PROCEDURES FOR GROUPING A SET
OF OBSERVED MEANS

By

JOHN RIFFE MURPHY

Bachelor of Science
Panhandle State College
Goodwell, Oklahoma
1964

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1967

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1973

PROCEDURES FOR GROUPING A SET

OF OBSERVED MEANS

Thesis Approved:

David L. Weeks
Thesis Adviser

Lyle D. Broemeling

Robert D. Morrison

John P. Chandler

J. Leroy Folks

D. D. Durham
Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like first to extend my sincere thanks and appreciation to Dr. David L. Weeks for suggesting the problem and for his invaluable guidance and assistance during the course of this research.

I wish also to thank each member of my advisory committee, Dr. J. Leroy Folks, Dr. Lyle Broemeling, Dr. J. P. Chandler, and Dr. Robert D. Morrison, for their interest and helpful suggestions.

I would also like to thank Dr. John Jewett for making a significant portion of the limited departmental allocation of computer time available to me for this research.

I wish to express appreciation to the National Aeronautics and Space Administration for the financial support in the form of a Traineeship during 1967 to 1970, and, also, to the Environmental Protection Agency for support in the form of a Grant during this final year of study.

To one whose friendship I value most highly, Cuthbert Daniel, I want to express sincere gratitude for taking an interest in my development as an applied statistician and for sharing with me his years of experience and rare insight.

A special thanks goes to my wife, Rosemary, and my family, who have had to do without for so many years.

Finally, to my mother and father for their continual encouragement and assistance, both material and otherwise, I owe more than I can ever repay, for college and graduate school would not have been possible without them.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1. Statement of the Problem

Suppose an experimenter has obtained as the result of an experiment, k means, each based upon n observations, and that he is willing to assume that they are samples from normal populations with common unknown variance. That is, X_{ij} are independently distributed as $N(\mu_i, \sigma^2)$, $i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n$. Having now $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k$, he desires a procedure which, in an unambiguous way will give him (with a reasonable chance of being correct) the pattern in the μ_i's. By this it is meant that the experimenter asks for some procedure to give him $q \leq k$ groups, such that the means in any one group are not appreciably different from each other.

The problem is not one of classification in the strictest sense, because the experimenter does not know *a priori* how many populations he must discriminate between; there can be anywhere from 1 to k of them.

Formally, let us take the following formulation as a reasonable approximation to the situation: Let $X_1, X_2, \ldots, X_k$ be independently distributed as $N(\mu_i, \sigma^2)$ and let s^2 be an estimate of σ^2, independent of the X_i's, with $\nu s^2/\sigma^2$ distributed as $\chi^2(\nu)$. For some $q \leq k$, there is a set $\lambda_1, \lambda_2, \ldots, \lambda_q$ such that each μ_i is some λ_j. A statistic $H(X_1, X_2, \ldots, X_k, s^2)$ is sought which will discern the

λ_j's. This is essentially the formulation given by Plackett (15); however, we shall not necessarily take as the primary goal the estimation of the λ_j as he does. The first objective is to determine q and which μ_i's belong together, after which it may be of interest to estimate perhaps some of the λ_j.

There are several procedures the experimenter might consider using, and he would probably first try some of the more commonly known multiple comparison techniques. Prior to applying any formal procedure, however, it seems likely that he would rank the k means according to their magnitude, perhaps even plotting them to get a look at their arrangement. Let us assume that this has been done, and now the experimenter wants to determine if the pattern observed (or tentatively hypothesized) is supported by further analysis.

2. Procedures Which Do Not Effect Unique Groupings

To implement the techniques to be discussed below, assume that the experimenter employs the ranking and underlining method, whereby series of underlinings are made under the ranked means with the interpretation that any group of means with a continuous line under them are to be considered not significantly different from one another.

2.1 Fisher's LSD

Originally proposed by R. A. Fisher (8) in 1935, this test consists of initially testing for non-homogeneity among the means with an F-test. If the F-test provides evidence of non-homogeneity of the means, one then proceeds to the second stage, otherwise all means are declared not significantly different from each other. At the second stage, the mean differences are tested pairwise with a series of t-tests. If the k means are all based on the same number of observations, say n, then the t-tests can be implemented by computing the least significant difference,

$$LSD = (s_{(\bar{X}_i - \bar{X}_j)}) \times t_{\alpha/2}(k(n-1)) = \sqrt{2/n} \times s \times t_{\alpha/2}(k(n-1)),$$

and comparing the observed mean differences with the LSD. Consider the following data presented as an example in Snedecor and Cochran (17) from a one-way classification experiment. Four classifications are used with six observations taken in each class. The data and pertinent statistics are summarized in Table I below.

TABLE I

DATA FROM A ONE-WAY CLASSIFICATION EXAMPLE TAKEN FROM SNEDECOR AND COCHRAN

Classes:	1	2	3	4
Responses:	64	78	75	55
	72	91	93	66
	68	97	78	49
	77	82	71	64
	56	85	63	70
	95	77	76	68
Means:	72	85	76	62

Mean Square Between Classes = 545.3 with 3 degrees of freedom.

Mean Square Within Classes = 100.9 with 20 degrees of freedom.

$F_{cal} = 5.40$ with 3 and 20 degrees of freedom; $F_{.01}(3,20) = 4.94$.

$LSD_{.05}(20) = \sqrt{100.9/3} \times 2.086 = 12.1$.

The F-test provides strong evidence that the observed means do not all belong to the same population, and one therefore proceeds to step 2. The means are ranked and plotted on a horizontal scale, and the differences computed and compared with the LSD. The significant differences according to the LSD criterion are indicated by the underlinings as in Figure 1. One concludes that Class 4 is not different from Class 1, but is different from Class 3 and Class 2; Class 1 is different from Class 2; and Class 3 is not different from Class 2.

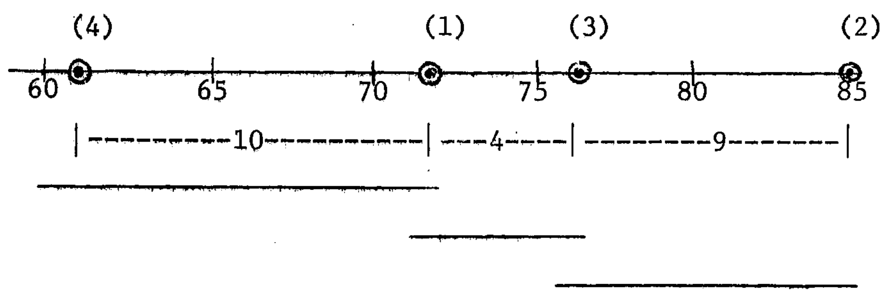


Figure 1. Results of the .05 level LSD test

It is now apparent that the original goal of the experimenter has not been met, in that an unambiguous grouping of the means has not been obtained; for example, from a grouping standpoint, the .05 level LSD would assign Class 1 to two different groups. In other words, the procedure tells us that $\overline{X}_4$ and $\overline{X}_3$ are probably samples from populations having different means, but it does not give any indication as to which of the two populations $\overline{X}_1$ belongs. One might consider overcoming this difficulty by testing only the differences between adjacent ranked means.

Any significant gaps found would be the breaking points between groups. However, a problem is also encountered here. Returning to the example, if only adjacent differences are tested, then at the .05 level of significance, no significant breaks would be declared, and only one group would result, the whole sample. Yet, there are differences within the group that are significant by the .05 level LSD, if further testing is done.

It appears that part of the problem arises from the fact that with the LSD procedure, all differences are tested with the same value, whereas all differences are not the same, on the average. Consider repeated sampling with samples of size 4 from a standard normal distribution. The expected differences of the ordered observations are (see Table XIX, Appendix)

$$E(X_{(2)}-X_{(1)}) = .7324$$

$$E(X_{(3)}-X_{(2)}) = .5940 \quad E(X_{(3)}-X_{(1)}) = 1.3264$$

$$E(X_{(4)}-X_{(3)}) = .7324 \quad E(X_{(4)}-X_{(2)}) = 1.3264 \quad E(X_{(4)}-X_{(1)}) = 2.0588$$

while in a sample of size 2, $E(X_{(2)}-X_{(1)}) = 1.1284$. Taking the expected differences as an indication of what happens on the average, it seems reasonable to conclude that the adjacent differences would be less likely to be found significant than other differences and that the difference $X_{(4)}-X_{(1)}$ would be found significant more often than it ought to be, even in null experiments. Thus, one would be faced with the dilemma above more often than he would like. In experiments where the true differences are large, of course, there is no problem; all differences would be almost certainly declared significant. Furthermore, by performing the α-level F-test first, one is allowing himself, on the average, in most $100\alpha\%$ of the null experiments to make the Type I error of declaring $\overline{X}_{max} - \overline{X}_{min}$ erroneously significant, and, moreover, in null experiments, each pair of the unranked means has the same chance of being $\overline{X}_{max} - \overline{X}_{min}$. The persisting question is, "In how many of the non-null experiments are real adjacent differences not being detected?"

2.2 Studentized Range Q-method

This method, called the Q-method by Snedecor and Cochran (17), and referred to by Steel and Torrie (18) as Tukey's W-procedure (20), is similar to Fisher's LSD method, but is based on the distribution of the studentized range. First the preliminary F-test is abandoned, and then the problem of making the Type I error of declaring $\overline{X}_{max} - \overline{X}_{min}$ significant too often is remedied by comparing all differences with the null sampling distribution of the studentized range. This is a two-parameter distribution, the parameters being the sample size n and the degrees of freedom ν of the studentizing statistic $s_{\overline{X}}$. To implement the procedure, one determines the critical α-level value of $Q = (\overline{X}_{max}-\overline{X}_{min})/s_{\overline{X}}$ for the particular values of n and ν. The criterion $W = Q \times s_{\overline{X}}$ is then computed, and all differences greater than W are declared significant at the level α. For the data of Table I, $W_{.05}(4) = Q_{.05}(4,20) \times 4.1 = 16.2$. The results of the .05-level test are shown in Figure 2, on the next page.

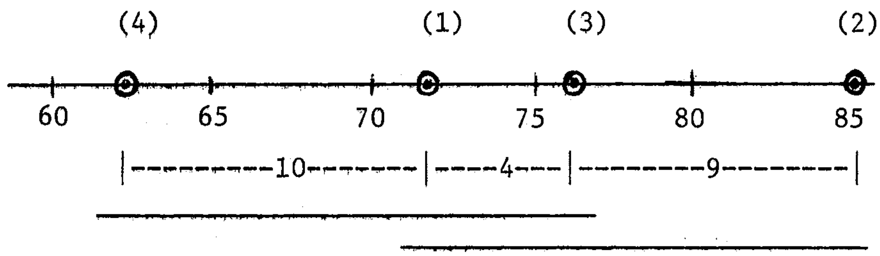


Figure 2. Results of the .05 level Q-test

Again, the original goal of the experimenter has not been attained; for example, "To which group does Class 3 belong?" It will not do to test only adjacent differences either, because, to offset the protection afforded by a preliminary F-test, the Q-method is more conservative than the LSD when testing adjacent differences, and, in fact, is a conservative test (with respect to Type I errors) when applied to any difference other than $\overline{X}_{max} - \overline{X}_{min}$. Hence, one is again faced with the suspicion that too many adjacent real differences are declared not significant.

Neither the LSD nor the Q-method deal with the logical inconsistency of declaring a significant difference and yet retaining the original null hypothesis for further testing. Consider a one-way experiment having four levels, and assume that the class means have been computed and ranked. Suppose one begins with some difference, say $\overline{X}_{(3)} - \overline{X}_{(2)}$ and finds this difference significant. By declaring significance, he is, in fact, discarding the null hypothesis that $\mu_1 = \mu_2 = \mu_3 = \mu_4$. For any further testing among $\overline{X}_{(1)}$, $\overline{X}_{(2)}$, $\overline{X}_{(3)}$, or $\overline{X}_{(4)}$, he must now take a new tentative null hypothesis of the form, "The two smaller means are a sample from the same population, and the two larger means are a sample from the same population, and the two populations are not the same." For larger group sizes, if significances are found at this stage, then the tentative null hypothesis should be changed again, and so on. No reference to this phenomenon has been found in the literature, and it is apparently not regarded as a serious problem by most, but it is difficult to see how one can avoid dealing with it whenever the situation requires more than one test and/or decision to be made. It may be argued that the LSD and the Q-method are nonsequential tests--that the testing at any stage does not depend on the results of previous tests, but this is invalid, for it is impossible to perform two or more tests simultaneously. They cannot be physically performed in any manner other than sequentially.

The next two procedures to be discussed do take into account the "changing null hypothesis" aspect of the problem, although this particular feature is not generally emphasized in discussions of them.

2.3 Student-Newman-Keuls Sequential Studentized Range Procedure (SNK)

This procedure, commonly called the Newman-Keuls procedure, is multistage and is currently carried out in a slightly different way than proposed by either Newman (14) or Keuls (11). The procedure is implemented by ranking the means and testing the observed studentized range $(\bar{X}_{(n)} - \bar{X}_{(1)})/s_{\bar{X}}$ via comparison with the null sampling distribution of this statistic in a sample of size n. If significance is found, one proceeds to test $(\bar{X}_{(n-1)} - \bar{X}_{(1)})/s_{\bar{X}}$ and $(\bar{X}_{(n)} - \bar{X}_{(2)})/s_{\bar{X}}$ with the appropriate distribution for $n - 1$ means. If both are declared significant, the three $n - 2$ ranges are tested, and so on.

If, however, a non-significant range is found at any stage, all ranges included within that range are declared not significant. The underlining technique is then usually applied with the usual interpretations.

Newman, at the suggestion of 'Student', was the first to tabulate the percentage points of the studentized range and to advocate its use in conjunction with the Analysis of Variance. The sequence of testing he proposed was determined by deleting the "most divergent" mean or means whenever the range was declared significant. Keuls, on the other hand, proposed that the sequence of testing be done in two main parts. In the first part, the sequence would be determined by eliminating the smallest mean of a ranked group each time. In the second part, the sequence would be determined by eliminating, each time, the largest mean of a ranked group. Presumably, if the underlining technique is applied, and the same stopping rule that is now used is applied, the same set of underlinings would result as those given by the current procedure. It is interesting to note that Keuls suggested that, in agricultural experiments, at least, the primary interest of the experimenter is often in answering the question, "What is the grouping?", and although he did not claim that the procedure provided the answer, he hinted that it could be useful in such an endeavor. The SNK procedure can be helpful in providing a partial answer, but it should be observed that the procedure can also give ambiguous results when applied as a grouping detection procedure. Applying the SNK procedure to the data of Table I, one computes $W_{.05}(4) = 16.2$ and $W_{.05}(3) = 14.7$, and obtains the results shown in Figure 3 on the next page.

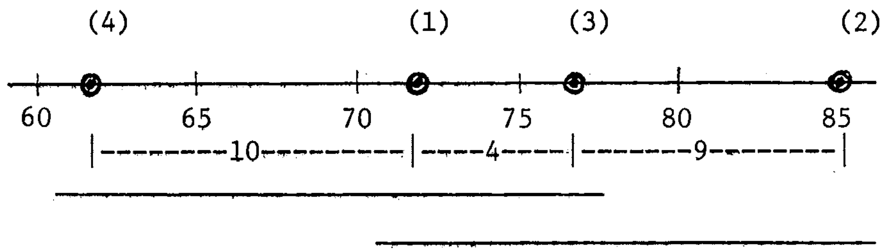


Figure 3. Results of the .05-level SNK Test

In this particular example, one arrives at the same conclusions as obtained with the Q-method, and consequently, with the same ambiguities with respect to the goal of determining the grouping.

2.4 Duncan's Multiple Range Procedure (DMR)

The DMR procedure proposed by Duncan (6) is identical to the SNK procedure, with the exception that a single α-level is not used throughout the test. Instead, at the kth stage when groups of $n - k + 1$ means are being tested, the α-level used is $\alpha(n - k + 1) = 1 - (1 - \alpha)^{n - k}$. The use of these "p-mean significance levels" as proposed by Duncan, has generated a great deal of debate, and the arguments pro and con will not be discussed here, since we are interested only in whether the procedure provides an answer to the grouping question. With regard to this, it is evident that the DMR procedure will also produce ambiguities.

Other procedures which would have to be contained in any discussion of multiple comparison procedures include, as a minimum: Tukey's Studentized Range Simultaneous Confidence Interval Procedure, Sheffe's

F-projections, Studentized Maximum Modules Procedures, Dunnett's Treatments vs Control Procedures, Duncan's Multiple F-test Procedure, the short-cut techniques of Kurtz, Link, Tukey, and Wallace, and the several nonparametric procedures. However, since this discussion of multiple comparisons is limited to the grouping problem aspect, descriptions of the above procedures will not be included, as they do not appear to shed any further light on the solution to the problem.

All of the procedures discussed to this point employ the technique of ranking the observed means. Very little thought is needed before one realizes that the observed ranking does not necessarily reflect the true ranking of the populations of which the means are a sample. The next procedures to be discussed deal with certain aspects of the ranking problem.

2.5 Bechhofer, Dunnett, and Sobel Procedures

Bechhofer (3) and, in a later paper, Bechhofer, Dunnett, and Sobel (4) attacked the problem of selecting the best k_s populations, the next best k_{s-1} populations, . . . , the worst k_1 populations for k means, ($k = k_1 + k_2 + . . . + k_s$), from (possibly) different normal (μ_i, $a_i\sigma^2$) distributions. In the first paper, the a_i and σ^2 were assumed known, and in the second paper, the a_i were assumed known and σ^2 was assumed unknown. For the latter case, a two-stage procedure is required; however, in both cases, the final step is that of ranking the observed means and declaring the largest k_s means best, the next largest k_{s-1} next best, and so forth. Thus, the procedures themselves consist essentially of stating a decision rule, and then evaluating the probabilities of incorrect decisions (rankings) for various alternatives.

It might appear at first glance that these procedures would be fertile ground for the solution of the grouping problem, but further reflection shows that they are applicable only if the experimenter knows the k's beforehand; here, it is being assumed that he does not.

3. Procedures Which Do Effect a Unique Grouping

Up to now, the discussion has dealt with procedures which can be readily discounted due to the fact that they fail to yield an unambiguous grouping. The procedures to be discussed below; however, all satisfy the unique grouping requirement. The first two procedures essentially non-statistical in nature.

3.1 A Graphical Procedure

This technique, proposed as an alternative to multiple comparison procedures by both Plackett and Nelder (15), consists of plotting the pairs $(\overline{X}_{(i)}, \xi_i)$, where ξ_i is the expected value of the ith order statistic in a sample from the standard normal distribution, and $\overline{X}_{(i)}$ is the ith of the ranked observed means, $i = 1, 2, \ldots, n$. A series of parallel lines with slope $1/s_{\overline{X}}$ are tried on the plotted values, and a judgement is made as to how many lines are needed for a reasonable fit. The number of parallel lines required is taken as the number of groups, and those points corresponding to a given line are taken to comprise that group. The use of the procedure was illustrated by Plackett with data from an example by Duncan quoted earlier in the paper by O'Neill and Wetherill (14). In the example, seven varieties of barley, A, B, C, D, E, F, and G were tested in a randomized block experiment with six blocks. The ranked variety mean responses were:

A	F	G	D	C	B	E
49.6	58.1	61.0	61.5	67.6	71.2	71.3

and $s_{\bar{X}}$, the standard deviation of a variety mean was 3.64 with 30 df. The graphing procedure yielded (in Nelder's opinion) the pattern shown in Figure 4. If it is assumed that the observed pattern is representative of the true grouping of the means, then the conclusion is that there are three groups: A by itself; F, G, and D together; and C, B, and E together.

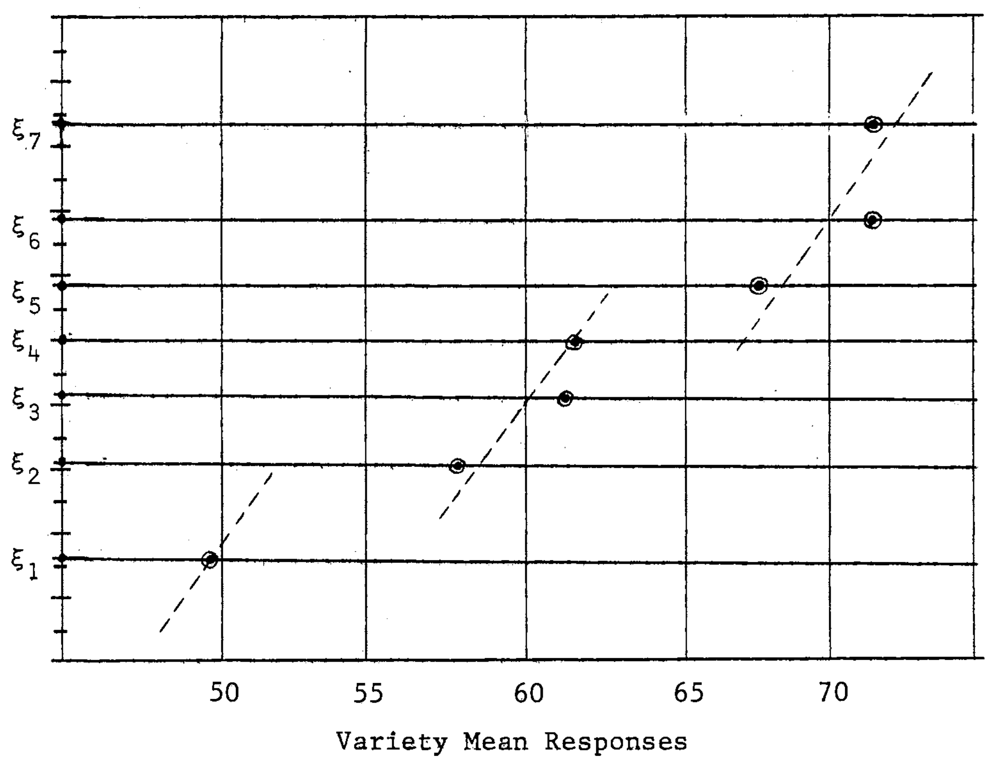


Figure 4. Results of the Graphical Procedure

This procedure has great appeal due to its simplicity; its chief disadvantage lies in its subjective nature. Nelder presents some pertinent results concerning the variances and covariances of the quantities $q_i = (\overline{X}_{(i+1)} - \overline{X}_{(i)})/(\xi_{i+1} - \xi_i)$ and of the residual quantities $e_i = (\overline{X}_i - \overline{X}) - \xi_i s_{\overline{X}}$, but in spite of his utilization of these results, the procedure remains largely subjective. It is conceivable that situations could arise where different conclusions would be obtained depending upon where one takes his starting point for breaking off the groups, for example, starting with the smallest means or starting with the largest means. Even if the null sampling distributions of the q_i were determined, one would be faced with both the starting point problem as well as the problem of the changing null hypothesis. If the procedure were expanded so as to overcome these difficulties, it would lose its greatest asset, that of simplicity.

3.2 <u>One-dimensional Cluster Analysis</u>

This procedure has been advanced by some, e.g. Plackett and Jolliffe (15), as a possible approach to the grouping problem. The essentials of a cluster analysis for one dimension appear to be:

(i) A set of observations.

(ii) A measure of distance between single points.

(iii) A rule for measuring distance from a single point to a group of points.

(iv) An algorithm for determining how the groups are to be built up or broken down.

Application of cluster analysis to the situation of multiple comparisons with the usual metrics appears to accomplish little more than would be accomplished by simply looking at the data. Jolliffe suggested that taking the observed significance levels of the gaps as the metric might be useful; however, one can see difficulties with item (iii) above. Moreover, after the data were subjected to the cluster analysis procedure, even with the significance level metric, it is not clear that definite conclusions would be reached with any of the cluster analysis algorithms in current use.

The following procedure could be termed semi-statistical, in that a preliminary F-test is to be used to determine if there is more than one group.

3.3 Mean Range Grouping Criterion

This technique proposed by Ottestad (16) is essentially a decision rule for grouping the means in the event the overall F-test gives significance. The criterion $V_n = E(W_n)/s_{\bar{X}}$, where $E(W_n)$ is the expected range of a sample of size n from the standard normal, is formed, and group boundaries are then taken to be $\bar{X}_{max} - V_n$ and $\bar{X}_{min} + V_n$. When this is done, one of four possibilities will occur:

(i) $\bar{X}_{min} + V_n \leq \bar{X}_{max} - V_n$, and none of the means fall between.

(ii) $\bar{X}_{max} - V_n \leq \bar{X}_{min} + V_n$, and none of the means fall between.

(iii) $\bar{X}_{min} + V_n < \bar{X}_{max} - V_n$, and none of the means fall between.

(iv) $\bar{X}_{max} - \bar{V}_n < X_{min} + V_n$, and none of the means fall between.

Presumably, in Cases (i) and (ii), two groups would be declared; in Case (iii), three groups would be declared; and in Case (iv), one is faced with a serious problem, to say the least.

The most striking feature of this procedure is its arbitrary character; no rationale is offered as to why V_n is a reasonable criterion, if indeed it is. The technique is not recommended, since it seems to offer little advantage over a completely subjective grouping and since no recommendations are given to the experimenter for resolving the difficulty represented by Case (iv).

The remainder of the procedures to be discussed could be classified as statistical procedures, in that they are presented in the form of sequential significance testing.

3.4 F-Test/Maximum Gap Sequential Test

Ottestad (16) also proposed that it may be possible to determine the grouping by performing the overall F-test, and, if significant differences are implied, to take the strongly marked gaps as indication of the grouping. In order for this approach to be useful in practice, it should be expanded somewhat to sequential testing, and more definite recommendations in the event of a significant F-test need to be given. A reasonable method might be to proceed as follows: perform the overall F-test and if significance is found, take the break to be at the largest gap. For each of the two groups, repeat the procedure for the appropriate sample sizes, and continue in this manner until no more significant F values are found. The grouping would then be determined by the breaks which were declared over all stages.

The chief disadvantage of a test such as this is the computation involved for forming the F ratios. Squares must be summed each time, and computation of sums of squares tends to be regarded as cumbersome. For that reason the procedure is not recommended because the Range/Gap test to be discussed next is simpler to apply.

3.5 Studentized Range/Maximum Gap Test

Tukey (20) suggests that a test combining the null sampling distribution properties of both the range and gaps would have merit. There are, undoubtedly, several ways of doing this, and one way would be to test the studentized range, and then test the studentized maximum gap for significance, if a significant range is found. A break would be declared only if the group "passes" both tests. This mode of testing is not recommended, however, because of the unsatisfactory state of affairs when one test gives significance and the other does not. What is proposed instead is a testing procedure similar to the F/Gap test discussed above. One proceeds in the same manner as before, except that the studentized range is used in the place of an F ratio. The procedure will be investigated further in later chapters for purposes of comparison with the studentized maximum gap procedure.

3.6 Tukey's Gap-Straggler-Variance Procedure

This procedure as proposed by Tukey (19) has been described in various places (13) as "rather messy", "a little lengthy", and "Rube Goldbergish", and has been pronounced "entirely obsolete" by its own author (20). There are, however, some points raised in the same reference which are not obsolete, and which have direct relevance to the grouping problem. In particular, Tukey noted that there are at least three ways that a group of (ranked) means can exhibit non-homogeneity:

(i) There are noticeable gaps between adjacent means.

(ii) When taken as a whole or in groups, some means straggle from the rest.

(iii) The group has excess variability.

Tukey designed the procedure to detect each of these three types of heterogeneity. A non-detailed description of the procedure is as follows:

Step 1. Rank the means.

Step 2. Test the differences of adjacent means (gaps) with the LSD. Any significances found are taken as group boundaries.

Step 3. In all the groups formed, test for stragglers (outliers), and break off those found from the groups. If any new groups are formed by the stragglers, reapply the straggler test until no new stragglers are found.

Step 4. Apply an F-test for homogeneity to each of the groups formed up to this point.

If at Step 4, some group is found to be significantly variable, a conclusion such as, "These means do not belong with the others, but neither do they all belong together." would be made.

Tukey's 1949 paper is significant for several reasons. First, it gave respectability to the idea that an experimenter's goal could be taken, in some cases, to be that of finding the pattern or determining the grouping and was directly addressed to that problem. Secondly, as he later points out (20), although the particular procedure is obsolete, the use of gaps may not be, and probably merits further investigation. Finally, he suggested that a procedure based on the distribution of the studentized maximum gap might show some promise.

His comments there and elsewhere have been taken as a starting point, and a sequential test procedure based upon approximations to the distribution of the studentized maximum gap has been devised. The procedure is outlined and briefly discussed in the next chapter. The mathematics and computing to arrive at the distributions are presented in Chapter III. In Chapter IV, the performance of the proposed procedure is studied, along with the Range/Gap procedure discussed earlier, and, also the LSD procedure (applied only to gaps).

CHAPTER II

GENERAL DISCUSSION OF THE PROCEDURE

Let $(X_1, X_2, \ldots, X_n)$ be a random sample from a standard normal distribution, and let $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ be the ordered sample values. Consider, now, the question, "In repeated sampling, what is the probability that $X_{(2)} - X_{(1)} > 1$?" First, let us ignore the fact that $X_{(1)}$ and $X_{(2)}$ are the two smallest observations in a larger sample. In this case, $\Pr(X_{(2)} - X_{(1)} > 1)$ is computed as

$$\begin{aligned} \Pr(X_{(2)} - X_{(1)} > 1) &= \Pr(X_2 - X_1 > 1 \text{ or } X_1 - X_2 > 1) \\ &= \Pr(X_2 - X_1 > 1) + \Pr(X_2 - X_1 < -1) \\ &= 2\Pr(X_2 - X_1 > 1) \\ &= 2(1 - F(1/\sqrt{2})) \\ &= 0.4795, \end{aligned}$$

where,

$$F(x) = (1/\sqrt{2\pi})\int_{-\infty}^{x} e^{-t^2/2}dt .$$

Now, consider the fact that $X_{(1)}$ and $X_{(2)}$ are part of a larger sample; then $X_{(2)} - X_{(1)}$ is gap 1 in a sample of size n. In Chapter III, a method of computing the probabilities for the individual gaps is shown, and given below are the probabilities that gap 1 is larger than 1 computed by that method for several values of n.

$\underline{n}$	$Pr(X_{(2)} - X_{(1)} > 1)$
2	0.4795
3	0.3396
4	0.2733
5	0.2339
6	0.2075
7	0.1884
8	0.1738
9	0.1622
10	0.1528

It is evident that two different answers to the question are obtained whenever the sample size is larger than 2.

For the case where σ^2 is not known, but is independently estimated, the foregoing discussion is analogous to two-sample t-tests versus studentized gap tests, and comparable results are found with respect to the significance levels. The implication is that a test based on the null sampling distribution of the studentized gaps would be more powerful than the t-test for this situation.

Suppose one is convinced that he wants to use studentized gaps for testing ranked means for grouping, and suppose he also has the necessary tables available to him; how does he proceed? Two problems can be anticipated. First, with a sample before him, one begins testing the studentized gaps for significance. Suppose, at first, he simply tests all gaps without regard at any stage as to whether any significant gaps have already been found. Before long, it becomes evident that when the first significant gap is found, the null hypothesis is, in effect, abandoned; that is, by declaring a gap significant, one is saying, "This gap is too large for all of these sample means to have come from one distribution." Thus, he decides that testing will be done in light of this fact, the appropriate adjustments being made for the sample size at each stage. But now, he finds that different groupings are obtained, depending upon where he begins with the procedure; for example, beginning with $X_{(n)}-X_{(n-1)}$ and working down or beginning with $X_{(2)}-X_{(1)}$ and working up. This is unacceptable, and it is clear, then, that unambiguous grouping cannot be achieved unless a starting rule is adopted, say, "Always start with gap 1.", or "Always begin in the middle." Yet, such a solution is unsatisfactory in that the grouping which results depends on the starting rule used. What is needed is a non-arbitrary starting place and a non-arbitrary sequence of testing which will work for any sample size or configuration. The studentized maximum gap procedure to be discussed below, satisfies these requirements and also utilizes sample size information as individual studentized gap testing does. Let us examine how such a procedure might be formulated.

Let $(X_1, X_2, \ldots, X_n)$ be a random sample from a normal distribution with mean μ and variance σ^2, and let $G = \max_{1 \leq i \leq n-1} \{X_{(i+1)} - X_{(i)}\}$, where $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ are the order statistics of the sample. If s^2 is an unbiased estimate of σ^2 based on ν degrees of freedom, then by the studentized maximum gap, it will be meant the statistic SMG = G/s. The null sampling distribution of SMG will also involve the parameter ν, and SMG will be said to have ν degrees of freedom. Given a sample, in order to compute the observed SMG, one needs an independent estimate s^2 of σ^2, and to carry out a test of significance, he needs the null sampling distribution of SMG tabulated for various sample sizes and degrees of freedom. Suppose these tables are available, then the testing procedure is as follows:

(i) Order the sample and compute the observed SMG = G/s.

(ii) Compare SMG with the appropriate tabulated distribution.

(iii) If the observed SMG is declared not significant, stop testing. If, however, significance is declared, break the sample into two groups and repeat the procedure for each group formed. Subsequent tests are made, of course, with the smaller sample sizes. Continue in this manner until no significant gaps remain.

Consider the barley data (Chapter I, p. 13). The ranked means and gaps were:

	A		F		G		D		C		B		E
Means:	49.6		58.1		61.0		61.5		67.6		71.2		71.3
Gaps:		8.5		2.9		0.5		6.1		3.6		0.1	

The estimate of the standard deviation of a barley variety mean was 3.64 with 30 df. Thus, the observed SMG_7 is 8.5/3.64 = 2.34. This is to be compared with the tabulated SMG(7, 30). Assume SMG_7 is judged significant; then the conclusion is that variety A does not belong with the others. Now, SMG_6 is computed for the six varieties excluding variety A. This is 6.1/3.64 = 1.68 and is compared with SMG(6, 30). Suppose this gap is also judged to be real. We now have the groups A, FGD, CBE. The next step is to compute SMG_3 = 2.9/3.64 and SMG_3' = 3.6/3.64. Those values are 0.80 and 0.99 respectively. Assume they are judged not significant; then the conclusion is that there are three groups: A, FGD, and CBE.

Let us turn now to a comparison of the LSD and the studentized maximum gap. To avoid confusion and to emphasize the fact that the LSD is being applied only to adjacent differences, let us call it the GLSD procedure (Gap LSD). It will be implemented in exactly the same manner as the SMG procedure, except that the LSD criterion will be used in place of the SMG criterion. The α-level studentized maximum gap analogue of the LSD is LSG = $s \times SMG_{\alpha}(n, \nu)$, where $SMG_{\alpha}(n, \nu)$ is the α-level critical value of the studentized maximum gap for sample size n and degrees of freedom ν. Let $\alpha = .05$ and $\nu = 9$, then,

$$LSD_{.05} = 3.20s \qquad LSG_{.05}(2,9) = 3.20s$$

$$LSG_{.05}(3,9) = 2.96s$$

$$LSG_{.05}(4,9) = 2.74s$$

$$LSG_{.05}(5,9) = 2.56s$$

$$LSG_{.05}(6,9) = 2.43s$$

$$LSG_{.05}(7,9) = 2.33s$$

It is observed that the .05 level critical values for the studentized maximum gap are uniformly less than or equal to the .05 level critical value for the LSD, and that the .05 level critical values for the studentized maximum gap decrease monotonically with increasing n. This pattern apparently holds for most α-levels and all degrees of freedom ν greater than 2. Hence, the studentized maximum gap test is generally more powerful than the GLSD for all cases except when there are only two means, in which case the two tests are identical. In other words, if any gap is declared significant by the GLSD, then it will also be declared significant by the studentized maximum gap test.

There is a trade-off, however, and it may be the following: Consider the case of two sets of observations from normal populations with different means, say $X_{1i} \sim N(\mu_1, \sigma^2)$ and $X_{2j} \sim N(\mu_2, \sigma^2)$ with $\mu_2 > \mu_1$. There is a positive probability, depending upon $\mu_2 - \mu_1$ and σ^2, that in a ranked sample for some i and j the event $X_{1i} > X_{2j}$ occurs. Whenever this happens, the studentized maximum gap test is more likely than the GLSD to declare $X_{1i} - X_{2j}$ significant. Let us call this type of error -- that of declaring a wrong way significance -- a Type III error. In Chapter IV, the probability of a Type III error will be investigated.

The phenomenon of decreasing critical values with increasing n was noted by Tukey (20), and a little thought reveals that it is intuitively logical and proper that they should. On the other hand, consider an experimenter applying the studentized maximum gap procedure using the LSG technique. At the first stage, he computes the observed maximum gap for the whole sample and compares it with LSG (n, ν). Suppose he declares it significant, then for each of the groups formed, he finds the respective maximum gaps and compares them with LSG (m, ν) and LSG (m', ν), where $m + m' = n$. An apparent logical inconsistency is now discovered, for the experimenter finds that in order for the maximum gaps in the subgroups to be declared significant at the same α-level, they must be larger than the largest gap of the whole sample. The first reaction is that there must be some mistake, but there is no mistake; this phenomenon is the very thing that gives the studentized maximum gap test greater power than the LSD. That is, under the null hypothesis that all observations are from the same normal distribution, adjacent differences become smaller on the average as the sample size becomes larger. This property is not shared by the sample range statistic $X_{max} - X_{min}$ for which, as the sample size increases the expected range also increases. Thus, as the sample size increases, although the observations spread out farther, at the same time, they get closer together. The maximum gap also follows this pattern, decreasing, on the average, with increasing sample size. Thus, one is faced with a two-sided coin -- the property of the maximum gap which allows the superior test to be devised but at the same time causes discomfort when the test procedure is first encountered.

Studying the null sampling distributions of the statistic G/s, where s^2 is an unbiased estimate of σ^2 based on ν degrees of freedom, reveals that, here also the α-level critical values exhibit the property of decreasing with increasing sample size n. It would be desirable but not necessary, to find a function of G/s whose critical values had the reverse trend.

Let $G' = G/\sigma$, then G' can be considered to be the maximum gap of a sample from the standard normal distribution. Denote Var (G') by κ_n^2, then, clearly, Var (G) $= \sigma^2\kappa_n^2$, and if s^2 is an estimate of σ^2, then $s^2\kappa_n^2$ is an estimate of Var (G). Preliminary examination of the null sampling distributions of $G/s\kappa_n$ indicates that the critical values for this statistic may exhibit the desired trend, but it cannot be definitely established at this time, due to the fact that precise estimates of the κ_n are required. Estimates of the κ_n are given in Table XXIV of the Appendix, and it is observed that the estimates are in the range 0.4 - 0.85 making the critical values of $G/s\kappa_n$ sensitive to errors in the $\hat{\kappa}_n$. It should be noted, of course, that comparing $G/s\kappa_n$ with its null sampling distribution is exactly equivalent to comparing G/s with its null sampling distribution, since, for any given sample size n, the two differ only by the constant κ_n. Thus, the use of the former would not change in any way the troublesome property of the maximum gap; it would merely obscure it. Notwithstanding, if it can be definitely established that the κ_n's in the denominator are sufficient to reverse the trend in critical values, it may be desirable to adopt the use of the statistic $G/s\kappa_n$ with the objective of making the procedure more esthetically acceptable to those who may use it.

In summary, the studentized maximum gap test procedure shows promise over any existing procedures for the purpose of detecting the pattern in a set of observations. It has several drawbacks and limitations, some of which have been discussed above. Some of the more apparent of these are:

(i) It must be assumed that the observations are normally distributed, all with the same variance.

(ii) If the procedure is applied to means, they must all be based on the same number of observations.

(iii) The frequency of Type III errors may be increased.

(iv) The maximum gap and studentized maximum gap have the property of decreasing average size and variability with increasing sample size. When applying the procedure, the consequences of this may go contrary to intuition, and additional explanation would thereby sometimes be required.

(v) It may be that after the testing has terminated, there still remain groups which would be judged significantly variable by the F-test or the studentized range test.

(vi) The procedure requires tables which do not exist at this time.

(vii) The exact null sampling distribution is difficult to calculate, so that, at present, this distribution must be approximated for all sample sizes other than sample sizes 3 and 4.

With respect to (i) and (ii), the restrictions mentioned are also shared by the majority of existing parametric procedures; the restrictions are not uniquely a property of the studentized maximum gap procedure. Item (iii) represents a compromise which may or may not be of great consequence, depending upon how serious Type III errors may be. With respect to (iv), it may be possible to find a simple and straightforward transformation of G/s which reverses the trend of the critical values; however, if not, it should be emphasized that the usefulness of the procedure is not affected either way. With respect to (v), the pertinent question is, "Can this indeed occur, and, if so, how often?" In Chapter IV the question will be further investigated. Finally, if the studentized maximum gap procedure proves as useful as it first appears, better and more efficient ways of calculating and tabulating the necessary tables will surely be devised.

CHAPTER III

THE DISTRIBUTION OF THE MAXIMUM GAP

1. Exact Distribution

Let $(X_1, X_2, \ldots, X_n)$ be random variables generated by sampling from a normal (μ, σ^2) distribution, and let $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ be the order statistics generated. Then, the joint probability density of $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is given by

$$g(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) = \frac{n!}{(2\pi)^{n/2}\,\sigma^n} \exp\left(-1/2\sigma^2\right) \sum_{i=1}^{n} (x_{(i)} - \mu)^2$$

$$\text{for } -\infty < x_{(1)} < x_{(2)} < \ldots < x_{(n)} < \infty$$

$$= 0, \text{ otherwise.} \qquad (1)$$

Let $w = x_{(1)}$ and let $g_i = x_{(i+1)} - x_{(i)}$, $i = 1, 2, \ldots, n-1$; then this defines a one-to-one transformation with Jacobian 1. Thus, the joint probability density of $(W, G_1, G_2, \ldots, G_{n-1})$ is

$$h(w, g_1, g_2, \ldots, g_{n-1}) = \frac{n!}{(2\pi)^{n/2}\,\sigma^n} \exp\left\{(-1/2\sigma^2)\left[(w - \mu)^2\right.\right.$$

$$\left.+ (w + g_1 - \mu)^2 + \ldots\right.$$

$$+ (w + g_1 + \dots + g_{n-1} - \mu)^2 \Big] \Big\}$$

for $-\infty < w < \infty$ and $0 < g_i < \infty$, $i = 1, 2, \dots, n-1$

$= 0$, otherwise. (2)

Define $s_0 = 0$, $s_i = \sum_{j=1}^{i} g_j$, $i = 1, 2, \dots, n - 1$, and define $z = w-\mu$, then the joint probability density of $(Z, G_1, G_2, \dots, G_{n-1})$ can be expressed as

$$k(z, g_1, g_2, \dots, g_{n-1}) = \frac{n!}{(2\pi)^{n/2} \sigma^n} \exp\left[(-1/2\sigma^2) \sum_{i=0}^{n-1} (z + s_i)^2\right]$$

for $-\infty < z < \infty$ and $0 < g_i < \infty$

$= 0$, otherwise.

Consider the expression $\sum_{i=0}^{n-1}(z + s_i)^2$ in the exponent. We can express it as

$$\sum_{i=0}^{n-1}(z + s_i)^2 = \sum_{i=0}^{n-1}(z^2 + 2zs_i + s_i^2)$$

$$= nz^2 + 2z \sum_{i=0}^{n-1} s_i + \sum_{i=0}^{n-1} s_i^2$$

$$= n\left[z^2 + 2z\sum_{i=0}^{n-1} s_i/n + (1/n)\left(\sum_{i=0}^{n-1} s_i\right)^2\right]$$

$$+ \left[\sum_{i=0}^{n-1} s_i^{\,2} - (1/n)\left(\sum_{i=0}^{n-1} s_i\right)^2\right]$$

$$= n\left[z + (1/n)\sum_{i=0}^{n-1} s_i\right]^2$$

$$+ \left[\sum_{i=0}^{n-1} s_i^{\,2} - (1/n)\left(\sum_{i=0}^{n-1} s_i\right)^2\right] ,$$

so that,

$$k(z, g_1, g_2, \ldots , g_n) = \frac{n!}{\sqrt{n}(2\pi)^{(n-1)/2}\sigma^{n-1}}$$

$$\times \exp(-1/2\sigma^2)\left[\sum_{i=0}^{n-1} s_i^{\,2} - (1/n)\left(\sum_{i=0}^{n-1} s_i\right)^2\right]$$

$$\times \exp(-n/2\sigma^2)\left[z + (1/n)\left(\sum_{i=0}^{n-1} s_i\right)\right]^2 .$$

The joint density of $(G_1, G_2, \ldots , G_{n-1})$ is then obtained by integrating out z. Thus we have,

$$f(g_1, g_2, \ldots, g_{n-1}) = \frac{n!}{\sqrt{n}(2\pi)^{(n-1)/2}\sigma^{n-1}}$$

$$\times \exp(-1/2\sigma^2)\left[\sum_{i=1}^{n-1} s_i^{\,2} - (1/n)\sum_{i=1}^{n-1} s_i\right]^2 . \quad (3)$$

But,

$$\sum_{i=1}^{n-1} s_i^{\,2} = \sum_{i=1}^{n-1}\left(\sum_{j=1}^{i} g_i^{\,2} + 2\sum_{j=1}^{i}\sum_{j'>j} g_j g_{j'}\right)$$

$$= (n-1)g_1^{\,2} + (n-2)g_2^{\,2} + \ldots + g_{n-1}^{\,2}$$

$$+ 2(n-2)g_1g_2 + 2(n-3)(g_1g_3 + g_2g_3)$$

$$+ 2(n-4)(g_1g_4 + g_2g_4 + g_3g_4) + \ldots$$

$$+ 2(g_1g_{n-1} + g_2g_{n-1} + \ldots g_{n-2}g_{n-1}),$$

and

$$(1/n)\left(\sum_{i=1}^{n-1} s_i\right)^2 = (1/n)\left[\sum_{i=1}^{n-1}(n-i)g_i\right]^2$$

$$= (1/n)\left[\sum_{i=1}^{n-1}(n-i)^2 g_i^{\,2} + 2\sum_{j=1}^{n-1}\sum_{j'>j}(n-j)(n-j')g_jg_{j'}\right]$$

$$= \frac{(n-1)^2}{n} g_1{}^2 + \frac{(n-2)^2}{n} g_2{}^2 + \ldots + \frac{1}{n} g_{n-1}{}^2$$

$$+ 2\left[\frac{(n-2)(n-1)}{n} g_1 g_2 + \frac{(n-3)}{n}((n-1)g_1g_3+(n-2)g_2g_3)\right.$$

$$+ \ldots (1/n)\Big((n-1)g_1g_{n-1}+(n-2)g_2g_{n-1}$$

$$\left. + \ldots + 2g_{n-2}g_{n-1}\Big)\right] ,$$

hence,

$$\sum_{i=1}^{n-1} s_i{}^2-(1/n)\left(\sum_{i=1}^{n-1} s_i\right)^2$$

$$= \frac{n-1}{n} g_1{}^2 + \frac{2(n-2)}{n} g_2{}^2 + \ldots + \frac{n-1}{n} g_{n-1}{}^2 + 2\left[\frac{(n-2)}{n} g_1g_2\right.$$

$$+ \frac{n-3}{n}(g_1g_3 + 2g_2g_3) + \ldots + \frac{1}{n}(g_1g_{n-1} + 2g_2g_{n-1}$$

$$\left. + \ldots + (n-2)g_{n-2}g_{n-1})\right]$$

$$= (1/n)\left[\sum_{i=1}^{n-1} i(n-i)g_i{}^2 + 2\sum_{j=1}^{n-1}\sum_{j'>j} j(n-j')g_jg_{j'}\right] .$$

The joint density of $(G_1, G_2, \ldots G_{n-1})$ is, therefore,

$$f(g_1, g_2, \ldots, g_{n-1}) = \frac{n!}{\sqrt{n}(2\pi)^{(n-1)/2}\,\sigma^{n-1}}$$

$$\times \exp(-1/2n\sigma^2)\left[\sum_{i=1}^{n-1} i(n-i)g_i{}^2\right.$$

$$\left. +2\sum_{j=1}^{n-1}\sum_{j'>j} j(n-j')g_jg_{j'}\right]$$

$$\text{for} \quad 0 < g_i < \infty, \; i = 1, 2, \ldots, n-1$$

$$= 0 \text{ , otherwise.} \qquad (4)$$

For $n = 3$ and $\sigma^2 = 1$, the density (4) reduces to

$$f(g_1, g_2) = \frac{3!}{\sqrt{3}(2\pi)} \exp(-1/6)(2g_1^2 + 2g_2^2 + 2g_1g_2)$$

$$= (\sqrt{3}/\pi)\exp(-1/3)(g_1^2 + g_1g_2 + g_2^2),$$

$$g_1 > 0, \; g_2 > 0 \; . \qquad (5)$$

Let $G = \max \{G_i\}$, then, for $n = 3$

$$H(g) = \Pr(G < g) = \Pr(G_1 < g, G_2 < g)$$

$$= (\sqrt{3}/\pi)\int_0^g\int_0^g \exp(-1/3)(g_1^2 + g_1g_2 + g_2^2)dg_1dg_2$$

$$= (\sqrt{3}/\pi)\int_0^g \exp(-g_2^2/4)(1/\sqrt{2\pi})\int_{g_2/\sqrt{6}}^{\sqrt{2}g/\sqrt{3} + g_2/\sqrt{6}} \exp(-t^2/2)dtdg_2$$

$$= (3\sqrt{6}/\pi)\int_0^{g/\sqrt{6}} \exp(-3u^2/2)[F(\sqrt{2}g/\sqrt{3} + u) - F(u)] \, du, \qquad (6)$$

where $F(z) = (1/\sqrt{2\pi})\int_{-\infty}^{z} \exp(-t^2/2)dt$. The density of G can be obtained by differentiating H(g) with respect to g.

$$h(g) = H'(g) = (3\sqrt{6}/\sqrt{\pi})(1/\sqrt{2\pi})\int_{0}^{g/\sqrt{6}} \exp(-3u^2/2)$$

$$\times (1/\sqrt{2\pi})(\sqrt{2}/\sqrt{3})\exp(-1/2)(\sqrt{2}g/\sqrt{3} + u)^2 du$$

$$+ (3\sqrt{6}/\sqrt{\pi})\exp(-g^2/4)[F(\sqrt{2}g/\sqrt{3} + g/\sqrt{6})]\cdot(1/\sqrt{6}) \qquad (7)$$

The differentiation step follows from Leibniz's Rule,

$$\frac{d}{dy}\int_{f(y)}^{g(y)} h(x,y)dx = \int_{f(y)}^{g(y)} \frac{\partial h(x,y)}{\partial y} dx + h[g(y), y]g'(y) - h[f(y), y]f'(y) .$$

Simplifying (7),

$$h(g) = (6/\sqrt{\pi})(1/\sqrt{2\pi})\int_{0}^{g/\sqrt{6}} \exp(-g^2/4)\exp(-1/2)(2u + g/\sqrt{6})^2 du$$

$$+ (3/\sqrt{\pi})\exp(-g^2/4)[F(\sqrt{3}g/\sqrt{2}) - F(g/\sqrt{6})]$$

$$= (3/\sqrt{\pi})\exp(-g^2/4)\left\{(1/\sqrt{2\pi})\int_{g/\sqrt{6}}^{\sqrt{3}g/\sqrt{2}} \exp(-t^2/2)dt + [F(\sqrt{3}g/\sqrt{2}) - F(g/\sqrt{6})]\right\}$$

$$= (6/\sqrt{\pi})\exp(-g^2/4)[F(\sqrt{3}g/\sqrt{2}) - F(g/\sqrt{6})], 0 < g < \infty \qquad (8)$$

Similarly, for $n = 4$ and $\sigma^2 = 1$,

$$H(g) = (3\sqrt{2}/\pi^{3/2})\int_0^g\int_0^g\int_0^g \exp(-1/8)(3g_1^2 + 4g_2^2 + 3g_3^2$$

$$+ 4g_1g_2 + 2g_1g_3 + 4g_2g_3)dg_3dg_2dg_1$$

$$= (4\sqrt{3}/\pi)\int_0^g\int_0^g \exp(-1/3)(g_1^2 + g_1g_2 + g_2^2)$$

$$\times \left\{F\left[(\sqrt{3}/2)\left(g + \frac{g_1 + 2g_2}{3}\right)\right] - F\left[(\sqrt{3}/2)\left(\frac{g_1 + 2g_2}{3}\right)\right]\right\} dg_2dg_1,$$

and,

$$h(g) = H'(g) = (4\sqrt{3}/\pi)(\sqrt{3}/2)\int_0^g \exp(-1/4)(g_1^2 + g^2)$$

$$\times \left[(1/\sqrt{2\pi})\int_0^g \exp(-1/2)\left(g_2 + \frac{g_1 + g}{2}\right)^2 dg_2\right] dg_1$$

$$+ (4\sqrt{3}/\pi)\int_0^g \exp(-1/3)(g_1^2 + g_1g + g^2)$$

$$\times \left\{F[(\sqrt{3}/2)(5g/3 + g/3)] - F[(\sqrt{3}/2)(2g/3 + g_1/3)]\right\} dg_1$$

$$+ (4\sqrt{3}/\pi)\int_0^g \exp(-1/3)(g^2 + gg_2 + g_2^{\,2})$$

$$\times\left\{F[(\sqrt{3}/2)(4g/3 + 2g_2/3)] - F[(\sqrt{3}/2)(g/3 + 2g_2/3)]\right\} dg_2$$

$$= (4\sqrt{3}/\pi)\int_0^g \Bigg[(\sqrt{3}/2)\exp(-1/4)(u^2 + g^2)$$

$$\times\left\{F[(1/2)(3g + u)] - F[(1/2)(g + u)]\right\}$$

$$+ \exp(-1/3)(u^2 + ug + g^2)$$

$$\times\Big\{F[(1/2\sqrt{3})(5g + u)] - F[(1/2\sqrt{3})(2g + u)]$$

$$+ F[(1/\sqrt{3})(2g + u)] - F[(1/\sqrt{3})(1/2g + u)]\Big\}\Bigg] du.$$

For sample sizes larger than 4, the method becomes very messy and unwieldy, and alternative approaches will be taken.

2. Utilization of the Distribution of the Individual Gaps

For the random gaps

$$(G_1, G_2, \ldots, G_{n-1}),$$

we have, of course, the identity

$$\Pr(G_{max} < g) = \Pr(G_1 < g, G_2 < g, \ldots, G_{n-1} < g).$$

If the G_i were independent, we would also have

$$Pr(G_{max} < g) = \prod_{i=1}^{n-1} Pr(G_i < g) \ . \tag{9}$$

We know that the G_i are not strictly independent, but how "nearly independent" are they? In other words, if one used (9), how closely would he approximate the distribution of the maximum gap? Let us examine the correlations between gaps for various sample sizes as a measure of the interdependence relationships. For sample size n, denote $Var(G_i)$ by θ_i^2 and $Cov(G_i, G_j)$ by θ_{ij}. Here, and elsewhere, there is no loss of generality to assume the G_i arise from sampling the standard normal distribution, because, if $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ are from a normal (μ, σ^2), then,

$$E(G_i) = E(X_{(i+1)} - X_{(i)}) = E[\sigma(\xi_{(i+1)} - \xi_{(i)})] = \sigma\, E(\xi_{(i+1)} - \xi_{(i)}) \ ,$$

where $\xi_{(j)}$ is the jth order statistic from a standard normal, and

$$Var(G_i) = \sigma^2\, Var(\xi_{(i+1)} - \xi_{(i)}) \ .$$

Assuming G_i is from the standard normal distribution, it is a simple and straightforward matter to obtain

$$\theta_i^2 = Var(G_i) = Var(\xi_{(i+1)}) + Var(\xi_{(i)}) - 2Cov(\xi_{(i+1)}, \xi_{(i)})$$

$$= \delta_{i+1}^2 + \delta_i^2 - 2\delta_{ij}, \text{ say} \tag{10}$$

and, similarly,

$$\theta_{ij} = \mathrm{Cov}(G_i, G_j) = \delta_{i+1,j+1} - \delta_{i+1,j} - \delta_{i,j+1} + \delta_{ij} \quad . \qquad (11)$$

Thus, $\rho_{ij} = \theta_{ij}/\theta_i\theta_j$ is easily determined from existing tables where the variances and covariances of the standard normal order statistics are tabulated. Table XX of the Appendix shows the correlation structure of the gaps for sample sizes 3 to 20. Let us consider the correlation matrices of the gaps for several selected sample sizes. (Only the first $[(n-1)/2]$ gaps need be considered, since the correlation structure is double symmetric. That is, $\rho_{ij} = \rho_{ji} = \rho_{i'j'} = \rho_{j'i'}$, where $i' = n-i-1$ and $j' = n-j-1$.)

It is observed that: (i) there are no correlations with absolute value greater than 0.136; (ii) The correlations between adjacent gaps are greatest; (iii) $\rho_{12} = \rho_{n,n-1}$ is the greatest correlation; and (iv) all correlations are reduced as the sample size increases. Thus, it appears that the gaps are "nearly independent" for the larger sample sizes, and that a product of individual gap probabilities may furnish an adequate approximation to the distribution of the maximum gap. We should note that justification for multiplying probabilities of non-independent events based on a correlation argument is not altogether convincing, and the final judgment of whether this should be done must rest upon how good we deem the approximation to be. The goodness of the approximation would be difficult to quantify since non-independence has no degree and, thus, "almost independence" is described here only in terms of "small" correlations. Rather than attempting to assess the goodness of approximation analytically, the sample size 7 will be arbitrarily taken as the dividing point, and a subjective evaluation of the goodness of approximation will be made for that case.

TABLE II

CORRELATION MATRICES OF GAPS FOR SOME SELECTED SAMPLE SIZES

(a) Sample Size 3

	1
1	1.000
2	-0.136

(b) Sample Size 5.

	1	2
1	1.000	
2	-0.113	1.000
3	-0.068	-0.110
4	-0.041	

(c) Sample Size 7.

	1	2	3
1	1.000		
2	-0.100	1.000	
3	-0.064	-0.093	1.000
4	-0.043	-0.064	-0.090
5	-0.029	-0.045	
6	-0.019		

(d) Sample Size 9.

	1	2	3	4
1	1.000			
2	-0.092	1.000		
3	-0.060	-0.083	1.000	
4	-0.042	-0.059	-0.078	1.000
5	-0.031	-0.043	-0.058	-0.076
6	-0.023	-0.032	-0.044	
7	-0.017	-0.024		
8	-0.011			

3. The Distribution of an Individual Gap

Let $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ be the order statistics generated by sampling from the standard normal distribution. Then, the joint density for $(X_{(i)}, X_{(i+1)})$ is given by

$$f(x_{(i)}, x_{(i+1)}) = \frac{n!}{(i-1)!(n-i-1)!} (1/2\pi)[F(x_{(i)})]^{i-1}$$

$$\times [1-F(x_{(i+1)})]^{n-i-1} \exp(-1/2)(x_{(i)}^2 + x_{(i+1)}^2)$$

$$\text{for } -\infty < x_{(i)} < x_{(i+1)} < \infty$$

$$= 0, \text{ otherwise.} \tag{12}$$

Let $w = x_{(i)}$, and $g_i = x_{(i+1)} - x_{(i)}$; this defines a one-to-one transformation with Jacobian 1. Thus, the joint density of W and G_i is

$$h(w,g_i) = \frac{n!}{(i-1)!(n-i-1)!} (1/2\pi)[F(w)]^{i-1}[1-F(w+g_i)]^{n-i-1}$$

$$\times \exp(-1/2)(w^2 + (w + g_i)^2)$$

$$\text{for } -\infty < w < \infty \quad , \quad 0 < g_i < \infty$$

$$= 0, \text{ otherwise}$$

and the marginal density of G_i is

$$f(g_i) = \frac{K(n,i)}{2\pi} \int_{-\infty}^{\infty} [F(w)]^{i-1} [1-F(w+g_i)]^{n-i-1} \exp(-1/2)(2w^2+2wg_i+g_i^2)dw \tag{13}$$

The integral (13) does not exist in closed form and must be evaluated by a suitable numerical quadrature procedure.

4. Studentization

In (9) and an earlier paper, H. O. Hartley showed how studentization may be accomplished. Let $(X_1, X_2, \ldots, X_n)$ be a random sample from a normal distribution with standard deviation σ, and let W be a statistic "proportional to σ" in the sense that if the X_i are measured in σ-units, W is transformed to W/σ. Let $P(w)$ be the cumulative probability integral for W for the case $\sigma = 1$, that is, $P(w) = \Pr(W < w)$. Let s^2 be an independent estimate of σ^2 based on ν degrees of freedom such that $\nu s^2/\sigma^2$ is distributed as $X^2(\nu)$, and let $R = W/s$. Then the cumulative probability function of R is given by

$$F(r) = \int_{-\infty}^{\infty} P(rs) \frac{s\nu^{-1}\nu^{\nu/2}}{\Gamma(\nu/2)\ 2^{\nu/2-1}} \exp(-s^2\nu/2)ds \quad . \tag{14}$$

In the case of obtaining the probability integral of the studentized maximum gap, this result is easily derived. Let $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ be the order statistics of a random sample of size n from a normal $(0, \sigma^2)$ distribution, and let $G = \max_i\{X_{(i+1)} - X_{(i)}\}$. Let s^2 be an unbiased estimate of σ^2 based on ν degrees of freedom, independent of $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ such that $\nu s^2/\sigma^2$ is distributed as $X^2(\nu)$, and form the studentized maximum gap G/s. Then,

$$\Pr(G/s < g) = \Pr(G < gs) = \Pr(G' < gs/\sigma)$$

where $G' = G/\sigma$ is the maximum gap in a sample of size n from the standard normal distribution. This probability can be evaluated by

$$Pr(G' < gs/\sigma) = \int_0^\infty Pr(G' < gs/\sigma | s) f(s) ds,$$

where $f(s)$ is the probability density of s,

$$= \int_0^\infty P(gs/\sigma) \frac{s^{\nu-1} \nu^{\nu/2}}{\Gamma(\nu/2)\; 2^{\nu/2-1} \sigma^{\nu}} \exp(-s^2\nu/2) ds$$

where $P(\cdot)$ is the probability integral of G'. Thus, the cumulative probability integral of the studentized maximum gap is given by

$$H(g) = \int_0^\infty P(gs) \frac{s^{\nu-1} \nu^{\nu/2}}{\Gamma(\nu/2)\; 2^{\nu/2-1}} \exp(-s^2\nu/2) ds \quad . \tag{15}$$

It is evident that if it were possible to determine P, then the distribution of the studentized maximum gap would be completely determined. The problem, as we have seen, is in the determination of P. The solution proposed is to calculate approximations to P for the several sample sizes, and to then apply the exact studentization formula (15) to these approximations for the various degrees of freedom for which tabulated critical values are desired.

In the same reference, Hartley also gives a reduction formula and an approximation formula, but unfortunately, neither of the two formulas lend themselves to straightforward calculation by computer whenever P is not an analytic function expressible in closed form.

5. Approximate Distribution Functions for the Maximum Gap

The approach taken to obtain the desired approximate distributions can be described as an unsophisticated frontal attack. All computer calculations were done on an IBM 360-65 computer with FORTRAN double-precision programs written by the author, with the standard FORTRAN Library and IBM SSP routines, and with two routines written by Dr. J. P. Chandler of Oklahoma State University's Department of Computing and Information Sciences. In the programs written by the author, a great deal of care was not given either to programming efficiency or computational efficiency. All numerical integrations were done with Simpson's rule integration and trapezoidal rule integration by means of the DQSF and DQTFE subroutines, respectively, in the IBM SSP Library. For the empirical distributions, a large number (25000 - 35000) of simulated normal samples of the appropriate size were taken, the maximum gap computed, and a frequency distribution generated for the interval 0.0 to 5.0 using subintervals of length 0.1. The adequacy of the generating procedure will be examined for sample sizes 3 and 4, where the exact distributions of the maximum gap can be derived.

The basic plan of attack to obtain P, the distribution of the maximum gap in standard normal samples will be:

(i) For sample sizes 3 and 4, calculate the exact distributions.

(ii) For sample sizes 5 through 7, generate empirical distributions.

(iii) For sample sizes larger than 7, calculate an approximation to the distribution by multiplying individual gap probabilities together. These are, of course, obtained by evaluation of (13) by numerical quadrature.

The agreement between the empirical distribution and the exact distribution will be examined for sample sizes 3 and 4, and the agreement between the empirical distribution and the product-approximated distribution will be examined for sample size 7.

In order to obtain simulated normal variates, a routine named GAUSF, written by J. P. Chandler was used. The routine is based on a generation algorithm proposed by Marsaglia and Bray (12), which utilizes mixing, in a specified way, functions of uniform variates. The uniform variables were obtained with a pseudo-random number generator called RANF, also written by Chandler. As a rough indication of CPU time requirements, it was possible to obtain samples and tabulate the frequency distributions of all gaps and the maximum gap for 25000 samples of size 3, in 1 minute, 10.864 seconds of CPU time, and to do the same for 25000 samples of size 6, in 2 minutes, 11.522 seconds of CPU time.

The empirical distributions of the maximum gap for sample sizes 5 through 7 are given in Tables XXI through XXIII of the Appendix. Table III and Figure 5 on the following pages give a comparison of the empirical distribution and the exact theoretical distribution for sample size 3. On the basis of the Chi-Square goodness of fit statistic calculated, the agreement between the two is judged to be adequate. The calculated Chi-Square 52.984 has a significance level of approximately 0.37, and is, therefore, not inconsistent with the hypothesis of no difference between the distributions.

TABLE III

EMPIRICAL AND THEORETICAL DISTRIBUTIONS OF THE MAXIMUM GAP, SAMPLE SIZE 3

INTERVAL	OBSERVED FREQUENCY	THEORETICAL FREQUENCY	CHI-SQUARE	OBSERVED CUMULATIVE FREQUENCY	THEORETICAL CUMULATIVE FREQUENCY
0.0 TO 0.1	131	137.40	0.298	0.005240	0.005496
0.1 TO 0.2	409	407.25	0.008	0.021600	0.021786
0.2 TO 0.3	691	662.40	1.235	0.049240	0.048282
0.3 TO 0.4	940	894.20	2.346	0.086840	0.084050
0.4 TO 0.5	1061	1095.42	1.082	0.129280	0.127867
0.5 TO 0.6	1292	1260.78	0.773	0.180960	0.178298
0.6 TO 0.7	1368	1387.02	0.261	0.235680	0.233779
0.7 TO 0.8	1441	1473.05	0.697	0.293320	0.292701
0.8 TO 0.9	1492	1519.88	0.511	0.353000	0.353496
0.9 TO 1.0	1576	1530.03	1.381	0.416040	0.414697
1.0 TO 1.1	1459	1507.62	1.568	0.474400	0.475002
1.1 TO 1.2	1454	1457.52	0.009	0.532560	0.533303
1.2 TO 1.3	1386	1385.20	0.000	0.588000	0.588711
1.3 TO 1.4	1306	1296.17	0.074	0.640240	0.640558
1.4 TO 1.5	1166	1195.77	0.741	0.686880	0.688389
1.5 TO 1.6	1058	1088.90	0.877	0.729200	0.731945
1.6 TO 1.7	990	979.68	0.109	0.768800	0.771132
1.7 TO 1.8	884	871.62	0.176	0.804160	0.805997
1.8 TO 1.9	722	767.53	2.700	0.833040	0.836698
1.9 TO 2.0	706	669.27	2.015	0.861280	0.863469
2.0 TO 2.1	594	578.35	0.423	0.885040	0.886603
2.1 TO 2.2	481	495.47	0.423	0.904280	0.906422
2.2 TO 2.3	431	421.05	0.235	0.921520	0.923264
2.3 TO 2.4	373	354.98	0.915	0.936440	0.937463
2.4 TO 2.5	284	297.10	0.578	0.947800	0.949347
2.5 TO 2.6	290	246.83	7.552	0.959400	0.959220
2.6 TO 2.7	199	203.65	0.106	0.967360	0.967366
2.7 TO 2.8	166	166.85	0.004	0.974000	0.974040
2.8 TO 2.9	116	135.80	2.887	0.978640	0.979472
2.9 TO 3.0	130	109.80	3.717	0.983840	0.983864
3.0 TO 3.1	98	88.20	1.089	0.987760	0.987392
3.1 TO 3.2	62	70.38	0.997	0.990240	0.990207
3.2 TO 3.3	64	55.78	1.213	0.992800	0.992439
3.3 TO 3.4	58	43.95	4.491	0.995120	0.994196
3.4 TO 3.5	29	34.40	0.847	0.996280	0.995572
3.5 TO 3.6	29	26.75	0.189	0.997440	0.996642
3.6 TO 3.7	18	20.67	0.346	0.998160	0.997469
3.7 TO 3.8	11	15.85	1.484	0.998600	0.998103
3.8 TO 3.9	7	12.10	2.149	0.998880	0.998587
3.9 TO 4.0	9	9.18	0.003	0.999240	0.998954
4.0 TO 4.1	3	6.90	2.204	0.999360	0.999230
4.1 TO 4.2	2	5.18	1.948	0.999440	0.999437
4.2 TO 4.3	4	3.85	0.006	0.999600	0.999591
4.3 TO 4.4	2	2.82	0.240	0.999680	0.999704
4.4 TO 4.5	3	2.10	0.385	0.999800	0.999786
4.5 TO 4.6	2	1.50	0.167	0.999880	0.999848
4.6 TO 4.7	1	1.10	0.009	0.999920	0.999892
4.7 TO 4.8	0	0.80	0.799	0.999920	0.999924
4.8 TO 4.9	1	0.58	0.314	0.999960	0.999947
4.9 TO 5.0	0	0.40	0.401	0.999960	0.999963

TOTAL: 25000 CHI-SQUARE: 52.984 WITH 50 DF

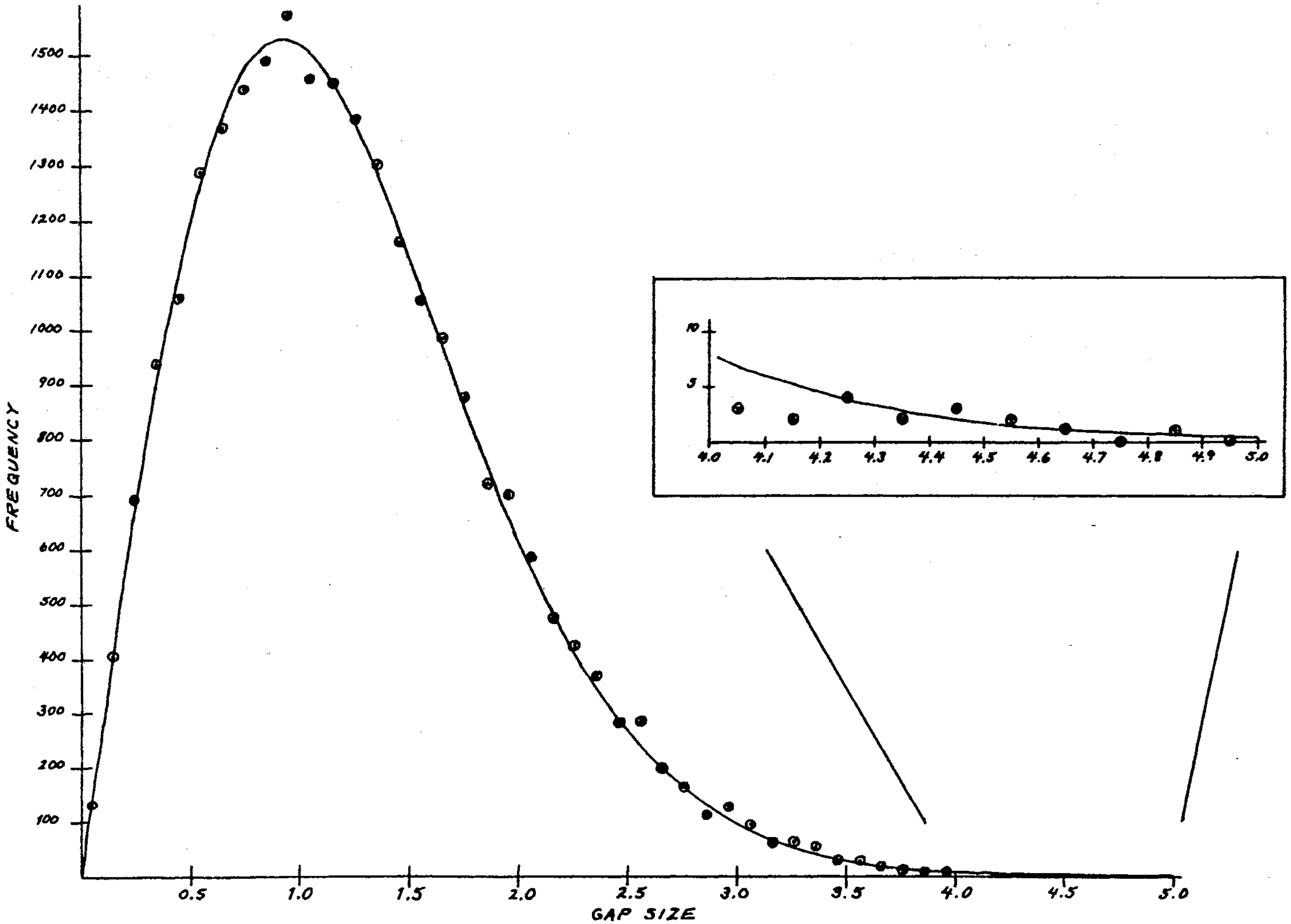


Figure 5. Comparison of Empirical and Theoretical Frequencies, Sample Size 3

TABLE IV

EMPIRICAL AND THEORETICAL DISTRIBUTIONS OF THE MAXIMUM GAP, SAMPLE SIZE 4

INTERVAL	OBSERVED FREQUENCY	THEORETICAL FREQUENCY	CHI-SQUARE	OBSERVED CUMULATIVE FREQUENCY	THEORETICAL CUMULATIVE FREQUENCY
0.0 TO 0.1	18	18.90	0.043	0.000720	0.000756
0.1 TO 0.2	122	129.12	0.393	0.005600	0.005921
0.2 TO 0.3	353	333.95	1.087	0.019720	0.019279
0.3 TO 0.4	592	605.37	0.295	0.043400	0.043494
0.4 TO 0.5	938	908.05	0.988	0.080920	0.079816
0.5 TO 0.6	1210	1205.70	0.015	0.129320	0.128044
0.6 TO 0.7	1427	1466.73	1.076	0.186400	0.186713
0.7 TO 0.8	1638	1668.13	0.544	0.251920	0.253438
0.8 TO 0.9	1740	1797.35	1.830	0.321520	0.325332
0.9 TO 1.0	1846	1851.97	0.019	0.395360	0.399411
1.0 TO 1.1	1864	1837.83	0.373	0.469920	0.472924
1.1 TO 1.2	1776	1766.27	0.054	0.540960	0.543575
1.2 TO 1.3	1679	1651.37	0.462	0.608120	0.609630
1.3 TO 1.4	1480	1507.72	0.510	0.667320	0.669939
1.4 TO 1.5	1327	1348.43	0.340	0.720400	0.723876
1.5 TO 1.6	1154	1184.47	0.784	0.766560	0.771255
1.6 TO 1.7	1049	1024.23	0.599	0.808520	0.812224
1.7 TO 1.8	886	873.43	0.181	0.843960	0.847161
1.8 TO 1.9	770	735.67	1.602	0.874760	0.876588
1.9 TO 2.0	615	612.85	0.008	0.899360	0.901102
2.0 TO 2.1	530	505.38	1.200	0.920560	0.921317
2.1 TO 2.2	402	412.95	0.290	0.936640	0.937835
2.2 TO 2.3	325	334.52	0.271	0.949640	0.951216
2.3 TO 2.4	272	268.85	0.037	0.960520	0.961970
2.4 TO 2.5	238	214.32	2.615	0.970040	0.970543
2.5 TO 2.6	173	169.65	0.066	0.976960	0.977329
2.6 TO 2.7	139	133.35	0.239	0.982520	0.982663
2.7 TO 2.8	117	104.02	1.618	0.987200	0.986824
2.8 TO 2.9	81	80.65	0.002	0.990440	0.990050
2.9 TO 3.0	56	62.05	0.590	0.992680	0.992532
3.0 TO 3.1	46	47.48	0.046	0.994520	0.994431
3.1 TO 3.2	34	36.03	0.114	0.995880	0.995872
3.2 TO 3.3	23	27.20	0.649	0.996800	0.996960
3.3 TO 3.4	26	20.37	1.553	0.997840	0.997775
3.4 TO 3.5	13	15.18	0.312	0.998360	0.998392
3.5 TO 3.6	11	11.23	0.005	0.998800	0.998831
3.6 TO 3.7	9	8.23	0.073	0.999160	0.999160
3.7 TO 3.8	2	6.02	2.688	0.999240	0.999401
3.8 TO 3.9	6	4.35	0.626	0.999480	0.999575
3.9 TO 4.0	4	3.15	0.229	0.999640	0.999701
4.0 TO 4.1	1	2.25	0.695	0.999680	0.999791
4.1 TO 4.2	0	1.60	1.600	0.999680	0.999855
4.2 TO 4.3	0	1.10	1.100	0.999680	0.999899
4.3 TO 4.4	2	0.80	1.799	0.999760	0.999931
4.4 TO 4.5	3	0.55	10.918	0.999880	0.999953
4.5 TO 4.6	2	0.38	7.028	0.999960	0.999968
4.6 TO 4.7	0	0.27	0.274	0.999960	0.999979
4.7 TO 4.8	0	0.18	0.176	0.999960	0.999986
4.8 TO 4.9	0	0.13	0.125	0.999960	0.999991
4.9 TO 5.0	0	0.07	0.075	0.999960	0.999994

TOTAL: 25000 CHI-SQUARE: 48.215 WITH 50 DF

Table IV gives a comparison of the empirical distribution and the theoretical distribution for sample size 4. Here, also, the agreement between the two is judged to be good.

Based on the results for sample sizes 3 and 4, the empirical distributions for sample sizes 5 through 7 will be taken as adequate approximations to the exact distributions; however, some smoothing may be used.

To obtain the distribution of the maximum gap for sample sizes larger than 7, approximation of the distribution will be made by obtaining exact distributions for the individual gaps in each case and forming a product of these. The exact probabilities for individual gaps are computed by the evaluation of (13) by numerical quadrature. The error of approximation comes from three sources:

(i) Computer round-off error

(ii) Integration error

(iii) Lack of independence of the gaps

Double precision computations help to minimize the effects of (i). With respect to (ii), there is some error introduced in setting up the function to be integrated: however, it is worthy of note that by using the IBM FORTRAN routine DERF to calculate values for the cumulative normal, one can obtain values which agree exactly with the fifteen-place values tabulated by Abramowitz (1). Also the IBM function DEXP is very accurate, so that errors in the integrand can probably be neglected. Error introduced by the integration routines is more difficult to assess, but for smooth functions, one can generally achieve sufficient accuracy with Simpson's Rule by taking small enough increments.

Smaller increments, of course, penalize one in terms of computing time, so that some experimentation is necessary to determine a reasonable increment value to take. For the integral (13) it was found that taking increments smaller than 0.05 or limits of integration outside of (-5.5, 5.5) changed the final answer only slightly (in the sixth decimal place), so that these values were taken as adequate working limits.

The FORTRAN programming necessary to perform the calculations is straightforward and will not be presented as part of this study. The product-approximated maximum gap probability distributions are given in Table XXV of the Appendix. A typical individual gap distribution is shown in Table V and Figure 6 for Gap 1 in a sample of size 7.

A comparison of the empirical distribution and the product-approximated distribution of the maximum gap for sample size 7 is shown in Table VI and Figure 7. The agreement between the two is good, especially in the tail areas. Therefore, if we assume that the empirical distribution for sample size 7 is as close to the exact distribution as it is for sample size 3, then the product-approximated distributions will be considered satisfactory for the majority of applications for sample sizes larger than 7. A point noted in passing is that the product-approximated distributions apparently have tail areas which are slightly smaller than the true tail areas, making any test based on them non-conservative with respect to Type I error rates.

TABLE V

DENSITY FUNCTION AND CUMULATIVE DISTRIBUTION FUNCTION OF GAP 1 IN A SAMPLE SIZE 7

GAP SIZE	PROBABILITY DENSITY	CUMULATIVE DISTRIBUTION
0.0	1.352178	0.0
0.1	1.230890	0.129139
0.2	1.112134	0.246263
0.3	0.997327	0.351697
0.4	0.887669	0.445900
0.5	0.784132	0.529436
0.6	0.687453	0.602955
0.7	0.598140	0.667172
0.8	0.516490	0.722839
0.9	0.442599	0.770729
1.0	0.376393	0.811615
1.1	0.317649	0.846256
1.2	0.266024	0.875382
1.3	0.221081	0.899684
1.4	0.182321	0.919805
1.5	0.149199	0.936336
1.6	0.121153	0.949814
1.7	0.097619	0.960717
1.8	0.078047	0.969469
1.9	0.061916	0.976441
2.0	0.048737	0.981951
2.1	0.038064	0.986272
2.2	0.029497	0.989634
2.3	0.022680	0.992229
2.4	0.017301	0.994218
2.5	0.013095	0.995729
2.6	0.009834	0.996868
2.7	0.007326	0.997721
2.8	0.005415	0.998353
2.9	0.003971	0.998819
3.0	0.002889	0.999160
3.1	0.002085	0.999406
3.2	0.001493	0.999584
3.3	0.001060	0.999710
3.4	0.000747	0.999800
3.5	0.000522	0.999862
3.6	0.000362	0.999906
3.7	0.000248	0.999936
3.8	0.000169	0.999957
3.9	0.000114	0.999971
4.0	0.000076	0.999980
4.1	0.000050	0.999987
4.2	0.000033	0.999991
4.3	0.000021	0.999993
4.4	0.000014	0.999995
4.5	0.000009	0.999996
4.6	0.000005	0.999997
4.7	0.000003	0.999997
4.8	0.000002	0.999998
4.9	0.000001	0.999998
5.0	0.000001	0.999998

MEAN: 0.595 VARIANCE: 0.25623

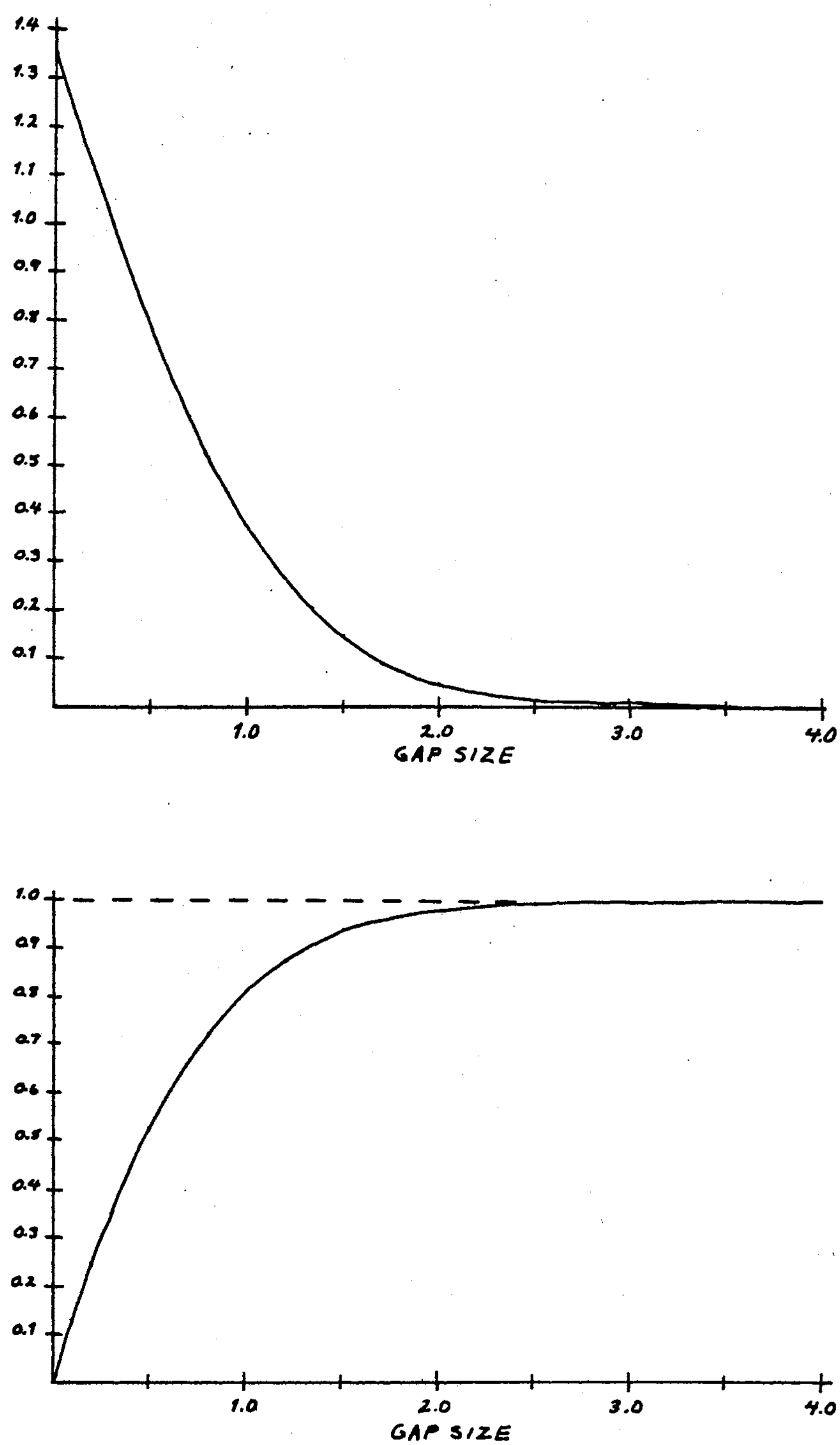


Figure 6. Density Function and Cumulative Distribution Function of Gap 1 in a Sample Size 7

TABLE VI

EMPIRICAL AND PRODUCT-APPROXIMATED DISTRIBUTIONS OF THE MAXIMUM GAP, SAMPLE SIZE 7

INTERVAL	OBSERVED FREQUENCY	RELATIVE FREQUENCY	EMPIRICAL CUMULATIVE FREQUENCY	PRODUCT-APPROXIMATED CUMULATIVE	SIGNED DIFFERENCE
0.0 TO 0.1	0	0.0	0.0	0.00003	0.00003
0.1 TO 0.2	16	0.00046	0.00046	0.00125	0.00079
0.2 TO 0.3	120	0.00343	0.00389	0.00892	0.00503
0.3 TO 0.4	495	0.01414	0.01803	0.03136	0.01333
0.4 TO 0.5	1200	0.03429	0.05231	0.07498	0.02267
0.5 TO 0.6	2022	0.05777	0.11009	0.14088	0.03079
0.6 TO 0.7	2771	0.07917	0.18926	0.22497	0.03571
0.7 TO 0.8	3354	0.09583	0.28509	0.32007	0.03498
0.8 TO 0.9	3492	0.09977	0.38486	0.41846	0.03360
0.9 TO 1.0	3554	0.10154	0.48640	0.51360	0.02720
1.0 TO 1.1	3204	0.09154	0.57794	0.60091	0.02297
1.1 TO 1.2	2921	0.08346	0.66140	0.67785	0.01645
1.2 TO 1.3	2463	0.07037	0.73177	0.74351	0.01174
1.3 TO 1.4	2034	0.05811	0.78989	0.79816	0.00827
1.4 TO 1.5	1754	0.05011	0.84000	0.84272	0.00272
1.5 TO 1.6	1227	0.03506	0.87506	0.87848	0.00342
1.6 TO 1.7	1049	0.02997	0.90503	0.90681	0.00178
1.7 TO 1.8	773	0.02209	0.92711	0.92901	0.00190
1.8 TO 1.9	608	0.01737	0.94449	0.94626	0.00177
1.9 TO 2.0	476	0.01360	0.95809	0.95955	0.00146
2.0 TO 2.1	359	0.01026	0.96834	0.96973	0.00139
2.1 TO 2.2	253	0.00723	0.97557	0.97747	0.00190
2.2 TO 2.3	216	0.00617	0.98174	0.98333	0.00159
2.3 TO 2.4	183	0.00523	0.98700	0.98773	0.00073
2.4 TO 2.5	125	0.00357	0.99054	0.99103	0.00049
2.5 TO 2.6	101	0.00289	0.99343	0.99348	0.00005
2.6 TO 2.7	70	0.00200	0.99543	0.99529	-0.00014
2.7 TO 2.8	54	0.00154	0.99697	0.99661	-0.00036
2.8 TO 2.9	35	0.00100	0.99797	0.99758	-0.00039
2.9 TO 3.0	22	0.00063	0.99860	0.99829	-0.00031
3.0 TO 3.1	15	0.00043	0.99903	0.99879	-0.00024
3.1 TO 3.2	9	0.00026	0.99929	0.99916	-0.00013
3.2 TO 3.3	4	0.00011	0.99940	0.99942	0.00002
3.3 TO 3.4	7	0.00020	0.99960	0.99960	0.0
3.4 TO 3.5	5	0.00014	0.99974	0.99972	-0.00002
3.5 TO 3.6	0	0.0	0.99974	0.99981	0.00007
3.6 TO 3.7	2	0.00006	0.99980	0.99987	0.00007
3.7 TO 3.8	1	0.00003	0.99983	0.99991	0.00008
3.8 TO 3.9	1	0.00003	0.99986	0.99994	0.00008
3.9 TO 4.0	5	0.00014	1.00000	0.99996	-0.00004
4.0 TO 4.1	0	0.0	1.00000	0.99997	-0.00003
4.1 TO 4.2	0	0.0	1.00000	0.99998	-0.00002
4.2 TO 4.3	0	0.0	1.00000	0.99999	-0.00001
4.3 TO 4.4	0	0.0	1.00000	0.99999	-0.00001
4.4 TO 4.5	0	0.0	1.00000	0.99999	-0.00001
4.5 TO 4.6	0	0.0	1.00000	0.99999	-0.00001
4.6 TO 4.7	0	0.0	1.00000	0.99999	-0.00001
4.7 TO 4.8	0	0.0	1.00000	0.99999	-0.00001
4.8 TO 4.9	0	0.0	1.00000	1.00000	0.0
4.9 TO 5.0	0	0.0	1.00000	1.00000	0.0
TOTAL:	35000				

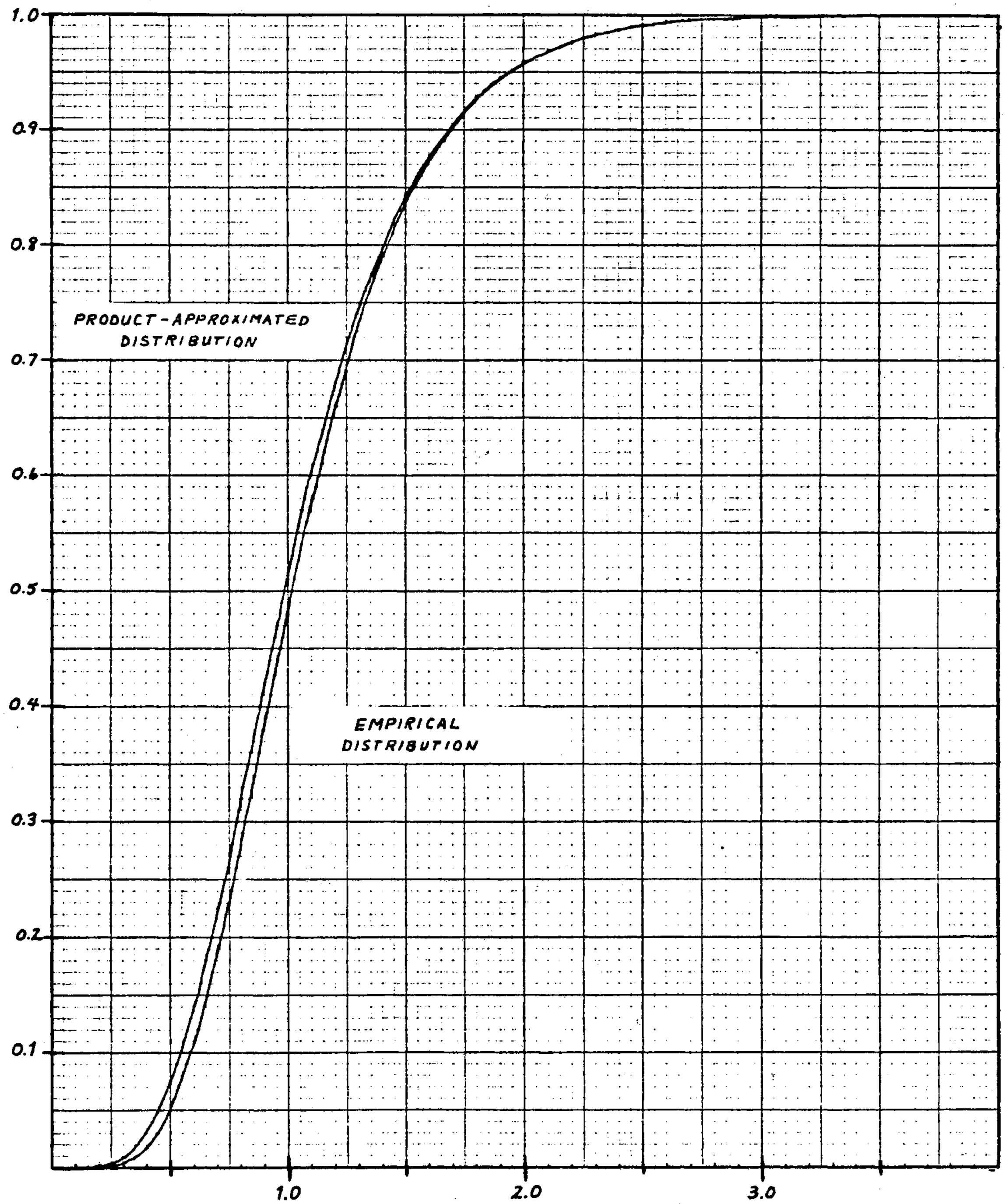


Figure 7. Comparison of the Empirical and Product-Approximated Distributions of the Maximum Gap, Sample Size 7

Having now approximate values for the cumulative probability integral of the maximum gap, the integral (15) may be evaluated by numerical quadrature to yield the distribution function of the studentized maximum gap and also the various critical values of this distribution for different sample sizes and denominator degrees of freedom. Table XXVI in the Appendix gives critical values for the studentized maximum gap for sample sizes 2 through 20.

CHAPTER IV

PERFORMANCE CHARACTERISTICS

1. General Discussion

For multistage procedures such as the studentized maximum gap test (SMG) or the studentized range/maximum gap test (SR/MG), computation and/or discussion of error rates is difficult and confusing, and is seldom done. Type I error rates are computed under the assumption that the null hypothesis is true and are thereby generally more manageable than others. However, for the procedures where further testing depends on rejection of the null hypothesis at the first stage, calculation of Type I error rates is complicated by the fact that the probabilities of such errors at successive stages are governed by the decision reached at the first stage. For example, in a sample of size 6, the probability of declaring another false significance, given that one has been declared at gap 1 is different from what it would be, given that the first false has been declared at gap 3. In the first case there is only one null hypothesis to be tested at the second stage--that the upper five observations are like a sample of five from a normal distribution; while in the second case, there are two null hypotheses to be tested at the second stage--that the smaller three observations are like a sample of three from one normal distribution and that the larger three observations are like a sample of three from another distribution. When there are three or four stages to be considered, the "complication factor" increases.

When the null hypothesis is not true, systematic study of multistage procedures like the SMG and SR/MG becomes all but impossible. First, consider the possible number of ways a random sample of size n may depart from the null hypothesis. There may be anywhere from 2 to n true populations involved and these may be configured in an infinity of ways. Secondly, recall that the SMG and SR/MG tests employ ranking of sample means as an integral part of the procedures--another possible source of error. When one obtains a set of means for which the sample ranking is incorrect, an erroneous grouping can occur in at least three ways: (i) A significant gap my be declared between two means that are reversed with respect to ranking. (This is a Type III error as defined in Chapter II, p. 25 and also as defined in (5).) (ii) A significant gap may be declared between two means from the same population. (This error also occurs in samples with correct ranking.) (iii) A significant gap may be declared between two means which are ranked correctly by themselves, but which have an erroneous ranking elsewhere in the sample. (When this error is made, both of the previous two errors are made also.)

If by "a correct inference" it is meant a correct grouping, then when the null hypothesis is not true, i.e., true gaps exist in the population, a correct inference will be made for a given sample only if the sample ranking is correct and all of the true gaps are detected. All other situations lead to an incorrect inference.

It appears, then, that hoping to draw general conclusions about the behavior of procedures such as the SMG or the SR/MG when the null hypothesis is not true, is probably being too optimistic. Certainly, examination of every potential situation is not possible, and apparently some have felt that the choice is between doing this or doing nothing and have consequently chosen the latter. In the writer's opinion, the choice is somewhere in between; it is felt that in the literature, too much attention is given to performance under the null hypothesis, and that performance under alternative hypotheses is all but ignored, due to the difficulty of studying such cases in a systematic way. A recent empirical evaluation of pairwise multiple comparison procedures by Cramer and Swanson (7) and an earlier empirical sampling experiment by Balaam (4) are a couple of the few cases known to the author where such studies have been attempted.

In this chapter, certain aspects of the performance of three procedures--the LSD applied only to gaps (GLSD), the Studentized Maximum Gap (SMG), and the Studentized Range/Maximum Gap (SR/MG)--will be studied, both under the null hypothesis and under selected alternatives, and some general conclusions will be drawn based on the special cases studied. Topics to be discussed include single stage Type I error rate of the SMG, multistage Type I error rates, probability of incorrect ranking, and comparisons of GLSD, SR/MG, and SMG with respect to error rates and power characteristics for three selected cases.

2. Single-stage Type I Error Frequency of the SMG Procedure

In Chapter III, it was conjectured that critical values for the studentized maximum gap based upon the product-approximated distribution of the maximum gap would yield tests which are non-conservative with respect to Type I error frequency. In this section, a particular case will be examined to determine if the degree of non-conservatism appears to be serious.

One thousand pairs of samples of size 7 and 11 respectively were drawn from a simulated standard normal population. For the first sample of each pair, the maximum gap G was determined, and for the second sample, the sample standard deviation s was computed. The statistic G/s was then formed and compared with the .05 level critical values computed by the studentization integral of the product-approximated distribution for sample size 7 and 10 degrees of freedom. In Table VII are tabulated the number of cases out of 1000 where the sample value of G/s exceeded the α-level critical value of the approximated null sampling distribution of this statistic for sample size 7 and 10 degrees of freedom. If this particular case is any indication of what happens in general, the most serious discrepancies appear to be in the α-levels .01 to .05, and the amount of disagreement might be considered quite unsatisfactory if the attitude were adopted that strict Type I error rates must be maintained. However, when the discrepancies are considered in the context of significance testing, they do not seem so serious. That is, when one computes the probability of obtaining a value greater than that observed, it seems likely to be of no great concern that the probability is actually, say, .056 rather than .05. Also, the approximation to the distribution of the maximum gap improves with larger sample sizes, so that the non-conservatism of the procedure will not be considered as a serious problem.

TABLE VII

EMPIRICAL CHECK OF TYPE I ERROR RATE FOR THE APPROXIMATED DISTRIBUTION OF THE STUDENTIZED MAXIMUM GAP, SAMPLE SIZE 7 AND 10 DEGREES OF FREEDOM

ALPHA LEVEL	CRITICAL VALUE	SIGNIFICANT VALUES		
		NUMBER	% OF TOTAL	STD. DEV.
.500	1.03	507	50.7 %	2.50 %
.400	1.17	412	41.2 %	2.42 %
.300	1.33	312	31.2 %	2.15 %
.200	1.55	215	21.5 %	1.69 %
.150	1.50	163	16.3 %	1.36 %
.100	1.91	109	10.9 %	0.97 %
.050	2.27	56	5.6 %	0.53 %
.040	2.38	45	4.5 %	0.43 %
.030	2.53	37	3.7 %	0.36 %
.025	2.63	33	3.3 %	0.32 %
.020	2.75	27	2.7 %	0.26 %
.015	2.90	20	2.0 %	0.20 %
.010	3.12	11	1.1 %	0.12 %
.005	3.52	6	0.6 %	0.06 %
.001	4.51	1	0.1 %	0.01 %
.0005	4.97	1	0.1 %	0.01 %

TOTAL SAMPLES 1000

3. Overall Error Rates under the Null Hypothesis

As noted earlier, in multi-staged procedures such as the SMG or the SR/MG, the calculation of Type I error rates is more complicated than for single stage procedures; however, in this section a few observations concerning Type I errors will be made for the SMG test. The conclusions will apply to any procedure structured like the SMG, GLSD, or the SR/MG tests.

To simplify the discussion, assume that some α-level is chosen and maintained for all stages of testing, and also assume that there is no error of approximation with respect to the α-level critical values.

The first point to be made is that in any procedure where further testing depends on significance at a previous stage, if an α-level test is consistently used at the first stage, then the probability of making one or more Type I errors in any experiment is exactly α. In other words, in any large set of null experiments, on the average only 100α % of them will have erroneous significances declared. This follows from the fact that the set of null experiments for which more than one erroneous significance is declared is a subset of the set of null experiments for which exactly one erroneous significance is declared. Stated another way, two or more Type I errors cannot be made in any null experiment unless one has already been made. If Tukey's (20) definition of experimentwise error rate is used:

$$\text{Experimentwise error rate} = \frac{\text{Number of [null] experiments with one or more erroneous inferences,}}{\text{Number of [null] experiments}}$$

then the above discussion condenses to: At α-level testing, the Type I experimentwise error rate of the SMG or the SR/MG is exactly α.

For questions such as, "What is the probability that two or more Type I errors will be made?" or, "Given that one Type I error has been made, what is the probability that at least one more will be made?", the answers are more difficult to obtain. Suppose that in an ordered set of means from a null experiment, the studentized maximum gap was erroneously declared significant. To assess the probability of now declaring yet another erroneous significance, it must be considered where the first significance was found. Suppose, for example, that the original sample size was 7 and that the third gap was declared significant, then at the second stage, one would be testing, respectively, the first two gaps out of a sample size 7 and the upper four gaps from a sample of 7. It is now clear that the conditional probability of declaring more erroneous significances, given that one has been declared is less than 2α, because, as has been noted earlier, the lower j gaps of a sample of size n are smaller, on the average, than the j gaps of a sample of size $j + 1$, and similarly for the upper $n - j - 1$ gaps. In any case, it is easy to see that the multiplicity of cases to be considered makes answering either of the two questions above a formidable task.

For purposes of illustration let us suppose that 1,000,000 null experiments are performed for which seven means are to be tested by the SMG procedure. Assume that $\alpha = .05$ is to be used throughout. Let E_i denote the event {Gap i is the largest gap}, then empirical sampling for sample size 7 has established that, approximately,

$$P(E_1) = P(E_6) = .2635$$

$$P(E_2) = P(E_5) = .1354$$

$$P(E_3) = P(E_4) = .1011$$

Let F_i denote the event {Gap i is erroneously declared significant.}, then at the .05 level,

$$P(F_1) = P(F_6) = .05 \times .2635 = .0132$$

$$P(F_2) = P(F_5) = .05 \times .1354 = .0068$$

$$P(F_3) = P(F_4) = .05 \times .1011 = .0050$$

To carry the calculations any further, it is necessary to be able to calculate probabilities of declaring significant gaps in the upper (lower) k observations of 7 when testing with the k-sample studentized maximum gap $(k < 7)$. Rather than actually calculating these probabilities, the assumption will be made that the probabilities are proportional to the ratio of the respective ranges for the case $\sigma^2 = 1$. No claim is being made for the validity of such an assumption; the purpose is to ascertain what sort of results one could obtain if he knew the probabilities in question. Let $R(k)$ denote the range of the upper (lower) k observations in a ranked sample of size 7, and let $S(k)$ denote the range of a sample of k observations. From Table XIX of the Appendix, ✓ we can obtain

$$E(R(2)) = 0.5948 \qquad E(S(2)) = 1.1284$$

$$E(R(3)) = 0.9995 \qquad E(S(3)) = 1.6926$$

$$E(R(4)) = 1.3522 \qquad E(S(4)) = 2.0688$$

$$E(R(5)) = 1.7049 \qquad E(S(5)) = 2.3258$$

$$E(R(6)) = 2.1096 \qquad E(S(6)) = 2.5345$$

Let $r_i = E(R(i))/E(S(i))$; then

$$r_2 = 0.5271 \qquad r_5 = 0.7330$$

$$r_3 = 0.5905 \qquad r_6 = 0.8324$$

$$r_4 = 0.6536$$

Let H_k denote the event {A significant gap is declared in the k upper (lower) observations of the ordered sample}. Then

$$P(H_2) = \alpha \times r_2 = .05 \times .5271 = .0264$$

$$P(H_3) = \alpha \times r_3 = .05 \times .5905 = .0295$$

$$P(H_4) = \alpha \times r_4 = .05 \times .6536 = .0327$$

$$P(H_5) = \alpha \times r_5 = .05 \times .7330 = .0367$$

$$P(H_6) = \alpha \times r_6 = .05 \times .8324 = .0416$$

Let S be the event {At least one gap is declared significant at the second stage} and let $F = F_1 \cup F_2 \cup . . . \cup F_6$, so that F is the event {a significance is declared at the first stage}. As noted earlier, S does not occur unless F occurs, i.e., $S \subseteq F$; hence,

$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(S)}{P(F)} ,$$

but

$$P(S|F) = \frac{P(S|F_1)P(F_1) + . . . + P(S|F_6)P(F_6)}{P(F)} ,$$

$$= \frac{P(H_6|F_1)P(F_1) + P(H_2 \cup H_5|F_2)P(F_2) + . . . + P(H_6|F_6)P(F_6)}{P(F)} ,$$

because $S \cap F_1 = H_6 \cap F_1$, $S \cap F_2 = (H_2 \cup H_5) \cap F_2$, etc.

$$= \frac{P(H_6)P(F_1) + P(H_2 \cup H_5)P(F_2) + . . . + P(H_6)P(F_6)}{P(F)} ,$$

since $(H_2, . . . , H_6)$ and $(F_1, . . . , F_6)$ are mutually independent,

$$= (2/.05)[(.0416)(.0312)+(.0264 + .0367)(.0068)+(.0295 + .0327)(.005)]$$

$$= (2/.05)(.00055 + .00043 + .00031)$$

$$= .00258/.05$$

$$= .0516$$

The answers to the two questions posed above are given by $P(S)$ and $P(S|F)$ respectively; hence, the unconditional probability that two or more Type I errors will be made is .00258, and the conditional probability that at least two Type I errors will be made, given that one has been made is .0516.

To further illustrate these ideas, assume that of the 1,000,000 null experiments, the proportion satisfying any condition is exactly equal to the expected proportion, that is, expected number of cases will be taken as the actual realization of cases satisfying some condition. With this agreement, 50,000 of the null experiments will be cases where false significances are declared at the first stage. Of these 50,000:

26.35 % or 13173 will have declared gap 1 significant (F_1)

13.54 % or 6772 will have declared gap 2 significant (F_2)

10.11 % or 5055 will have declared gap 3 significant (F_3)

10.11 % or 5055 will have declared gap 4 significant (F_4)

13.54 % or 6772 will have declared gap 5 significant (F_5)

26.35 % or 13173 will have declared gap 5 significant (F_6)

At the second stage, the numbers of cases where at least one additional significance will be declared are:

4.16 % of the 13173 in F_1, or 548

6.31 % of the 6772 in F_2, or 427

6.22 % of the 5055 in F_3, or 314

6.22 % of the 5055 in F_4, or 314

6.31 % of the 6772 in F_5, or 427

4.16 % of the 13173 in F_6, or 548

Thus, a total of 2580 of the null experiments will be cases where two or more false significances are declared. Unconditionally this is 2580 out of 1,000,000 or .258 %. Conditionally, it is 2580 out of 50,000 or 5.16 %. These values coincide with the probabilities $P(S)$ and $P(S|F)$ obtained earlier.

One other point may be inferred from this exercise, and that is that of the 50,000 expected cases where at least one Type I error is made, 2580 are expected to be cases where more than one Type I error is made, therefore, 47,420 cases where exactly one Type I error is made are expected. Hence, the probability of exactly one Type I error is .04742.

The validity of the specific probabilities obtained rests, of course, upon how realistic the approximations for the probabilities of the H_k events are. The assumptions made in order to compute those probabilities are not altogether unreasonable, and it is conjectured that the final answers computed with the exact probabilities would not differ materially from those obtained. In any case, it is felt that the expenditure of computer time to calculate the exact probabilities is not warranted at this time.

4. Probability of Correct Rankings

Any procedure which employs the practice of ranking sample means must contend with the possibility that, when the means are from different populations, the sample ranking may not reflect the true ranking of the means. Consider the case of two normal populations with common variance σ^2 and whose means differ by a positive quantity measured in σ-units. Specifically, let $(\overline{X}_{(1)}, \overline{X}_{(2)}, \ldots, \overline{X}_{(k)})$ be k means, each based on n observations randomly selected from a normal $(0, \sigma^2)$ population, and let $(\overline{Y}_{(1)}, \overline{Y}_{(2)}, \ldots, \overline{Y}_{(m)})$ be m means, each based on n observations randomly selected from a normal $(\Delta\sigma, \sigma^2)$ population. Note that there is no loss in generality by assuming $\mu_x = 0$, $\mu_y = \Delta\sigma$, since the controlling quantity is $\mu_y - \mu_x = \Delta\sigma$. The problem is to determine the probability that all the $\overline{Y}_{(j)}$ are larger than all the $\overline{X}_{(i)}$. This probability can be evaluated by writing it as $E_x\{\Pr(\overline{Y}_{(1)} > x|\overline{X}_{(k)} = x)\}$. Let us evaluate the slightly more general expression $\Pr(\overline{Y}_{(1)} > \overline{X}_{(k)} - c)$ where c is some positive constant, which is the probability that the X's and the Y's do not overlap by more than c.

$$\Pr(\overline{Y}_{(1)} > \overline{X}_{(k)} - c) = \int_{-\infty}^{\infty} \Pr(\overline{Y}_{(1)} > x-c|\overline{X}_{(k)} = x) f(x)dx$$

$$= \int_{-\infty}^{\infty} \Big[(\sqrt{n}m/\sqrt{2\pi}\sigma) \int_{x-c}^{\infty} [1-F(\sqrt{n}(t/\sigma-\Delta))]^{m-1}$$

$$\times \exp(-n/2\sigma^2)(t-\sigma\Delta^2)dt\Big]$$

$$\times\ k[F(\sqrt{n}x/\sigma)]^{k-1}(\sqrt{n}/\sqrt{2\pi}\sigma)\exp(-nx^2/2\sigma^2)dx$$

$$= (kn/2\pi)(\sqrt{n}/\sigma)\int_{-\infty}^{\infty}\left[\int_{\sqrt{n}(x/\sigma-c/\sigma-\Delta)}^{\infty}[1-F(w)]^{m-1}\exp(-w^2/2)dw \times [F(\sqrt{n}x/\sigma)]^{k-1}\exp(-nx^2/2\sigma^2)\right]dx$$

$$= (kn/2\pi)\int_{-\infty}^{\infty}\int_{v-\sqrt{n}(c/\sigma+\Delta)}^{\infty}[1-F(w)]^{n-1}[F(v)]^{k-1} \times \exp(-1/2)(w^2+v^2)dwdv \quad . \qquad (1)$$

Note that (1) is a function of k, m, $\sqrt{n}c/\sigma$, and $\sqrt{n}\Delta$, and for $c = 0$, is a function of k, m, and $\sqrt{n}\Delta$ only. Bechhofer (3) evaluated probabilities of this sort for the case $c = 0$ for the purpose of determining the number of observations n necessary to achieve a given probability p of correct ranking for given k, m, and Δ, and his tables can be utilized with minor modifications. The integral (1) was evaluated by numerical integration for selected cases and the results checked against Bechhofer's tables; the two agreed in every case.

To obtain exact probabilities of correct ranking in cases where more than two populations are involved, numerical evaluation of multiple integrals is unfortunately necessary. To see that this is true, consider the general case. For each i, let $\overline{X}_i$ be a mean based on n observations chosen at random from a normal $(\Delta_i\sigma, \sigma^2)$ population,

$i = 1, 2, \ldots, k$, and assume that $0 = \Delta_1 \le \Delta_2 \le \ldots \le \Delta_k$. The case where only two populations are involved is obtained when, for some $q < k$, $\Delta_1 = \Delta_2 = \ldots = \Delta_q = 0$ and $\Delta_{q+1} = \ldots = \Delta_k > 0$. Let $\lambda_1, \ldots, \lambda_r$ represent the set of distinct Δ_i other than those equal to $\Delta_1 = 0$. Then the first k_1 of the $\overline{X}_i$ will be from the population with mean 0; the next k_2 of the $\overline{X}_i$ will be from the population with mean $\lambda_1\sigma$; . . . ; and the remaining k_{r+1} of the $\overline{X}_i$ will be from the population with mean $\lambda_r\sigma$. A correct sample ranking occurs if and only if the event

$$\{\overline{X}_{max_1} < \overline{X}_{min_2}, \overline{X}_{max_2} < \overline{X}_{min_3}, \ldots, \overline{X}_{max_r} < \overline{X}_{min_{r+1}}\}$$

occurs, where

$$\overline{X}_{max_1} = \max\{\overline{X}_1, \ldots, \overline{X}_{k_1}\},\ \overline{X}_{min_2} = \min\{\overline{X}_{k_1+1}, \ldots, \overline{X}_{k_1+k_2}\}, \text{ etc.}$$

Denote the probability of correct ranking by $P(k_1, k_2, \ldots, k_{r+1}; \lambda_1, \ldots, \lambda_r)$. Then, if the events $\{\overline{X}_{max_i} < \overline{X}_{min_{i+1}}\}$ $i = 1, 2, \ldots, r$, were mutually independent events, we would have

$$P(k_1, k_2, \ldots, k_{r+1}; \lambda_1, \lambda_2, \ldots, \lambda_r) = \prod_1^r \Pr(\overline{X}_{max_i} < \overline{X}_{min_{i+1}})$$

$$= \prod_1^r P(k_i, k_{i+1}; \lambda_i - \lambda_{i-1}) \tag{2}$$

so that products of probabilities of the form (1) could be employed. Unfortunately, the events $\{\overline{X}_{max_i} < \overline{X}_{min_{i+1}}\}$ are not mutually independent. Thus, $P(k_1, k_2, \ldots, k_{r+1}; \lambda_1, \lambda_2, \ldots, \lambda_r) \neq \prod_1^r P(k_i, k_{i+1}; \lambda_i - \lambda_{i-1})$. This brings up the question of what to do when it is desired to study cases involving more than only two populations. There are at least three alternative approaches, all of which are approximate solutions:

(i) One could form the appropriate multiple integrals and estimate their solution by Monte Carlo techniques.

(ii) The product probabilities in (2) could be multiplied together as if the events involved were mutually independent.

(iii) Empirical sampling from the appropriate simulated populations could be done.

Alternative (i) will not be considered here, but a "one shot" comparison of (ii) and (iii) will be made. Consider a situation where in a sample of size 7, there are four true populations involved. Let each of the seven means be based on the same number of observations, say n, and let each population mean be separated from the rest by a constant amount measured in units of σ, say $\Delta\sigma$. Suppose the subsample sizes from the respective populations are 2, 2, 2, and 1. Then the probability sought is $P(2, 2, 2, 1; \Delta\sigma, 2\Delta\sigma, 3\Delta\sigma)$, and the appropriate product would give,

$$P(2, 2, 2, 1; \Delta\sigma, 2\Delta\sigma, 3\Delta\sigma) \doteq P(2, 2; \Delta\sigma) \times P(2, 2; \Delta\sigma) \times P(2, 1; \Delta\sigma) \quad (3)$$

Shown on the following page in Table VIII is a comparison of results obtained by methods (ii) and (iii) for the estimation of the probability (3). The estimated probabilities are given as a function of $n\Delta^2$.

TABLE VIII

COMPARISON OF EMPIRICAL SAMPLING AND PRODUCT APPROXIMATION FOR THE PROBABILITY OF CORRECT RANKING, SAMPLE SIZE 7

$n\Delta^2$	$P(2,2;\ \Delta\sigma)$	$P(2,1;\ \Delta\sigma)$	$P(2,2,2,1;\ \Delta\sigma,2\Delta\sigma,3\Delta\sigma)$ Product-approx.	Empirical*
0	0.1667	0.3333	0.0093	0.0095**
1	0.4625	0.6337	0.1356	0.100
2	0.5527	0.7449	0.2276	0.226
4	0.7749	0.8658	0.5199	0.479
6	0.8702	0.9260	0.7012	0.674
10	0.9562	0.9760	0.8925	0.886
16	0.9913	0.9996	0.9783	0.983

* Each number is based on 1000 samples.
** Exact value, computed as (2!)(2!)(2!)/(7!)

The results of the two methods are in relative agreement, and for rough calculations, multiplication of Bechhofer's probabilities is adequate. If more precise estimates are needed, then either empirical sampling or Monte Carlo estimation of multiple integrals should probably be used.

5. Comparison of the GLSD, SMG, and SR/MG with Respect to Correct Detection of Grouping and Certain Error Rates

5.1 <u>Special Case 1, Group Size 4, One True Gap</u>

Let $\overline{X}_1$ and $\overline{X}_2$ be means based on n observations each from a normal (μ, σ^2) population, and let $\overline{X}_3$ and $\overline{X}_4$ be means based on n observations each from a normal $(\mu+\Delta\sigma, \sigma^2)$ population. Since it is not necessary to distinguish between $\overline{X}_1$ and $\overline{X}_2$, nor between $\overline{X}_3$ and

$\overline{X}_4$, denote, for any given sample, $\overline{X}_1$ and $\overline{X}_2$ by A and $\overline{X}_3$ and $\overline{X}_4$ by B.

Recall that by a Type III error is meant declaring a wrong-way significance. A Type III error would occur for the situation above if, for example, a sample ranking resulted in the order ABAB and a significance was declared at the second gap. Suppose, on the other hand, that the sample ranking is correct, i.e. AABB is the order. A correct grouping is still not guaranteed, because it would be possible for the first gap or the third gap to be the largest and be declared significant. An error of this type would be a form of a Type I error; however, let us reserve the name Type I error for erroneous significances under the null hypothesis and call this error a Type IV error. Thus, a Type IV error is simply a Type I error when the null hypothesis is not true.

There is yet a further distinction to be made. Consider, again, the incorrect sample ranking ABAB. If the second gap is declared significant, then, as was noted above, an explicit Type III error would be made, but, in addition, two Type IV errors would be implicitly committed also. Notationally, let this distinction be made by the terms Type III(EX) and Type IV(IM). For the erroneous ordering ABBA with gap 2 being declared significant, along with the Type IV(EX) error committed, one Type IV(IM) error and one Type III(IM) error would also result.

For the situation under consideration, let us employ the technique of constructing a two-way table of all possible relevant states of nature versus all possible decisions, and in each cell indicate the result of each decision with respect to Type III errors and Type IV errors for the given state of nature. For the letters A,A,B,B, there are six possible orderings, one of which represents a correct ranking, so that there are six possible states of nature. Also, there are eight distinct decisions possible: declare no gaps significant; declare one gap significant at gap 1, gap 2, or gap 3; declare two gaps significant at gaps 1 and 2, gaps 1 and 3, or gaps 2 and 3; and declare all three gaps significant. Table IX shows that a correct grouping occurs in only 1 out of 48 possible outcomes for this case of two sample means from two populations, the correct grouping occurring, of course, when a correct sample ranking is obtained and a significant gap is declared at gap 2 alone. Each of the other decision/ranking combinations represents an error in grouping, certainly, and involves errors of the sort discussed above in varying number and form.

The construction of a table such as Table IX is a valuable exercise because it illustrates very plainly the difficulty in calculating specific Type III and Type IV error rates. It shows, for example, that the type and number of errors made depend upon what decision is made for a given sample ranking configuration. When one considers construction of such a table for other cases, he realizes that the size and complexity of the table increase very rapidly as the sample size or the number of true gaps increases, and the business of assessing Type III and Type IV error rates is all but hopeless, except on a case-by-case basis for a few simple cases.

For the case of sample size 4, one true gap, empirical sampling was done to estimate Type III and Type IV error rates and also to compare performance characteristics for the GLSD, SMG, and SR/MG procedures. A constant α-level of .05 was chosen; similar results would be obtained for other levels. As was shown in Section 4 of this chapter, the amount of separation of the populations in relative σ-units can be represented by

TABLE IX

POSSIBLE OUTCOMES, GROUP SIZE 4, ONE TRUE GAP AT GAP 2

		Correct Ranking	Incorrect Ranking				
		AABB	ABAB	ABBA	BAAB	BABA	BBAA
Declare No Gaps Significant							
Declare One Gap Significant	Gap 1	1 Type IV(EX)	1 Type IV(IM)	1 Type IV(IM)	1 Type IV(IM) 1 Type III(EX)	1 Type IV(IM) 1 Type III(EX)	1 Type IV(EX) 1 Type III(IM)
	Gap 2	Correct Grouping	2 Type IV(IM) 1 Type III(EX)	1 Type IV(EX) 1 Type III(IM)	2 Type IV(EX) 1 Type III(IM)	2 Type IV(IM) 1 Type III(IM)	1 Type III(EX) 3 Type III(IM)
	Gap 3	1 Type IV(EX)	1 Type IV(IM)	1 Type IV(IM) 1 Type III(EX)	1 Type IV(IM)	1 Type IV(IM) 1 Type III(EX)	1 Type IV(EX) 2 Type III(IM)
Declare Two Gaps Significant	Gap 1 and Gap 2	1 Type IV(EX) 1 Correct	2 Type IV(IM) 1 Type III(EX)	1 Type IV(EX) 1 Type IV(IM) 1 Type III(IM)	1 Type IV(EX) 1 Type IV(IM) 1 Type III(EX) 1 Type III(IM)	2 Type IV(IM) 1 Type III(EX) 1 Type III(IM)	1 Type IV(EX) 1 Type III(EX) 3 Type III(IM)
	Gap 1 and Gap 3	2 Type IV(EX)	2 Type IV(IM)	1 Type IV(IM) 1 Type III(EX) 2 Type III(IM)	1 Type IV(IM) 1 Type III(EX) 2 Type III(IM)	2 Type IV(IM) 2 Type III(EX) 1 Type III(IM)	2 Type IV(EX) 3 Type III(IM)
	Gap 2 and Gap 3	1 Type IV(EX) 1 Correct	2 Type IV(IM) 1 Type III(IM)	1 Type IV(EX) 1 Type IV(IM) 1 Type III(EX) 1 Type III(IM)	1 Type IV(EX) 1 Type IV(IM) 1 Type III(IM)	2 Type IV(IM) 1 Type III(EX) 1 Type III(IM)	1 Type IV(EX) 1 Type III(EX) 3 Type III(IM)
Declare Three Gaps Significant		2 Type IV(EX) 1 Correct	2 Type IV(IM) 1 Type III(EX)	1 Type IV(EX) 1 Type IV(IM) 1 Type III(EX) 1 Type III(IM)	1 Type IV(EX) 1 Type IV(IM) 1 Type III(EX) 1 Type III(IM)	2 Type IV(IM) 2 Type III(EX) 1 Type III(IM)	2 Type IV(EX) 1 Type III(EX) 3 Type III(IM)

the quantity $\sqrt{n}\Delta$, where $\Delta\sigma$ is the distance between the unknown true means of the populations and n is the number of observations upon which each mean is based. For convenience, let the parameter $\sqrt{n}\Delta$ be called the separation index. For each of the eight values $\sqrt{n}\Delta = 1, \sqrt{2}, 2, \sqrt{6}, \sqrt{10}, 4, 5,$ and 6, 700 random samples of size 4 were drawn, with X_1 and X_2 from a simulated normal (0, 1) population, and X_3 and X_4 from a simulated normal ($\sqrt{n}\Delta$, 1) population. It should be noted that sampling in this manner simulates the general formulation of $\overline{X}_1$ and $\overline{X}_2$ from a normal (μ, σ^2) population and $\overline{X}_3$ and $\overline{X}_4$ from a normal ($\mu+\Delta$, σ^2) population, with each $\overline{X}_i$ based on n observations. For each of the 700 samples of size 4, an independent sample of size 10 was also drawn from a simulated normal (0, 1) population in order to obtain an independent estimate of σ^2 (=1). The observed t-values, studentized maximum gaps, and studentized ranges were therefore based on 9 degrees of freedom. For each of the 700 pairs of samples, the GLSD, SR/MG, and SMG tests were performed, and the results tabulated in the format of Table IX. Table X shows the empirical percentages obtained for the events identified in Table IX. From the table, it is concluded that, for this case, the SR/MG and SMG procedures give comparable results with respect to obtaining a correct grouping, (These are the numbers in the (3,1) cells; see Table IX), with the SR/MG obtaining slightly more correct groupings in most cases. Both, however, are superior to the GLSD in that respect. It is also observed that the SR/MG and SMG procedures have higher frequencies of Type III and Type IV errors than the GLSD (an expected result). From Table IX and Table X, the specific frequencies of Type III(EX), Type III(IM), Type IV(EX), and Type IV(IM) errors may be calculated. Table XI contains the results of such calculations.

TABLE X

RESULTS OF EMPIRICAL SAMPLING, GROUP SIZE 4, ONE TRUE GAP AT GAP 2

	Gaps Declared	GLSD						SR/MG						SMG					
		AABB	ABAB	ABBA	BAAB	BABA	BBAA	AABB	ABAB	ABBA	BAAB	BABA	BBAA	AABB	ABAB	ABBA	BAAB	BABA	BBAA
$n\Delta^2 = 1$	None	45.1	22.3	8.4	11.4	5.4	2.6	42.0	21.3	8.3	11.1	5.1	2.7	43.1	21.4	7.1	11.0	5.1	2.6
	1 at Gap 1	1.4	0.6	0.3	0	0	0.1	2.3	1.1	0.4	0.1	0	0	1.9	0.9	0.4	0.1	0	0.1
	1 at Gap 2	0.9	0	0.1	0	0	0	1.4	0	0.1	0.1	0.3	0	1.4	0	0.1	0	0.3	0
	1 at Gap 3	0.7	0.3	0	0.1	0	0.1	2.4	0.7	0	0.1	0	0.1	1.7	0.9	0	0.4	0	0.1
	2 at Gaps 1,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 at Gaps 1,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 at Gaps 2,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	All Gaps	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$n\Delta^2 = 2$	None	57.7	20.0	6.1	5.6	2.0	2.0	51.0	18.6	5.7	5.6	2.1	2.1	53.7	18.0	6.1	5.3	2.0	2.1
	1 at Gap 1	1.1	0.7	0.1	0	0	0	3.3	0.7	0.4	0	0	0	2.0	1.0	0.1	0	0	0
	1 at Gap 2	1.7	0	0	0	0.1	0	4.0	0.1	0.1	0	0	0	3.7	0.3	0	0	0.1	0
	1 at Gap 3	1.1	1.0	0	0.1	0	0	3.3	2.3	0	0.1	0	0	2.3	2.4	0	0.4	0	0
	2 at Gaps 1,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 at Gaps 1,3	0.1	0	0	0	0	0	0.3	0	0	0	0	0	0.1	0	0	0	0	0
	2 at Gaps 2,3	0.1	0	0	0	0	0	0.1	0	0	0	0	0	0.1	0	0	0	0	0
	All Gaps	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$n\Delta^2 = 4$	None	66.0	12.7	3.3	2.9	1.0	0.4	56.3	12.9	3.3	2.9	1.0	0.4	58.3	11.9	3.1	2.9	1.0	0.4
	1 at Gap 1	1.9	1.9	0	0	0	0	4.3	1.7	0	0	0	0	3.3	2.1	0.1	0	0	0
	1 at Gap 2	4.3	0	0.3	0	0	0	7.9	0	0.3	0	0	0	8.1	0.1	0.3	0	0	0
	1 at Gap 3	3.4	1.9	0	0.1	0	0	7.1	1.4	0	0.1	0	0	5.9	1.9	0	0.1	0	0
	2 at Gaps 1,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 at Gaps 1,3	0.3	0	0	0	0	0	0.3	0.1	0	0	0	0	0.3	0.1	0	0	0	0
	2 at Gaps 2,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	All Gaps	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$n\Delta^2 = 6$	None	72.1	8.4	0.7	1.3	0.4	0.3	57.7	8.3	0.7	1.1	0.4	0.3	60.1	7.7	0.4	1.1	0.4	0.3
	1 at Gap 1	3.1	0.3	0	0	0	0	4.1	0.4	0	0	0	0	4.7	0.6	0.3	0	0	0
	1 at Gap 2	8.7	0	0	0	0	0	17.3	0	0	0.1	0	0	16.6	0	0	0	0	0
	1 at Gap 3	2.9	0.9	0	0.1	0	0	7.1	0.9	0	0.1	0	0	5.4	1.3	0	0.3	0	0
	2 at Gaps 1,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 at Gaps 1,3	0.7	0	0	0	0	0	1.1	0	0	0	0	0	0.7	0	0	0	0	0
	2 at Gaps 2,3	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
	All Gaps	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

		AABB	ABAB	ABBA	BAAB	BABA	BBAA	AABB	ABAB	ABBA	BAAB	BABA	BBAA	AABB	ABAB	ABBA	BAAB	BABA	BBAA
$n\Delta^2 = 10$	None	68.3	2.0	0.1	0.4	0	0	48.1	1.3	0.1	0.4	0	0	55.4	1.6	0.1	0.3	0	0
	1 at Gap 1	1.9	0.1	0	0	0	0	6.6	0.6	0	0	0	0	3.3	0.4	0	0	0	0
	1 at Gap 2	20.9	0	0	0	0	0	31.0	0	0	0	0	0	29.6	0	0	0	0	0
	1 at Gap 3	4.9	0.3	0	0	0	0	9.4	0.6	0	0	0	0	7.0	0.4	0	0.1	0	0
	2 at Gaps 1,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 at Gaps 1,3	0.7	0	0	0	0	0	0.9	0	0	0	0	0	0.9	0	0	0	0	0
	2 at Gaps 2,3	0.3	0	0	0	0	0	0.9	0	0	0	0	0	0.7	0	0	0	0	0
	All Gaps	0.1	0	0	0	0	0	0.1	0	0	0	0	0	0.1	0	0	0	0	0
$n\Delta^2 = 16$	None	48.1	0.6	0.1	0	0	0	26.9	0.4	0.1	0	0	0	33.6	0.4	0.1	0	0	0
	1 at Gap 1	1.6	0	0	0	0	0	4.3	0	0	0	0	0	3.4	0	0	0	0	0
	1 at Gap 2	42.3	0	0	0	0	0	56.6	0	0	0	0	0	53.9	0	0	0	0	0
	1 at Gap 3	4.9	0	0	0	0	0	8.6	0.1	0	0	0	0	5.9	0.1	0	0	0	0
	2 at Gaps 1,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 at Gaps 1,3	1.4	0	0	0	0	0	1.6	0	0	0	0	0	1.4	0	0	0	0	0
	2 at Gaps 2,3	0.9	0	0	0	0	0	1.3	0	0	0	0	0	1.0	0	0	0	0	0
	All Gaps	0.1	0	0	0	0	0	0.1	0	0	0	0	0	0.1	0	0	0	0	0
$n\Delta^2 = 25$	None	25.1	0.1	0	0	0	0	12.3	0.1	0	0	0	0	13.7	0	0	0	0	0
	1 at Gap 1	1.7	0	0	0	0	0	2.0	0	0	0	0	0	1.9	0	0	0	0	0
	1 at Gap 2	66.3	0	0	0	0	0	75.3	0	0	0	0	0	76.6	0	0	0	0	0
	1 at Gap 3	2.6	0	0	0	0	0	5.6	0	0	0	0	0	3.3	0.1	0	0	0	0
	2 at Gaps 1,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 at Gaps 1,3	2.3	0	0	0	0	0	2.4	0	0	0	0	0	2.3	0	0	0	0	0
	2 at Gaps 2,3	1.9	0	0	0	0	0	2.3	0	0	0	0	0	2.1	0	0	0	0	0
	All Gaps	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$n\Delta^2 = 36$	None	7.6	0	0	0	0	0	2.7	0	0	0	0	0	3.6	0	0	0	0	0
	1 at Gap 1	0.6	0	0	0	0	0	0.9	0	0	0	0	0	0.9	0	0	0	0	0
	1 at Gap 2	83.0	0	0	0	0	0	86.9	0	0	0	0	0	86.6	0	0	0	0	0
	1 at Gap 3	4.7	0	0	0	0	0	4.9	0	0	0	0	0	4.6	0	0	0	0	0
	2 at Gaps 1,2	0.1	0	0	0	0	0	0.1	0	0	0	0	0	0.1	0	0	0	0	0
	2 at Gaps 1,3	2.7	0	0	0	0	0	2.9	0	0	0	0	0	2.9	0	0	0	0	0
	2 at Gaps 2,3	1.1	0	0	0	0	0	1.6	0	0	0	0	0	1.3	0	0	0	0	0
	All Gaps	0.1	0	0	0	0	0	0.1	0	0	0	0	0	0.1	0	0	0	0	0

* Note: All quantities are expressed as percent of the total 700

TABLE XI

ERROR FREQUENCIES OF THE GLSD, SR/MG, AND SMG
OBTAINED BY EMPIRICAL SAMPLING, (N = 700)

$n\Delta^2$	Type IV(EX)			Type IV(IM)			Type III(EX)			Type III(IM)		
	GLSD	SR/MG	SMG	GLSD	SR/MG	SMG	GLSD	SR/MG	SMG	GLSD	SR/MG	SMG
1	18	37	28	9	22	23	0	1	1	5	6	7
2	19	52	33	16	27	34	0	1	2	1	1	1
4	43	86	70	25	25	34	0	0	1	2	2	2
6	52	98	81	9	10	17	0	0	0	0	1	0
10	61	132	91	3	8	7	0	0	0	0	0	0
16	73	123	94	0	1	1	0	0	0	0	0	0
25	75	103	83	0	0	1	0	0	0	0	0	0
36	86	94	90	0	0	0	0	0	0	0	0	0

Several observations may be made from Table XI. First, the occurrence of Type III errors does not appear to be of major concern. The explanation for this is that Type III errors occur only when incorrect rankings occur, and the cases for which incorrect rankings are more likely are also cases where the breaks are less likely to be declared significant. Another thing to note is that the frequency of Type IV errors is greater than might have been suspected, and that the absolute frequency of explicit Type IV errors increases with the separation index $\sqrt{n}\Delta$. However, when the change in frequency of explicit Type IV errors with increasing $\sqrt{n}\Delta$ is considered relative to the number of significances declared in each case, then the percentage of declared significances resulting in explicit Type IV errors decreases on the average with $\sqrt{n}\Delta$ (with GLSD and SMG yielding the smaller percentages) as shown in Table XII on the following page.

TABLE XII

PERCENTAGE OF DECLARED SIGNIFICANCES RESULTING IN TYPE IV(EX) ERRORS

$n\Delta^2$	GLSD	SR/MG	SMG
1	55 %	56 %	47 %
2	40 %	49 %	36 %
4	44 %	52 %	44 %
6	43 %	43 %	38 %
10	29 %	37 %	30 %
16	20 %	24 %	20 %
25	14 %	17 %	14 %
36	13 %	14 %	13 %

Recall that explicit and implicit Type IV errors occur when means from the same population are put into different groups. Is this a serious type error? The answer, of course, depends (as with all other errors) upon the particular situation. If he is faced with a situation where Type IV(EX) errors are a major consideration, then one should be aware that in the case under study, at least, the absolute percentage of such errors was observed to go as high as 19 %.

Of greater interest, however, is the relative performance of the tests with respect to obtaining a correct grouping. Figure 8 gives a comparison of the power characteristics of the three procedures with respect to estimated probability of correct grouping. The curves are given as functions of the squared separation index $n\Delta^2$, and graphs of this type could be used in at least two ways. One might formulate his requirements as, "If the true distance between the populations is Δ sigma units, then I want to have probability of at least p of correct

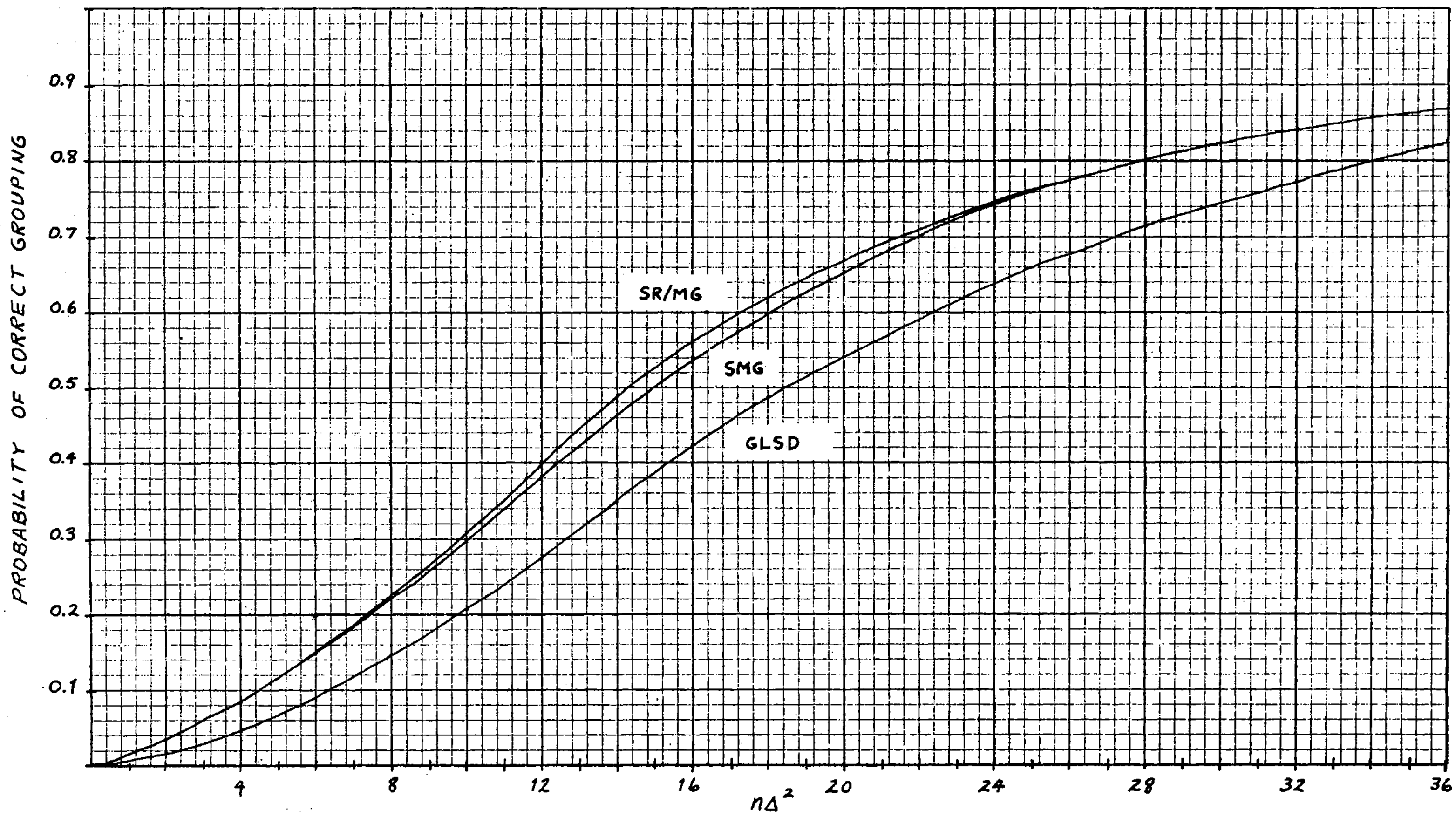


Figure 8. Comparison of Estimated Probability of Correct Grouping for the GLSD, SR/MG, and SMG Procedures, Case 1

grouping." The minimum sample size n for each mean which is required in order to attain the objective is then determined by $n = k/\Delta^2$, where k is the value of $n\Delta^2$ corresponding to the ordinate p for the particular procedure under consideration. Alternatively, one might ask, "With a sample like the one I have, what chances do I have of obtaining a correct grouping if the true separation is Δ sigma units?" The answer is obtained, of course, as the ordinate p corresponding to $n\Delta^2$. A word of caution here--the term "correct grouping" requires special interpretation in the cases $n\Delta^2 = 0$. Literally, a correct grouping when $n\Delta^2 = 0$ is obtained when all means are grouped together into one group, and the probability of doing this is $1 - \alpha$ in accordance with Type I error specifications. However, consider what correct grouping means when the true separation is some small quantity. A correct grouping for the case under study is obtained when the sample ranking configuration is AABB and the second gap is declared significant. Taking the separation index to 0, it is seen that the event being examined for $n\Delta^2 = 0$ is the simultaneous occurrence of the sample ranking AABB and the (erroneous) significance of gap 2. The probabilities of these events are, respectively, $1/6 = .1667$ and approximately .0181. Thus, p for $n\Delta^2 = 0$ is about .003 for all three procedures when they are conducted at an α-level of .05.

Figure 8 clearly shows that the SR/MG and SMG procedures have uniformly greater probability of detecting the correct grouping than the GLSD. It also indicates that there is essentially no difference between the performances of SR/MG and SMG in this regard.

5.2 Special Case 2, Group Size 7, One True Gap at Gap 4

Let $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_3$, and $\overline{X}_4$ be four means, each based on n observations from a normal (μ,σ^2) population, and let $\overline{X}_5$, $\overline{X}_6$, and $\overline{X}_7$ be three means, each based on n observations from a normal $(\mu+\Delta\sigma, \sigma^2)$ population. Using the same convention as earlier, a correct grouping is obtained when the sample ranking is AAAABBB and a gap is declared between the A's and B's.

A study of this case similar to that of the previous case is not at all practical, as there are $(7!)/(4!)(3!) = 35$ incorrect sample rankings possible and a similar expansion in the number of cases with regard to the possible decisions. Re-examination of Case 1 reveals the following trend: For the smaller values of the separation index $\sqrt{n}\Delta$, where incorrect sample rankings are more likely, the most prevalent errors are Type II errors (failure to detect the true gaps), whereas, for larger values of $\sqrt{n}\Delta$, where correct sample rankings have higher probabilities, the most common errors are Type II and Type IV(EX). This suggests that for Case 2, it may be possible to extract most of the relevant performance information about the procedures under consideration by examining only the conditional sampling space of correct sample rankings. Without studying incorrectly ranked samples, we can still assert that for those samples, it would be expected that the majority of errors would be Type II errors with a smattering of Type III and Type IV errors. For Case 1, it was found that Type IV(EX) errors were the most common of the four: Type IV(EX), Type IV(IM), Type III(EX), and Type III(IM). A little further calculation reveals that of the Type IV(EX) errors committed, for each value of $\sqrt{n}\Delta$ examined, no less than 75% and, generally, 95-100% of them were made when the sample ranking was correct. Therefore, for Case 2, correctly ranked samples only will be analyzed with the GLSD, SR/MG, and SMG procedures; when an incorrectly ranked sample is drawn, it will be counted and discarded, and another sample drawn.

For the correctly ranked samples, the results of testing in each case can be cross-classified by how many gaps were declared significant, versus how many were correctly declared. The number and type of errors resulting for each case are given in Table XIII on the next page.

For each of the seven values of the separation index, $\sqrt{n}\Delta = 2, \sqrt{10}, 4, 5, \sqrt{35}, \sqrt{45}, \sqrt{55}$, samples of size 4 and 3 were drawn from simulated normal (0, 1) and normal $(\sqrt{n}\Delta, 1)$ populations, respectively, until 100 samples with correct sample ranking had been drawn and analyzed by the GLSD, SR/MG, and SMG procedures. The studentizing statistic s was based, as before, on an independent sample of size 10 from a simulated normal (0, 1) population, making each of the tests based on 9 degrees of freedom. The results are tabulated in Table XIV using the format of Table XIII.

TABLE XIII

BREAKDOWN OF CORRECTLY RANKED SAMPLES, TYPE II AND TYPE IV(EX) ERRORS, GROUP SIZE 7, ONE TRUE GAP AT GAP 4

Number of Gaps Declared Significant		Number of Gaps Correctly Declared: 0	1
0	Type II	1	Impossible Event
	Type IV(EX)	0	
1	Type II	1	0
	Type IV(EX)	1	0
2	Type II	1	0
	Type IV(EX)	2	1
3	Type II	1	0
	Type IV(EX)	3	2
4	Type II	1	0
	Type IV(EX)	4	3
5	Type II	1	0
	Type IV(EX)	5	4
6	Type II	Impossible Event	0
	Type IV(EX)		5

* Cell Entries are: Number of Type II Errors / Number of Type IV(EX) Errors

TABLE XIV

RESULTS OF EMPIRICAL SAMPLING, GROUP SIZE 7, ONE TRUE GAP AT GAP 4

$n\Delta^2 = 4.0$ — 182 Total Samples, 82 Incorrectly Ranked

$n\Delta^2 = 10.0$ — 111 Total Samples, 11 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	GLSD 1	SR/MG Gaps Correct 0	SR/MG 1	SMG Gaps Correct 0	SMG 1	GLSD Gaps Correct 0	GLSD 1	SR/MG Gaps Correct 0	SR/MG 1	SMG Gaps Correct 0	SMG 1
	$n\Delta^2 = 4.0$						$n\Delta^2 = 10.0$					
0	97	-	73	-	81	-	84	-	48	-	64	-
1	1	2	19	6	14	5	2	12	24	25	11	20
2	0	0	1	1	0	0	1	1	1	2	2	2
3	0	0	0	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	-	0	-	0	-	0	-	0	-	0	-	0

$n\Delta^2 = 16.0$ — 100 Total Samples, 0 Incorrectly Ranked

$n\Delta^2 = 25.0$ — 101 Total Samples, 1 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	GLSD 1	SR/MG Gaps Correct 0	SR/MG 1	SMG Gaps Correct 0	SMG 1	GLSD Gaps Correct 0	GLSD 1	SR/MG Gaps Correct 0	SR/MG 1	SMG Gaps Correct 0	SMG 1
	$n\Delta^2 = 16.0$						$n\Delta^2 = 25.0$					
0	77	-	28	-	47	-	48	-	18	-	23	-
1	1	21	22	42	8	42	2	45	10	58	7	64
2	0	1	3	5	0	3	0	3	3	10	0	4
3	0	0	0	0	0	0	0	2	0	1	0	2
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	-	0	-	0	-	0	-	0	-	0	-	0

$n\Delta^2 = 35.0$ — 100 Total Samples, 0 Incorrectly Ranked

$n\Delta^2 = 45.0$ — 100 Total Samples, 0 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	GLSD 1	SR/MG Gaps Correct 0	SR/MG 1	SMG Gaps Correct 0	SMG 1	GLSD Gaps Correct 0	GLSD 1	SR/MG Gaps Correct 0	SR/MG 1	SMG Gaps Correct 0	SMG 1
	$n\Delta^2 = 35.0$						$n\Delta^2 = 45.0$					
0	18	-	4	-	6	-	6	-	1	-	0	-
1	0	79	1	89	0	89	1	87	0	88	1	88
2	0	3	0	6	0	3	0	5	0	8	0	10
3	0	0	0	0	0	2	0	1	0	3	0	1
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	-	0	-	0	-	0	-	0	-	0	-	0

$n\Delta^2 = 55.0$ — 100 Total Samples, 0 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	GLSD 1	SR/MG Gaps Correct 0	SR/MG 1	SMG Gaps Correct 0	SMG 1
0	2	-	0	-	1	-
1	0	95	0	89	0	91
2	0	3	0	10	0	6
3	0	0	0	1	0	2
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	-	0	-	0	-	0

In an attempt to summarize some of the results presented in Table XVI, let us define a loss function and examine what the expected loss is for each test for each value of $n\Delta^2$, where estimated probabilities are taken from Table XVI. To further simplify, let us consider only expected losses within the conditional sample subspace of correct sample rankings, since, if this is done, the numbers in Table XVI can be used as probability estimates directly without modification.

To devise a loss function, let it be based on relative costs of Type II and Type IV(EX) errors. Suppose 100 units of cost are to be divided so that the cost of a Type II error is c_1 units and the cost of a Type IV(EX) error is c_2 units. ($c_1 + c_2 = 100$). Let

t = Number of true gaps,

i = Number of gaps declared significant,

j = Number of gaps correctly declared significant.

From Table XIII, it is easily determined that

$t-j$ = Number of Type II errors in the $(i+1, j+1)$ cell,

$i-j$ = Number of Type IV(EX) errors in the $(i+1, j+1)$ cell.

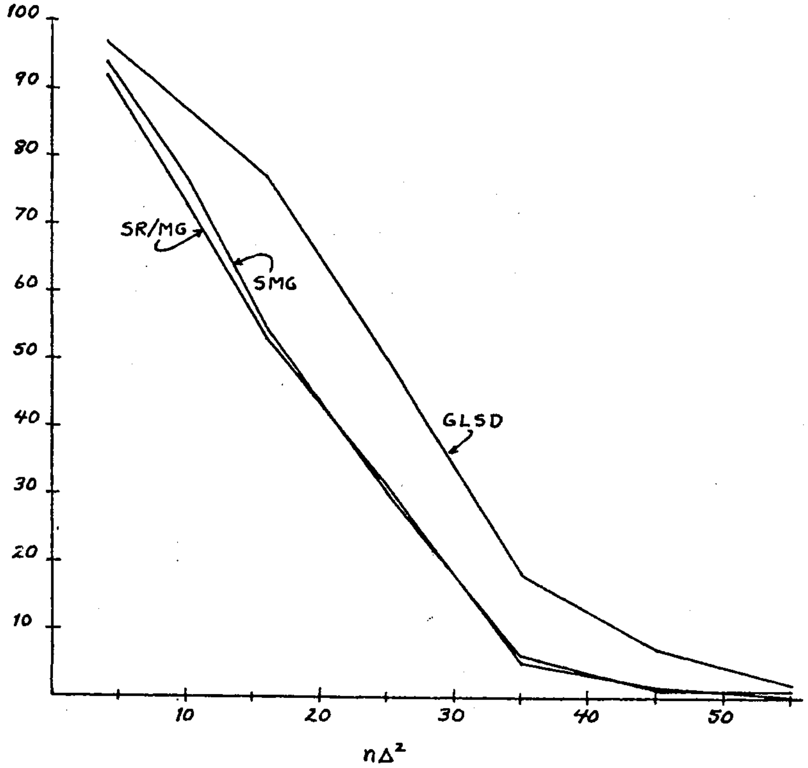
The loss associated with condition (t, i, j) could then be taken as

$$L(t,i,j) = c_1(t-j) + c_2(i-j), \text{ for } i = 0, 1, \ldots, 6 \text{ and } j = 0,1 \quad (3)$$

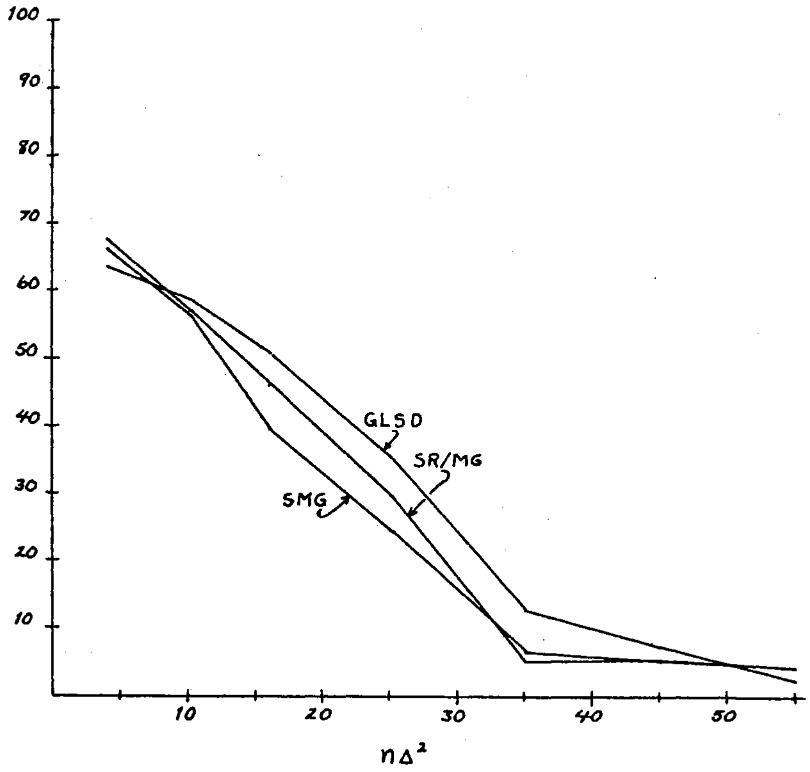
which is simply a weighted sum of the number of Type II and Type IV(EX) errors. The estimated average loss with respect to this loss function for a particular procedure at a specified value of $n\Delta^2$ would be

$$\widehat{E(L)} = \sum_{ij} L(t,i,j)\hat{p}_{ij} = \frac{1}{100}\sum_{ij}\hat{n}_{ij}[c_1(t-j) + c_2(i-j)] \quad (4)$$

where the $\hat{p}_{ij}$ or the $\hat{n}_{ij}$ are taken directly from Table XIII.



(a) $c_1 : c_2 = 100:0$



(b) $c_1 : c_2 = 65:35$

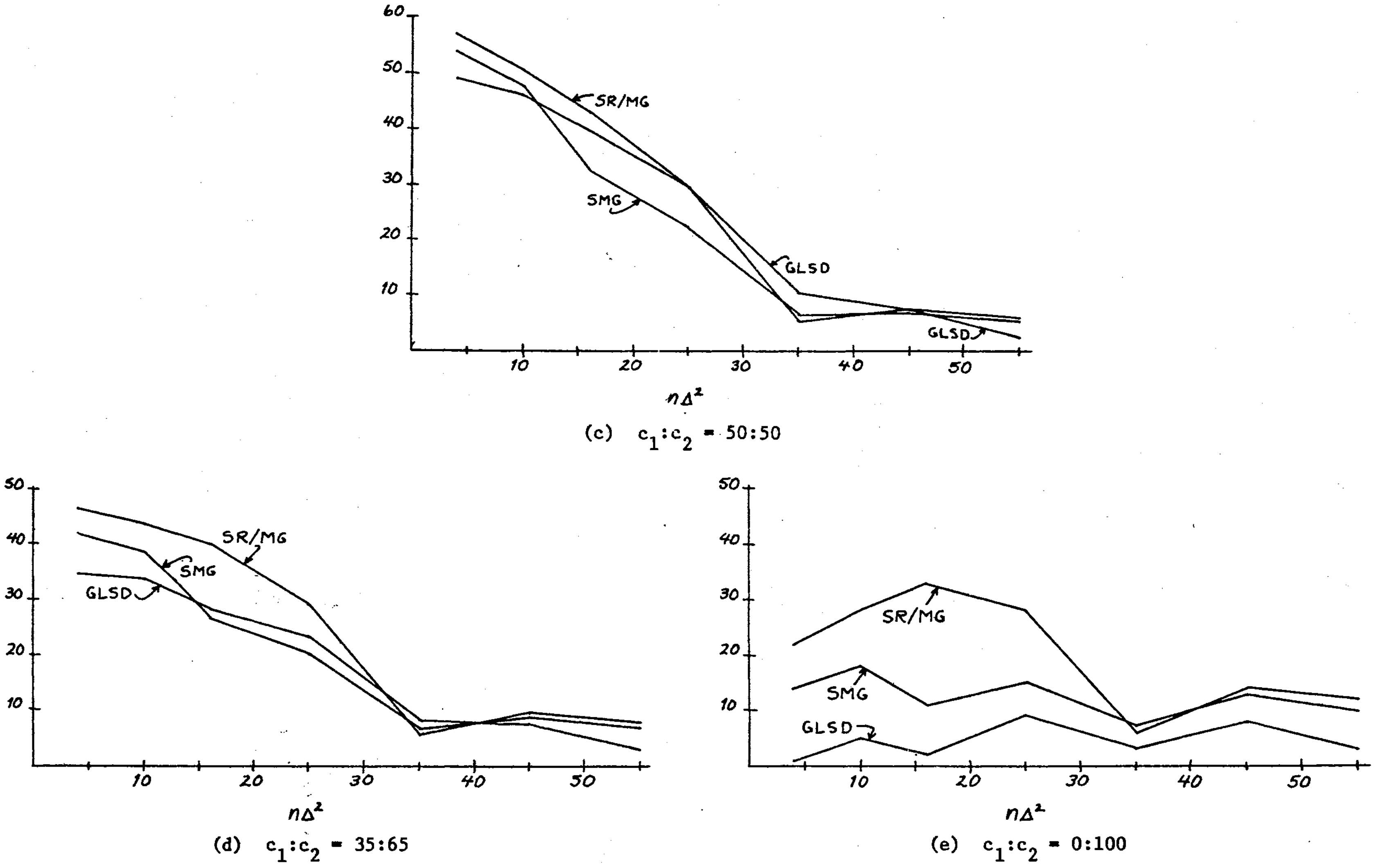


Figure 9. Comparison of Estimated Average Loss, Case 2

The estimated average loss for each of the three procedures GLSD, SR/MG and SMG are shown in Figure 9 as a function of $n\Delta^2$. The costs are varied so as to also study the effects of changing the relative seriousness of Type II and Type IV(EX) errors; five cases are taken: (a) $c_1:c_2$ = 100:0, (b) $c_1:c_2$ = 65:35, (c) $c_1:c_2$ = 50:50, (d) $c_1:c_2$ = 35:65, and (e) $c_1:c_2$ = 0:100. Thus, (a) represents loss from Type II errors only, and (e) represents loss from Type IV(EX) errors only.

The graphs in Figure 9 and table illustrate the (see Table , Appendix) essential features of the three procedures with respect to Type II and Type IV(EX) errors, given correct sample ranking. The GLSD is a less sensitive test than SR/MG or SMG, declaring fewer gaps significant in every case. The SR/MG, on the other hand, declares more significances than the other two, but has the greatest tendency to declare wrong gaps significant. The SMG is seen as a compromise of sorts; it declares more significances than the GLSD, but declares fewer wrong gaps than the SR/MG. Thus, the GLSD suffers most when Type II errors are heavily penalized (Figure 9(a)), while the SR/MG suffers most when Type IV(EX) errors are heavily penalized (Figure 9(e)). Across all conditions, the SMG yields the smallest average loss.

The performance of the three procedures with respect to probability of correct grouping may also be estimated from Table XIV with the aid of Bechhofer's ranking probabilities (3). That is, the values in the (2,2) cells of Table XIV are estimated probabilities of correct grouping, given correct sample ranking. Multiplying these by the probability of correct ranking taken from Bechhofer's tables yields the unconditional probability of correct grouping. A comparison of the GLSD, SR/MG, and SMG for the case under study is given in Figure 10 on the following page.

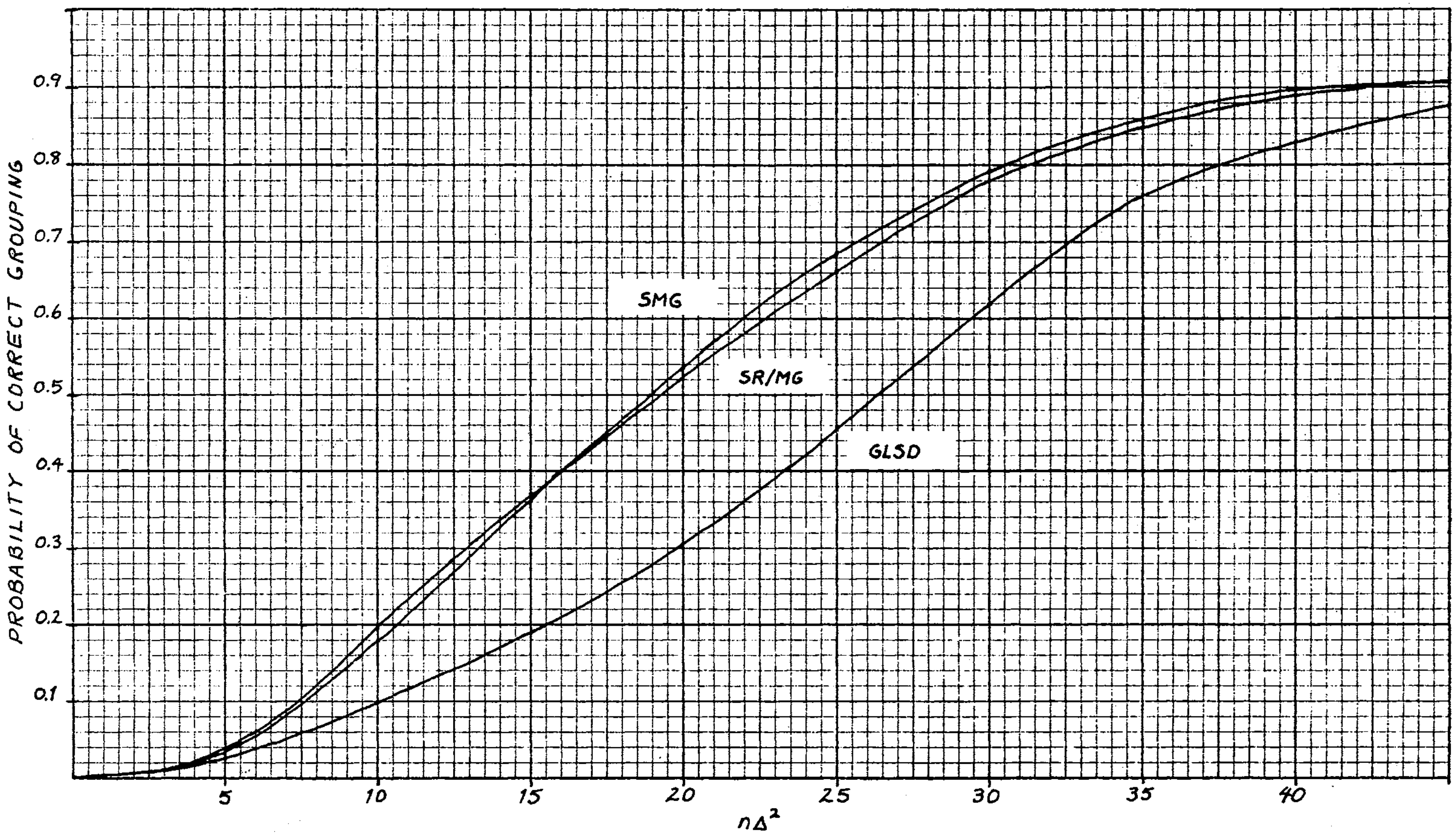


Figure 10. Comparison of Estimated Probability of Correct Grouping for the GLSD, SR/MG, and SMG Procedures, Case 2

The variation in the estimates precludes making any distinction between SMG and SR/MG, but it is possible to assert that both SMG and SR/MG have uniformly greater probability of correct grouping than GLSD for $n\Delta^2$ between 10 and 40. Some estimated standard deviations for the curves in Figure 10 for the three procedures are in the range of .03-.04 at $n\Delta^2 = 10$, .05 at $n\Delta^2 = 25$, and .03 - .038 at $n\Delta^2 = 40$.

In summary, as far as Case 2 is concerned, SMG is the preferred procedure because it apparently yields the least average loss the majority of the time, and it appears to have power greater than GLSD and at least as great as SR/MG. Whenever Type IV(EX) errors are not serious, the SR/MG would also be preferable to the GLSD.

5.3 Special Case 3, Group Size 7, Three True Gaps at Gaps 2, 4, and 6

This is a case where it might be expected that the SR/MG procedure would have an advantage, since more real gaps would tend to make the sample range larger. It will be seen that, although the SR/MG does tend to identify more of the true gaps than the other procedures, it does not appear to give a correct grouping any more frequently than the SMG because of the tendency to make Type IV(EX) errors.

For each of the seven values of the separation index, $\sqrt{n}\Delta = \sqrt{6}$, $\sqrt{10}$, 4, 5, 6, 7, and $\sqrt{65}$, samples of size 2, 2, 2, and 1 were taken from simulated normal (0,1), normal $(\sqrt{n}\Delta,1)$, normal $(2\sqrt{n}\Delta,1)$, and normal $(3\sqrt{n}\Delta,1)$ populations, respectively, until 100 groups of such samples were obtained for which the sample ranking was correct, that is, the sample configuration AABBCCD, where the A's are from the normal (0,1) distribution, etc. Note that as a simplification, all the true gaps are taken to be the same size. As with Case 2, for each correctly ranked group of samples drawn, an additional sample of size 10 was drawn from a simulated normal (0,1) population to obtain an independent estimate s^2 based on 9 degrees of freedom; $\alpha = .05$ is used throughout for all three procedures. For the correctly ranked samples, a cross-classification similar to Table XIII can be made, and is shown below in Table XV.

TABLE XV

BREAKDOWN OF CORRECTLY RANKED SAMPLES, TYPE II AND TYPE IV(EX) ERRORS, GROUP SIZE 7, THREE TRUE GAPS AT GAPS 2, 4, AND 6

Number of Gaps Declared Significant	Number of Gaps Correctly Declared			
	0	1	2	3
0	3 / 0			
1	3 / 1	2 / 0		
2	3 / 2	2 / 1	1 / 0	
3	3 / 3	2 / 2	1 / 1	0 / 0
4		2 / 3	1 / 2	0 / 1
5			1 / 3	0 / 2
6				0 / 3

* Cell Entries are: Number of Type II Errors / Number of Type IV(EX) Errors

TABLE XVI

RESULTS OF EMPIRICAL SAMPLING, GROUP SIZE 7, THREE TRUE GAPS AT GAPS 2, 4, 6

$n\Delta^2 = 6.0$

142 Total Samples
42 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	1	2	3	SR/MG Gaps Correct 0	1	2	3	SMG Gaps Correct 0	1	2	3
0	67	-	-	-	3	-	-	-	34	-	-	-
1	3	20	-	-	12	51	-	-	9	36	-	-
2	1	5	4	-	3	13	13	-	1	9	9	-
3	0	0	0	0	0	3	2	0	0	1	1	0
4	-	0	0	0	-	0	0	0	-	0	0	0
5	-	-	0	0	-	-	0	0	-	-	0	0
6	-	-	-	0	-	-	-	0	-	-	-	0

$n\Delta^2 = 10.0$

106 Total Samples
6 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	1	2	3	SR/MG Gaps Correct 0	1	2	3	SMG Gaps Correct 0	1	2	3
0	40	-	-	-	0	-	-	-	9	-	-	-
1	4	30	-	-	4	37	-	-	9	42	-	-
2	0	4	18	-	1	17	31	-	0	8	26	-
3	1	0	1	2	1	3	3	2	1	0	2	2
4	-	0	0	0	-	0	0	1	-	0	0	1
5	-	-	0	0	-	-	0	0	-	-	0	0
6	-	-	-	0	-	-	-	0	-	-	-	0

$n\Delta^2 = 16.0$

101 Total Samples
1 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	1	2	3	SR/MG Gaps Correct 0	1	2	3	SMG Gaps Correct 0	1	2	3
0	17	-	-	-	0	-	-	-	1	-	-	-
1	1	33	-	-	0	13	-	-	2	22	-	-
2	1	1	26	-	2	9	40	-	1	7	39	-
3	0	0	6	13	0	2	15	17	0	1	8	16
4	-	0	2	0	-	0	2	0	-	0	2	1
5	-	-	0	0	-	-	0	0	-	-	0	0
6	-	-	-	0	-	-	-	0	-	-	-	0

$n\Delta^2 = 25.0$

100 Total Samples
0 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	1	2	3	SR/MG Gaps Correct 0	1	2	3	SMG Gaps Correct 0	1	2	3
0	2	-	-	-	0	-	-	-	0	-	-	-
1	0	13	-	-	0	7	-	-	0	7	-	-
2	0	2	30	-	0	2	31	-	0	2	30	-
3	0	0	8	39	0	0	14	40	0	0	10	45
4	-	0	0	4	-	0	0	4	-	0	0	4
5	-	-	0	2	-	-	0	2	-	-	0	2
6	-	-	-	0	-	-	-	0	-	-	-	0

$n\Delta^2 = 36.0$

100 Total Samples
0 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	1	2	3	SR/MG Gaps Correct 0	1	2	3	SMG Gaps Correct 0	1	2	3
0	0	-	-	-	0	-	-	-	0	-	-	-
1	0	1	-	-	0	0	-	-	0	0	-	-
2	0	0	12	-	0	0	13	-	0	0	10	-
3	0	0	2	69	0	0	2	69	0	0	2	72
4	-	0	0	15	-	0	0	15	-	0	0	15
5	-	-	0	1	-	-	0	1	-	-	0	1
6	-	-	-	0	-	-	-	0	-	-	-	0

$n\Delta^2 = 49.0$

100 Total Samples
0 Incorrectly Ranked

Gaps Declared	GLSD Gaps Correct 0	1	2	3	SR/MG Gaps Correct 0	1	2	3	SMG Gaps Correct 0	1	2	3
0	0	-	-	-	0	-	-	-	0	-	-	-
1	0	0	-	-	0	0	-	-	0	0	-	-
2	0	0	1	-	0	0	4	-	0	0	1	-
3	0	0	0	80	0	0	0	77	0	0	0	80
4	-	0	0	18	-	0	0	18	-	0	0	18
5	-	-	0	1	-	-	0	1	-	-	0	1
6	-	-	-	0	-	-	-	0	-	-	-	0

$n\Delta^2 = 65.0$

100 Total Samples
0 Correctly Ranked

Gaps Declared	GLSD Gaps Correct 0	1	2	3	SR/MG Gaps Correct 0	1	2	3	SMG Gaps Correct 0	1	2	3
0	0	-	-	-	0	-	-	-	0	-	-	-
1	0	0	-	-	0	0	-	-	0	0	-	-
2	0	0	0	-	0	0	0	-	0	0	0	-
3	0	0	0	82	0	0	0	82	0	0	0	82
4	-	0	0	14	-	0	0	14	-	0	0	14
5	-	-	0	4	-	-	0	4	-	-	0	4
6	-	-	-	0	-	-	-	0	-	-	-	0

The results of empirical sampling are given in Table XVI for each of the procedures. To assess the performance of the procedures with respect to Type II and Type IV(EX) errors, a loss function of the form of (3) can be defined, and average loss may be estimated as in (4). Unfortunately, the comparison of average loss does not lend itself readily to graphical presentation in this instance, because all three procedures have very similar average loss curves. Instead, the results are summarized in Table XVII on the next page. The same trend as in Case 2 is observed here also, although not as pronounced. That is, the GLSD procedure has greatest average loss except when Type II errors are lightly penalized or not penalized at all. The SMG and SR/MG procedures have similar average loss, both less than GLSD for the most part, and SR/MG suffers most when Type IV(EX) errors are penalized. The overall average loss for SR/MG and SMG is about the same, however, it appears that for smaller values of $n\Delta^2$, the SMG average loss is generally greater than that of SR/MG, while for larger values of $n\Delta^2$, the average losses are about the same with that of SMG being, perhaps, slightly less than that of SR/MG.

The estimated probability of a correct grouping may be computed for the three procedures from Table XVI and from Table VIII, p. 71. To form the probability curves, an eyeball smoothing of the values in Table VIII and the (4,4) cells of Table XVI is made, and the curves "multiplied" together to yield unconditional probability of correct grouping. As with the average loss curves, the procedures are so nearly the same as to make graphical comparison of power characteristics impractical, hence the results are summarized in Table XVIII on page 96.

TABLE XVII

ESTIMATED AVERAGE LOSS FROM TYPE II AND TYPE IV(EX) WHEN THE SAMPLE RANKING IS CORRECT

	$c_1 = 100$, $c_2 = 0$			$c_1 = 65$, $c_2 = 35$		
$n\Delta^2$	GLSD	SR/MG	SMG	GLSD	SR/MG	SMG
6.0	264.0	200.0	231.0	175.1	143.3	158.2
10.0	220.0	165.0	183.0	147.2	119.8	127.0
16.0	157.0	111.0	121.0	106.9	84.7	87.7
25.0	74.0	63.0	58.0	54.4	49.3	44.7
36.0	16.0	15.0	12.0	16.7	16.0	14.1
49.0	1.0	4.0	1.0	7.7	9.6	7.7
65.0	0.0	0.0	0.0	7.7	7.7	7.7

	$c_1 = 50$, $c_2 = 50$			$c_1 = 35$, $c_2 = 65$		
$n\Delta^2$	GLSD	SR/MG	SMG	GLSD	SR/MG	SMG
6.0	137.0	119.0	127.0	98.9	94.7	95.8
10.0	116.0	100.5	103.0	84.8	81.1	79.0
16.0	85.5	73.5	73.5	64.0	62.2	59.2
25.0	46.0	43.5	39.0	37.6	37.6	33.3
36.0	17.0	16.5	15.0	17.3	16.9	15.9
49.0	10.5	12.0	10.5	13.3	14.4	13.3
65.0	11.0	11.0	11.0	14.3	14.3	14.3

	$c_1 = 0$, $c_2 = 100$		
$n\Delta^2$	GLSD	SR/MG	SMG
6.0	10.0	38.0	23.0
10.0	12.0	36.0	23.0
16.0	14.0	36.0	26.0
25.0	18.0	24.0	20.0
36.0	18.0	18.0	18.0
49.0	20.0	20.0	20.0
65.0	22.0	22.0	22.0

TABLE XVIII

COMPARISON OF ESTIMATED PROBABILITY OF CORRECT GROUPING FOR THE GLSD, SR/MG, AND SMG PROCEDURES

$n\Delta^2$	Probability of Correct Ranking	Conditional Probability of Correct Grouping			Unconditional Probability of Correct Grouping		
		GLSD	SR/MG	SMG	GLSD	SR/MG	SMG
2	.226	.000	.000	.000	.000	.000	.000
5	.580	.005	.005	.005	.003	.003	.003
10	.886	.020	.030	.030	.018	.027	.027
15	.975	.100	.130	.120	.098	.127	.117
20	.995	.230	.255	.270	.229	.254	.269
25	1.000				.380	.405	.440
30	1.000				.525	.540	.575
35	1.000				.660	.660	.710
40	1.000				.740	.740	.755
45	1.000				.770	.765	.780
50	1.000				.790	.775	.790
55	1.000				.805	.785	.805
60	1.000				.810	.800	.810

It is somewhat surprising to find that the SR/MG and SMG procedures do not appear to be much more powerful than GLSD for correct grouping in this case. A possible explanation for this phenomenon is that Case 3 deals with more true gaps in the group, and, hence, smaller true subgroups of means result, i.e., groups of 2, 2, 2, and 1. As was noted earlier, when smaller groups of means are being analyzed, the difference between Student's t and the SMG or SR criteria becomes less and less, until with group size 2, all three criteria are the same, except for a constant ($\sqrt{2}t$ = SR = SMG). Another surprising observation about the power characteristics of the procedures is that they all three seem to level off to around .80. Examination of Table XVI p. 93, shows that for

$n\Delta^2$ greater than 50, the errors are mostly Type IV(EX) errors, so it may be that a 19-20% Type IV error rate prevents any greater power being attained. Recall that Type IV error rates of 19-20% were observed in Case 1.

To sum up, Case 3 was thought beforehand to be a case where the SR/MG procedure would have a distinct advantage over the GLSD and SMG procedures, but no support for that conjecture has been found. The SMG and SR/MG procedures proved to be about the same with respect to average losses from Type II and Type IV(EX) errors and with respect to power characteristics for grouping with the GLSD not far behind in either category.

6. Brief Summary of the Chapter

The studies in this chapter were undertaken with the objectives of determining some of the performance characteristics of the SMG, as well as a comparison of the SMG and the competing procedures GLSD and SR/MG. An analytic treatment would have been desirable; however, the multiplicity of ways that a sample may depart from the null hypothesis appears to preclude such a study. Even when the null hypothesis is true, it was shown that systematic study of multistage procedures can be complicated. The discussion in Section 2 pointed out, however, that Type I error rates under the null hypothesis should not be considered a major problem. The special cases studied also indicate that Type III error rates should be de-emphasized. The unexpected result coming out of the case studies was that the SMG and SR/MG procedures are apparently more susceptible to Type IV errors than previously thought. Recall that Type IV errors occur when gaps are erroneously declared significant, but, perhaps a distinction should be made between (i) the situation where the true gap is not detected and another gap is erroneously declared significant in its place, and (ii) the situation where the true gap is detected but an additional gap is also declared significant. The two procedures mentioned are more prone toward the latter.

Before any sweeping generalities are made concerning the power characteristics of the three procedures studied, their performance should be systematically studied over a wider range of alternatives and sample sizes. The preliminary indications are, however, that both the SMG and SR/MG procedures are superior to the GLSD procedure with respect to grouping detection. From a performance standpoint, the results in this chapter indicate that the SMG is the preferred procedure, as it apparently tends to make fewer Type IV errors than the SR/MG without severe sacrifice of power for correct grouping.

CHAPTER V

SUMMARY, CONCLUSIONS, AND SUGGESTIONS FOR FURTHER INVESTIGATION

1. Summary

The common denominator of the majority of statistical procedures commonly referred to as multiple inference and/or multiple decision procedures is the fact that whenever the experimental objective is to determine the underlying grouping pattern in a set of observed means, none of them can guarantee an unambiguous answer. The reason is, of course, that they were designed for a different purpose, namely inferences about some or all of the contrasts among the means. The first procedure designed specifically toward grouping detection was put forward by Tukey (19) and has since become known as the Tukey Gap-Straggler-Variance Procedure. In developing the procedure, Tukey noted that departures from homogeneity in means could be partially characterized by large gaps between observed means.

The objective of this study has been to investigate some procedures which can be used to detect underlying grouping patterns in a set of observed means. The primary emphasis has been on the development of a procedure based on the distribution of the largest gap in a set of k ordered observations from a normal distribution. From a practical standpoint, this distribution can be calculated exactly only for three or four of the smallest values of k. For larger values, approximations to the distribution have been proposed. The approximations are of two types: empirical approximations, subject to sampling variation; and approximations by multiplication of probabilities of non-independent events. To eliminate the necessity of knowing σ when the observations are taken from normal distributions with standard deviation σ, the distribution of the maximum gap is "studentized" by calculating the distribution of the ratio (maximum gap)/s, where s^2 is an independent estimate of σ^2 based on ν degrees of freedom. This, of course, has the effect of replacing the unknown parameter σ^2 in the distribution with a set of parameters $\{\nu\}$. The situation is completely analogous to the relation between the normal distribution and Student's t-distribution (the original studentization problem).

Using the distributions calculated by the above methods, a sequential multi-staged procedure was devised, the studentized maximum gap procedure (SMG). Two other procedures, based on well-known distributions were also constructed along the same lines, the maximum gap LSD procedure (GLSD) and the studentized range/maximum gap procedure (SR/MG). In developing these procedures, it was noted that although some of the multiple inference procedures are referred to by some authors as simultaneous inference procedures, they are not simultaneous at all. In a significance-testing context, (or, perhaps more properly, a hypothesis-testing with variable α-level context) the procedures are carried out in practice as multiple decision procedures in which the decisions are made sequentially, not simultaneously. Thus, it was discovered that when testing is done this way, a single tentative null hypothesis that all means are the same will not suffice, for as soon as the decision that two means are not the same is made, the null hypothesis cannot still be that all means are the same if any further testing is to be done. We have referred to this phenomenon as the changing null hypothesis aspect and have constructed the SMG, GLSD, and SR/MG procedures in such a way as to incorporate it into their logic. The essential steps in each procedure can be described as follows:

(i) Rank the observed means.

(ii) Examine the group as a whole for departure from homogeneity.

(iii) If evidence of non-homogeneity is found, break the means into two groups.

(iv) Repeat steps (ii) and (iii) for each new group formed until no more evidence of non-homogeneity within the groups can be found.

Included in the study is an evaluation of the comparative performances of the GLSD, SR/MG, and SMG procedures with respect to power for grouping detection and certain error tendencies.

The evaluation was carried out by means of computer simulated sampling from normal distributions for which the true grouping pattern was known. Of necessity, the performance study was limited in scope, due to the fact that funds for computer time were limited and extensive tabulation of critical values for the SMG would have been unwise while its performance relative to the GLSD and SR/MG procedures remained in question. The latter two procedures, of course, have tabulated critical values which are readily available, so that if it had turned out that the SMG was clearly inferior, it would likely have been discarded in favor of some other procedure. The conclusions about the performance of the three procedures is therefore based on results obtained for three special cases. Since a general treatment of the procedure performances for all possible departures from the null hypothesis would be formidable, if not impossible, the case-study approach represents a reasonable and informative alternative.

2. Conclusions

Of the existing multiple inference procedures, the LSD, Studentized Range, and Multiple-F procedures can be adapted for use as statistical grouping procedures. The performance of an F/Maximum Gap procedure would be expected to be very similar to the SR/MG procedure studied. For the SMG procedure, although exact distributions cannot be obtained in general, reasonable approximations can be made (and improved upon) yielding a procedure adequate to most situations encountered in practice. This study has shown that viable solutions to the grouping problem exist. On the basis of the evaluations of the performances of the three procedures treated, it is concluded that the two competitors are the SMG and SR/MG procedures. If forced to a decision at this time, the SMG procedure would probably be chosen as the preferred procedure. However, the SR/MG has the distinct advantage of having readily accessible critical values extensively tabulated.

3. Further Study

A more extensive study of the performances of the SMG and SR/MG under alternative hypotheses should be made with a more systematic choice of alternative hypotheses. If such studies indicate that SMG is still the preferred procedure, then better and more efficient methods of approximating the null sampling distribution should be devised. The present study should be considered only a first-cut attempt in both respects. More accurate approximations should also help establish whether or not studentizing with $s_G = s\kappa_n$, rather than s only, will reverse the trend with respect to the critical values. It was conjectured in Chapter II that this may be the case, but it has not been established at this time.

The procedures have been formulated and studied under assumptions of normality, homogeneous variances, and equal sample sizes. Many cases in practice do not conform to these assumptions, and therefore, should not be ignored if workable solutions can be found.

It is the author's opinion that the sort of experimental objective which has been here referred to as the grouping detection problem is a very real problem encountered frequently in research and which has heretofore received too little attention. It has been shown that existing procedures are inadequate and that new procedures are needed to fill the void. The alternatives presented in this study represent a beginning.

A SELECTED BIBLIOGRAPHY

(1) Abramowitz, M. and Stegun, I. A., eds. Handbook of Mathematical Functions. New York: Dover, 1965.

(2) Balaam, L. N. "Multiple Comparisons--A Sampling Experiment." The Australian Journal of Statistics, Vol. 5 (1963), 62-84.

(3) Bechhofer, R. E. "A Single-sample Multiple Decision Procedure for Ranking Means of Normal Populations With Known Variances." Annals of Mathematical Statistics, Vol. 25 (1954), 16-39.

(4) Bechhofer, R. E., Dunnett, C. W., and Sobel, M. "A Two-sample Multiple Decision Procedure for Ranking Means of Normal Populations With a Common Unknown Variance." Biometrika, Vol. 41 (1954), 170-176.

(5) Cramer, S. G. and Swanson, M. R. "An Evaluation of Ten Pairwise Multiple Comparison Procedures by Monte Carlo Methods." Journal of the American Statistical Association, Vol. 68 (1973), 66-74.

(6) Duncan, D. B. "Multiple Range and Multiple F Tests." Biometrics, Vol. 11 (1955), 1-42.

(7) Duncan, D. B. "A Bayesian Approach to Multiple Comparisons." Technometrics, Vol. 7 (1965), 171-222.

(8) Fisher, R. A. The Design of Experiments, rev. ed. New York: Hafner, 1960.

(9) Hartley, H. O. "Studentization." Biometrika, Vol. 33 (1944), 173-180.

(10) Hogg, R. V. and Craig, A. T. Introduction to Mathematical Statistics, 2nd ed. New York: Macmillan, 1965.

(11) Keuls, M. "The Use of the Studentized Range in Connection With an Analysis of Variance." Euphytica, Vol. 1 (1952), 112-122.

(12) Marsaglia, G. and Bray, T. A. "A Convenient Method for Generating Normal Variables." S.I.A.M. Review, Vol. 6 (1964), 260-264.

(13) Miller, R. G. Simultaneous Statistical Inference. New York: McGraw-Hill, 1966.

(14) Newman, D. "The Distribution of the Range in Samples from a Normal Population Expressed in Terms of an Independent Estimate of the Standard Deviation." Biometrika, Vol. 31 (1939), 20-30.

(15) O'Neill, R. and Wetherill, G. B. "The Present State of Multiple Comparison Methods.", with discussion by S. D. Silvey, R. L. Plackett, J. A. Nelder, D. V. Lindley, I. T. Jolliffe, and M. R. Novick, Journal of the Royal Statistical Society, Vol. 33 (1971), 218-250.

(16) Ottestad, P. Statistical Models and Their Experimental Application, Griffin's Statistical Monographs and Courses, No. 25. London: 1970, 46-50.

(17) Snedecor, G. W. and Cochran, W. G. Statistical Methods, 6th ed. Ames, Iowa: Iowa State University Press, 1969, 271-275.

(18) Steele, R. G. D. and Torrie, J. H. Principles and Procedures of Statistics. New York: McGraw-Hill, 1960, 106-112.

(19) Tukey, J. W. "Comparing Individual Means in the Analysis of Variance." Biometrics, Vol. 5 (1949), 99-114.

(20) Tukey, J. W. "The Problem of Multiple Comparisons." (Unpublished Manuscript, Princeton, 1953).

APPENDIXES

TABLE XIX

EXPECTED VALUES AND VARIANCES OF GAPS

N	GAP	EXPECTED VALUE	VARIANCE	N	GAP	EXPECTED VALUE	VARIANCE
2	1	1.12838	0.72676045	15	1	0.48797	0.18390352
3	1	0.84628	0.45680940		2	0.30025	0.07501835
4	1	0.73237	0.36098516		3	0.23281	0.04653739
	2	0.59402	0.24902284		4	0.19918	0.03463794
5	1	0.66794	0.31039107		5	0.18040	0.02869533
	2	0.49502	0.18148172		6	0.17000	0.02562169
6	1	0.62545	0.27849883		7	0.16530	0.02428635
	2	0.44021	0.14781833	16	1	0.48125	0.17962754
	3	0.40310	0.12587947		2	0.29447	0.07245076
7	1	0.59481	0.25625259		3	0.22710	0.04446297
	2	0.40466	0.12748778		4	0.19316	0.03271478
	3	0.35271	0.09904873		5	0.17379	0.02675138
8	1	0.57138	0.23968339		6	0.16247	0.02352342
	2	0.37940	0.11377805		7	0.15646	0.02188758
	3	0.32031	0.08323812		8	0.15458	0.02138436
	4	0.30502	0.07602143	17	1	0.47516	0.17579073
9	1	0.55271	0.22676331		2	0.28932	0.07019007
	2	0.36033	0.10384679		3	0.22208	0.04266913
	3	0.29744	0.07278550		4	0.18792	0.03108319
	4	0.27453	0.06267738		5	0.16813	0.02513692
10	1	0.53739	0.21634215		6	0.15614	0.02182330
	2	0.34530	0.09628218		7	0.14920	0.02000042
	3	0.28030	0.06533766		8	0.14599	0.01918269
	4	0.25309	0.05397496	18	1	0.46962	0.17232269
	5	0.24534	0.05090987		2	0.28468	0.06818116
11	1	0.52452	0.20771545		3	0.21761	0.04110022
	2	0.33308	0.09030253		4	0.18333	0.02967971
	3	0.26686	0.05974436		5	0.16321	0.02377333
	4	0.23709	0.04784516		6	0.15074	0.02041691
	5	0.22489	0.04331355		7	0.14310	0.01847683
12	1	0.51386	0.20042622		8	0.13894	0.01745639
	2	0.32253	0.08543921		9	0.13760	0.01713711
	3	0.25600	0.05537699	19	1	0.46454	0.16916782
	4	0.22459	0.04328711		2	0.28049	0.06638151
	5	0.20966	0.03800908		3	0.21359	0.03971455
	6	0.20518	0.03648642		4	0.17925	0.02845817
13	1	0.50391	0.19416386		5	0.15890	0.02260520
	2	0.31425	0.08139360		6	0.14607	0.01923335
	3	0.24698	0.05186326		7	0.13790	0.01722061
	4	0.21452	0.03975897		8	0.13302	0.01606692
	5	0.19781	0.03409432		9	0.13072	0.01553783
	6	0.19052	0.03175169	20	1	0.45988	0.16628140
14	1	0.49548	0.18870938		2	0.27665	0.06475788
	2	0.30677	0.07796615		3	0.20997	0.03848029
	3	0.23937	0.04896850		4	0.17560	0.02738415
	4	0.20619	0.03694233		5	0.15508	0.02159238
	5	0.18827	0.03108419		6	0.14197	0.01822285
	6	0.17914	0.02827829		7	0.13340	0.01616664
	7	0.17632	0.02743644		8	0.12797	0.01492449
					9	0.12496	0.01425397
					10	0.12400	0.01404163

TABLE XX

CORRELATIONS OF GAPS

N	I	J	VALUE	N	I	J	VALUE	N	I	J	VALUE
3	1	2	-0.1361849	10	2	3	-0.0791782	12	3	7	-0.0322475
4		2	-0.1231622			4	-0.0568594			8	-0.0257677
		3	-0.0681794			5	-0.0425047			9	-0.0204548
5		2	-0.1133406			6	-0.0323949		4	5	-0.0538093
		3	-0.0679069			7	-0.0247623			6	-0.0504139
		4	-0.0406390			8	-0.0186011			7	-0.0403196
	2	3	-0.1096829		3	4	-0.0736256			8	-0.0323472
6	1	2	-0.1060835			5	-0.0555529		5	6	-0.0622217
		3	-0.0657781			6	-0.0426471			7	-0.0500295
		4	-0.0428443			7	-0.0327936	13	1	2	-0.0824982
		5	-0.0269995		4	5	-0.0711908			3	-0.0543442
	2	3	-0.0998312			6	-0.0550910			4	-0.0392223
		4	-0.0666367	11	1	2	-0.0867541			5	-0.0297770
7	1	2	-0.1005064			3	-0.0567240			6	-0.0233066
		3	-0.0635514			4	-0.0405002			7	-0.0185803
		4	-0.0429828			5	-0.0302047			8	-0.0149537
		5	-0.0295039			6	-0.0233857			9	-0.0120525
		6	-0.0193011			7	-0.0180906			10	-0.0096349
	2	3	-0.0926231			8	-0.0140350			11	-0.0075175
		4	-0.0638961			9	-0.0106709			12	-0.0054852
		5	-0.0445219			10	-0.0075997		2	3	-0.0714552
	3	4	-0.0901780		2	3	-0.0761949			4	-0.0521293
8	1	2	-0.0960636			4	-0.0550781			5	-0.0398709
		3	-0.0615200			5	-0.0417381			6	-0.0313790
		4	-0.0425276			6	-0.0319831			7	-0.0251227
		5	-0.0303351			7	-0.0251064			8	-0.0202892
		6	-0.0215994			8	-0.0195615			9	-0.0164003
		7	-0.0145392			9	-0.0149288			10	-0.0131437
	2	3	-0.0871151		3	4	-0.0701922			11	-0.0102789
		4	-0.0612459			5	-0.0534262		3	4	-0.0648929
		5	-0.0442302			6	-0.0415642			5	-0.0499720
		6	-0.0318059			7	-0.0326308			6	-0.0395308
	3	4	-0.0831705			8	-0.0255337			7	-0.0317773
		5	-0.0608509		4	5	-0.0671192			8	-0.0257486
9	1	2	-0.0924211			6	-0.0525722			9	-0.0208717
		3	-0.0597200			7	-0.0415027			10	-0.0167686
		4	-0.0418816		5	6	-0.0661702		4	5	-0.0610532
		5	-0.0305530	12	1	2	-0.0844886			6	-0.0485518
		6	-0.0226039			3	-0.0554696			7	-0.0391937
		7	-0.0165399			4	-0.0398425			8	-0.0318689
		8	-0.0113883			5	-0.0300584			9	-0.0259101
	2	3	-0.0827465			6	-0.0233366		5	6	-0.0590179
		4	-0.0589013			7	-0.0184060			7	-0.0478622
		5	-0.0434308			8	-0.0145973			8	-0.0390677
		6	-0.0323981			9	-0.0115130		6	7	-0.0583785
		7	-0.0238703			10	-0.0088774	14	1	2	-0.0807309
	3	4	-0.0778400			11	-0.0064052			3	-0.0533280
		5	-0.0580182		2	3	-0.0736535			4	-0.0386410
		6	-0.0436527			4	-0.0535138			5	-0.0294845
	4	5	-0.0763550			5	-0.0406955			6	-0.0232252
10	1	2	-0.0893654			6	-0.0317829			7	-0.0186659
		3	-0.0581313			7	-0.0251847			8	-0.0151827
		4	-0.0411873			8	-0.0200493			9	-0.0124158
		5	-0.0304975			9	-0.0158644			10	-0.0101394
		6	-0.0230760			10	-0.0122684			11	-0.0081973
		7	-0.0175366		3	4	-0.0673280			12	-0.0064607
		8	-0.0131075			5	-0.0515829			13	-0.0047603
		9	-0.0091928			6	-0.0405138		2	3	-0.0695292

TABLE XX (Continued)

N	I	J	VALUE	N	I	J	VALUE	N	I	J	VALUE
14	2	4	-0.0508943	15	3	9	-0.0210776	16	4	7	-0.0363234
		5	-0.0391062			10	-0.0175317			8	-0.0302108
		6	-0.0309628			11	-0.0145261			9	-0.0253296
		7	-0.0249836			12	-0.0118945			10	-0.0213193
		8	-0.0203864		4	5	-0.0566967			11	-0.0179383
		9	-0.0167154			6	-0.0455044			12	-0.0150127
		10	-0.0136817			7	-0.0371966		5	6	-0.0521522
		11	-0.0110834			8	-0.0307610			7	-0.0429620
		12	-0.0087517			9	-0.0256008			8	-0.0358290
	3	4	-0.0627905			10	-0.0213371			9	-0.0301084
		5	-0.0485517			11	-0.0177108			10	-0.0253910
		6	-0.0386230		5	6	-0.0540966			11	-0.0214009
		7	-0.0312798			7	-0.0443813		6	7	-0.0504928
		8	-0.0256007			8	-0.0368127			8	-0.0422331
		9	-0.0210436			9	-0.0307150			9	-0.0355780
		10	-0.0172619			10	-0.0256558			10	-0.0300681
		11	-0.0140109		6	7	-0.0526810		7	8	-0.0497143
	4	5	-0.0587138			8	-0.0438410			9	-0.0419964
		6	-0.0469301			9	-0.0366819	17	1	2	-0.0764171
		7	-0.0381516		7	8	-0.0522306			3	-0.0508906
		8	-0.0313220	16	1	2	-0.0777172			4	-0.0369926
		9	-0.0258141			3	-0.0515612			5	-0.0286224
		10	-0.0212235			4	-0.0375896			6	-0.0228396
	5	6	-0.0563544			5	-0.0289028			7	-0.0186475
		7	-0.0459992			6	-0.0229814			8	-0.0154657
		8	-0.0378920			7	-0.0186837			9	-0.0129629
		9	-0.0313186			8	-0.0154168			10	-0.0109355
	6	7	-0.0552654			9	-0.0128414			11	-0.0092506
		8	-0.0456967			10	-0.0107484			12	-0.0078167
15	1	2	-0.0791473			11	-0.0090003			13	-0.0065658
		3	-0.0524047			12	-0.0075001			14	-0.0054426
		4	-0.0380975			13	-0.0061725			15	-0.0043899
		5	-0.0291912			14	-0.0049451			16	-0.0033101
		6	-0.0231126			15	-0.0037033		2	3	-0.0647613
		7	-0.0186940		2	3	-0.0662991			4	-0.0480604
		8	-0.0153282			4	-0.0487810			5	-0.0371376
		9	-0.0126666			5	-0.0377442			6	-0.0297642
		10	-0.0104932			6	-0.0301494			7	-0.0243824
		11	-0.0086636			7	-0.0245977			8	-0.0202759
		12	-0.0070715			8	-0.0203537			9	-0.0170315
		13	-0.0056222			9	-0.0169926			10	-0.0143939
		14	-0.0041783			10	-0.0142505			11	-0.0121952
	2	3	-0.0678235			11	-0.0119527			12	-0.0103188
		4	-0.0497847			12	-0.0099752			13	-0.0086782
		5	-0.0383988			13	-0.0082205			14	-0.0072017
		6	-0.0305502			14	-0.0065944			15	-0.0058151
		7	-0.0248020		3	4	-0.0593259		3	4	-0.0578751
		8	-0.0203971			5	-0.0461577			5	-0.0451369
		9	-0.0168968			6	-0.0370213			6	-0.0363153
		10	-0.0140267			7	-0.0303005			7	-0.0298381
		11	-0.0116020			8	-0.0251367			8	-0.0248721
		12	-0.0094854			9	-0.0210302			9	-0.0209335
		13	-0.0075531			10	-0.0176682			10	-0.0177211
	3	4	-0.0609517			11	-0.0148425			11	-0.0150356
		5	-0.0472890			12	-0.0124040			12	-0.0127383
		6	-0.0377884			13	-0.0102352			13	-0.0107252
		7	-0.0307831		4	5	-0.0549349			14	-0.0089098
		8	-0.0253857			6	-0.0442397		4	5	-0.0533791

TABLE XX (Continued)

N	I	J	VALUE	N	I	J	VALUE	N	I	J	VALUE
17	4	6	-0.0431089	18	3	8	-0.0246027	19	2	6	-0.0290457
		7	-0.0355245			9	-0.0208070			7	-0.0239450
		8	-0.0296826			10	-0.0177194			8	-0.0200628
		9	-0.0250316			11	-0.0151481			9	-0.0170056
		10	-0.0212258			12	-0.0129608			10	-0.0145309
		11	-0.0180354			13	-0.0110611			11	-0.0124807
		12	-0.0152994			14	-0.0093740			12	-0.0107471
		13	-0.0128965			15	-0.0078341			13	-0.0092529
	5	6	-0.0504557		4	5	-0.0519923			14	-0.0078630
		7	-0.0417053			6	-0.0420905			15	-0.0068319
		8	-0.0349334			7	-0.0347916			16	-0.0056780
		9	-0.0295206			8	-0.0291809			17	-0.0046362
		10	-0.0250764			9	-0.0247242		3	4	-0.0553879
		11	-0.0213400			10	-0.0210879			5	-0.0433641
		12	-0.0181274			11	-0.0180519			6	-0.0350612
	6	7	-0.0486105			12	-0.0154634			7	-0.0289817
		8	-0.0408252			13	-0.0132107			8	-0.0243348
		9	-0.0345763			14	-0.0112065			9	-0.0206627
		10	-0.0294272		5	6	-0.0489591			10	-0.0176816
		11	-0.0250845			7	-0.0405834			11	-0.0152059
	7	8	-0.0475862			8	-0.0341164			12	-0.0131080
		9	-0.0404012			9	-0.0289609			13	-0.0112965
		10	-0.0344574			10	-0.0247414			14	-0.0098026
	8	9	-0.0472574			11	-0.0212091			15	-0.0081810
18	1	2	-0.0752279			12	-0.0181902			16	-0.0069487
		3	-0.0500723			13	-0.0155574		4	5	-0.0507462
		4	-0.0366691		6	7	-0.0469696			6	-0.0411673
		5	-0.0283517			8	-0.0395805			7	-0.0341175
		6	-0.0226923			9	-0.0336673			8	-0.0287068
		7	-0.0185935			10	-0.0288118			9	-0.0244168
		8	-0.0154861			11	-0.0247355			10	-0.0209244
		9	-0.0130459			12	-0.0212429			11	-0.0180169
		10	-0.0110737		7	8	-0.0457572			12	-0.0155480
		11	-0.0094403			9	-0.0390067			13	-0.0134122
		12	-0.0080575			10	-0.0334443			14	-0.0115296
		13	-0.0068615			11	-0.0287600			15	-0.0098356
		14	-0.0058035		8	9	-0.0451815		5	6	-0.0476263
		15	-0.0048412			10	-0.0388199			7	-0.0395744
		16	-0.0039284	19	1	2	-0.0741343			8	-0.0333691
		17	-0.0029806			3	-0.0494104			9	-0.0284324
	2	3	-0.0636799			4	-0.0362512			10	-0.0244020
		4	-0.0470321			5	-0.0280913			11	-0.0210383
		5	-0.0365741			6	-0.0225429			12	-0.0181758
		6	-0.0293962			7	-0.0185274			13	-0.0156947
		7	-0.0241634			8	-0.0154860			14	-0.0135041
		8	-0.0201762			9	-0.0131005		6	7	-0.0455233
		9	-0.0170319			10	-0.0111758			8	-0.0384709
		10	-0.0144820			11	-0.0095856			9	-0.0328403
		11	-0.0123640			12	-0.0082442			10	-0.0282295
		12	-0.0105663			13	-0.0070905			11	-0.0243715
		13	-0.0090082			14	-0.0060790			12	-0.0210809
		14	-0.0076270			15	-0.0051730			13	-0.0182229
		15	-0.0063685			16	-0.0043396		7	8	-0.0441642
		16	-0.0051725			17	-0.0035405			9	-0.0377756
	3	4	-0.0565701			18	-0.0027014			10	-0.0325274
		5	-0.0442102		2	3	-0.0625430			11	-0.0281238
		6	-0.0356639			4	-0.0462639			12	-0.0243584
		7	-0.0293983			5	-0.0360494		8	9	-0.0433989

TABLE XX (Continued)

N	I	J	VALUE	N	I	J	VALUE
19	8	10	-0.0374395	20	4	12	-0.0155795
		11	-0.0324240			13	-0.0135385
	9	10	-0.0431516			14	-0.0117505
20	1	2	-0.0731241			15	-0.0101568
		3	-0.0487950			16	-0.0087081
		4	-0.0358581		5	6	-0.0464297
		5	-0.0278414			7	-0.0386609
		6	-0.0223935			8	-0.0326832
		7	-0.0184531			9	-0.0279355
		8	-0.0154708			10	-0.0240667
		9	-0.0131338			11	-0.0208456
		10	-0.0112507			12	-0.0181130
		11	-0.0096977			13	-0.0157546
		12	-0.0083910			14	-0.0136853
		13	-0.0072714			15	-0.0118382
		14	-0.0062954		6	7	-0.0442363
		15	-0.0054292			8	-0.0374740
		16	-0.0046450			9	-0.0320854
		17	-0.0039164			10	-0.0276822
		18	-0.0032108			11	-0.0240074
		19	-0.0024625			12	-0.0208833
	2	3	-0.0614997			13	-0.0181819
		4	-0.0455544			14	-0.0158078
		5	-0.0355597		7	8	-0.0427611
		6	-0.0287125			9	-0.0366793
		7	-0.0237299			10	-0.0316950
		8	-0.0199410			11	-0.0275247
		9	-0.0169606			12	-0.0239713
		10	-0.0145516			13	-0.0208927
		11	-0.0125596		8	9	-0.0418475
		12	-0.0108797			10	-0.0362223
		13	-0.0094375			11	-0.0315027
		14	-0.0081780			12	-0.0274715
		15	-0.0070585		9	10	-0.0414099
		16	-0.0060435			11	-0.0360732
		17	-0.0050991				
		18	-0.0041834				
	3	4	-0.0543104				
		5	-0.0425876				
		6	-0.0345018				
		7	-0.0285876				
		8	-0.0240720				
		9	-0.0205082				
		10	-0.0176196				
		11	-0.0152256				
		12	-0.0132026				
		13	-0.0114628				
		14	-0.0099410				
		15	-0.0085865				
		16	-0.0073567				
		17	-0.0062111				
	4	5	-0.0496185				
		6	-0.0403256				
		7	-0.0334953				
		8	-0.0282600				
		9	-0.0241151				
		10	-0.0207466				
		11	-0.0179485				

TABLE XXI

EMPIRICAL DISTRIBUTION OF THE MAXIMUM GAP, SAMPLE SIZE 5

INTERVAL	OBSERVED FREQUENCY	SMOOTHED FREQUENCY	EMPIRICAL CUMULATIVE	SMOOTHED CUMULATIVE
0.0 TO 0.1	9	9.38	0.000257	0.000268
0.1 TO 0.2	73	71.50	0.002343	0.002311
0.2 TO 0.3	292	294.25	0.010686	0.010718
0.3 TO 0.4	642	668.12	0.029029	0.029806
0.4 TO 0.5	1186	1157.09	0.062914	0.062866
0.5 TO 0.6	1676	1689.63	0.110800	0.111140
0.6 TO 0.7	2188	2205.65	0.173314	0.174158
0.7 TO 0.8	2664	2620.52	0.249429	0.249028
0.8 TO 0.9	2860	2862.57	0.331143	0.330815
0.9 TO 1.0	2927	2943.27	0.414771	0.414907
1.0 TO 1.1	2915	2913.22	0.498057	0.498140
1.1 TO 1.2	2789	2776.79	0.577743	0.577476
1.2 TO 1.3	2539	2533.15	0.650286	0.649851
1.3 TO 1.4	2197	2213.59	0.713057	0.713095
1.4 TO 1.5	1915	1899.84	0.767771	0.767375
1.5 TO 1.6	1576	1616.46	0.812800	0.813559
1.6 TO 1.7	1414	1368.64	0.853200	0.852663
1.7 TO 1.8	1100	1115.28	0.884629	0.884527
1.8 TO 1.9	878	889.16	0.909714	0.909931
1.9 TO 2.0	706	697.23	0.929886	0.929852
2.0 TO 2.1	553	547.77	0.945686	0.945502
2.1 TO 2.2	410	425.62	0.957400	0.957663
2.2 TO 2.3	348	338.03	0.967343	0.967320
2.3 TO 2.4	276	274.08	0.975229	0.975151
2.4 TO 2.5	207	217.44	0.981143	0.981364
2.5 TO 2.6	186	166.56	0.986457	0.986122
2.6 TO 2.7	103	121.68	0.989400	0.989599
2.7 TO 2.8	99	94.30	0.992229	0.992293
2.8 TO 2.9	78	71.73	0.994457	0.994342
2.9 TO 3.0	48	50.00	0.995829	0.995771
3.0 TO 3.1	26	31.14	0.996571	0.996660
3.1 TO 3.2	29	27.15	0.997400	0.997436
3.2 TO 3.3	29	26.47	0.998229	0.998192
3.3 TO 3.4	22	21.71	0.998857	0.998813
3.4 TO 3.5	11	13.77	0.999171	0.999206
3.5 TO 3.6	11	8.35	0.999486	0.999445
3.6 TO 3.7	3	5.03	0.999571	0.999588
3.7 TO 3.8	4	3.59	0.999686	0.999691
3.8 TO 3.9	4	2.58	0.999800	0.999765
3.9 TO 4.0	0	1.91	0.999800	0.999819
4.0 TO 4.1	3	2.00	0.999886	0.999876
4.1 TO 4.2	2	2.14	0.999943	0.999937
4.2 TO 4.3	2	1.64	1.000000	0.999984
4.3 TO 4.4	0	0.50	1.000000	0.999998
4.4 TO 4.5	0	0.06	1.000000	1.000000
4.5 TO 4.6	0	0.0	1.000000	1.000000
4.6 TO 4.7	0	0.0	1.000000	1.000000
4.7 TO 4.8	0	0.0	1.000000	1.000000
4.8 TO 4.9	0	0.0	1.000000	1.000000
4.9 TO 5.0	0	0.0	1.000000	1.000000

TOTAL: 35000

TABLE XXII

EMPIRICAL DISTRIBUTION OF THE MAXIMUM GAP, SAMPLE SIZE 6

INTERVAL	OBSERVED FREQUENCY	SMOOTHED FREQUENCY	EMPIRICAL CUMULATIVE	SMOOTHED CUMULATIVE
0.0 TO 0.1	2	4.01	0.000080	0.000160
0.1 TO 0.2	23	14.97	0.001000	0.000759
0.2 TO 0.3	117	129.04	0.005680	0.005919
0.3 TO 0.4	380	393.20	0.020880	0.021643
0.4 TO 0.5	795	804.32	0.052680	0.053807
0.5 TO 0.6	1313	1295.07	0.105200	0.105595
0.6 TO 0.7	1757	1773.74	0.175480	0.176526
0.7 TO 0.8	2153	2141.87	0.261600	0.262177
0.8 TO 0.9	2374	2324.57	0.356560	0.355134
0.9 TO 1.0	2279	2329.25	0.447720	0.448279
1.0 TO 1.1	2238	2236.32	0.537240	0.537707
1.1 TO 1.2	2093	2059.65	0.620960	0.620070
1.2 TO 1.3	1800	1819.92	0.692960	0.692847
1.3 TO 1.4	1516	1535.77	0.753600	0.754261
1.4 TO 1.5	1327	1274.34	0.806680	0.805220
1.5 TO 1.6	964	1037.94	0.845240	0.846726
1.6 TO 1.7	908	860.54	0.881560	0.881138
1.7 TO 1.8	692	697.86	0.909240	0.909045
1.8 TO 1.9	540	550.97	0.930840	0.931078
1.9 TO 2.0	427	415.80	0.947920	0.947705
2.0 TO 2.1	307	316.16	0.960200	0.960348
2.1 TO 2.2	244	239.85	0.969960	0.969940
2.2 TO 2.3	184	185.59	0.977320	0.977361
2.3 TO 2.4	148	148.76	0.983240	0.983310
2.4 TO 2.5	120	117.07	0.988040	0.987992
2.5 TO 2.6	89	87.49	0.991600	0.991490
2.6 TO 2.7	57	61.54	0.993880	0.993951
2.7 TO 2.8	45	43.76	0.995680	0.995701
2.8 TO 2.9	34	30.32	0.997040	0.996914
2.9 TO 3.0	16	20.06	0.997680	0.997716
3.0 TO 3.1	14	14.72	0.998240	0.998305
3.1 TO 3.2	18	13.68	0.998960	0.998852
3.2 TO 3.3	7	10.80	0.999240	0.999284
3.3 TO 3.4	10	6.92	0.999640	0.999561
3.4 TO 3.5	1	3.09	0.999680	0.999684
3.5 TO 3.6	2	1.77	0.999760	0.999755
3.6 TO 3.7	2	1.36	0.999840	0.999810
3.7 TO 3.8	1	1.67	0.999880	0.999876
3.8 TO 3.9	2	1.19	0.999960	0.999924
3.9 TO 4.0	0	0.54	0.999960	0.999945
4.0 TO 4.1	0	0.11	0.999960	0.999950
4.1 TO 4.2	0	0.0	0.999960	0.999950
4.2 TO 4.3	0	0.0	0.999960	0.999950
4.3 TO 4.4	0	0.08	0.999960	0.999953
4.4 TO 4.5	0	0.30	0.999960	0.999965
4.5 TO 4.6	1	0.47	1.000000	0.999984
4.6 TO 4.7	0	0.30	1.000000	0.999996
4.7 TO 4.8	0	0.10	1.000000	0.999999
4.8 TO 4.9	0	0.0	1.000000	0.999999
4.9 TO 5.0	0	0.02	1.000000	1.000000

TOTAL: 25000

TABLE XXIII

EMPIRICAL DISTRIBUTION OF THE MAXIMUM GAP, SAMPLE SIZE 7

INTERVAL	OBSERVED FREQUENCY	SMOOTHED FREQUENCY	EMPIRICAL CUMULATIVE	SMOOTHED CUMULATIVE
0.0 TO 0.1	0	3.56	0.0	0.000102
0.1 TO 0.2	16	1.75	0.000457	0.000152
0.2 TO 0.3	120	141.37	0.003886	0.004189
0.3 TO 0.4	495	533.93	0.018029	0.019436
0.4 TO 0.5	1200	1199.94	0.052314	0.053702
0.5 TO 0.6	2022	2017.20	0.110086	0.111305
0.6 TO 0.7	2771	2767.78	0.189257	0.190343
0.7 TO 0.8	3354	3304.36	0.285086	0.284703
0.8 TO 0.9	3492	3532.12	0.384857	0.385568
0.9 TO 1.0	3554	3500.16	0.486400	0.485519
1.0 TO 1.1	3204	3252.96	0.577943	0.578412
1.1 TO 1.2	2921	2890.06	0.661400	0.660941
1.2 TO 1.3	2463	2477.75	0.731771	0.731696
1.3 TO 1.4	2034	2068.64	0.789886	0.790769
1.4 TO 1.5	1754	1677.69	0.840000	0.838678
1.5 TO 1.6	1227	1305.49	0.875057	0.875958
1.6 TO 1.7	1049	1012.28	0.905029	0.904864
1.7 TO 1.8	773	780.95	0.927114	0.927165
1.8 TO 1.9	608	613.66	0.944486	0.944689
1.9 TO 2.0	476	467.62	0.958086	0.958043
2.0 TO 2.1	359	355.01	0.968343	0.968181
2.1 TO 2.2	253	266.53	0.975571	0.975792
2.2 TO 2.3	216	212.16	0.981743	0.981850
2.3 TO 2.4	183	173.23	0.986971	0.986797
2.4 TO 2.5	125	133.27	0.990543	0.990603
2.5 TO 2.6	101	97.94	0.993429	0.993400
2.6 TO 2.7	70	71.65	0.995429	0.995446
2.7 TO 2.8	54	52.43	0.996971	0.996943
2.8 TO 2.9	35	35.24	0.997971	0.997949
2.9 TO 3.0	22	23.00	0.998600	0.998606
3.0 TO 3.1	15	14.09	0.999029	0.999008
3.1 TO 3.2	9	8.62	0.999286	0.999255
3.2 TO 3.3	4	6.00	0.999400	0.999426
3.3 TO 3.4	7	5.35	0.999600	0.999579
3.4 TO 3.5	5	4.08	0.999743	0.999695
3.5 TO 3.6	0	2.00	0.999743	0.999752
3.6 TO 3.7	2	0.78	0.999800	0.999775
3.7 TO 3.8	1	1.24	0.999829	0.999810
3.8 TO 3.9	1	2.24	0.999857	0.999874
3.9 TO 4.0	5	2.58	1.000000	0.999948
4.0 TO 4.1	0	1.51	1.000000	0.999991
4.1 TO 4.2	0	0.33	1.000000	1.000000
4.2 TO 4.3	0	0.0	1.000000	1.000000
4.3 TO 4.4	0	0.0	1.000000	1.000000
4.4 TO 4.5	0	0.0	1.000000	1.000000
4.5 TO 4.6	0	0.0	1.000000	1.000000
4.6 TO 4.7	0	0.0	1.000000	1.000000
4.7 TO 4.8	0	0.0	1.000000	1.000000
4.8 TO 4.9	0	0.0	1.000000	1.000000
4.9 TO 5.0	0	0.0	1.000000	1.000000

TOTAL: 35000

TABLE XXIV

MEANS, VARIANCES, AND STANDARD DEVIATIONS OF THE MAXIMUM GAP

Sample Size	Mean	Variance	Standard Deviation
* 2	1.128	0.7262	0.8522
3	1.239	0.4648	0.6817
4	1.218	0.3344	0.5783
5	1.174	0.2688	0.5185
6	1.130	0.2268	0.4762
7	1.087	0.2038	0.4514
8	1.025	0.1945	0.4410
9	0.995	0.1810	0.4254
10	0.969	0.1709	0.4134
11	0.945	0.1631	0.4039
12	0.925	0.1570	0.3962
13	0.906	0.1520	0.3898
14	0.890	0.1479	0.3845
15	0.875	0.1444	0.3800
16	0.861	0.1414	0.3761
17	0.849	0.1389	0.3727
18	0.838	0.1367	0.3697
19	0.827	0.1347	0.3670
20	0.818	0.1328	0.3645

* Taken from Standard Tables

TABLE XXV

PROBABILITY INTEGRAL OF THE MAXIMUM GAP

GAP SIZE	SAMPLE SIZE 3	4	5	6	7	8	9	10	11
0.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1	0.0055	0.0008	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
0.2	0.0218	0.0059	0.0023	0.0008	0.0002	0.0008	0.0006	0.0005	0.0004
0.3	0.0483	0.0193	0.0107	0.0059	0.0042	0.0075	0.0066	0.0062	0.0059
0.4	0.0841	0.0435	0.0298	0.0216	0.0194	0.0297	0.0294	0.0300	0.0312
0.5	0.1279	0.0798	0.0629	0.0538	0.0537	0.0762	0.0795	0.0844	0.0905
0.6	0.1783	0.1280	0.1111	0.1056	0.1113	0.1484	0.1586	0.1708	0.1843
0.7	0.2338	0.1867	0.1742	0.1765	0.1903	0.2409	0.2594	0.2793	0.2999
0.8	0.2927	0.2534	0.2490	0.2622	0.2847	0.3442	0.3699	0.3958	0.4209
0.9	0.3535	0.3253	0.3308	0.3551	0.3856	0.4489	0.4791	0.5079	0.5344
1.0	0.4147	0.3994	0.4149	0.4483	0.4855	0.5476	0.5793	0.6081	0.6335
1.1	0.4750	0.4729	0.4981	0.5377	0.5784	0.6356	0.6664	0.6930	0.7157
1.2	0.5333	0.5436	0.5775	0.6201	0.6609	0.7110	0.7392	0.7626	0.7819
1.3	0.5887	0.6096	0.6499	0.6928	0.7317	0.7737	0.7983	0.8181	0.8340
1.4	0.6406	0.6699	0.7131	0.7543	0.7908	0.8246	0.8454	0.8616	0.8744
1.5	0.6884	0.7239	0.7674	0.8052	0.8387	0.8652	0.8823	0.8953	0.9055
1.6	0.7319	0.7713	0.8136	0.8467	0.8760	0.8971	0.9109	0.9213	0.9293
1.7	0.7711	0.8122	0.8527	0.8811	0.9049	0.9220	0.9330	0.9411	0.9473
1.8	0.8060	0.8472	0.8845	0.9090	0.9272	0.9412	0.9498	0.9561	0.9509
1.9	0.8367	0.8766	0.9099	0.9311	0.9447	0.9559	0.9626	0.9675	0.9712
2.0	0.8635	0.9011	0.9299	0.9477	0.9580	0.9671	0.9723	0.9761	0.9789
2.1	0.8866	0.9213	0.9455	0.9603	0.9682	0.9756	0.9796	0.9825	0.9846
2.2	0.9064	0.9378	0.9577	0.9699	0.9758	0.9820	0.9851	0.9872	0.9889
2.3	0.9233	0.9512	0.9673	0.9774	0.9818	0.9868	0.9891	0.9908	0.9920
2.4	0.9375	0.9620	0.9752	0.9833	0.9868	0.9904	0.9921	0.9934	0.9943
2.5	0.9493	0.9705	0.9814	0.9880	0.9905	0.9930	0.9943	0.9953	0.9959
2.6	0.9592	0.9773	0.9861	0.9915	0.9934	0.9950	0.9960	0.9966	0.9971
2.7	0.9674	0.9827	0.9896	0.9940	0.9954	0.9964	0.9971	0.9976	0.9980
2.8	0.9740	0.9868	0.9923	0.9957	0.9969	0.9974	0.9980	0.9983	0.9986
2.9	0.9795	0.9900	0.9943	0.9969	0.9979	0.9982	0.9986	0.9988	0.9990
3.0	0.9839	0.9925	0.9958	0.9977	0.9986	0.9987	0.9990	0.9992	0.9993
3.1	0.9874	0.9944	0.9967	0.9983	0.9990	0.9991	0.9993	0.9995	0.9996
3.2	0.9902	0.9959	0.9974	0.9989	0.9993	0.9994	0.9995	0.9995	0.9997
3.3	0.9924	0.9970	0.9982	0.9993	0.9994	0.9996	0.9997	0.9998	0.9998
3.4	0.9942	0.9978	0.9988	0.9996	0.9996	0.9997	0.9998	0.9998	0.9999
3.5	0.9956	0.9984	0.9992	0.9997	0.9997	0.9998	0.9999	0.9999	0.9999
3.6	0.9966	0.9988	0.9994	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999
3.7	0.9975	0.9992	0.9996	0.9998	0.9998	0.9999	0.9999	1.0000	1.0000
3.8	0.9981	0.9994	0.9997	0.9999	0.9998	0.9999	1.0000	1.0000	
3.9	0.9986	0.9996	0.9998	0.9999	0.9999	1.0000	1.0000		
4.0	0.9990	0.9997	0.9998	0.9999	0.9999	1.0000			
4.1	0.9992	0.9998	0.9999	0.9999	1.0000				
4.2	0.9994	0.9999	0.9999	1.0000					
4.3	0.9996	0.9999	1.0000						
4.4	0.9997	0.9999							
4.5	0.9998	1.0000							
4.6	0.9998								
4.7	0.9999								
4.8	0.9999								
4.9	0.9999								
5.0	1.0000								

TABLE XXV (Continued)

GAP SIZE	SAMPLE SIZE 12	13	14	15	16	17	18	19	20
0.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
0.3	0.0059	0.0060	0.0062	0.0065	0.0068	0.0073	0.0078	0.0084	0.0091
0.4	0.0330	0.0352	0.0378	0.0408	0.0441	0.0476	0.0514	0.0554	0.0596
0.5	0.0975	0.1053	0.1137	0.1225	0.1316	0.1409	0.1502	0.1596	0.1688
0.6	0.1986	0.2134	0.2283	0.2432	0.2577	0.2719	0.2855	0.2985	0.3109
0.7	0.3205	0.3405	0.3598	0.3781	0.3953	0.4113	0.4262	0.4401	0.4529
0.8	0.4447	0.4669	0.4874	0.5061	0.5230	0.5385	0.5524	0.5651	0.5767
0.9	0.5585	0.5801	0.5993	0.6163	0.6314	0.6449	0.6569	0.6677	0.6774
1.0	0.6556	0.6749	0.6916	0.7061	0.7188	0.7300	0.7398	0.7487	0.7566
1.1	0.7349	0.7512	0.7651	0.7770	0.7873	0.7963	0.8042	0.8113	0.8175
1.2	0.7979	0.8112	0.8223	0.8319	0.8401	0.8472	0.8535	0.8590	0.8639
1.3	0.8469	0.8575	0.8664	0.8739	0.8804	0.8860	0.8909	0.8952	0.8990
1.4	0.8847	0.8931	0.9001	0.9060	0.9110	0.9153	0.9192	0.9225	0.9255
1.5	0.9136	0.9202	0.9256	0.9302	0.9341	0.9375	0.9404	0.9430	0.9454
1.6	0.9356	0.9407	0.9449	0.9484	0.9515	0.9541	0.9564	0.9584	0.9602
1.7	0.9522	0.9562	0.9594	0.9621	0.9645	0.9665	0.9682	0.9698	0.9712
1.8	0.9647	0.9678	0.9702	0.9723	0.9741	0.9757	0.9770	0.9782	0.9792
1.9	0.9741	0.9764	0.9783	0.9799	0.9813	0.9825	0.9835	0.9844	0.9852
2.0	0.9811	0.9829	0.9843	0.9855	0.9866	0.9874	0.9882	0.9889	0.9895
2.1	0.9863	0.9876	0.9887	0.9896	0.9904	0.9911	0.9916	0.9921	0.9926
2.2	0.9901	0.9911	0.9920	0.9925	0.9932	0.9937	0.9941	0.9945	0.9948
2.3	0.9929	0.9937	0.9943	0.9948	0.9952	0.9956	0.9959	0.9962	0.9964
2.4	0.9950	0.9955	0.9960	0.9964	0.9967	0.9969	0.9972	0.9973	0.9975
2.5	0.9965	0.9969	0.9972	0.9975	0.9977	0.9979	0.9980	0.9982	0.9983
2.6	0.9975	0.9978	0.9981	0.9983	0.9984	0.9985	0.9987	0.9988	0.9989
2.7	0.9983	0.9985	0.9987	0.9988	0.9989	0.9990	0.9991	0.9992	0.9992
2.8	0.9988	0.9990	0.9991	0.9992	0.9993	0.9993	0.9994	0.9994	0.9995
2.9	0.9992	0.9993	0.9994	0.9995	0.9995	0.9996	0.9996	0.9996	0.9997
3.0	0.9994	0.9995	0.9995	0.9996	0.9997	0.9997	0.9997	0.9998	0.9998
3.1	0.9996	0.9997	0.9997	0.9998	0.9998	0.9998	0.9998	0.9998	0.9999
3.2	0.9998	0.9998	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999
3.3	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.4	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	1.0000	1.0000	1.0000
3.5	0.9999	0.9999	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	
3.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		
3.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000			
3.8	1.0000	1.0000	1.0000	1.0000	1.0000				
3.9	1.0000	1.0000	1.0000	1.0000					
4.0	1.0000	1.0000	1.0000						
4.1	1.0000	1.0000							
4.2	1.0000								

TABLE XXVI

APPROXIMATE CRITICAL VALUES OF THE STUDENTIZED MAXIMUM GAP

	SAMPLE SIZE 2					SAMPLE SIZE 3					SAMPLE SIZE 4				
df	.1	.05	.025	.01	.005	.1	.05	.025	.01	.005	.1	.05	.025	.01	.005
1	8.93	17.97	36.0	90.0		9.8	20.0	39.1	82		9.7	19.5	38.8	78	
2	4.12	6.08	8.78	14.0	19.9	4.2	6.2	8.8	14.1	20	4.1	5.9	8.4	13.4	19
3	3.33	4.50	5.91	8.26	10.54	3.30	4.39	5.70	7.92	10.07	3.14	4.14	5.38	7.36	8.95
4	3.02	3.93	4.94	6.51	7.92	2.95	3.76	4.67	6.09	7.37	2.78	3.52	4.35	5.66	6.85
5	2.85	3.64	4.47	5.70	6.75	2.76	3.44	4.17	5.26	6.19	2.59	3.20	3.87	4.86	5.71
6	2.75	3.46	4.20	5.24	6.11	2.64	3.25	3.88	4.79	5.54	2.47	3.01	3.59	4.41	5.09
7	2.68	3.34	4.02	4.95	5.70	2.57	3.12	3.69	4.49	5.14	2.39	2.89	3.40	4.12	4.71
8	2.63	3.26	3.89	4.75	5.42	2.51	3.03	3.56	4.28	4.86	2.34	2.80	3.28	3.93	4.45
9	2.59	3.20	3.80	4.60	5.22	2.47	2.96	3.46	4.13	4.66	2.30	2.74	3.18	3.78	4.26
10	2.56	3.15	3.73	4.48	5.06	2.43	2.91	3.39	4.02	4.51	2.26	2.69	3.11	3.68	4.12
11	2.54	3.11	3.67	4.39	4.95	2.41	2.87	3.33	3.93	4.39	2.24	2.65	3.05	3.59	4.01
12	2.52	3.08	3.62	4.32	4.85	2.39	2.84	3.28	3.86	4.30	2.21	2.61	3.00	3.52	3.92
13	2.50	3.06	3.58	4.26	4.77	2.37	2.81	3.24	3.80	4.22	2.20	2.58	2.97	3.45	3.85
14	2.49	3.03	3.55	4.21	4.70	2.35	2.79	3.20	3.75	4.16	2.18	2.56	2.93	3.42	3.79
15	2.48	3.01	3.52	4.17	4.65	2.34	2.77	3.18	3.70	4.10	2.17	2.54	2.90	3.38	3.74
16	2.47	3.00	3.50	4.13	4.60	2.33	2.75	3.15	3.67	4.06	2.16	2.53	2.88	3.35	3.69
18	2.45	2.97	3.46	4.07	4.52	2.31	2.72	3.11	3.61	3.98	2.14	2.50	2.84	3.29	3.62
20	2.44	2.95	3.43	4.02	4.46	2.29	2.70	3.08	3.57	3.92	2.12	2.48	2.81	3.25	3.57
25	2.42	2.91	3.37	3.94	4.35	2.27	2.66	3.02	3.48	3.82	2.10	2.44	2.76	3.17	3.47
30	2.40	2.89	3.34	3.89	4.29	2.25	2.63	2.99	3.43	3.76	2.08	2.41	2.73	3.12	3.41
40	2.38	2.86	3.29	3.82	4.20	2.23	2.60	2.94	3.37	3.68	2.06	2.38	2.68	3.06	3.34
50	2.37	2.84	3.27	3.79	4.15	2.22	2.58	2.92	3.33	3.63	2.05	2.36	2.66	3.03	3.30
100	2.35	2.80	3.22	3.71	4.06	2.19	2.54	2.87	3.26	3.54	2.02	2.33	2.61	2.96	3.21
∞	2.33	2.77	3.17	3.64	3.97	2.17	2.51	2.82	3.19	3.46	2.00	2.29	2.56	2.90	3.14

TABLE XXVI (Continued)

	SAMPLE SIZE 5					SAMPLE SIZE 6					SAMPLE SIZE 7				
df	.1	.05	.025	.01	.005	.1	.05	.025	.01	.005	.1	.05	.025	.01	.005
1	9.3	18.8	37.3	73		9.0	18.1	36.2	69		8.6	17.4	34.6	66	
2	3.9	5.6	8.0	12.8	18	3.7	5.4	7.7	12.2	17	3.6	5.2	7.4	11.7	17
3	2.97	3.92	5.08	7.03	8.93	2.84	3.73	4.83	6.68	8.52	2.72	3.58	4.63	6.41	8.18
4	2.63	3.32	4.10	5.34	6.44	2.50	3.15	3.89	5.06	6.12	2.39	3.02	3.74	4.85	5.86
5	2.44	3.01	3.64	4.57	5.37	2.32	2.86	3.45	4.33	5.09	2.23	2.74	3.31	4.15	4.88
6	2.33	2.83	3.37	4.14	4.78	2.21	2.69	3.19	3.92	4.52	2.12	2.58	3.06	3.76	4.35
7	2.25	2.71	3.19	3.87	4.42	2.14	2.57	3.02	3.66	4.17	2.05	2.47	2.90	3.52	4.01
8	2.20	2.63	3.07	3.68	4.17	2.09	2.49	2.91	3.48	3.94	2.00	2.39	2.79	3.35	3.79
9	2.16	2.57	2.98	3.55	4.00	2.05	2.43	2.82	3.35	3.77	1.96	2.33	2.71	3.22	3.63
10	2.12	2.52	2.91	3.45	3.86	2.02	2.38	2.75	3.25	3.64	1.93	2.28	2.64	3.13	3.51
11	2.10	2.48	2.86	3.36	3.76	1.99	2.35	2.70	3.17	3.54	1.91	2.25	2.59	3.06	3.42
12	2.08	2.45	2.81	3.30	3.67	1.97	2.32	2.66	3.11	3.47	1.89	2.22	2.56	3.00	3.43
13	2.06	2.42	2.77	3.25	3.61	1.96	2.29	2.62	3.06	3.40	1.87	2.19	2.52	2.95	3.28
14	2.04	2.39	2.74	3.20	3.56	1.94	2.27	2.59	3.02	3.35	1.86	2.17	2.49	2.91	3.23
15	2.03	2.38	2.72	3.16	3.51	1.93	2.25	2.57	2.98	3.30	1.84	2.16	2.47	2.88	3.18
16	2.02	2.36	2.69	3.13	3.47	1.92	2.24	2.55	2.96	3.26	1.83	2.14	2.45	2.85	3.15
18	2.00	2.33	2.66	3.08	3.40	1.90	2.21	2.51	2.90	3.19	1.81	2.12	2.42	2.80	3.09
20	1.99	2.31	2.63	3.04	3.35	1.89	2.19	2.49	2.87	3.15	1.80	2.10	2.39	2.77	3.04
25	1.96	2.27	2.58	2.97	3.26	1.86	2.15	2.44	2.80	3.06	1.78	2.07	2.35	2.70	2.97
30	1.95	2.25	2.55	2.93	3.20	1.85	2.13	2.40	2.76	3.01	1.76	2.04	2.32	2.66	2.91
40	1.92	2.22	2.51	2.87	3.14	1.83	2.10	2.37	2.70	2.95	1.74	2.02	2.28	2.62	2.85
50	1.91	2.20	2.49	2.84	3.10	1.81	2.08	2.35	2.67	2.91	1.73	2.00	2.26	2.59	2.82
100	1.88	2.17	2.44	2.78	3.02	1.79	2.05	2.30	2.61	2.83	1.70	1.97	2.22	2.54	2.74
∞	1.86	2.14	2.40	2.71	2.94	1.77	2.02	2.27	2.56	2.78	1.68	1.94	2.19	2.48	2.68

TABLE XXVI (Continued)

df	SAMPLE SIZE 8 .1	.05	.025	.01	.005	SAMPLE SIZE 9 .1	.05	.025	.01	.005	SAMPLE SIZE 10 .1	.05	.025	.01	.005
1	8.1	16.5	32.5	61		7.9	16.0	31.6	59		7.7	15.6	30.7	57	
2	3.4	4.9	7.0	11.1	16	3.3	4.8	6.8	10.8	15	3.2	4.6	6.6	10.5	15
3	2.58	3.40	4.40	6.11	7.79	2.50	3.30	4.27	5.92	7.56	2.43	3.21	4.15	5.77	7.36
4	2.28	2.88	3.56	4.63	5.60	2.20	2.79	3.46	4.50	5.44	2.14	2.71	3.36	4.38	5.30
5	2.12	2.61	3.16	3.97	4.67	2.05	2.53	3.06	3.86	4.54	1.99	2.47	2.98	3.76	4.42
6	2.02	2.46	2.93	3.60	4.17	1.95	2.38	2.84	3.50	4.05	1.90	2.32	2.76	3.41	3.95
7	1.95	2.35	2.78	3.37	3.86	1.89	2.28	2.69	3.27	3.75	1.84	2.22	2.62	3.19	3.66
8	1.90	2.28	2.67	3.21	3.65	1.84	2.21	2.59	3.12	3.54	1.79	2.15	2.52	3.04	3.46
9	1.87	2.23	2.59	3.09	3.49	1.81	2.16	2.51	3.01	3.39	1.76	2.10	2.45	2.93	3.32
10	1.84	2.18	2.53	3.01	3.38	1.78	2.12	2.46	2.92	3.29	1.73	2.06	2.39	2.85	3.21
11	1.82	2.15	2.49	2.94	3.29	1.76	2.08	2.41	2.86	3.20	1.71	2.03	2.35	2.79	3.96
12	1.80	2.12	2.45	2.88	3.22	1.74	2.06	2.38	2.80	3.13	1.69	2.00	2.32	2.74	3.06
13	1.78	2.10	2.42	2.84	3.16	1.73	2.04	2.35	2.76	3.08	1.68	1.98	2.29	2.69	3.01
14	1.77	2.08	2.39	2.80	3.11	1.71	2.02	2.32	2.72	3.03	1.67	1.97	2.26	2.66	2.96
15	1.76	2.07	2.37	2.77	3.08	1.70	2.00	2.30	2.69	2.99	1.66	1.95	2.24	2.63	2.93
16	1.75	2.05	2.35	2.74	3.04	1.69	1.99	2.28	2.67	2.96	1.65	1.94	2.22	2.60	2.89
18	1.73	2.03	2.32	2.70	2.98	1.68	1.97	2.25	2.62	2.90	1.63	1.91	2.19	2.56	2.84
20	1.72	2.01	2.29	2.66	2.94	1.67	1.95	2.23	2.59	2.87	1.62	1.90	2.17	2.53	2.80
25	1.70	1.98	2.25	2.60	2.87	1.64	1.92	2.19	2.53	2.79	1.60	1.87	2.13	2.48	2.73
30	1.68	1.96	2.22	2.57	2.82	1.63	1.90	2.16	2.50	2.75	1.58	1.85	2.11	2.44	2.69
40	1.66	1.93	2.19	2.52	2.76	1.61	1.87	2.13	2.45	2.69	1.57	1.83	2.08	2.40	2.63
50	1.65	1.92	2.17	2.49	2.73	1.60	1.86	2.11	2.43	2.66	1.56	1.81	2.06	2.37	2.60
100	1.63	1.89	2.13	2.44	2.66	1.58	1.83	2.07	2.38	2.59	1.54	1.78	2.02	2.32	3.01
∞	1.61	1.86	2.09	2.39	2.60	1.56	1.80	2.03	2.33	2.54	1.52	1.76	1.99	2.28	2.49

TABLE XXVI (Continued)

	SAMPLE SIZE 11					SAMPLE SIZE 12					SAMPLE SIZE 13				
df	.1	.05	.025	.01	.005	.1	.05	.025	.01	.005	.1	.05	.025	.01	.005
1	7.1	14.3	28.2	53		6.9	14.1	27.8	52		6.8	13.9	27.3	51	
2	2.9	4.2	6.1	9.6	14	2.9	4.2	6.0	9.5	13	2.9	4.1	5.9	9.3	13
3	2.24	2.95	3.83	5.32	6.80	2.22	2.92	3.78	5.23	6.64	2.17	2.87	3.72	5.17	6.60
4	1.97	2.50	3.11	4.05	4.90	1.95	2.47	3.07	3.99	4.84	1.91	2.43	3.02	3.94	4.77
5	1.83	2.28	2.76	3.48	4.10	1.80	2.24	2.72	3.43	4.05	1.78	2.21	2.68	3.39	3.99
6	1.75	2.14	2.56	3.17	3.67	1.72	2.11	2.52	3.12	3.62	1.70	2.08	2.49	3.08	3.58
7	1.69	2.05	2.43	2.97	3.40	1.66	2.02	2.40	2.93	3.36	1.64	1.99	2.37	2.89	3.32
8	1.65	1.99	2.34	2.83	3.22	1.62	1.96	2.31	2.79	3.18	1.60	1.93	2.28	2.76	3.15
9	1.62	1.94	2.28	2.73	3.10	1.59	1.91	2.24	2.70	3.06	1.57	1.89	2.22	2.67	3.02
10	1.59	1.91	2.22	2.66	3.00	1.57	1.88	2.19	2.62	2.96	1.55	1.85	2.17	2.59	2.93
11	1.57	1.88	2.19	2.60	2.92	1.55	1.85	2.16	2.57	2.89	1.53	1.83	2.13	2.54	2.86
12	1.56	1.85	2.15	2.55	2.86	1.53	1.83	2.12	2.52	2.83	1.51	1.80	2.10	2.49	2.80
13	1.55	1.84	2.13	2.52	2.82	1.52	1.81	2.10	2.48	2.78	1.50	1.79	2.07	2.46	2.75
14	1.53	1.82	2.10	2.48	2.77	1.51	1.79	2.08	2.45	2.74	1.49	1.77	2.05	2.43	2.71
15	1.52	1.81	2.09	2.46	2.74	1.50	1.78	2.06	2.43	2.71	1.48	1.76	2.03	2.40	2.68
16	1.52	1.79	2.07	2.43	2.71	1.49	1.77	2.04	2.40	2.68	1.47	1.75	2.02	2.38	2.65
18	1.50	1.77	2.04	2.40	2.66	1.48	1.75	2.02	2.37	2.63	1.46	1.73	1.99	2.34	2.60
20	1.49	1.76	2.02	2.37	2.63	1.47	1.73	2.00	2.34	2.60	1.45	1.71	1.97	2.31	2.57
25	1.47	1.73	1.99	2.32	2.56	1.45	1.71	1.96	2.29	2.53	1.43	1.69	1.94	2.26	2.51
30	1.46	1.71	1.96	2.29	2.52	1.44	1.69	1.94	2.26	2.49	1.42	1.67	1.92	2.23	2.47
40	1.45	1.69	1.94	2.25	2.47	1.42	1.67	1.91	2.22	2.45	1.40	1.65	1.89	2.19	2.42
50	1.44	1.68	1.92	2.22	2.45	1.41	1.66	1.89	2.20	2.42	1.40	1.64	1.87	2.17	2.39
100	1.42	1.66	1.89	2.18	2.39	1.40	1.63	1.86	2.15	2.36	1.38	1.61	1.84	2.13	2.34
∞	1.40	1.63	1.86	2.14	2.34	1.38	1.61	1.83	2.11	2.31	1.36	1.59	1.81	2.09	2.29

TABLE XXVI (Continued)

	SAMPLE SIZE 14					SAMPLE SIZE 15					SAMPLE SIZE 16				
df	.1	.05	.025	.01	.005	.1	.05	.025	.01	.005	.1	.05	.025	.01	.005
1	7.1	14.3	28.2	53		6.9	14.1	27.8	52		6.8	13.9	27.3	51	
2	2.9	4.2	6.1	9.6	14	2.9	4.2	6.0	9.5	13	2.9	4.1	5.9	9.3	13
3	2.24	2.95	3.83	5.32	6.80	2.22	2.92	3.78	5.23	6.64	2.17	2.87	3.72	5.17	6.60
4	1.97	2.50	3.11	4.05	4.90	1.95	2.47	3.07	3.99	4.84	1.91	2.43	3.02	3.94	4.77
5	1.83	2.28	2.76	3.48	4.10	1.80	2.24	2.72	3.43	4.05	1.78	2.21	2.68	3.39	3.99
6	1.75	2.14	2.56	3.17	3.67	1.72	2.11	2.52	3.12	3.62	1.70	2.08	2.49	3.08	3.58
7	1.69	2.05	2.43	2.97	3.40	1.66	2.02	2.40	2.93	3.36	1.64	1.99	2.37	2.89	3.32
8	1.65	1.99	2.34	2.83	3.22	1.62	1.96	2.31	2.79	3.18	1.60	1.93	2.28	2.76	3.15
9	1.62	1.94	2.28	2.73	3.10	1.59	1.91	2.24	2.70	3.06	1.57	1.89	2.22	2.67	3.02
10	1.59	1.91	2.22	2.66	3.00	1.57	1.88	2.19	2.62	2.96	1.55	1.85	2.17	2.59	2.93
11	1.57	1.88	2.19	2.60	2.92	1.55	1.85	2.16	2.57	2.89	1.53	1.83	2.13	2.54	2.86
12	1.56	1.85	2.15	2.55	2.86	1.53	1.83	2.12	2.52	2.83	1.51	1.80	2.10	2.49	2.80
13	1.55	1.84	2.13	2.52	2.82	1.52	1.81	2.10	2.48	2.78	1.50	1.79	2.07	2.46	2.75
14	1.53	1.82	2.10	2.48	2.77	1.51	1.79	2.08	2.45	2.74	1.49	1.77	2.05	2.43	2.71
15	1.52	1.81	2.09	2.46	2.74	1.50	1.78	2.06	2.43	2.71	1.48	1.76	2.03	2.40	2.68
16	1.52	1.79	2.07	2.43	2.71	1.49	1.77	2.04	2.40	2.68	1.47	1.75	2.02	2.38	2.65
18	1.50	1.77	2.04	2.40	2.66	1.48	1.75	2.02	2.37	2.63	1.46	1.73	1.99	2.34	2.60
20	1.49	1.76	2.02	2.37	2.63	1.47	1.73	2.00	2.34	2.60	1.45	1.71	1.97	2.31	2.57
25	1.47	1.73	1.99	2.32	2.56	1.45	1.71	1.96	2.29	2.53	1.43	1.69	1.94	2.26	2.51
30	1.46	1.71	1.96	2.29	2.52	1.44	1.69	1.94	2.26	2.49	1.42	1.67	1.92	2.23	2.47
40	1.45	1.69	1.94	2.25	2.47	1.42	1.67	1.91	2.22	2.45	1.40	1.65	1.89	2.19	2.42
50	1.44	1.68	1.92	2.22	2.45	1.41	1.66	1.89	2.20	2.42	1.40	1.64	1.87	2.17	2.39
100	1.42	1.66	1.89	2.18	2.39	1.40	1.63	1.86	2.15	2.36	1.38	1.61	1.84	2.13	2.34
∞	1.40	1.63	1.86	2.14	2.34	1.38	1.61	1.83	2.11	2.31	1.36	1.59	1.81	2.09	2.29

TABLE XXVI (Continued)

df	SAMPLE SIZE 17 .1	.05	.025	.01	.005	SAMPLE SIZE 18 .1	.05	.025	.01	.005	SAMPLE SIZE 19 .1	.05	.025	.01	.005
1	6.7	13.7	26.9	51		6.7	13.7	26.5	50		6.6	13.3	26.1	49	
2	2.8	4.0	5.8	9.2	13	2.8	4.0	5.7	9.1	13	2.8	4.0	5.7	9.0	13
3	2.14	2.83	3.67	5.11	6.52	2.11	2.79	3.63	5.05	6.44	2.09	2.76	3.59	4.99	6.37
4	1.87	2.40	2.98	3.89	4.71	1.86	2.37	2.95	3.85	4.66	1.84	2.34	2.91	3.81	4.61
5	1.75	2.18	2.65	3.35	3.95	1.73	2.16	2.62	3.31	3.91	1.71	2.13	2.59	3.28	3.87
6	1.67	2.05	2.46	3.05	3.54	1.65	2.03	2.43	3.02	3.50	1.64	2.01	2.41	2.99	3.47
7	1.62	1.97	2.34	2.86	3.28	1.60	1.95	2.31	2.83	3.25	1.58	1.93	2.29	2.80	3.22
8	1.58	1.91	2.25	2.73	3.11	1.56	1.89	2.23	2.70	3.08	1.54	1.87	2.21	2.68	3.05
9	1.55	1.87	2.19	2.64	2.99	1.53	1.85	2.17	2.61	2.96	1.52	1.83	2.15	2.59	2.93
10	1.53	1.83	2.14	2.57	2.90	1.51	1.81	2.12	2.54	2.87	1.49	1.79	2.10	2.52	2.84
11	1.51	1.81	2.11	2.51	2.83	1.49	1.79	2.08	2.49	2.80	1.48	1.77	2.06	2.46	2.77
12	1.49	1.78	2.08	2.47	2.77	1.48	1.76	2.05	2.44	2.74	1.46	1.75	2.03	2.42	2.72
13	1.48	1.77	2.05	2.43	2.72	1.46	1.75	2.03	2.41	2.70	1.45	1.73	2.01	2.39	2.67
14	1.47	1.75	2.03	2.40	2.68	1.45	1.73	2.01	2.38	2.66	1.44	1.71	2.00	2.36	2.64
15	1.46	1.74	2.01	2.37	2.65	1.45	1.72	1.99	2.35	2.63	1.43	1.70	1.97	2.33	2.60
16	1.45	1.73	2.00	2.35	2.62	1.44	1.71	1.98	2.33	2.60	1.42	1.69	1.96	2.31	2.58
18	1.44	1.71	1.97	2.32	2.58	1.42	1.69	1.95	2.29	2.55	1.41	1.67	1.93	2.27	2.53
20	1.43	1.69	1.95	2.29	2.54	1.41	1.68	1.93	2.27	2.52	1.40	1.66	1.91	2.25	2.50
25	1.41	1.67	1.92	2.24	2.48	1.40	1.65	1.90	2.22	2.46	1.38	1.63	1.88	2.20	2.44
30	1.40	1.65	1.89	2.21	2.44	1.39	1.63	1.88	2.19	2.42	1.37	1.62	1.86	2.17	2.40
40	1.39	1.63	1.87	2.17	2.40	1.37	1.61	1.85	2.15	2.37	1.36	1.60	1.83	2.13	2.36
50	1.38	1.62	1.85	2.15	2.37	1.36	1.60	1.83	2.13	2.35	1.35	1.59	1.82	2.11	2.33
100	1.36	1.60	1.82	2.11	2.31	1.35	1.58	1.80	2.09	2.29	1.33	1.56	1.79	2.07	2.27
∞	1.34	1.57	1.79	2.07	2.27	1.33	1.56	1.78	2.05	2.25	1.32	1.54	1.76	2.03	2.23

TABLE XXVI (Continued)

SAMPLE SIZE 20					
df	.1	.05	.025	.01	.005
1	6.5	13.2	25.6	49	
2	2.8	3.9	5.6	8.9	13
3	2.06	2.72	3.54	5.02	6.54
4	1.82	2.32	2.88	3.75	4.55
5	1.70	2.11	2.57	3.24	3.82
6	1.62	1.99	2.39	2.96	3.44
7	1.57	1.91	2.27	2.78	3.19
8	1.53	1.85	2.19	2.65	3.03
9	1.50	1.81	2.13	2.56	2.91
10	1.48	1.78	2.08	2.50	2.82
11	1.46	1.75	2.05	2.44	2.75
12	1.45	1.73	2.02	2.40	2.70
13	1.43	1.71	1.99	2.37	2.65
14	1.42	1.70	1.97	2.34	2.62
15	1.42	1.69	1.96	2.31	2.58
16	1.41	1.68	1.94	2.29	2.56
18	1.40	1.66	1.92	2.26	2.51
20	1.39	1.64	1.90	2.23	2.48
25	1.37	1.62	1.87	2.18	2.42
30	1.36	1.60	1.84	2.15	2.38
40	1.34	1.58	1.81	2.09	2.29
50	1.33	1.57	1.80	2.10	2.31
100	1.32	1.55	1.77	2.05	2.26
∞	1.30	1.53	1.74	2.02	2.21

VITA

John Riffe Murphy

Candidate for the Degree of

Doctor of Philosophy

Thesis: PROCEDURES FOR GROUPING A SET OF OBSERVED MEANS

Major Field: Statistics

Biographical:

Personal Data: Born in Hooker, Oklahoma, April 12, 1942, the son of Robert and Ethel Murphy.

Education: Attended elementary school in Goodwell, Oklahoma; graduated from Goodwell High School, Goodwell, Oklahoma, in 1960; attended Oklahoma Baptist University, Shawnee, Oklahoma, 1960-1961; received the Bachelor of Science in Mathematics from Panhandle State College, Goodwell, Oklahoma, in August, 1964; received the Master of Science in Mathematics from Oklahoma State University, Stillwater, Oklahoma, in May, 1967; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in July, 1973.

Professional Experience: Graduate teaching assistant at Oklahoma State University, Stillwater, Oklahoma, 1964-1969; Systems Engineer, IBM, Tulsa, Oklahoma, 1967; Statistician, Environmental Protection Agency, Research Triangle Park, North Carolina, 1971-1972.