivalent. If we accept $H_0$, we can conclude that the $\theta$ values are all equal; if we reject $H_0$, we can conclude that they are not all equal but we have no conclusion about which is the best. Of course, some modifications and extensions of the test of homogeneity, like multiple comparisons, are available to provide additional information. However, no modification of that test can be appropriate if we assume at the outset that differences must surely exist. Thus, ranking and selection procedures are also used for comparisons, but their objective is different. In particular, if some selection or ordering of the populations is desired, then a homogeneity test procedure is not the appropriate statistical tool.

2. **Historical Perspective**

The early origin of the theory of selecting populations is in the 1940's when Wald developed sequential analysis and Girshick (1946) modified Wald's technique and adapted it to the problem of ranking two populations. In the late 1940's and early 1950's several authors dealt with the so-called shift model (that is, the movement or shift of one parameter value to the right or left while all the other parameter values remain equal); these include Mosteller (1948), Mosteller and Tukey (1950), Bahadur (1950), Bahadur and Robbins (1950), Paulson (1952a, b), and Truax (1953). These papers were forerunners to the present formulation which was first de-
veloped in Bechhofer (1954). Other early landmark papers were Bechhofer, Dunnett, and Sobel (1954), Bechhofer and Sobel (1954), Gupta (1956), and Gupta and Sobel (1957). The development given in this paper depends heavily on these last five papers.

3. The Philosophy of the Indifference Zone Approach to Ranking and Selection

In this section we discuss the philosophy of the indifference zone approach to selection problems.

We use bracketed subscripts to denote the ordered \( \theta \) values, that is,

\[
\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}.
\]

Thus the notation adopted is that \( \theta_{[j]} \) is the parameter of the \( j \)th population, but \( \theta_{[j]} \) is the \( j \)th from the smallest \( \theta \) value. In order to make the present discussion explicit, we assume that the best population is defined as the one with the largest \( \theta \) value, \( \theta_{[k]} \), and the problem is to identify that one population, that is, to identify the value of \( j \), among \( j = 1, 2, \ldots, k \), for which the parameter value \( \theta_{[j]} \) is equal to \( \theta_{[k]} \). The selection procedure for this problem is to take a random sample from each of the \( k \) populations, compute some estimator \( \hat{\theta}_{[j]} \) of each \( \theta_{[j]} \) from the corresponding set of sample data, and assert that the best population is the one for which the value of the sample estimate is largest. (In the Becker (1961) example, we would calculate the sample mean of egg production for each stock, and assert that the best stock is the one that produces the largest sample mean.)

As is always the case when we have incomplete (or sample) information, the possibility of an error must be considered. A selection is defined to be correct if the \( \theta \) value of the population selected is the true largest \( \theta \) value. Hence, whenever \( \theta_{[k]} > \theta_{[k-1]} \), there is only one selection that is correct. Note that our goal is not to estimate the value of \( \theta_{[k]} \), but
only to select or identify the population with \( \theta \) value equal to \( \theta[k] \). Hence an error can occur only if the selection is incorrect. The probability of making an incorrect selection when \( \theta[k] > \theta[k-1] \) might be thought of as the analogue of the probability \( \beta \) of a Type II error in classical hypothesis testing, and the probability of making a correct selection when \( \theta[k] > \theta[k-1] \) as the counterpart of \( 1 - \beta \) or the power of the test. A correct selection corresponds to a decision where we reject an \( H_0 \) of homogeneity when it is false. We define a good selection procedure as one for which the probability of a correct selection is large. Note that only one kind of error is possible in a selection procedure. We have no counterpart here for the Type I error; there is no need to consider any error when the parameter values are all equal since the choice of any one is equally acceptable.

The probability of a correct selection (CS) depends on the true set (or configuration) of \( \theta \) values; hence we call it the \( P\{CS|\theta\} \) function (here \( \theta \) denotes a vector of the \( k \) parameters). If we had two different selection rules, say \( R_1 \) and \( R_2 \), with corresponding \( P\{CS|\theta\} \) functions \( P_1 \) and \( P_2 \), it is likely that \( P_1 \geq P_2 \) for some sets of \( \theta \) values and \( P_1 \leq P_2 \) for other sets of \( \theta \) values. Hence in order to set up criteria for a good selection rule, we can divide the set of all possible \( \theta \) values (the total parameter space or TPS) into two parts, those regions of \( \theta \) values where we have a strong preference for making a correct selection, called the preference zone \( PZ \), and those where we are indifferent about the selection, called the indifference zone \( IZ \). (The TPS is the union of the \( PZ \) and \( IZ \).) This procedure is called the indifference zone approach to ranking and selection.

Suppose that \( k = 2 \) and \( \theta_1, \theta_2 \) can be any real numbers so that the total parameter space is the entire two-dimensional plane. Then if \( \theta_1 \) and \( \theta_2 \) are close in value, we may be indifferent about which one is asserted
to be larger; but if $\theta_1$ and $\theta_2$ differ greatly, then we want a very high probability of making a correct selection. Suppose that some specified $\delta^* > 0$ is the smallest difference worth bothering about; then an appropriate preference zone is $|\theta_2 - \theta_1| \geq \delta^*$, as shown in Figure 1. Figure 2 shows the same preference zone on a graph with ordered $\theta$ values for the axes.

For arbitrary $k$, the entire parameter space is a region with $k$ dimensions. Since we wish to choose the population corresponding to $\theta_{[k]}$, we primarily need be concerned with how much $\theta_{[k]}$ is larger than $\theta_{[k-1]}$, and only to a lesser extent with their values relative to $\theta_{[1]}, \ldots, \theta_{[k-2]}$. Thus the definition of the PZ and IZ need depend only on the two parameters $\theta_{[k]}$ and $\theta_{[k-1]}$ even for $k > 2$. In general, we need to define a function $\delta$ of $\theta_{[k]}$ and $\theta_{[k-1]}$ to measure the distance between these two parameters, and a threshold value $\delta^*$ to separate the IZ from the PZ. Since we are mostly concerned about the mistake of selecting $\theta_{[k-1]}$ as best when $\theta_{[k]}$ and $\theta_{[k-1]}$ are not close in value, the PZ should be such that these configurations are included.

Thus, for arbitrary $k$, if the distance measure is $\delta = \theta_{[k]} - \theta_{[k-1]}$ and the parameter space is unlimited so that the $\theta$ values vary on the entire real line, the preference zone should be as shown in Figure 3, where

$$\text{PZ is } \theta_{[k]} - \theta_{[k-1]} \geq \delta^*. $$

Now suppose the range of possible $\theta$ values is limited, as, for example, in the binomial distribution. Then the $\theta$ values are in the closed interval $[0, 1]$ and we have the preference zone shown in Figure 4, where

$$\text{PZ is } \theta_{[k]} - \theta_{[k-1]} \geq \delta^*. $$

The function $\delta$ of $\theta_{[k]}$ and $\theta_{[k-1]}$ need not be their difference. If the distributions are binomial, $\delta$ might be defined as the so-called odds
ratio, the odds for success in the best population divided by the odds for success in the next best population. The values are again in the closed interval $[0, 1]$ but now we have the preference zone shown in Figure 5, where

$$\text{PZ is } \frac{\theta[k](1 - \theta[k-1])}{(1 - \theta[k])\theta[k-1]} \geq \delta^*.$$ 

In general, we want the probability of a correct selection to be large when the parameter configuration is in the preference zone; otherwise we are indifferent. That is, we do not mind if we make an incorrect selection in the indifference zone because $\theta[k-1]$ and $\theta[k]$ are close together there. Hence we can restrict primary consideration to the preference zone. The preference zone generally has an infinite number of points. However, in many cases there is some special configuration in the preference zone for which the probability of a correct selection is a minimum over all configurations in the preference zone. This configuration is called the least favorable configuration and denoted by $\theta_{LF}$. In other words, the probability of a correct selection for any configuration in the preference zone is always at least as large as the probability of a correct selection for the configuration $\theta_{LF}$, or, in symbols,

$$P(\text{CS} | \theta) \geq P(\text{CS} | \theta_{LF}) \text{ for all } \theta \in \text{PZ}.$$ 

If $P(\text{CS} | \theta_{LF}) = P^*$ say, then we know that $P(\text{CS} | \theta) \geq P^*$ for all $\theta$ in the preference zone, and thus $P^*$ is the minimum probability of a correct selection for all configurations of concern. In many cases the least favorable configuration is a simple function of the distance measure $\delta$. For example, in the problem of selecting the normal population with the largest mean, the least favorable configuration for all sample sizes is where all but one of the population means are equal to the second largest mean, or, in symbols,

$$\theta[1] = \theta[2] = \ldots = \theta[k-1] = \theta' \text{ and } \delta = \theta[k] - \theta' = \delta^*.$$
4. Aspects of Selection Problems Using the Indifference Zone Approach

One aspect of the problem of selecting the best population arises when the experiment is in the process of being designed and the problem is to determine the common sample size \( n \) per population needed to satisfy the requirement

\[ P(\text{CS}|\theta) \geq P^* \text{ for all } \delta \geq \delta^*, \]

that is, for all \( \theta \) in the PZ, where \( \delta^* \) and \( P^* \) are specified by the experimenter. Tables are available for many common distributions, like the normal, binomial, multinomial, gamma, etc., and also for some nonparametric problems. After we determine \( n \) by this procedure, we can generally make a confidence statement about the selection such as the following:

"With confidence level \( P^* \), the \( \theta \) value of the population selected, \( \theta_s \), satisfies

\[ \theta[k] - \delta^* \leq \theta_s \leq \theta[k], \]

This type of confidence statement can be made in the problem of selecting normal means, normal variances, binomial probabilities, etc. (In

Remark 1. In general it can be made whenever we deal with a known family involving a sufficient statistic or a stochastically increasing family for which the least favorable configuration is attained by the so-called slippage configuration in the preference zone. This covers most of the families of practical interest in the parametric case. A proof of this remark is beyond the scope of this expository paper. In nonparametric problems various pathologies can and do occur but we shall not deal with them here.

To illustrate this aspect, consider the problem of determining \( n \) for the poultry stock example in Becker (1961); here we use \( \mu \) for a normal mean instead of the general parameter \( \theta \) above. Since Becker claimed that a difference of 2\( \frac{1}{2} \) eggs (2 dozen) between the average production of the best and
second-best stock (in 500 days) is significant, suppose we assert that we would like to have a $P(\text{CS} \mid \mu) \geq P^* = .90$ when the difference between the best and second-best means of the $k=10$ stocks is $\delta^* = 2^\frac{1}{4}$ or more. Then we need to determine $n$ such that $P(\text{CS} \mid \mu) \geq P^* = .90$ whenever

$$\delta = \mu_{[10]} - \mu_{[9]} \geq \delta^* = 2^\frac{1}{4}.$$ 

It can be shown that $n = 81$ chickens of each stock are needed for a 500-day experiment. Suppose the experiment is carried out with this sample size and stock A has the largest sample mean egg production. Then we can state with confidence level .90 that the mean for stock A, $\mu_A$, satisfies the inequality

$$\mu_{[10]} - 2^\frac{1}{4} \leq \mu_A \leq \mu_{[10]},$$

that is, $\mu_A$ is within $2^\frac{1}{4}$ units of the mean of the best stock.

In other cases where the problem is to select the best population, the sample size may be determined by extra-statistical considerations or the data may have already been obtained. Then we can calculate the operating characteristic curve as the locus of all pairs ($\delta^*$, $P^*$) that are satisfied for the fixed sample size. In many cases, we can also estimate the true configuration and use this estimate to estimate the true probability of a correct selection. More information, more details, and the necessary tables for many such problems are all given in Gibbons, Olkin, and Sobel (1977).

We now give some examples of the operating characteristic curve for some specific problems where the goal is to select the one best population.

Figure 6 shows the locus of ($\delta^*$, $P^*$) pairs that are satisfied for the problem of selecting the best one of $k = 4$ normal populations with known variance $\sigma^2 = 1$ based on sample sizes (1) $n = 20$, and (2) $n = 50$ from each of the four populations. Here best is defined as the population with the largest mean and the distance measure is $\delta = \mu_{[4]} - \mu_{[3]}$. Notice that the curve for $n = 50$ is always higher than for $n = 20$, which means that the probability of a correct selection is always larger for a greater $n$. Both
curves approach $1/k = .25$ as $\delta^* \to 0$ (all means equal) and approach 1 as $\delta^* \to \infty$ (the two largest means very disparate). To interpret a specific point for $n = 50$, say $\delta^* = .40$, $p^* = .94$, this means that (1) the probability of a correct selection is at least .94 whenever $\mu[4] - \mu[3] \geq .40$, and (2) with confidence level $p^* = .94$, the mean $\mu_s$ of the population selected is within .40 of the true largest mean.

Figure 7 shows two operating characteristic curves for the problem of selecting the best one of $k = 4$ binomial populations where best is defined as the population with the largest probability of success. The distance measure here is also the difference of the two largest parameter values. Notice that here again the curve for $n = 50$ is always above the curve for $n = 20$, that they both approach 1 as $\delta^* \to 1$, and both approach $1/k = .25$ as $\delta^* \to 0$ (all probabilities of success are equal).

Figure 8 shows two operating characteristic curves for the problem of selecting the best one of $k = 4$ normal populations where best is defined as the population with the smallest variance. The distance measure here is the ratio $\Delta = \sigma[1]/\sigma[2] \leq 1$, and thus here the preference zone must include small values of $\Delta$, i.e., $\Delta \leq \Delta^*$. As a result, the curves are reversed; they approach $1/k = .25$ as $\Delta^* \to 1$. The curves shown apply for (1) $v = 20$ degrees of freedom, which means that $n = 20$ if the population means are known and $n = 21$ if the means are unknown, and (2) $v = 50$ degrees of freedom so that $n = 50$ or 51 depending on whether the means are known or unknown, respectively. The curve for $v = 50$ is always above the curve for $v = 20$. 
5. Examples of Application

As a first example of application, we consider the problem of Kleijnen, Naylor, and Seaks (1972), where a firm's output is the result of a sequential production process. The management of the firm is considering five different production processes, and wishes to select the one most profitable process to adopt. Data on profit will be obtained by simulation experiments over a ninety day period with n replications for each process. From previous experience with similar processes, it is known that profit data are normally distributed with common variance approximately $\sigma^2 = 52,000$. Further, it is known that the range of mean profit is roughly $500$, and hence an appropriate specification of $\delta^* \sigma$ is $\delta^* = $100; that is, $100$ is the smallest difference worth detecting. The value of $n$ is found from the expression

$$n = \frac{\sigma^2 \tau}{\delta^*}$$

where $\tau$ is given in Table 1 as a function of $k$ and $P^*$. Here $k = 5$, and if we specify $P^* = .95$, then $\tau = 3.0552$ and $n$ is calculated as

$$n = 52,000(3.0552/100)^2 = 48.54.$$ 

We round upward for a conservative result and hence $n = 49$ replications are required. When the data are collected, the sample mean profit is calculated for each production process, and the process with the largest sample mean profit is asserted to be the best; we have confidence .95 that the mean profit from the process selected is within $100$ of the true best process.

As another example of a practical application, consider the winter sport of snow skiing. Each ski has a binding and this binding is supposed to separate from the ski if the skier falls. If the binding fails to release, an accident is likely to result. A consulting service has been asked to design
an experiment to determine which one of \( k \) popular types of ski bindings is the safest to use, and safest is defined as the type with the smallest probability of an accident.

The plan is to obtain samples of \( n \) bindings of each type, give them to equipment rental agencies at various ski resorts, and have the agencies collect data on the number of accidents that occur. For each type of binding, the accidents are assumed to occur independently and with a constant probability so that the binomial distribution applies for the number of accidents. This probability \( p_j \) for the jth binding is estimated by the respective number of accidents per skier-day, i.e., the total number of accidents divided by the number of "skier days" (the number of ski days multiplied by the average number of skiers per skier day).

Suppose we have \( k = 2 \) types of bindings; we want to choose the safer one, that is, the one with the smallest \( p \) value, and we specify that the probability of a correct selection is to be at least \( P^* \) when \( \delta = p_{[2]} - p_{[1]} > \delta^* \) for some specified values of \( \delta^* \) and \( P^* \). Values of \( n \) required under the least favorable configuration for various \((\delta^*, P^*)\) pairs are given in Table 2. For example, if \( P^* = .90 \), we need samples of \( n = 83 \) if \( \delta^* = .10 \), and samples of size \( n = 329 \) if \( \delta^* = .05 \). How shall we choose \( \delta^* \)?

From extensive records of insurance companies it is known that the average rate of ski accidents is about .006, with a range of approximately .001 to .010. Thus the true difference between \( p_{[1]} \) and \( p_{[2]} \) cannot be very large and a realistic value of \( \delta^* \) must be very small; then the corresponding value of \( n \) for \( P^* = .90 \) under the least favorable configuration is so large that for practical purposes the experiment is impossible. Table 3 gives the very large \( n \) values required under the least favorable configuration for \( \delta^* = .001 \) and \( \delta^* = .010 \) and selected \( P^* \) values.
Table 3

Common n Value Required with the
Least Favorable Configuration

\[ p[1] = .5 - \delta*/2, \quad p[2] = .5 + \delta*/2 \]

<table>
<thead>
<tr>
<th>( \delta* )</th>
<th>.75</th>
<th>.90</th>
<th>.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>227,468</td>
<td>821,184</td>
<td>1,352,765</td>
</tr>
<tr>
<td>.010</td>
<td>2,275</td>
<td>8,211</td>
<td>13,527</td>
</tr>
</tbody>
</table>

The least favorable configuration for this problem when \( k = 2 \) is where the \( p \) values are centered about .5, specifically,

\[ p[1] = .5 - \delta*/2, \quad p[2] = .5 + \delta*/2 \]

Since the approximate range of \( p \) values is known to be .001 to .010, \( p[1] \) and \( p[2] \) might be pairs like (.001, .004), (.002, .006), and they cannot be more discrepant than (.001, .010). As a result, the true \( p \) values cannot be centered about .5, and the true configuration cannot be even close to the least favorable configuration; this means that any \( n \) value calculated under the least favorable configuration is highly conservative.

In order to make use of the information available, we should restrict the least favorable configuration to the range (.001, .010) of \( p[1] \) and \( p[2] \). With this alternative formulation the least favorable configuration is

\[ p[1] = .010 - \delta*, \quad p[2] = .010. \]

In other words, we are now restricted to the appropriate range of \( p \) values less than .010, rather than allowing the \( p \) values to be near .5, where the previous (unrestricted) least favorable configuration is located. Table 4 gives the common sample size \( n \) for \( k = 2 \) binomial populations under this restricted configuration (i.e., for \( p[2] = .010 \) for \( \delta* = .001 \) and \( \delta* = .010 \) and selected
P* values. The reader should note that for this formulation the investigator must specify p[2] (or p[1]) in addition to P* and δ*.

Table 4

Common n Value Required with the Configuration

<table>
<thead>
<tr>
<th>δ*</th>
<th>.75</th>
<th>.90</th>
<th>.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>8,562</td>
<td>30,908</td>
<td>50,916</td>
</tr>
<tr>
<td>.010</td>
<td>45</td>
<td>163</td>
<td>268</td>
</tr>
</tbody>
</table>

A comparison of the results in Table 4 with the corresponding results in Table 3 shows that for δ* = .001 the n values differ by a factor of about 25, whereas for δ* = .010 they differ by a factor of about 50. This shows how very important it is to make use of any information available about the true p values.

To complete this example, suppose that we give 31 ski bindings of each of the two types to each of 10 rental agencies. Since there are approximately 100 ski days per season, this plan will generate n = (10)(31)(100) = 31,000 observations on each type of binding in one season. The results given in Table 4 indicate that for n = 31,000, the probability of a correct selection is approximately (slightly over) .90 whenever p[2] is at most .010 and

6. The Philosophy of Subset Selection

Subset selection procedures may provide a useful and practical alternative to the indifference zone approach for the problem of (ultimately) selecting the best population. In this section we consider some typical subset selection problems with their associated procedures. We again define the best population as the one with the largest θ-value θ[k] and assume at the outset that θ[k] > θ[k-1] so that the best population is well defined.

Selecting a Subset that Contains the Best Population

The basic procedure selects a subset of random size and the goal is to have a high probability (at least P*, where P* is prespecified) of dividing the set of k populations into two identifiable groups or subsets of random size, one (the selected subset) which contains the best population and the other (the eliminated subset) which does not. Since no assertion is made about which population is best within the selected subset, a correct selection now occurs if the subset selected merely contains the population with parameter value θ[k], and an error occurs if the best population is contained in the eliminated group.

With this approach we first take a random sample from each of the k populations and compute some estimator \( \hat{\theta}_j \) of each \( \theta_j \) from the corresponding set of sample data. Then for each \( j = 1, 2, \ldots, k \), the selection rule is to place the \( j \)th population in the selected subset if and only if \( \hat{\theta}_j \) is included in a certain region \( I \). This region \( I \) is usually a closed interval of the form

\[
I = [\hat{\theta}_k - c, \hat{\theta}_k]
\]  

(6.1)

where \( c \) (to be determined) satisfies \( c > 0 \). Note that \( \hat{\theta}_k \) is
always contained in I and hence the subset selected cannot be empty.
(In some cases, especially in scale parameter problems, the region
I is a closed interval of the form
\[ I = [\hat{\theta}_k, \hat{\theta}_k] \text{ where } 0 < d < 1. \]

In order to determine the constant \( c \) for (6.1), we must first specify
a value \( P^* \) for the minimum probability of a correct selection, that is,
the probability that the subset selected contains the population with
parameter \( \theta_k \), regardless of the configuration of \( \theta \) values. The specified
\( P^* \) should satisfy \( .5 < P^* < 1 \) since the probability \( .5 \) can be achieved
by flipping a coin for each population to decide whether or not it goes into
the selected subset (without taking any observations). For \( k > 2 \) this
is better than randomly selecting one population for the selected subset
which has the lower bound \( 1/k \) for the probability of a correct result.
Based on a purely empirical rule-of-thumb, we recommend that \( P^* \geq .5 + (.5/k) \)
(i.e., that \( P^* \) be somewhat greater than \( .5 \) depending on \( k \)) in order to avoid
uninteresting degenerate situations. Since the experimenter will usually
want \( P^* \) around \( .90 \) or \( .95 \), this is not a restrictive condition. Here \( P^* \)
can also be interpreted as the joint confidence level for the statements that
the \( \theta \) value for each eliminated population is smaller than the \( \theta \) value of
the best population. Values of \( c \) are tabulated as a function of \( k \) and \( P^* \) for
some specific distributions and goals; these can also be obtained from
Table 1 at the end of this paper.

To illustrate this approach suppose we specify \( P^* = .90 \) in the poultry
stock example (of Becker) used above, where \( k = 10 \) and \( \sigma^2 = 5225 \). Suppose
chicks come in packages of 50 and we use \( n = 50 \) chicks from each of the
10 stocks. Then entering Table 1 at the end of this paper with \( k = 10 \) and
\[ P^* = 0.90, \text{ the appropriate } c\text{-value is given by} \]

\[ c = \frac{(2.9829) \sqrt{5225}}{\sqrt{50}} = 30.49, \quad (6.2) \]

and we put in the selected subset all stocks whose mean production for the 500-day experiment is in the closed interval \([\bar{x}_{[10]} - 30.49, \bar{x}_{[10]}]\).

Note that the \(c\)-value in this formulation is analogous to \(\delta^*\) in the indifference-zone formulation. Hence a reduction from 30.49 to 24 increases the \(n\)-value from 50 to 81; the smaller value of \(c\) (or \(\delta^*\)) corresponds to a larger sample size (or a more severe requirement).

The subset selection procedure described above has the property that the selected subset cannot be empty but the eliminated group can be empty. Hence to keep the efficiency of the procedure high we must keep the size of the selected group small. This is generally done by computing the expected size of the selected group (which is at least 1) and, as a subsequent separate problem, determining the common sample size \(n\) needed to keep the size of the selected subset less than \(1 + \varepsilon^*\) for specified \(\varepsilon^* > 0\) and for some particular configuration of interest. Thus although \(n\) is fixed and arbitrary for the main problem and for the principal applications, very small values of \(n\) may not give satisfactory results since we wish to have a small expected size of the selected subset and also a negligible probability of the trivial result, namely of putting all the populations into the selected subset.

Clearly the latter difficulties are associated with the fact that the selected subset is of random size. In those formulations where we select a subset of fixed size these considerations will not arise. Here the procedure is, of course, different and we are again back to determining
a least favorable configuration and a common sample size \( n \) needed to have probability at least \( P^* \) (under that configuration) that the selected subset of fixed size contains the one with parameter \( \theta[k] \).

**Comparison with a control or standard**

Two related problems are those of selecting a subset of \( k \) populations that includes all populations (i) better than a control population with unknown parameter \( \theta_o \), and (ii) better than a known standard value \( \theta_o \). The selection rule for (i) is to put population \( \Pi_1 \) into the selected subset if and only if

\[
\hat{\theta}_1 \geq \hat{\theta}_o - c,
\]

where \( c > 0 \) has to be determined (as a function of \( k \), \( \delta^* \) and \( P^* \)) and \( \hat{\theta}_o \) is an estimate of \( \theta_o \) based on a random sample from the control population. For problem (ii) we merely replace \( \hat{\theta}_o \) by the known value \( \theta_o \) for the standard. Here a correct selection occurs if all populations with \( \theta \) values greater than or equal to \( \theta_o \) are included in the selected subset, and \( P^* \) can be interpreted as the joint confidence level for the statements that the \( \theta \) value of each eliminated population is smaller than \( \theta_o \). In each of these two problems the subset selected can be empty; if this occurs the interpretation is that \( \theta_o \) is the largest \( \theta \) value, or equivalently that none of the \( k \) populations is better than (i) the control population, and (ii) the known standard value.

In this problem the specified \( P^* \) should satisfy \( (.5)^k < P^* < 1 \) since we can attain \( (.5)^k \) by flipping a fair coin for each of the \( k \) populations to decide whether it goes into the selected subset or the eliminated subset. However, we recommend keeping \( P^* \) above .5 and more precisely that \( P^* > .5 + (.5/k) \).
General Remarks

The subset selection approach is particularly useful in screening problems, e.g., drug screening. It is also appropriate when \( k \) is very large and we want to select a smaller number of populations to test further or to compare for secondary properties. Of course, the statistical control described here applies only to a single subset selection stage of such sequential problems.

The primary difference between the subset selection approach to selecting a subset containing the best population and the approach covered in Sections 3-5 is that in subset selection we have no indifference zone and the least favorable configuration is simply the worst configuration with all parameters equal. As a result, it is almost impossible to compare the two approaches analytically. However, we do point out that the indifference zone approach is useful at the experimental design stage where a common sample size is to be determined, whereas the subset selection approach (in the main formulation) assumes that the sample sizes may be fixed arbitrarily or by other considerations. When a subset is selected, no single population within that subset is asserted to be the best one (except by implication if it happens that the subset selected is of size one). Hence subset procedures give less precise information, but they do provide more flexibility and we may not have to specify the threshold value associated with an indifference zone.

The original papers for subset selection are Gupta (1956, 1965), Gupta and Sobel (1957, 1960), and Gupta, Huyett, and Sobel (1957). Generalized procedures that combine the indifference zone and subset selection approaches to ranking and selection have been considered in the literature. In particular, Sobel (1969) and Santner (1975) develop more general frameworks that cover both ideas.
7. **Description of Problems and Primary References**

In the last ten to fifteen years, a great number of publications in the statistical literature have been devoted to various theoretical aspects of many kinds of ranking and selection procedures. Some of the kinds of problems that have been solved using these procedures are as follows:

1. Selecting the **one** best population.

2. Selecting a **random** number of populations such that all populations better than a control population or standard are included in the selected subset.

3. Selecting the **t** best populations for \( t \geq 2 \),
   (a) in an ordered manner, or
   (b) in an unordered manner.

4. Selecting a **random** number of populations, say \( r \), that include the t best populations (for \( 1 \leq t \leq r \leq k \)).

5. Selecting a **fixed** number of populations, say \( r \), that include the t best populations (for \( 1 \leq t \leq r \leq k \)).

6. Ranking all the \( k \) populations from best to worst (or vice versa).

7. Ranking a **fixed-size** subset from best to worst (or vice versa).

The major references for the primary solution to some of the more important problems are listed below, grouped according to the type of problem. Unless otherwise noted, these solutions are one-stage procedures. Sequential or Bayesian procedures are not included in this list.
LIST OF PROBLEMS AND PRIMARY REFERENCES FOR SOLUTION
(USING ONE-STAGE PROCEDURES UNLESS OTHERWISE NOTED)

I. Selecting the one best population

A. Selecting the normal distribution with the largest (smallest) mean for
   1. variances common and known (Bechhofer, 1954),
   2. variances common and unknown (two-stage procedure) (Bechhofer, Dunnett, and Sobel, 1954; Dunnett and Sobel, 1954),
   3. variances unequal and known (Bechhofer, 1954),
   4. variances unequal and unknown (two-stage procedure) (Dudewicz and Dalal, 1975).

B. Selecting the binomial (or Bernoulli) distribution with the largest
   (smallest) probability of success (Sobel and Huyett, 1957).

C. Selecting the category with the largest probability for the multi-
   nomial distribution (Bechhofer, Elmaghraby, and Morse, 1959).

D. Selecting the normal distribution with the smallest variance
   (Bechhofer and Sobel, 1954).

E. Nonparametric procedures for selecting the distribution with the
   largest quantile of order q (Sobel, 1967).

F. Selecting the best object in a design with paired comparisons (David,
   1963).

G. Selecting the normal distribution with the largest (smallest)
   regression value.

H. Selecting the Gamma distribution with the smallest (or largest) value
   of the scale parameter θ (Gupta, 1963).

I. Selecting the multivariate normal distribution with the largest
   multiple correlation (Rizvi and Solomon, 1973).

J. Selecting the multivariate normal distribution with the largest
   weighted average of means.
II. Selecting a subset of populations that contains the one best population

A. Selecting a random size subset with respect to

1. Means for normal distributions with common known variance (Gupta, 1956, 1965) and with common unknown variance (Gupta, 1956; Gupta and Sobel, 1957).

2. Probabilities for binomial (or Bernoulli) distributions (Gupta and Sobel, 1960; Gupta, Huyett, and Sobel, 1957).

B. Selecting a fixed size subset with respect to means for normal distributions with a common known variance (Desu and Sobel, 1968).

III. Selecting a subset of populations that includes all populations better than a control with respect to

A. Means for normal distributions with known variances (not necessarily common) (Tong, 1969).

B. Probabilities for binomial (or Bernoulli) distributions (Gupta and Sobel, 1960; Gupta, Huyett, and Sobel, 1957).

IV. Selecting all populations better than a control with respect to

A. Means for normal distributions with a common known variance and with a common unknown variance (two-stage procedure) (Tong, 1969).

B. Variances for normal distributions (Schafer, 1976).

V. Selecting the t best populations for normal distributions with a common known variance (Bechhofer, 1954).

VI. Complete ordering (or ranking) of normal populations with respect to

A. Means for a common known variance (Carroll and Gupta, 1975), and with a common unknown variance (two-stage procedure) (Freeman, Kuzmack, and Maurice, 1967).

B. Variances (Schafer, 1974; Schafer and Rutemiller, 1975).

Details and examples for most of these problems are given in Gibbons, Olkin, and Sobel (1977); this reference also includes the tables needed for applications.
Basic Journal References on the Theory of Ranking and Selection


Books


Journal References on the Applications of Ranking and Selection


Figure 1. PZ is \(|\theta_2 - \theta_1| \geq \delta^*\); TPS is \(-\infty < \theta_1, \theta_2 < \infty\) (axes unordered)

Figure 2. PZ is \(|\theta_2 - \theta_1| = \theta_2 - \theta_1 \geq \delta^*\); TPS is \(-\infty < \theta_1 < \theta_2 < \infty\)
(axes ordered)
Figure 3. PZ is $\theta[k] - \theta[k-1] \geq \delta^*$; TPS is $-\infty < \theta[k-1] < \theta[k] < \infty$

Figure 4. PZ is $\theta[k] - \theta[k-1] \geq \delta^*$; TPS is $0 < \theta[k-1] < \theta[k] < 1$

Figure 5. PZ is \[ \frac{\theta[k](1 - \theta[k-1])}{(1 - \theta[k])\theta[k-1]} \geq \delta^* \]; TPS is $0 < \theta[k-1] < \theta[k] < 1$
Figure 6. Operating characteristic curves for the problem of selecting the one normal population with the largest mean when $k = 4$, $n = 20$ and $n = 50$; values of $(\delta^*, p^*)$ such that the confidence level is $P^*$ that $\mu_{[4]} - \delta^* \leq \mu_s \leq \mu_{[4]}$, or values of $(\delta^*, p^*)$ such that $P(\text{CS}|\theta) \geq P^*$ for all $\delta = \mu_{[4]} - \mu_{[3]} \geq \delta^*$.

Figure 7. Operating characteristic curves for the problem of selecting the one binomial population with the largest probability of success when $k = 4$, $n = 20$ and $n = 50$; values of $(\delta^*, p^*)$ such that the confidence level is $P^*$ that $p_{[4]} - \delta^* \leq p_s \leq p_{[4]}$, or values of $(\delta^*, p^*)$ such that $P(\text{CS}|\theta) \geq P^*$ for all $p_{[4]} - p_{[3]} \geq \delta^*$. 
Figure 8. Operating characteristic curve for the problem of selecting the one normal population with the smallest variance when \( k = 4 \), \( \nu = 20 \) and \( \nu = 50 \) (\( \nu = n \) for means known, \( \nu = n-1 \) for means unknown; values of \( (\Delta^*, p^*) \) such that the confidence level is \( p^* \) that \( \frac{\sigma_{[1]}}{\sigma_{[1]}} \leq \frac{\sigma}{\Delta^*} \) or equivalently that

\[
0 \leq \frac{\sigma - \sigma_{[1]}}{\sigma_{[1]}} \leq \sigma_{[1]} \left( \frac{1}{\Delta^*} - 1 \right),
\]

or values of \( (\Delta^*, p^*) \) such that

\[
P(\text{CS}|0) \geq p^* \text{ for all } \Delta = \frac{\sigma_{[1]}}{\sigma_{[2]}} \leq \delta^*.
\]
Table 1

Smallest Value of \( \tau \) Needed to Satisfy the \((\delta^*, P^*)\) Requirement in Selecting the One Population with the Largest Mean in the Case of \( k \) Normal Populations with Known Common Variance \( \sigma^2 \)

<table>
<thead>
<tr>
<th>( P^* )</th>
<th>(.750)</th>
<th>(.900)</th>
<th>(.950)</th>
<th>(.975)</th>
<th>(.990)</th>
<th>(.999)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.2</td>
<td>0.9539</td>
<td>1.8124</td>
<td>2.3262</td>
<td>2.7718</td>
<td>3.2900</td>
<td>4.3702</td>
</tr>
<tr>
<td>3</td>
<td>1.4338</td>
<td>2.2302</td>
<td>2.7101</td>
<td>3.1284</td>
<td>3.6173</td>
<td>4.6450</td>
</tr>
<tr>
<td>4</td>
<td>1.6822</td>
<td>2.4516</td>
<td>2.9162</td>
<td>3.2220</td>
<td>3.7970</td>
<td>4.7987</td>
</tr>
<tr>
<td>5</td>
<td>1.8463</td>
<td>2.5997</td>
<td>3.0552</td>
<td>3.4532</td>
<td>3.9196</td>
<td>4.9048</td>
</tr>
<tr>
<td>6</td>
<td>1.9674</td>
<td>2.7100</td>
<td>3.1591</td>
<td>3.5517</td>
<td>4.0121</td>
<td>4.9855</td>
</tr>
<tr>
<td>7</td>
<td>2.0626</td>
<td>2.7972</td>
<td>3.2417</td>
<td>3.6303</td>
<td>4.0860</td>
<td>5.0504</td>
</tr>
<tr>
<td>8</td>
<td>2.1407</td>
<td>2.8691</td>
<td>3.3099</td>
<td>3.6953</td>
<td>4.1475</td>
<td>5.1046</td>
</tr>
<tr>
<td>9</td>
<td>2.2067</td>
<td>2.9301</td>
<td>3.3679</td>
<td>3.7507</td>
<td>4.1999</td>
<td>5.1511</td>
</tr>
<tr>
<td>10</td>
<td>2.2637</td>
<td>2.9829</td>
<td>3.4182</td>
<td>3.7989</td>
<td>4.2456</td>
<td>5.1916</td>
</tr>
<tr>
<td>15</td>
<td>2.4678</td>
<td>3.1734</td>
<td>3.6004</td>
<td>3.9738</td>
<td>4.4121</td>
<td>5.3407</td>
</tr>
<tr>
<td>20</td>
<td>2.6009</td>
<td>3.2986</td>
<td>3.7207</td>
<td>4.0899</td>
<td>4.5230</td>
<td>5.4409</td>
</tr>
<tr>
<td>25</td>
<td>2.6987</td>
<td>3.3911</td>
<td>3.8099</td>
<td>4.1761</td>
<td>4.6057</td>
<td>5.5161</td>
</tr>
</tbody>
</table>

Table 2

Smallest Integer Sample Size \( n \) Needed to Satisfy the \((\delta^*, P^*)\) Requirement in Selecting the Binomial Population with the Largest Probability from \( k = 2 \) Populations

<table>
<thead>
<tr>
<th>( \delta^* )</th>
<th>( P^* )</th>
<th>(.50)</th>
<th>(.60)</th>
<th>(.75)</th>
<th>(.80)</th>
<th>(.85)</th>
<th>(.90)</th>
<th>(.95)</th>
<th>(.99)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.05</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>15</td>
<td>29</td>
<td>1082</td>
</tr>
<tr>
<td>.10</td>
<td>0</td>
<td>4</td>
<td>23</td>
<td>36</td>
<td>54</td>
<td>83</td>
<td>135</td>
<td>270</td>
<td></td>
</tr>
<tr>
<td>.15</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>16</td>
<td>24</td>
<td>37</td>
<td>60</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>.20</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>21</td>
<td>34</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>.25</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>13</td>
<td>22</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>.30</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>15</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>.35</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>11</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>.40</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>.45</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>.50</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>