THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MODELS FOR MULTIPLE COMPARISONS IN STATISTICS AND SOME NONPARAMETRIC TECHNIQUES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

BAHRAM ZEIGHAMI

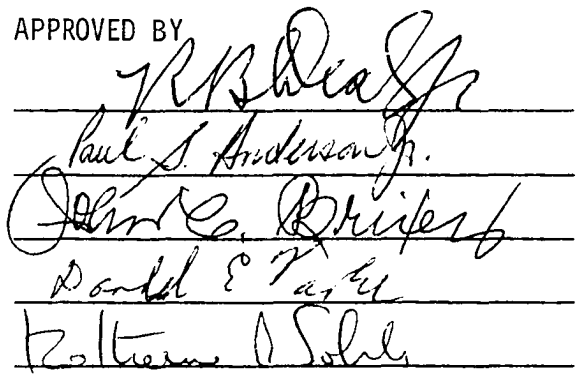
Oklahoma City, Oklahoma

1973

MODELS FOR MULTIPLE COMPARISONS IN STATISTICS

AND SOME NONPARAMETRIC TECHNIQUES

APPROVED BY



DISSERTATION COMMITTEE

ACKNOWLEDGMENTS

I wish to extend my appreciation and gratitude to the many people who have provided assistance and encouragement during the course of this research.

I am especially indebted to Dr. Roy B. Deal, Jr. for his great patience and unparalleled ability in directing this dissertation.

I would also like to thank Dr. Paul S. Anderson, Jr., Dr. Donald Parker, Dr. John C. Brixey and Dr. Katherine Sohler for their assistance in serving as members of my dissertation committee.

I would also like to thank my fellow students, particularly Mr. Jim Goin, for his interest and his assistance.

Credit also goes to Mrs. Rose Titsworth for her very able typing abilities.

My deepest gratitude goes to my Mother and Father for their support and encouragement through the years of my education, and to my wife, Elaine.

Lastly, I would like to acknowledge the financial assistance given to me through the National Institute of General Medical Sciences Grant No. 5 T01 GM00001 and the Department of Biostatistics and Epidemiology, College of Health and Allied Health Professions, University of Oklahoma Health Sciences Center.

TABLE OF CONTENTS

		Page
LIST OF TABLES		v
Chapter		
I.	INTRODUCTION	1
II.	REVIEW OF EXISTING METHODS	17
III.	MATHEMATICAL FORMULATION OF THE PROBLEM	26
IV.	DERIVATION OF METHODS	42
V.	PRACTICAL SAMPLES	53
VI.	SUMMARY	64
LIST OF REFERENCES		67
APPENDIX		69

LIST OF TABLES

Table		Page
1.	A Crude Estimate of the Comparison Between α and α^*.....	8
2.	An Estimate of the Comparison Between α and α^* for Given g.....	10
3.	An Estimate of the Comparison Between α and α^* for Given g with Sample Size = 25.....	11
4.	Comparison Between α^*'s Obtained in Tables 1, 2 and 3.....	12
5.	1,469 White Females Classified by Age and Calcium Level of Blood.....	13
6.	1,565 White Females Classified by Age and Phosphorus Level in Blood.....	15
7.	The Cumulative Sampling Distributions of Blood Phosphorus Level of the Age Groupings Employed in the Contrasts.....	55
8.	The Maximum Observed Differences of the Cumulative Sampling Distributions for Various Specific Contrasts Corrected for the Sample Size.....	57
9.	The Cumulative Sampling Distributions of DNCB Concentration of the Groups Composing the Contrasts.....	59
10.	The Maximum Observed Difference of the Cumulative Sampling Distributions for Various Specific Contrasts Corrected for the Sample Sizes.....	60
11.	The Cumulative Sampling Distributions of DNCB Concentration of the Special Groups Composing the Contrasts.....	61
12.	Maximum Observed Differences of the Cumulative Sampling Distributions for Various Specific Contrasts Corrected for the Sample Sizes.....	61

LIST OF TABLES--Continued

Table		Page
13.	Values Required to Show Significant Differences of Two Rates for Given λ at α = .05	63
14.	Table for Estimating the Goodness of Fit of Empirical Distributions by Smirnov [19]	70

MODELS FOR MULTIPLE COMPARISONS IN STATISTICS AND SOME NONPARAMETRIC TECHNIQUES

CHAPTER I

INTRODUCTION

In recent years multiple comparisons methods and problems associated with their use have attracted a great deal of attention. Many experimental problems are amenable to the use of multiple comparisons. The majority of practicing statisticians are beginning to realize that in some cases analysis of variance techniques and their results are only a first step toward the eventual decision. Multiple comparisons are quite often a proper and desirable second step in making optimum use of the results.

In the early days of experimentation and collection of data, inference was limited to inspecting the data and making certain statements about the data and what it implied, based on subjective judgments alone. But the work of early statisticians in introducing more sophisticated techniques greatly improved the methods of decision-making. This does not mean, however, that we should rely on these techniques alone and forget the procedures of those who preceded them. The ability to inspect data and draw reasonable conclusions is a valuable gift which people possess in varying degrees. Experience is obviously an aid in making such subjective judgments.

At this time it is again essential to understand and improve the theory of multiple comparisons testing in order to be able to make more scientific and useful decisions in many research designs. However, in spite of the importance of this subject, and the great number of publications on the subject, there remains a great deal of disagreement among statisticians about the quality of the various multiple comparisons tests. This is expected in a developing area of the science and hopefully there will be methods developed eventually with which every statistician can agree. It is also to be hoped that techniques can be developed which do not have such stringent assumptions that their applicability is severely limited.

Another problem in the use of multiple comparisons is that of knowing when such a test provides the best choice. T. E. Kurtz [10] states that not every situation in which multiple comparison is possible is a situation in which multiple comparison is wise. A second error in the use of multiple comparisons is the failure to choose the proper multiple comparison test because of a lack of statistical knowledge. There are too many researchers who do not realize that their results are meaningless if the assumptions with the test are not satisfied. Perhaps the most common error by practicing statisticians, however, is that of making several statements at a given α level without being sufficiently aware of the fact that under an overall null hypothesis the probability is much greater than α that at least one of these statements would be made from the observations.

A common experimental situation in medical or health related research is the case in which the researcher applies a set of different treatments to g groups, not necessarily of equal sizes. The researcher

often wishes to decide which is the better treatment, or determine that the evidence does not justify making a decision either way. If N = 2, this can be answered by a t-test, or some nonparametric test such as the Mann-Whitney U-test. In this case one can make one of four possible decisions: (a) treatment one is better than treatment two, (b) treatment two is better than treatment one, (c) there is not enough evidence to make a decision, and (d) on rare occasions there is sufficient data to conclude with reasonable safety that there is no essential difference. In an experiment in which g is greater than two, the researcher may want to make the same sort of decision about every pair of treatments.

The problem now is that there are several alternatives to choose from. An example illustrating these alternatives is given below. If a series of two-sample tests, such as a t-test or Mann-Whitney U-test, are to be applied to a collection of treatment means, each at significance level α, then the a priori probability of obtaining at least one significant result is higher than α. This dissertation is concerned with extending the development of tests to detect any existing differences, and to apply the techniques to some existing problems in biomedical research. In this area, however, there are so many situations where there is no standard procedure available that it seems worthwhile to provide a detailed statement of the problem and the results of various approaches. The level of this treatment is that of S. S. Wilks, [25], C. R. Rao, [15], and D. A. S. Fraser, [6].

It is enlightening to give a few illustrations to clarify some of the above discussion and to show the difficulties of computing the

significance level associated with multiple comparisons.

Let g, the number of groups, be three. The hypothesis to be tested is that all means are equal.

The sample space can be partitioned in a variety of ways depending on the types of decisions to be made. If only individual paired comparisons are to be made, and those are not dependent on order, the decision function induces the following partition.

$$W_1 = \left\{ X \middle| \begin{array}{l} |\bar{X}_1 - \bar{X}_2| > k_{1,2} \\ |\bar{X}_1 - \bar{X}_3| > k_{1,3} \\ |\bar{X}_2 - \bar{X}_3| > k_{2,3} \end{array} \right\} \quad \text{implies} \left\{ \begin{array}{l} \text{reject } \mu_1 = \mu_2, \\ \text{reject } \mu_1 = \mu_3, \\ \text{reject } \mu_2 = \mu_3 \end{array} \right\}$$

$$W_2 = \left\{ X \middle| \begin{array}{l} |\bar{X}_1 - \bar{X}_2| \leq k_{1,2} \\ |\bar{X}_1 - \bar{X}_3| > k_{1,3} \\ |\bar{X}_2 - \bar{X}_3| > k_{2,3} \end{array} \right\} \quad \text{implies} \left\{ \begin{array}{l} \text{not reject } \mu_1 = \mu_2, \\ \text{reject } \mu_1 = \mu_3, \\ \text{reject } \mu_2 = \mu_3 \end{array} \right\}$$

$$W_3 = \left\{ X \middle| \begin{array}{l} |\bar{X}_1 - \bar{X}_2| > k_{1,2} \\ |\bar{X}_1 - \bar{X}_3| \leq k_{1,3} \\ |\bar{X}_2 - \bar{X}_3| > k_{2,3} \end{array} \right\} \quad \text{implies} \left\{ \begin{array}{l} \text{reject } \mu_1 = \mu_2 \\ \text{not reject } \mu_1 = \mu_3, \\ \text{reject } \mu_2 = \mu_3 \end{array} \right\}$$

$$W_4 = \left\{ X \middle| \begin{array}{l} |\bar{X}_1 - \bar{X}_2| > k_{1,2} \\ |\bar{X}_1 - \bar{X}_3| > k_{1,3} \\ |\bar{X}_2 - \bar{X}_3| \leq k_{2,3} \end{array} \right\} \quad \text{implies} \left\{ \begin{array}{l} \text{reject } \mu_1 = \mu_2, \\ \text{reject } \mu_1 = \mu_3, \\ \text{not reject } \mu_2 = \mu_3 \end{array} \right\}$$

$$W_5 = \left\{ X \left| \begin{array}{l} |\bar{X}_1 - \bar{X}_2| > k_{1,2} \\ |\bar{X}_1 - \bar{X}_3| \leq k_{1,3} \\ |\bar{X}_2 - \bar{X}_3| \leq k_{2,3} \end{array} \right. \right\} \quad \text{implies} \left\{ \begin{array}{l} \text{reject } \mu_1 = \mu_2 \\ \text{not reject } \mu_1 = \mu_3, \\ \text{not reject } \mu_2 = \mu_3 \end{array} \right\}$$

$$W_6 = \left\{ X \left| \begin{array}{l} |\bar{X}_1 - \bar{X}_2| \leq k_{1,2} \\ |\bar{X}_1 - \bar{X}_3| > k_{1,3} \\ |\bar{X}_2 - \bar{X}_3| \leq k_{2,3} \end{array} \right. \right\} \quad \text{implies} \left\{ \begin{array}{l} \text{not reject } \mu_1 = \mu_2 \\ \text{reject } \mu_1 = \mu_3 \\ \text{not reject } \mu_2 = \mu_3 \end{array} \right\}$$

$$W_7 = \left\{ X \left| \begin{array}{l} |\bar{X}_1 - \bar{X}_2| \leq k_{1,2} \\ |\bar{X}_1 - \bar{X}_3| \leq k_{1,3} \\ |\bar{X}_2 - \bar{X}_3| > k_{2,3} \end{array} \right. \right\} \quad \text{implies} \left\{ \begin{array}{l} \text{not reject } \mu_1 = \mu_2 \\ \text{not reject } \mu_1 = \mu_3 \\ \text{reject } \mu_2 = \mu_3 \end{array} \right\}$$

$$W_8 = \left\{ X \left| \begin{array}{l} |\bar{X}_1 - \bar{X}_2| \leq k_{1,2} \\ |\bar{X}_1 - \bar{X}_3| \leq k_{1,3} \\ |\bar{X}_2 - \bar{X}_3| \leq k_{2,3} \end{array} \right. \right\} \quad \text{implies} \left\{ \begin{array}{l} \text{not reject } \mu_1 = \mu_2 \\ \text{not reject } \mu_1 = \mu_3 \\ \text{not reject } \mu_2 = \mu_3 \end{array} \right\}$$

Even this simple situation leads to a variety of levels of type I and type II errors. Suppose, for example, the following standard, but naive, partition of the parameter space is assumed. The term naive is used because the fact is ignored that, even in a deterministic model, if the underlying distributions are continuous, experimental error would make it impossible to ascertain whether two means are equal or simply differ by amounts too small to distinguish. If the researcher knows the level at which he considers two measurements to be distinguishably different it is possible to partition the parameter space in a manner analogous to that of the sample space. More is said on this subject later.

For this discussion the standard partition is the following.

$$\Omega_1 = \{(\mu_1, \mu_2, \mu_3) \mid \mu_1 = \mu_2 = \mu_3\}$$

$$\Omega_2 = \{(\mu_1, \mu_2, \mu_3) \mid \mu_1 = \mu_2, \mu_1 \neq \mu_3, \mu_2 \neq \mu_3\}$$

$$\Omega_3 = \{(\mu_1, \mu_2, \mu_3) \mid \mu_1 \neq \mu_2, \mu_1 = \mu_3, \mu_2 \neq \mu_3\}$$

$$\Omega_4 = \{(\mu_1, \mu_2, \mu_3) \mid \mu_1 \neq \mu_2, \mu_1 \neq \mu_3, \mu_2 = \mu_3\}$$

$$\Omega_5 = \{(\mu_1, \mu_2, \mu_3) \mid \mu_1 \neq \mu_2, \mu_1 \neq \mu_3, \mu_2 \neq \mu_3\}$$

If the sample is drawn from a population whose means fall in Ω_1 (all means are equal) then type I errors will be committed if the sample falls in any of the first seven regions. In W_1 three errors of type I would be made, in W_2, W_3, W_4, two errors in each, and in W_5, W_6, W_7 one each. This leads to the concept, discussed later, of error rate. If the parameter means are in Ω_2 there would be type I errors if the sample falls in W_1, W_3, W_4, or W_5, one in each case. However, the situation in W_5 is different from the other three. There is no corresponding region in the parameter space in which $\mu_2 = \mu_3$, $\mu_3 = \mu_1$ but $\mu_1 \neq \mu_2$. Also, the probability that $|\bar{X}_1 - \bar{X}_2| > k_{1,2}$, while $|\bar{X}_1 - \bar{X}_3| \leq k_{1,3}$ and $|\bar{X}_2 - \bar{X}_3| \leq k_{2,3}$, is certain to be small with any reasonable α-levels. Similar statements hold for regions W_6 and W_7 with similar conditions in the parameter space. For this reason it is only slightly more conservative, and certainly easier to manage, to make the decision to reject none of the three equalities when the sample falls in $W_5 \cup W_6 \cup W_7 \cup W_8$.

In many experimental situations the statistician is confronted with the following type of problem. Let $t_1, t_2, \ldots, t_g$ be g different treatments applied to as many groups. The researcher may be only interested in a portion of the possible comparisons. One very common situation

occurs when g-1 of the treatments are compared with a control. Here, the researcher wishes to make all pairwise comparisons in which the pair consists of the mean effect of the control, and the mean effect of one of the experimental groups.

If there are no additional contrasts, and the data comes from normally distributed random variables with homogeneous variance, Dunnett [3] has solved the problem. Certain nonparametric tests are available with strict assumptions, and available for more general contrasts under normality assumptions. These are discussed in the review section along with some new tests, developed here, for handling other nonparametric situations and more general contrasts.

Suppose again that the researcher has g groups. In testing the null hypothesis

$$H_o: \mu_1 = \mu_2 = \ldots = \mu_g$$

with an "overall" F-test we are risking a probability α of rejecting the hypothesis when it is true. Assuming the researcher is interested in making N individual comparisons, the risk of obtaining at least one falsely significant result is much higher than the risk of committing a type I error in every single test. For example, let us suppose that there are g means to be compared. Then there are $\frac{g(g-1)}{2}$ possible pairwise comparisons. To obtain a crude estimate (simulation indicates it is sometimes close) of the comparison between α and the level α^* at which individual comparisons could be made to produce an overall level α, the computation is made as if the individual tests were independent. That is,

$$\alpha = 1 - (1 - \alpha^*)^{g(g-1)/2} \quad .$$

These values of α and α^* have been computed for several values of g and tabulated in Table 1.

TABLE 1

A CRUDE ESTIMATE OF THE COMPARISON BETWEEN α AND α^*

g	α	α^*	α	α^*	α	α^*
5	.01	.00101	.05	.00512	.1	.01048
10	.01	.00023	.05	.00113	.1	.00232
20	.01	.00005	.05	.00025	.1	.00051

Let us consider another example. Each K_{ij} is determined so that $|\overline{X}_i - \overline{X}_j| > K_{ij}$ would be an α^* rejection for $\mu_i = \mu_j$ under H_0: $\mu_1 = \mu_2 = \ldots = \mu_g$.

When $g = 2$, α^* is the usual α. As g increases α^* must decrease, or the probability will be large that for some i,j, $|\overline{X}_i - \overline{X}_j| > K_{ij}$. The probability of not rejecting $\mu_i = \mu_j$ for any $i \neq j$ is given by

$$\text{probability}\{|\overline{X}_i - \overline{X}_j| \leq K_{ij} \text{ for every } i,j\}.$$

Hence the probability of rejecting at least one is given by

$$1 - \text{probability}\{|\overline{X}_i - \overline{X}_j| \leq K_{ij} \text{ for every } i,j\}.$$

Now let $K \geq K_{ij}$ for every i,j. We have

$$\text{probability}\{|\overline{X}_i - \overline{X}_j| \leq K \text{ for every } i,j\} \geq$$
$$\text{probability}\{|\overline{X}_i - \overline{X}_j| \leq K_{ij} \text{ for every } i,j\}.$$

From the above inequality we conclude that

$$1 - \text{probability}\{|\overline{X}_i - \overline{X}_j| \leq K_{ij} \text{ for every } i,j\} \geq$$
$$1 - \text{probability}\{|\overline{X}_i - \overline{X}_j| \leq K \text{ for every } i,j\}.$$

An example is presented here for purely illustrative purposes. Although it is clear that sample means cannot be distributed uniformly, this example is used because it has the rare quality of having the exact probabilities easy to compute and display.

Let $f(X) = \begin{cases} 1 & 0 \le X \le 1 \\ 0 & \text{elsewhere} \end{cases}$

and $\bar{X}_{(1)} < \bar{X}_{(2)} < \ldots < \bar{X}_{(g)}$

where $\bar{X}_{(i)}$ is the i^{th} order mean

$$1 - \text{probability } \{|X_i - X_j| < K \text{ for every } i,j\} =$$

$$1 - \text{probability } \{|X_{(1)} - X_{(g)}| < K\}$$

$$= 1 - g(g-1) \int_0^k \int_{-\infty}^{\infty} [F(X+Y) - F(Y)]^{g-2} f(Y) f(X+Y) \, d_Y \, d_X$$

$$= 1 - g(g-1)\left[\frac{K^{g-1}}{g-1} - \frac{K^g}{g} \right] = 1 - gK^{g-1} + (g-1) K^g \le \alpha .$$

Then $gK^{g-1} - (g-1) K^g \le 1 - \alpha.$

Now let us for given α and g compute the proper value of K in our equation. Let g = 2 in the equation above. Substituting our value of K in the resulting equation we have $\alpha^* = (1-K)^2$. The values are tabulated in Table 2.

It is easy to see that there is one and only one value of K, say K_*, for which $0 < K_* < 1$ and $gK_*^{g-1} - (g-1) K_*^g = 1 - \alpha$. The level α^* can be computed by letting g = 2. That is $\alpha^* = (1 - K_*)^2$.

The following illustration is a more practical one, where it is possible to compute the relation between α and α^*. Suppose there are g groups, with each group having n observations. The observations y_{ij} are independently and identically distributed (i.i.d.) normally with mean μ and variance σ^2. Let $\bar{Y}_{i.}$ be the mean of the i^{th} group.

TABLE 2

AN ESTIMATE OF THE COMPARISON BETWEEN α AND α^* FOR GIVEN g

	$\alpha = .01$		$\alpha = .025$		$\alpha = .05$		$\alpha = .1$	
g	K_*	α^*	K_*	α^*	K_*	α^*	K_*	α^*
2	.90000	.01	.84188	.02500	.77639	.05	.68377	.1
3	.94109	.00347	.90570	.00889	.86464	.01831	.80419	.03833
4	.95800	.00176	.93241	.00456	.90238	.00927	.85744	.02032
5	.96731	.00106	.94725	.00278	.92355	.00584	.88776	.01259
6	.97323	.00071	.95672	.00187	.93715	.00395	.90740	.00857
7	.97733	.00051	.96330	.00134	.94662	.00284	.92117	.00621
8	.98034	.00038	.96814	.00101	.95361	.00215	.93137	.00470
9	.98264	.00030	.97185	.00079	.95897	.00168	.93923	.00369
10	.98446	.00024	.97478	.00063	.96322	.00135	.94547	.00297
11	.98593	.00019	.97716	.00052	.96668	.00111	.95054	.00244
12	.98715	.00016	.97913	.00043	.96953	.00092	.95475	.00204
13	.98817	.00013	.98074	.00036	.97194	.00078	.95830	.00174
14	.98904	.00011	.98220	.00031	.97400	.00067	.96134	.00149
15	.98980	.00010	.98342	.00027	.97577	.00058	.96396	.00129

Suppose $\bar{Y}_{(1.)} \leq \bar{Y}_{(2.)} \leq \bar{Y}_{(3.)} \leq \cdots \leq \bar{Y}_{(g.)}$ where $\bar{Y}_{(i.)}$ is the i^{th} order mean.The statistic $|\bar{Y}_{(g.)} - \bar{Y}_{(1.)}|/(\hat{\sigma}/\sqrt{n})$ is the studentized range which for given α is denoted by $q_{\alpha,(n-1)g}$. In order to apply the studentized range for given α to determine the individual protection level, α^*, the following steps are taken:

(1) find $q_{\alpha,(n-1)g}$

(2) $(q_{\alpha,(n-1)g}) / \sqrt{2} = t_{\alpha^*,(n-1)g}$

(3) consult the student-t table to find α^*.

Let n = 25. Table 3 gives the value of α^* for given values of α.

TABLE 3

AN ESTIMATE OF THE COMPARISON BETWEEN α AND α^*
FOR GIVEN g WITH SAMPLE SIZE = 25

	g = 5	g = 10	g = 20
α = .05	α^* = .007	α^* = .0016	α^* = .0005
α = .01	α^* = .0013	α^* = .0004	α^* = .00085

An estimate of the comparison between α^*'s obtained in Tables 1, 2, and 3 for given α and g are given in Table 4.

Two examples are now given which illustrate the need for the new methods discussed in this dissertation. The information is given in Table 5. In the table the columns represent the age groups as shown and the rows represent calcium levels of the blood. The number in each cell represents the number of persons who fell into that category. The groups studied consisted of white females only.

TABLE 4

COMPARISON BETWEEN α^*'s OBTAINED IN TABLES 1, 2 AND 3

	g = 5	g = 10	g = 20
α = .05	α_1^* = .00512	.00113	.00025
	α_2^* = .00584	.00168	NA
	α_3^* = .007	.00160	.0005
α = .01	α_1^* = .00101	.00023	.0005
	α_2^* = .00106	.00024	NA
	α_3^* = .0013	.0004	.00085

NA = Not available

TABLE 5

1,469 WHITE FEMALES CLASSIFIED BY AGE AND CALCIUM LEVEL OF BLOOD

Calcium Blood Level mg%	Age Group 10-20	20-30	30-40	40-50	50-60	60→
8.00	1	0	1	2	0	0
8.25	0	2	1	0	0	0
8.50	0	5	3	9	2	1
8.75	1	13	6	12	8	1
9.00	11	63	35	60	43	11
9.25	15	77	38	70	60	17
9.50	37	127	72	96	113	21
9.75	23	72	35	43	57	9
10.00	17	71	28	30	47	9
10.25	6	17	12	11	1	2
10.50	3	2	1	9	5	1
10.75	0	0	0	1	2	0
11.00	0	0	0	1	0	0
> 11.00	1	1	0	0	0	0

The researcher is interested in the following questions:

(1) Is $F_i = F_j$ for each pair i and j as i, j = 1, ..., 6, or in words, do group i and group j come from the same distribution for each i and j?

(2) If disjoint pairs of subsets of the groups can be considered to be identically distributed with each of the subsets, and therefore pooled to yield a distribution for each subset, the question is whether these two distributions are significantly different.

The questions asked in (1) and (2) are questions about the contrasts which will be discussed in CHAPTER III.

The second example to be considered is concerned with the analysis of the information given in Table 6. Table 6 is identical to the previous one except that it represents phosphorus levels of the blood. The questions to be asked would be identical to those of Table 5.

The questions asked above cannot be answered by existing multiple comparisons methods. One of the purposes of this dissertation is to present a method for dealing with this type of comparison.

The two following paragraphs summarize the related principal contributions of this dissertation to the field of multiple comparisons in statistics. Both have been briefly mentioned before. Because of the extremely complex intrinsic nature of the problems, the analysis is only brought to the point where it is clear as to how to proceed with the use of large scale computers.

In surveying the literature and in examining some applied problems, it is evident that there is a need for specific definitions and

TABLE 6

1,565 WHITE FEMALES CLASSIFIED BY AGE
AND PHOSPHORUS LEVEL IN BLOOD

Phosphorus mg%	Age Group 10-20	20-30	30-40	40-50	50-60	60→
1.	1	2	1	0	0	0
1.733333	0	0	0	0	0	0
2.066666	1	4	1	2	0	0
2.399999	5	9	1	4	0	0
2.733333	6	23	15	26	11	1
3.066666	8	51	27	25	24	5
3.399999	16	79	47	66	55	4
3.733333	38	120	65	99	94	31
4.066666	18	87	43	56	75	14
4.399999	7	45	14	35	52	16
4.733333	12	21	12	25	27	1
5.066666	2	9	2	3	6	0
5.399999	6	0	2	3	2	0
>5.399999	1	0	2	0	0	0

statements of the problems. To be complete, of course, it is necessary to discuss the appropriate measure theory and topology on the class of measures. It is possible for those who are only interested in applications to extract the essential information. However, those who hope to understand the development in its entirety, or who plan to contribute to the creation of much needed new techniques, should find it helpful to have the accurate definitions and statements of the problems.

The new techniques discussed here center around problems arising from having univariate observations on members from several groups of possibly unequal sizes. The observations are considered to be independent and within each group they are considered to have a common distribution function. The contrasts which are natural to the problem are discussed. The application of the general discussion on types of errors and α and β levels to this specific type of problem entails a new look at the Kolmogorov-Smirnov statistics for pairs and the consideration of a new statistic for these purposes.

CHAPTER II

REVIEW OF EXISTING METHODS

In this chapter a description will be given of the principal methods of multiple comparisons. The use of these methods in uncomplicated situations, and more detailed information about multiple comparison tests can be found in books by Miller [11], Seeger [18] and papers by Duncan [1] and R. O'Neill and G. B. Wetherill [14]. Because the well known parametric techniques are presented only to provide a background for the general models, the discussion, here, of parametric methods will be very brief.

The methods to be discussed can be classified in two groups according to whether the same critical level is used for all contrasts, as in

(1) Multiple t-method

(2) Fisher Method

(3) LSD Method

(4) Tukey's T Method (1951)

(5) Scheffé's S-method,

or variable critical levels are used, as in

(1) Newman-Keuls Method

(2) Tukey's Method (1953)

(3) Duncan's Method (1955).

Some parametric methods will be outlined briefly.

1. Multiple t-method. This method consists of making a number of ordinary t-tests instead of using an F-test.
2. LSD and FSD. Whenever the overall F-test is significant; then all the differences between pairs of means are compared with the least significant difference (LSD) which is

$$D_{LSD} = t_{\alpha}^{v} \hat{\sigma}_{\bar{y}} \cdot \sqrt{2}.$$

We are not always interested in all comparisons between all pairs of means. Consequently Fisher [5] suggested in 1935 that if one is interested, say, in m comparisons, we use α/m for our significance level. This is called the FSD method.

3. The S-method. The Scheffé [16] S-method is a method which is based on the F-distribution in such a way that at least one linear contrast will be different from zero if and only if the overall F-test gives a significant result. Suppose $\bar{y}_i$ has a variance of $K_i \sigma^2$. Then with probability $1 - \alpha$ all the contrasts $\Psi = \Sigma c_i \mu_i$ will satisfy

$$|\hat{\Psi} - \Psi| \leq \hat{\sigma} \sqrt{(p-1) F(\alpha,p-1,v) \sum_{1}^{p} c_i^2 K_i}$$

where $F(\alpha,p-1,v)$ is the 100α percent of the F distribution with parameter $(p-1,v)$.

4. T-method. The Tukey [23] method is based on the studentized-range method. It states that with probability $1 - \alpha$ all possible contrasts simultaneously satisfy

$$|\hat{\Psi} - \Psi| \leq q(\alpha,a,v) \cdot \hat{\sigma}_{\bar{y}} \left(\frac{1}{2} \sum_{i=1}^{a} |c_i|\right)$$

where $q(\alpha,a,v)$ is the upper α-point in the distribution of the studentized range.

(5) Newman-Keuls test. This is a test proposed by Newman [13] and Keuls [8]. In this test the critical differences depend on the number of means under comparison. The observed treatment means are arranged in order of magnitude and one compares the largest with the smallest. If $|\bar{y}_{max} - \bar{y}_{min}| > q(\alpha,p,v)\hat{\sigma}_{\bar{y}}$ holds then one compares two sets of p-1 means, that is $\bar{y}_1,\ldots,\bar{y}_{p-1}$ and $\bar{y}_2,\ldots,\bar{y}_p$ with the statistic $q(\alpha,p-1,v)$. This process is continued until no significances are obtained.

(6) Tukey method. Tukey [14] has suggested a method identical to the Newman-Keuls method except for the use of the statistic $\frac{1}{2}\{q(\alpha,a,v) + q(\alpha,p,v)\}\cdot\hat{\sigma}_{\bar{y}}$, rather than the one given in the Newman-Keuls tests.

(7) Duncan new multiple range method. Duncan [2] proposed a method wherein he uses a statistic whose significance level depends on the number of means under comparison. His statistic for k means is $q(\alpha_k,K,v)\hat{\sigma}_{\bar{y}}$ where $\alpha_k = 1-(1-\alpha)^{K-1}$.

(8) Dunnett's method. When researchers include a standard treatment or control treatment as one of p treatments, he may be interested in all comparisons with the standard or control. Dunnett proposed a method for this purpose. Dunnett's significant difference for a two-sided comparison between one of p treatment means and the control is

$$|d|^{\alpha}_{(p,v)}\hat{\sigma}_{\bar{y}} \cdot \sqrt{2}$$

where $|d|^{\alpha}_{(p,v)}$ is 100α percent point of $\dfrac{\max_{1\le i\le p}\{|z_i - z_0|/\sqrt{2}\}}{\chi^2_v/v}$,

where z_k, i=0, 1, ...,p are independent unit normal deviates and the χ^2_v are independent χ^2 variables with v degrees of freedom.

There are problems in medical research in which the assumptions of normality underlying the analysis of variance are not valid. For this research non-parametric, or distribution free, procedures have been developed. Unfortunately most of these procedures do not provide for multiple comparisons.

Non-parametric methods are used when there are enough reasons for the researcher to believe that the observations are not normally distributed, nor follow some other standard parametric distribution. In this section, five important non-parametric methods will be discussed. More detail can be found in Miller.

1. Many-one sign statistic [20]. Let $(x_{0j},x_{1j},\ldots x_{kj})$ be the results of a single trial with the subscript 0 associated with a control group, let $d_{ij} = x_{1j} - x_{0j}$, i = 1, 2, ...k, j = 1, 2, ...n. Then

$$\bar{d}_{ij} = \begin{cases} 1 \text{ if } d_{ij} < 0 \\ 0 \text{ if } d_{ij} > 0 \end{cases} \qquad \text{and}$$

$$r_i = \sum_j \bar{d}_{ij},\ i = 1, 2, \ldots k.$$

To test for significance, compare each r_i with a single tabulated critical value. To achieve a significance statement, the value of r_i should be equal or smaller than the tabulated critical value. For locating treatments which give a smaller response

than the control everything is done as above with the exception that d_{ij} is now taken to be $x_{0j} - x_{ij}$.

In this test the appropriate critical region is one-sided. For a two-tailed test we let $r_i^* = \min(n-r_i, r_i)$. If r_i^* is equal or smaller than the tabulated critical value we can make a significance statement.

2. Many-one rank statistic [21]. This test is another non-parametric analogue of Dunnett's procedure. This was the first simultaneous technique based on ranking observations.

Let $\{x_{ij}: i = 0, 1, \ldots k, j = 1, \ldots n_i\}$ be k+1 groups with $n_0, n_1, \ldots n_k$ independent observations, where i = 0 represents a control group. The x_{ij}'s are random variables measuring some characteristic of a control group and k treatment groups and having continuous cumulative distribution functions $F_0, F_1, \ldots F_k$. We wish to test the null hypothesis

$$H_0: F_0 = F_1 = \ldots = F_k.$$

An alternative hypothesis could be one of the following:

$$H_1: F_0 < F_i, \text{ at least one } i$$

$$H_1: F_0 > F_i, \text{ at least one } i$$

$$H_1: F_0 \neq F_i, \text{ at least one } i$$

The test procedure for the i^{th} treatment sample $(x_{i1} \ldots x_{in})$ and the control sample $(x_{01}, \ldots, x_{0m})$ is as follows. Combine the two samples into one sample of size n+m. Rank the observations (1 to the smallest, 2 to the next smallest, ..., n+m to the last observation). Let $R_{i1}, \ldots R_{in}$ be the rank of the observation in the i^{th} group and $R_{01}, \ldots R_{0m}$ be the rank of the observations in

the control group. Let $T_i = \sum_j R_{ij}$ and $T_i' = [n(m+n+1) - T_i]$. Then compare $\min(T,T')$ with the appropriate critical value. If the minimum is smaller than the critical value we reject the null hypothesis. For the one-sided alternative observe whether $\min(T_i,T_i')$ is associated with the control or a treatment group.

3. k-sample rank statistics [22]. This procedure is concerned with pairwise testing of treatments in a one-way classification, or completely randomized design. The model is the same as the previous test.

4. Kruskal-Wallis rank statistic [12]. This method was proposed and analyzed by Nemenyi for a one-way classification. Let x_{ij}: $i = 1, 2, \ldots k$, $j = 1, \ldots n_i$ be k independent samples of size $n_1, \ldots n_k$ of independent observations. Let $\sum_i n_i = N$ and rank the N observations x_{ij} in order of size. Let R_{ij} be the rank of x_{ij} and $\bar{R}_{i.} = i/n_i \sum_j R_{ij}$, $i = 1, \ldots, k$.

With the probability greater than $1 - \alpha$, the inequalities

$$|\bar{R}_{i.} - \bar{R}_{i'.}| \leq (h_{k-1}^{\alpha})^{\frac{1}{2}} \left[\frac{N(N+1)}{12}\right]^{\frac{1}{2}} \left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)^{\frac{1}{2}}$$

hold simultaneously for $\binom{K}{2}$ pairs of the population (i,i') where for large N, h_{k-1}^{α} is approximately $\chi_{k-1}^{2\alpha}$. In the special case where $n_i = n$ for all i and n is sufficiently large, then with probability approximately $1-\alpha$ the inequalities

$$|\bar{R}_{1.} - \bar{R}_{i'.}| \leq q_k^{\alpha}(\infty) \left[\frac{K(KN+1)}{12}\right]^{\frac{1}{2}} \qquad i \neq i'$$

hold simultaneously, where $q_k^{\alpha}(\infty)$ is the upper percentile point of the range k independent unit normal random variables.

5. Friedman Rank statistic [12]. This procedure was proposed by Nemenyi for two-way classifications with one observation per cell in which the statistician is interested in comparing all treatment populations pairwise or in comparing all treatment populations with a control.

Let x_{ij}: i = 1, ...k, j = 1, ...n be k independent samples of n independent observations. Let R_{ij} be the rank of x_{ij} relative to the order of $x_{(1)j} < x_{(2)j} < \ldots < x_{(k)j}$ in block j, and

$$R_i = 1/n \sum_{j=1}^{n} R_{ij}$$

with probability greater than $1-\alpha$. The inequalities

$$|\bar{R}_{i.} - \bar{R}_{i'.}| \leq (\chi_r^{2\alpha})^{\frac{1}{2}} \left[\frac{K(K+1)}{6n}\right]^{\frac{1}{2}} \quad i, i' = 1, 2, \ldots, k$$

holds simultaneously. When n is large then with probability (approximately) $1 - \alpha$, the inequality

$$|\bar{R}_{i.} - \bar{R}_{i'.}| \leq q_k^{\alpha}(\infty) \left[\frac{K(K+1)}{12n}\right]^{\frac{1}{2}} \quad i, i' = 1, \ldots, k$$

holds simultaneously.

It would be advantageous to mention something about applicability of each of these tests and see how they compare. The parametric methods will be discussed first. For a fuller discussion of these methods, the interested reader should consult books by Miller and Seeger referred to in this dissertation and the unpublished paper by Tukey [23]. Some statisticians, in particular, Federer [4] suggest that significant

F-values for treatments, or some prior knowledge that treatment differences exist, may be required before performing a multiple comparisons test. The majority of statisticians long ago replaced the multiple-t method with the LSD method; however, some are not satisfied with the LSD method since in some instances it is considered to give too many significances. The LSD method applies the same procedure as the S- and T-methods, but the least significant differences are computed differently.

The T-method is more powerful than the S-method in making pairwise comparisons. However, the S-method is better for contrasts involving more than two means. Application of the Newman-Keuls or Duncan multiple range tests is limited to testing pairwise mean differences in a balanced one-way classification. The choice between these two tests depends on whether one agrees with Duncan's use of error rate per degree of freedom concept.

The application of the many-one sign method is restricted by its design. The number of observations must be the same for each treatment and control group and the observations in each of the n blocks must be k+1. The model for this test and the Steel many-one rank test is the same as Dunnett's with the exception of normality. If the observations do not happen in blocks then the many-one rank statistic test can be used. Another disadvantage of the many-one sign test is that control and treatment groups must have an equal number of observations, which handicaps the researcher who would like to have more in the control group than in the treatment groups. The Steel-Dwass k-sample test is analogous to the Tukey studentized test. The Kruskal-Wallis rank test is the

non-parametric rank analogue of Scheffé's test for one-way analysis of variance, Tukey's studentized range, and Dunnett's test and is a competitor to the many-one rank statistic and the k-sample rank statistic. When the assumptions hold, the many-one rank test and the k-sample rank test are preferable to the Krushkal-Wallis Rank statistic. One of the disadvantages of this test is that the test statistic i versus i' depends upon observations from other groups. The advantage of this test is that it can be applied in some situations where the many-one rank and k-sample rank test could not, that is when sample sizes are not small.

CHAPTER III

MATHEMATICAL FORMULATION OF THE PROBLEM

The object of this chapter is to present a complete discussion, with definitions, of certain general aspects of multiple comparisons and to give some details of the development in a few practical cases which have not been given before.

A standard basic unit for building a statistical theory of inference is a triple (M, Γ, p) where M is a set called the model, or collection of simple events; Γ is a σ-field of subsets of M (i.e. a collection of subsets which contains M and is closed under complementation and countable unions); and p is a probability function on the collection Γ of events to the closed interval [0,1]. A probability function on a σ-field to [0,1] is one which is countably additive and $p(M) = 1$. Probability functions are countably additive measures, so the pertinent theorems of measure theory apply.

A topology τ on a set S is a collection of subsets of S, called the open sets of the topology, which contains the empty set ϕ, the set S, and is closed under arbitrary unions and finite intersections. The collection of topologies, like the collection of σ-fields, on a set are closed under arbitrary intersections. Hence any collection of subsets of a set has a smallest topology containing it and a smallest σ-field containing it, namely the intersection of all topologies, or σ-fields,

containing the collection. The Borel sets of a topology are the sets in the smallest σ-field containing the topology. Because the collection of all subsets of a set is both a topology and a σ-field, it follows that any nonempty collection of subsets has a nonempty smallest topology and a nonempty smallest σ-field containing it. The closed sets of a topology are those sets whose complements are open.

The Cartesian product of a collection $\{S_i\}_{i\in I}$ of sets on some index set I is written $\underset{i\in I}{\times} S_i$, or if $I = \{1,2,\ldots,n\}$, $\underset{i\in I}{\times} S_i = S_1 \times S_2 \times \ldots \times S_n$, or $S_1 \times S_2 \times S_3 \ldots$ if $I = \{1,2,3,\ldots\}$ and defined to be the set of all functions $x: i \to x_i$ where $x_i \in S_i$ for all $i \in I$. Such an element is also sometimes written $\{x_i\}_{i\in I}$ or simply $\{x_i\}$. If each S_i has a topology τ_i and $A = \{i_1,\ldots,i_K\}$ is any finite nonempty subset of I, $\{U_i\}_{i\in A}$ is a collection $\{U_{i_1},\ldots,U_{i_K}\}$ of nonempty open sets, where $U_{i_j} \in \tau_{i_j}$ for $j = 1,2,\ldots,K$, then an open box $[\{U_i\}_{i\in A}]$ is defined $\{\{x_i\}_{i\in I} \mid$ for every $i \in A$, $x_i \in U_i\}$. The product topology, written $\underset{i\in I}{\times} \tau_i$ is simply the smallest topology on the Cartesian product which contains all of the open boxes.

For the real numbers R it is convenient to use the following notations for the nine types of intervals. It is assumed that a and b are real numbers and $a<b$. The first three are called open intervals, the next two are called half-open intervals and the next three are closed intervals. The last is both open and closed.

$$(-\infty,a) = \{X \mid X < a\}$$

$$(a,b) = \{X \mid a < X < b\}$$

$$(b,\infty) = \{X \mid b < X\}$$

$$(a,b] = \{X \mid a < X \leq b\}$$

$$[a,b) = \{X \mid a \leq X < b\}$$

$$[a,b] = \{X \mid a \leq X \leq b\}$$

$$[b,\infty) = \{X \mid b \leq X\}$$

$$(-\infty,a] = \{X \mid X \leq a\}$$

$$(-\infty,\infty) = \{X \mid X \in R\} = R$$

The natural topology for R, and the one used henceforth, as it is that which is basic to statistics, is the smallest topology containing the open intervals. The Borel sets of R, B(R), are those for the natural topology. Because of the theorem that every open set in the natural topology of R is the union of a countable collection of disjoint open intervals, it is easy to see that any one of the collections of intervals of one of the above nine types generates the Borel sets of R.

In order to put this problem in its proper mathematical context, as promised in CHAPTER I, it is necessary to examine the parameter space more closely. Without Bayesian statistics it is customary to put some kind of topology on the parameter space for determining when an estimate is close to a parameter. This is usually done by assigning a metric, and, if possible, it is convenient to have the metric be a norm on a vector space. In the vector space of all functions on the two-point compactification $R^{\#} = R \cup \{\infty\} \cup \{-\infty\}$ to R, the smallest subspace which contains the distribution functions is the space V of functions of bounded variation on $R^{\#}$ which take the value 0 at $-\infty$ and are continuous at this point as well as being continuous at ∞. The subspace V_0 of V, consisting of those functions which are continuous on $R^{\#}$, is the smallest subspace containing the continuous distributions. Certainly for many purposes a natural norm

for this space is the sup norm. That is, $||f|| = \sup_{x \in R^{\#}} |f(x)|$. Then the metric, or distance, between two functions is $||f-g||$. This is the norm used by Kolmogorov. On occasion, when the null hypotheses specifies a continuous distribution function F, it is convenient to use one of the Minkowski norms. Since all functions in V are Riemann-Stieltzes or Lebesque-Stieltzes p^{th} power integrable with respect to F for any $p > 1$ it is possible to define a norm by

$$||f||_p = [\int_{-\infty}^{\infty} |f(x)|^p \, dF(x)]^{1/p}$$

The most common value used is when $p = 2$, or less frequently when $p = 1$.

In considering g sets of random variables with distribution functions $F_1, F_2, \ldots, F_g$ the parameter space is a subset of $V_g = V \times V \times \ldots \times V$, g times. Norms on V_g will be defined by combining a norm on $R_g = R \times R \times \ldots \times R$, g times with a norm on V as follows. Any norm on R_g of interest in this problem will have the property that for all $X = (X_1, \ldots, X_g)'$,

$$||X|| = ||(|X_1|, |X_2|, \ldots |X_g|)'|| = \phi(|X_1|, \ldots |X_g|)$$

where ϕ is so defined. This property assures that if $0 \leq y \leq X$ in the usual partial order on R_g, then $\phi(y) \leq \phi(X)$, which in turn is sufficient to show that

$$||(f_1, f_2, \ldots, f_g)'|| = \phi(||f_1||, ||f_2||, \ldots ||f_g||)$$

defines a norm on V_g for any such norm on R_g and any norm on V.

A random variable is a function from M to R, or to a topological vector space V, such that inverse images of Borel sets are in the σ-field Γ. If the image is in R, the random variables are univariate, if in $R \times R \times \ldots \times R$ they are multivariate, and if in an infinite dimensional Banach

space they are random variables in a stochastic process.

Let $E_1, E_2, \ldots, E_K \in (M, \Gamma)$ be K events. They are said to be independent if and only if $p(\cap E_i) = \Pi\, p(E_i)$ where p is the probability measure on (M, Γ).

Let $X_1, X_2, \ldots, X_m$ be random variables on (M,Γ,p) to V. The set of random variables is said to be independent if for any collection $B_1, \ldots, B_m$ of Borel sets in V, the inverse images.

$$E_i = X_i^{-1}(B_i) = \{\omega | X_i(\omega) \in B_i, \omega \in M\}$$

are independent sets in Γ relative to p. That is

$$p(\bigcap_{i=1}^{m} E_i) = \prod_{i=1}^{m} p(E_i)$$

except for the cases where the random variables are different quadratic forms or other composite functions of a single random variable, independence rarely ever appears in this context. In the common case of sampling from a set $X_1, \ldots, X_m$ of random variables it is technically easier to describe a new set of random variables

$$X_1^*, \ldots, X_n^* \text{ by } X_i^*: M_1 \times M_2 \times \ldots \times M_n \to V, \text{ where}$$

$$X_i: M_i \to V, \text{ and } X_i^*: (\omega_1, \ldots, \omega_n) \to X_i(\omega_i).$$

Independence is then assured by using the product probabilities for $M_1 \times M_2 \times \ldots \times M_n$. That is, if $B_1, \ldots, B_m$ are Borel sets in V, $E_1, \ldots, E_m$ are the inverse images

$$E_i = \{\omega_i | X(\omega_i) \in B_i\}, \quad \text{then}$$

$$p(E_1 \times E_2 \times \ldots \times E_m) = \prod_{i=1}^{m} p(E_i).$$

If $E_i^* = \{(\omega_1,\ldots,\omega_n) | X_i^*(\omega_1,\ldots,\omega_n) \in B_i\}$ then

$E_1 \times E_2 \times \ldots \times E_m = E_1^* \cap E_2^* \cap \ldots \cap E_m^*$ and this concept of independence is a special case of the above definition.

Now suppose X is a random variable defined on the space (M,Γ,p). The point function

$$F(x) = p\{X(\omega) \leq x\} = p\{\omega | X(\omega) \leq x\} = p\{X^{-1}(-\infty, x)\}$$

defined on R is called a distribution function and has several important properties:

(1) $F(-\infty) = 0, \quad F(+\infty) = 0$

(2) $F(x_1) \geq F(x_2)$ if and only if $x_1 \geq x_2$

(3) F is right-hand continuous.

It is necessary to exercise some care in discussing the sampling distribution when it is to be used as a statistic for estimating a distribution. If $\{X_i\}_{i=1}^n$ is a set of independent and identically distributed random variables with distribution function F where $X_i: M \to R$ and $X_i: \omega_i \to X_i(\omega_i) = x_i$ is a particular sample, then the statistic

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} U(x - x_i)$$

is an unbiased estimate of F(x) for each x. Here $U(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$ gives the distribution function for a random variable which takes on only the value zero. In order to consider the function $\hat{F}: x \to \hat{F}(x)$ as a distribution function itself, it is necessary to look at a new set of random variables. If M is a set $\{a_1, \ldots, a_n\}$ of n elements with each point having probability 1/n and $Y: a_i \to x_i$ then this discrete random variable has $\hat{F}$ for a distribution function.

The expression, distribution of the statistic, also needs to be more explicitly stated in this context. In considering a set of random variables $\{X_1, \ldots, X_n\}$, each of which maps M to R, again, it helps to use the functions X_i^*: $(\omega_1, \ldots, \omega_n) \to X_i\ (\omega_i)$. That is, X_t^*: $M \times M \times \ldots \times M \to R$. Then $X_1^* + X_2^* + \ldots + X_n^*$ is given by ordinary addition of functions. If g is a function on $R \times R \times \ldots \times R$ to R it is called a statistic when its values are realizations of a random variable Z: $\omega = ((\omega_1, \ldots, \omega_n) \to g(X_1^*\ (\omega), X_2^*\ (\omega), \ldots X_n^*\ (\omega))$. The distribution of this random variable is called the distribution of the statistic. The statistic given by $\hat{F}(x)$, with corresponding random variable Z_x, is a stochastic process with index set the interval [0, 1], where Z_x is defined by

$$Z_x = \frac{1}{n} \sum_{i=1}^{n} U(x - X_i^*).$$

If $\{X_1, \ldots, X_n\}$ are independent and identically distributed, F, then for each x, Z_x has a binomial distribution in which

$$\text{prob}(\{\omega | Z_x(\omega) \leq t\}) = \sum_{0 \leq K \leq nt} \binom{n}{K} [F(x)]^K [1 - F(x)]^{n-K}.$$

Multiple comparisons also multiply the number of subjective judgments to be made. The additional types of errors which can be made from making multiple inferences increases the difficulty of communicating results in such a way that it is clear as to what the probabilities are of the outcomes being due only to the random nature of the phenomena. For this reason the subject of error rates is often discussed in connection with problems where many conclusions are to be reached from one set of data. There is no attempt here to discuss hypothetical error rates of the type involved with imagining an experiment to be repeated a certain number of times and dividing the number of errors by the number

of trials. Nor, for reasons which follow, is the concept of experiment-wise error rate discussed, in which the statistic used is the number of experiments involved, having at least one erroneous inference, divided by the number of experiments. The position taken here is that if inferences are to be made between outcomes of several experiments, they shall be considered as part of one experiment. If no cross inferences are to be made the experiments shall be treated individually.

In CHAPTER I the variety of Type I errors was discussed. If an experiment has a finite set of at least two inferential statements to be made, and if regions have been determined for each in the sample space for which the hypothesis is to be rejected, then type one errors occur with many possible conditions in the parameter space. Because of this complexity, and the position taken that one overall experiment is being considered, it is assumed here that errors refer to the situation in which all hypotheses are considered to hold. In this context each rejection becomes an error, relative to the hypothesis, and the error rate statistic is the ratio of the number of rejections to the total number of statements. The expected value of this statistic depends on the distribution and it is not often that requiring all hypotheses to hold simultaneously determines a unique distribution for the statistic. It is desired, in communicating the outcomes of an experiment, to establish that reasonable attempts are being made to hold down possible errors. For this reason the least upper bound of the expected values of the error rate statistic under the assumption that all hypotheses hold is referred to here as the expected error rate.

Suppose for example that K hypotheses of the form H_i: $\Theta \epsilon \omega_i$,

$i = 1, 2, \ldots, K$, and R_i, $i = 1, 2, \ldots, K$ are the corresponding rejection regions in the sample space. Let X_i be the binary random variable which is the composition of the function $\chi_{R_i} : x \to \begin{cases} 1 & x \in R_i \\ 0 & x \notin R_i \end{cases}$, and the total sample random variable. If x_i is a sample for X_i then $\bar{x} = \frac{1}{K}(x_1+\ldots+x_K)$ is the error rate statistic. For any $\theta \in \omega = \bigcap_{i=1}^{K} \omega_i$ the probability $\tilde{p}_\theta(R_i)$ is the probability that $X_i = 1$ and $E(X_i) = \tilde{p}_\theta(R_i) \leq \alpha_i = \sup_{\theta\in\omega} \tilde{p}_\theta(R_i)$, the level for the i^{th} individual inference. The expected error rate

$$\sup\left(\frac{1}{K}\sum_{i=1}^{K} E_\theta(X_i)\right) = \sup_{\theta\in\omega}\left(\frac{1}{K}\sum_{i=1}^{K} \tilde{p}_\theta(R_i)\right) \leq \frac{1}{K}\sum_{i=1}^{K} \sup_{\theta\in\omega} \tilde{p}_\theta(R_i) = \frac{1}{K}\sum_{i=1}^{K} \alpha_i^* .$$

It is reasonable to set $\alpha_i^* = \alpha^*$, $i = 1, 2, \ldots, K$. Then $\frac{1}{K}\sum_{i=1}^{K} \alpha_i^* = \alpha^*$.

The term contrast is used classically in three different ways. If the parameter space is an n-dimensional vector space then a contrast is an element of the dual space which maps the constant vectors to zero. That is, if $C' = (c_1, \ldots, c_n)$, $J' = (1, 1, \ldots, 1)$, $\theta = (\theta_1, \ldots, \theta_n)$, and C is used for the linear function, as well as its vector of coefficients, then $C: \theta \to C'\theta$ is a contrast if $C'J = 0$, or $\sum_{i=1}^{n} c_i = 0$. It is common and not confusing to call either of the linear combinations $C'\theta$ or $C'\hat{\theta}$ a contrast, where $\hat{\theta}$ is an estimate of θ.

The situation in nonparametrics is similar, but somewhat different. In the problem considered here of sampling from g independent random variables $X_1, X_2, \ldots, X_g$, having continuous distribution functions $F_1, F_2, \ldots, F_g$, respectively, the parameter space is the product of g convex subsets of the infinite dimensional function space $V_g = V \times V \times \ldots \times V$, discussed on page 29. If $\theta' = (F_1, \ldots, F_g)$ then for every x,

$\theta'(x) = (F_1(x), \ldots, F_g(x))$ is an element in the g-dimensional space $R \times R \times \ldots \times R$, g times. General contrasts have no statistical meaning in this setting. However, the researcher is often interested in comparing various pairs of pooled subgroups of the treatment groups. If $I_g = \{1, 2, \ldots, g\}$ is the index set for the groups and H is a subset of the set of all pairs (A,B) of non-empty disjoint subsets of I_g, it is reasonable to ask the following question. If the random variables $\{X_i\}$ $i\varepsilon A$ are considered to have a common distribution function and if $\{X_j\}$ $j\varepsilon B$ have a common distribution function, are the two the same? Suppose there are n_i samples from $\{X_{ij}\}$, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, g$, where $X_{ij} \sim F_i$, $i = 1, 2, \ldots, g$. It is assumed that the variables are independent and, in the Kolmogorov-Smirnov case, that the distribution functions are continuous. For each x, the sampling distribution function $\hat{F}_i(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} U(x - x_{ij})$ is an unbiased estimate of $F_i(x)$, where x_{ij} is a sample from X_{ij}, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, g$. For a non-empty subset $A \subseteq I_g$, the sampling distribution for the pooled data $\{x_{ij}\}$, $j = 1, 2, \ldots, n_i$, $i\varepsilon A$, is

$$\hat{F}_A(x) = \frac{1}{n_A} \sum_{i\varepsilon A} \sum_{j=1}^{n_i} U(x - x_{ij}) = \sum_{i\varepsilon A} \frac{n_i}{n_A} \frac{1}{n_i} \sum_{j=1}^{n_i} U(x - x_{ij}) = \sum_{i\varepsilon A} \frac{n_i}{n_A} \hat{F}_i(x),$$

where $n_A = \sum_{i\varepsilon A} n_i$. A statistic to be used in comparing the subgroups determined by $(A,B) \varepsilon H$, is $\hat{F}_A(x) - \hat{F}_B(x)$. If C_{AB} is a vector whose i^{th} component is $\frac{n_i}{n_A}$ if $i\varepsilon A$, $\frac{-n_i}{n_B}$ if $i\varepsilon B$, 0 if $i \notin A \cup B$, then $C'_{A,B}\hat{\theta}(x) = \sum_{i\varepsilon A} \frac{n_i}{n_A} \hat{F}_i(x) - \sum_{i\varepsilon B} \frac{n_i}{n_B} \hat{F}_i(x) = \hat{F}_A(x) - \hat{F}_B(x)$. The $C_{A,B}$ is a special

contrast.

To clarify the above definition of this special contrast, a simple example will be given. For further examples the reader should refer to CHAPTER V which is devoted to several applied problems. Suppose the researcher has four groups with five observations in the first group, four in the second, six in the third and five in the fourth. He proposes to see if the distribution of group 1 is the same as that of groups 2 and 4 combined. In this example, $J = \{1, 2, 3, 4\}$, $A = \{1\}$, $B = \{2,4\}$. The appropriate contrast is

$$C_{AB} = \Sigma \frac{n_i}{n_A} \hat{F}_i - \Sigma \frac{n_K}{n_B} \hat{F}_K = \frac{5}{5} F_1 - \{\frac{4}{9} F_2 + \frac{5}{9} F_4\} = \hat{F}_1 - \frac{1}{9}\{4\hat{F}_2 + 5\hat{F}_4\} .$$

Any multiple of a contrast is essentially the same contrast. There are two natural ways to normalize such a contrast. Let Z_A be the random variable for the sampling distribution $\hat{F}_A$. That is,

$$Z_A(x) = \frac{1}{n_A} \sum_{i \in A} \sum_{j=1}^{n_i} U(x - X_{ij}) = \sum_{i \in A} \frac{n_i}{n_A} Z_i(x).$$

Then if the range of $\{Z_A(x) - Z_B(x)\}$, $(A,B) \in H$, is to be considered it is convenient to leave $C_{A,B}$ as is, because when $\{F_i(x)\}$ are ordered $F_{(1)}(x) \leq F_{(2)}(x) \leq \ldots \leq F_{(g)}(x)$, it follows from the nature of convex combinations that

$$|C'_{AB}\, \Theta\, (x)| = \left| \sum_{i \in A} \frac{n_i}{n_A} \hat{F}_i(x) - \sum_{j \in B} \frac{n_i}{n_B} \hat{F}_J(x) \right| \leq F_{(g)}(x) - F_{(1)}(x).$$

For Kolmogorov-Smirnov statistics it is better to multiply $C_{A,B}$ by $\sqrt{\frac{1}{\frac{1}{n_A} + \frac{1}{n_B}}}$ in the large sample case. Let $D_{A,B} = \sqrt{\frac{1}{\frac{1}{n_A} + \frac{1}{n_B}}}\, C_{A,B}$.

Then for large samples $\sup_x |D'_{AB}\,\Theta(x)|$ is approximately distributed Λ, where

$$\Lambda(x) = 1 - 2\sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2x^2}$$

Two theorems on nonparametric contrasts will be given below.

Theorem 3.1. Let g be the number of groups. Then the number of possible C-contrasts is $\frac{3^g-2^{g+1}+1}{2}$.

Proof: The proof will be by induction. Let g = 2. Then there is only one possible C contrast. In this case $(3^g-2^{g+1}+1)/2 = (9 - 8 + 1)/2 = 1$.

Suppose the theorem is true for g = n. Let (A,B) be any pair in H_n, where H_n is the set of C contrasts for n groups. There are four possible types of C contrasts in H_{n+1} corresponding to the pair (A,B), namely

(1) (A,B)

(2) (A U {n+1}B)

(3) (A,B U {n+1}), and

(4) (C, {n+1}),

where n+1 represents the additional group, and C is any set in H_{n+1} not containing n+1. There are $(3^n-2^{n+1}+1)/2$ of type one by the induction hypothesis. Likewise there are $(3^n-2^{n+1}+1)/2$ of types two and three. Now there are 2^n-1 nonempty subsets of H_n. Hence there are 2^n-1 of these. Hence the number of elements in H_{n+1} is

$$3\left[\frac{3^n-2^{n+1}+1}{2}\right] + [2^n - 1]$$

$$= \frac{3^{n+1} - (2+1)2^{n+1} + 3}{2} + \frac{2^{n+1} - 2}{2}$$

$$= \frac{3^{n+1} - 2^{n+2} - 2^{n+1} + 3 + 2^{n+1} - 2}{2}$$

$$= \frac{3^{n+1} - 2^{n+2} + 1}{2}$$

Hence the theorem is true for $g = n+1$. Q.E.D.

Theorem 3.2. Let A and B be subsets of contiguous elements of I_g, the index set. The number of C_{AB} contrasts is $\binom{g}{2} + 2\binom{g}{3} + \binom{g}{4} = \binom{g+2}{4}$.

Proof: Let $R_g = \{(i,j,k,\ell) \mid 1 \leq i \leq j < k \leq \ell \leq g\}$. There is a one-to-one correspondence between the set R_g and the set of all pairs of subsets of contiguous elements of I_g. For any pair (i,j,k,ℓ), there are four possibilities,

(1) $i < j < k < \ell$,

(2) $i = j < k < \ell$,

(3) $i = j < k = \ell$, or

(4) $i < j < k = \ell$.

There are $\binom{g}{4}$ of the first type, $\binom{g}{3}$ of the second type, and fourth type, $\binom{g}{2}$ of the third type. Therefore the number of subsets of contiguous elements of I_g is $\binom{g}{2} + 2\binom{g}{3} + \binom{g}{4}$. Now

$$\binom{g}{2} + 2\binom{g}{3} + \binom{g}{4} = \frac{g!}{4!(g-4)!} + 2\frac{g!}{3!(g-3)!} + \frac{g!}{2!(g-2)!}$$

$$= \frac{g!}{4!(g-2)!}\,[g^2 + 3g + 2] = \frac{g!(g+1)(g+2)}{4!(g-2)!} = \binom{g+2}{4}$$

The following four theorems were published by Kolmogorov in 1933

and a discussion of them in English is found in a 1939 paper by Kolmogorov [9]. They will be presented here for later use without proof.

Theorem 3.3. If the function $F(\xi)$ is continuous then the distribution law of the quantities $D_n = \sup |F(\xi) - F_n(\xi)| \sqrt{n}$ does not depend on $F(\xi)$.

Let $\phi_n(\lambda)$ be the value of the probability $P(D_n \leq \lambda)$. The next theorem and Theorem 4.6 are basic to the subject.

Theorem 3.4. For n tending to infinity the distribution function $\phi_n(\lambda)$ tends to

$$\phi(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2}$$

uniformly with respect to λ.

Theorem 3.5. For any distribution function $F(\xi)$, $P\{D_n \leq \lambda\} \geq \phi_n(\lambda)$.

Theorem 3.6. If the probability law $F(\xi)$ is continuous, then the probability

$$P\{\sup |F_{n_1}(\xi) - F_{n_2}(\xi)| < \lambda\sqrt{\frac{n_1 + n_2}{n_1 n_2}}\} = \phi_{n_1,n_2}(\lambda)$$

is independent of the function $F(\xi)$. If n_1 and n_2 are indefinitely increased subject to the restriction that the ratio n_1/n_2 remains between two fixed numbers a_1 and a_2

$$0 < a_1 \leq n_1/n_2 \leq a_2 < +\infty$$

then

$$\phi_{n_1,n_2}(\lambda) \to \phi(\lambda).$$

In the general case, where the probability law $F(\xi)$ is absolutely arbitrary, we have

$$P\{\sup |F_{n_1}(\xi) - F_{n_2}(\xi)| \leq \lambda\sqrt{\frac{n_1 + n_2}{n_1 n_2}} \} \leq \phi_{n_1,n_2}(\lambda).$$

The definitions and remarks or theorems have been presented for the purpose of stating the most general problems in multiple comparisons to be presented here. If $\{X_{ij}\}_{j=1,2,\ldots,n_i;\ i=1,2,\ldots,g}$ are independent random variables, distributed F_i, i=1,2,...,g, H is a subset of the set of all pairs (A,B) of non-empty disjoint subsets of $I_g = 1,2,\ldots,g$ with $\{C_{(A,B)} | (A,B)\varepsilon H\}$ being the contrasts of the types discussed above which are of interest to the researcher, how are the rejection regions to be determined? Specifically, how are the $\lambda_{A,B}$ and α^* to be determined, for a given α, so that if W_{AB} is the rejection region $\{x| \ ||C'_{AB}\hat{\theta}||>\lambda_{A,B}\}$ for the contrast $C_{A,B}$ where $\sup_{\Theta\varepsilon\Omega_o} p_\Theta(W_{A,B}) = \alpha^*$ for $\Omega_o=\{\Theta|\Theta'=(F,F,\ldots,F)=FJ$ for some F in V}, then $\sup_{\Theta\varepsilon\Omega_o} p_\Theta(\bigcup_{(A,B)\varepsilon H} W_{A,B}) = \alpha$. This problem is discussed in more detail in the next two chapters.

The other class of problems is only alluded to here for the sake of completeness. Scheffé [17] has done some excellent work in the parametric case on simultaneous confidence intervals. A confidence set $\psi(x)$ of confidence $1-\alpha$ is the image of a function ψ on the sample space to the parametric space such that if $\psi^{-1}(\Theta) = \{x \ |\Theta\varepsilon\psi(x)\}$ then $p_\Theta(\psi^{-1}(\Theta)\ ) = 1 - \alpha$. When a metric $\delta(\Theta_1,\Theta_2)$ is introduced in the parameter space and $\hat{\Theta}$ is a statistic being used to estimate Θ it is common to define ψ by determining λ so that $\psi(x) = \{\Theta|\delta(\Theta,\hat{\Theta}(x)) \leq \lambda\}$ is a confidence set of confidence $1 - \alpha$. It is also, often only of interest to know whether certain parameters are covered. For example in the multiple comparison case with contrasts $\{C_{(A,B)} | (A,B)\varepsilon H\}$, where

$\psi_{A,B,\lambda}(x) = \{\Theta \mid ||C'_{A,B}[\Theta - \hat{\Theta}(x)]|| < \lambda\}$ the interest may be only in whether $\psi_{A,B,\lambda}(x)$ covers $\Omega_o = \{\Theta|\Theta \epsilon FJ\}$. If the confidence is set so that $1 - \alpha^* = \inf_{\Theta \epsilon \Omega_o} p_\Theta(\psi^{-1}_{A,B,\lambda}(\Theta))$ then $1 - \alpha = \inf_{\Theta \epsilon \Omega_o} p_\Theta(\cap \psi^{-1}_{A,B,\lambda}(\Theta))$ as in the case of hypothesis testing it is desired to find α^* to insure a given α.

CHAPTER IV

DERIVATION OF METHODS

This chapter will deal with the development of two methods for making multiple comparisons. The use of each will be restricted by the availability of tables, and by assumptions which are necessary in developing the method.

The first method to be discussed here is the F.S.D. test first proposed by Fisher in 1935. Although Fisher suggested the method, no record of his development of it could be found in the literature.

Let $H_1, H_2, \ldots, H_k$ be k hypotheses in a multiple comparison test where H_i represents the hypothesis that $\Theta \epsilon \omega_i$ with a corresponding rejection region R_i such that if $x \epsilon R_i$, H_i would be rejected. The probability, $\alpha_i^* = \alpha^*$, as mentioned in CHAPTER III, of a Type I error, corresponding to H_i is given as $\alpha^* = \sup_{\Theta \epsilon \omega_i} \tilde{p}_\Theta(R_i)$. Let

$$\chi_{R_i}(x) = \begin{cases} 1 \text{ if } x \in R_i \\ 0 \text{ if } x \notin R_i \end{cases}, \qquad \chi_{\omega_i}(\Theta) = \begin{cases} 1 \text{ if } \Theta \epsilon \omega_i \\ 0 \text{ if } \Theta \notin \omega_i \end{cases}$$

and $\Pi_{i,\Theta} = \tilde{p}_\Theta(R_i)$. Under the null hypothesis $\Theta \epsilon \omega_i$ which implies that $\Pi_{i,\Theta} \leq \alpha^*$. Let A be a subset of the set $I_k = \{1, 2, \ldots, k\}$. Then

$$\omega_A = [\bigcap_{i \epsilon A} \omega_i] \cap [\bigcap_{i \notin A} \complement(\omega_i)], \text{ where } \complement \text{ is complement}$$

A Type I error occurs when H_i is rejected and $\Theta \epsilon \omega_i$. This can be

expressed by the product $\chi_{R_i}(x) \cdot \chi_{\omega_i}(\theta)$. This product is equal to one if $x \in R_i$ and $\theta \in \omega_i$. Otherwise the product would be zero. Therefore by the definition of error rate, we obtain the following functional formula.

$$ER(\theta,x) = 1/k \sum_i \chi_{R_i}(x)\, \chi_{\omega_i}(\theta) .$$

From the definition of the characteristic functions for R_i and ω_i, we observe that

$$\text{Prob}\{\chi_{R_i}(x)\, \chi_{\omega_i}(\theta) = 1\} = \begin{cases} \Pi_{i,\theta}, & \theta \in \omega_i \\ 0, & \theta \notin \omega_i \end{cases}$$

We can therefore write the expected value of the product as

$$E(\chi_{R_i}(x)\, \chi_{\omega_i}(\theta)) = \begin{cases} \Pi_{i,\theta}, & \theta \in \omega_i \\ 0, & \theta \notin \omega_i \end{cases} .$$

Hence $E(ER(\theta,x)) = 1/k\, E(\sum_{i \in A} \chi_{R_i}(x)\, \chi_{\omega_i}(\theta)) = 1/k \sum_{i \in A} E(\chi_{R_i}(x)\, \chi_{\omega_i}(\theta))$

$$= 1/k \sum_{i \in A} \Pi_{i,\theta}, \theta \in \omega_A$$

If $\theta \in \bigcap_{i \in I_k} \omega_i$, then for all $i \in I_k$, $\chi_{\omega_i}(\theta) = 1$, and

$$E(ER(\theta,x)) = 1/k \sum_i \Pi_{i,\theta} \leq 1/k \sum \alpha_i^* .$$

Therefore $ER(\theta,x) = \dfrac{\sum \chi_{R_i}(x)\, \chi_{\omega_i}(\theta)}{k} = \dfrac{\sum \chi_{R_i}(x)}{k}$.

Then the number of type one errors, when $\theta \in \bigcap_{i \in I_k} \omega_i$, is $k \cdot ER(\theta) = \sum \chi_{R_i}(x)$, and $E(k \cdot ER(\theta,x)) = E[\sum \chi_{R_i}(x)] = \sum E(\chi_{R_i}(x)) = \sum \Pi_{i,\theta}$. The variance of $k \cdot ER(\theta,x)$ is given as $\text{Var}(k \cdot ER(\theta)) = E[k\, ER(\theta,x) - E(k \cdot ER(\theta,x))]^2 =$

$$E[\Sigma\ \chi_{R_i}(x)\ \chi_{R_j}(x)] - \Sigma\ \Pi_{i,\Theta},\Pi_{j\Theta} = \Sigma(\Pi_{ij,\Theta} - \Pi_{i,\Theta}\Pi_{j,\Theta})$$

where $\Pi_{ij,\Theta} = \tilde{p}_\Theta(R_i \cap R_j) \leq \sqrt{\Pi_{i,\Theta}\ \Pi_{j,\Theta}}$, therefore

$$\text{Var}(k\ ER(\Theta)) \leq \sum_{i,j} [\sqrt{\Pi_{i,\Theta}\Pi_{j,\Theta}}\ (1 - \sqrt{\Pi_{i,\Theta}\ \Pi_{j,\Theta}})].$$

Since $1 > \alpha^* \geq \Pi_{i,\Theta} > 0$ and $1 > \alpha^* \geq \Pi_{j,\Theta} > 0$, we have $\sqrt{\Pi_{i,\Theta}\ \Pi_{j,\Theta}} \leq \alpha^*$. The value $\alpha^*\ (1-\alpha^*)$ is an increasing function between 0 and 1/2 but the value of α^* for all practical purposes lies between 0 and 1/2 so it is clear that $\text{Var}(k \cdot ER(\Theta)) \leq \sum_{ij} \alpha^*\ (1-\alpha^*) = k^2\alpha^*\ (1-\alpha^*)$.

A reasonable explanation of Fisher's admittedly intuitive use of α/k for α^* is the following: If, for example, $\alpha = .05$ is to be the overall α-level then the expected number of type one errors should be no more than one in twenty. Thus, if $\alpha^* = \alpha/k$, $E(k\ ER(\Theta,x)) \leq k\alpha^* = \alpha$ when $\Theta \varepsilon \bigcap_{i \varepsilon I_k} \omega_i$.

From the definitions and theorems in CHAPTER III, one can proceed with development of the second method, which is a new multiple comparisons test of a type designed to help test one particular type of hypothesis presented in CHAPTER I. However the test to be developed here is also applicable, with some difficulty, for the other types of hypotheses presented there.

Consider g groups of independent random variables $\{X_{ij}\}$, $j = 1, \ldots, n_i$, $i = 1, 2, \ldots, g$ where $X_{ij} \sim F_i$, $i = 1, 2, \ldots, g$, and each F_i is continuous. Let $Y_{ij} = F_i(X_{ij})$, $Y_i = F_i(X)$, $y_{ij} = F_i(x_{ij})$, $y_i = F_i(x)$, and $\hat{G}(y_i) = \hat{F}_i(x)$. Because of the continuity and monotonicity of F_i, $y_{ij} \leq y_i$ if and only if $x_{ij} \leq x$, so $U(x - x_{ij}) = U(y - y_{ij})$.

Therefore, $\hat{F}_i(x) = 1/n_i \sum_{j=1}^{n_i} U(x - x_{ij}) = 1/n_i \sum_{j=1}^{n_i} U(y_i - y_{ij}) = \hat{G}_i(y_i)$,

and $\hat{F}_A(x) = \sum_{i\in A} (n_i/n_A)\hat{F}_i(x) = \sum_{i\in A} (n_i/n_A)\hat{G}_i(y_i)$. In the case of the null hypothesis where $F_1 = F_2 = \ldots = F_g = F$, and $Y = F(X)$, $y = F(x)$, $\hat{F}_A(x) = \hat{G}_A(y)$. Therefore

$$\sup_x \left| \frac{\hat{F}_A(x) - \hat{F}_B(x)}{\sqrt{1/n_A + 1/n_B}} \right| = \sup_y \left| \frac{\hat{G}_A(y) - \hat{G}_B(y)}{\sqrt{1/n_A + 1/n_B}} \right| .$$

The random variables Y_{ij} and Y are each distributed uniformly in the interval $[0,1]$. For fixed y, these random variables can be transformed into Bernoulli - type random variables by considering

$$Z_{ij}(y) = U(y - Y_{ij}) = \begin{cases} 1, & Y_{ij} < y \\ 0, & Y_{ij} \geq y \end{cases} .$$

Clearly for $y < 0$, $Z_{ij}(y) = 0$ and for $y > 1$, $Z_{ij}(y) = 1$.

Let A be a subset of $J = \{1, \ldots, g\}$. Then the sampling distribution of combinations of observations over the set A is given as follows:

$Z_A(y) = 1/n_A \sum_{i\in A} \sum_{j=1}^{n_i} Z_{ij}(y) = \Sigma(n_i/n_A)Z_i(y)$, where $Z_i(y) = (1/n_i) \sum_{j=1}^{n_i} Z_{ij}(y)$.

The mean of the random variable $Z_A(y)$ is given by

$$E(Z_A(y)) = E(\sum_{i\in A} (n_i/n_A)Z_i(y)) = \sum_{i\in A} (n_i/n_A)E(Z_i(y)) = \Sigma(n_i/n_A)y$$

$$= (y/n_A) \sum_{i\in A} n_i = (n_A/n_A) \cdot y = y.$$

To determine the covariance of $Z_C(y_p)$ and $Z_D(y_q)$, when $C \cap D$ may not be empty, it is helpful to consider the following:

$$E(Z_{ij}(y_p)\ Z_{k\ell}(y_q)) = \begin{cases} y_p\ y_q, & i \neq k \text{ or } i = k,\ j \neq \ell \\ y_p \wedge y_q, & i = k \text{ and } j = \ell \end{cases},$$

where $y_p \wedge y_q$ is the minimum of y_p and y_q. Therefore,

$$E(Z_C(y_p)\ Z_D(y_q)) = \sum_{\substack{i\in C \\ k\in D}} (1/n_C\ 1/n_D) \sum_{j=1}^{n_i} \sum_{\ell=1}^{n_k} E(Z_{ij}(y_p)\ Z_{k\ell}(y_q))$$

$$= \sum_{\substack{i\in C\cap D \\ i=k}} (1/n_C n_D)(n_i y_p \wedge y_q + n_i(n_i-1)y_p y_q) + \sum_{\substack{i\in C \\ k\in D}} (1/n_C n_D) n_i n_k y_p y_q -$$

$$\sum_{i\in C\cap D} (1/n_C n_D) n_i^2\ y_p y_q = \sum_{i\in C\cap D} (1/n_C n_D) n_i (y_p \wedge y_q - y_p y_q) + (1/n_C n_D) \sum_{\substack{i\in C \\ k\in D}} n_i n_k y_p y_q$$

$$= (n_{C\cap D}/n_C n_D)(y_p \wedge y_q - y_p y_q) + y_p\ y_q.$$

In addition,

$\mathrm{Cov}(Z_C(y_p), Z_D(y_q)) = E[Z_C(y_p) - E(Z_C(y_p))\ (Z_D(y_q) - E(Z_D(y_q))]$

$= E(Z_C(y_p)Z_D(y_q)) - y_p y_q = (n_{C\cap D}/n_C n_D)(y_p \wedge y_q - y_p y_q$. For the variance, it follows that $\mathrm{Var}\ Z_A(y_p) = (n_{A\cap A}/n_A n_A)(y_p \wedge y_p - y_p y_p) = (1/n_A)(y_p - y_p^2)$. If $A \cap B = \emptyset$ the variance of the random variable $Z_A(y_p) - Z_B(y_p)$ is

$(1/n_A)(y_p - y_p^2) + (1/n_B)(y - y^2) = (1/n_A + 1/n_B)\ (y_p - y_p^2)$.

Let $V_i(y_p) = \sqrt{n_i}\ (Z_i(y_p) - y_p)$. We can now compute the covariance of $V_i(y_p)$ and $V_k(y_q)$ as follows:

$$\mathrm{Cov}\ (V_i(y_p), V_k(y_q)) = \sqrt{n_i}\ \sqrt{n_k}\ E[(Z_i(y_p) - y_p)\ (Z_k(y_q) - y_q)].$$

By the assumption of independence between groups, if $i \neq k$ then $\mathrm{Cov}\ (V_i(y_p), V_k(y_q)) = 0$. If $i = k$, then $\mathrm{Cov}\ (V_i(y_p), V_k(y_q)) = y_p \wedge y_q - y_p y_q$, or using the Kroenecker delta notation

$$\mathrm{Cov}\ (V_i(y_p), V_k(y_q)) = \delta_{ik}(y_p \wedge y_q - y_p y_q).$$

The relation $V_i(y_p) = \sqrt{n_i}\ (Z_i(y_p) - y_p)$ will be used in computing the variance-covariance of

$$\frac{Z_A(y_p) - Z_B(y_p)}{\sqrt{1/n_A + 1/n_B}} .$$

To do so, it is helpful to use the notation

$$V(p) = \begin{bmatrix} V_1(y_p) \\ \vdots \\ V_g(y_p) \end{bmatrix} , \text{ and}$$

$$W = \begin{bmatrix} V(1) \\ \vdots \\ V(m) \end{bmatrix} = \begin{bmatrix} V_1(y_1) \\ V_2(y_1) \\ \vdots \\ V_g(y_1) \\ \vdots \\ V_1(y_p) \\ \vdots \\ V_g(y_p) \\ \vdots \\ V_1(y_m) \\ \vdots \\ V_g(y_m) \end{bmatrix}$$

where W is the vector having the V_i's as components.

Now $E(V_i(y_p)) = 0$, $i = 1, \ldots, g$, $p = 1, \ldots, m$, so $E(W) = 0$.

The matrix WW' can be blocked so that $(WW')_{pq} = (V(p) \cdot V'(q))$ is the p,q

block. Let $y_{pq} = y_p \wedge y_q - y_p y_q$. Then

$$E(V(p)\ V'(q)) = \begin{bmatrix} y_{pq} & 0 & \cdots & 0 \\ 0 & y_{pq} & \cdots & 0 \\ \cdot & 0 & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0.. & y_{pq} \end{bmatrix} = y_{pq} I_g,$$

$$E(WW') = \begin{bmatrix} y_{11}I_g & y_{12}I_g & \cdots & y_{1m}I_g \\ y_{21}I_g & y_{22}I_g & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ y_{m1}I_g & y_{m2}I_g & \cdots & y_{mm}I_g \end{bmatrix}.$$

By using the well-known notation of tensor products, the above can be written $E(WW') = T \otimes I_g$ where $\otimes$ is tensor product and T is a matrix whose pq^{th} element is y_{pq}. let

$$\left|K(A,B)V(p)\right| = \left| \frac{Z_A(y_p) - Z_B(y_p)}{\sqrt{1/n_A + 1/n_B}} \right| = \left| \frac{\sum_{i\varepsilon A}(n_i/n_A)[Z_i(y_p) - y_p] - \sum_{j\varepsilon B}(n_j/n_B)[Z_j(y_p) - y_p]}{\sqrt{1/n_A + 1/n_B}} \right|$$

$$= \left| \frac{\left(1/n_A\right)\sum_{i\varepsilon A} \sqrt{n_i/n_A}\ \sqrt{n_i}(Z_i(y_p) - y_p) - \left(1/\sqrt{n_B}\right)\sum_{j\varepsilon B} \sqrt{n_j/n_B}\sqrt{n_j}(Z_j(y_p) - y_p)}{\sqrt{1/n_A + 1/n_B}} \right|$$

where the j^{th} element of the K_{AB} is

$$K_{AB}(j) = \begin{cases} 1/\sqrt{n_A} \cdot \sqrt{n_i/n_A} \,/\, \sqrt{1/n_A + 1/n_B}\,, & j \in A \\ -1/n_B \sqrt{n_k/n_B} \,/\, \sqrt{1/n_A + 1/n_B}\,, & j \in B \\ 0, \text{ otherwise} \end{cases}$$

let

$$K = \begin{bmatrix} K'_{A_1B_1} \\ K'_{A_2B_2} \\ \vdots \\ K'_{A_kB_k} \end{bmatrix}$$

Therefore one may write

$$(I_m \otimes K)W = \begin{bmatrix} KV(1) \\ \vdots \\ KV(m) \end{bmatrix} = W^*$$

$$E(W^* W^{*\prime}) = E[(I_m \otimes K')\, WW' (I_m \otimes K')']$$

$$= (I_m \otimes K')\, E(WW')\, (I_m \otimes K)$$

$$= (I_m \otimes K')\, T \otimes I_g (I_m \otimes K) = T \otimes KK'$$

The desired probability that each contrast will have norm less than or equal to λ can be approximated as follows:

$$P\left\{\left|\frac{\hat{Z}_A(y_j) - \hat{Z}_B(y_j)}{\sqrt{1/n_A + 1/n_B}}\right| \leq \lambda \text{ for all } A,B,y_j\right\} \simeq$$

$$1/(\sqrt{2\pi})^m \sqrt{|T \otimes KK'|}\int_{-\lambda}^{\lambda}\int_{-\lambda}^{\lambda}\cdots\int_{-\lambda}^{\lambda} e^{-1/2[X'(T \otimes KK')^{-1}X]}dx_1\cdots dx_m$$

Now $(T \otimes KK')^{-1}$ can be computed as $T^{-1} \otimes (KK')^{-1}$, so it is necessary to find T^{-1} and $(KK')^{-1}$.

Suppose $y_i = i/N$ for $i = 1, \ldots, m = N-1$. Then $y_{i\ell} = (i/N \wedge \ell/N - i\ell/N^2)$. For $i < \ell$, $y_{i\ell}=(i/N - i\ell/N^2)=(i/N^2)(N-\ell)$ and since T is symmetric it can be represented as

$$T =(1/N^2)\begin{bmatrix} 1(N-1) & 1(N-2) & . & . & . & 1(N-m) \\ & 2(N-2) & 2(N-3) & . & . & 2(1) \\ & & . & & & . \\ & & & & & . \\ & & & & & . \\ & & & & & (N-1)1 \end{bmatrix}$$

The inverse of this matrix is computed in a paper by Greenberg and Sarhan [7] and is reproduced below.

$$T^{-1} = N^2A^{-1}=N^2\begin{bmatrix} 2/N & -1/N & 0 & 0 & 0 & . & . & . & . & 0 \\ -1/N & 2/N & -1/N & 0 & 0 & . & . & . & . & 0 \\ 0 & -1/N & 2/N & -1/N & 0 & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ 0 & . & . & . & . & . & . & -1/N & 2/N & -1/N \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1/N & 2/N \end{bmatrix}$$

For the case in which the only contrasts are those comparing each experimental group with a control KK' is nonsingular and has the following form:

$$KK' = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & -1 & \cdots & 0 \\ \vdots & & & & & \\ 1 & 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & & 1 \\ -1 & 0 & \cdots & & 0 \\ 0 & -1 & \cdots & & 0 \\ \vdots & & & & \\ 0 & 0 & \cdots & & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ 1 & 1 & 2 & \cdots & 1 \\ \vdots & & & & \\ 1 & 1 & 1 & \cdots & 2 \end{bmatrix}$$

$$(KK')^{-1} = \begin{bmatrix} \frac{g}{g+1} & \frac{-1}{g+1} & \frac{-1}{g+1} & \cdots \\ \frac{-1}{g+1} & \frac{g}{g+1} & \frac{-1}{g+1} & \cdots \\ \vdots & & & \\ \frac{-1}{g+1} & \frac{-1}{g+1} & \cdots & \frac{g}{g+1} \end{bmatrix}$$

For other designs KK' may not be a nonsingular matrix. This method may be extended by use of the generalized inverse for KK'.

Since the integral

$$\int_{-\lambda}^{\lambda} \int_{-\lambda}^{\lambda} \cdots \int_{-\lambda}^{\lambda} e^{-1/2[X'(T \otimes KK')^{-1}X]} dx_1 \ldots dx_m$$

must be evaluated by numerical procedures and high speed computers it is left in this form.

CHAPTER V

PRACTICAL EXAMPLES

This chapter is devoted to applications and examples utilizing the method developed in CHAPTER IV. Emphasis is placed upon non-parametric problems. The first two examples are chosen because existing methods do not allow for multiple comparisons in Kolmogorov-Smirnov statistics. The third example treats a parametric problem involving Poisson processes for which existing methods are not appropriate.

The first example is concerned with the analysis of the information given in Table 6 on page 15. The 1,565 females are divided into the age groups 10-20, 20-30, 30-40, 40-50, 50-60 and 60^{+}. The number of people falling into each category of phosphorus levels in each age group is tabulated. The hypothesis to be tested is that $F_i = F_j$ for i,j=1,...6. That is, the object is to test whether group i and group j come from the same distribution for all i and j, and whether disjoint subsets of contiguous elements of the groups are identically distributed for all subsets. Simply stated, the object is to ask if $F_A=F_B$ when A and B are subsets of contiguous elements of the set {1,2,3,4,5,6} and $A \cap B=\emptyset$. Let $\alpha = .05$. According to Theorem 3.2 the number of subsets of contiguous elements of {1,2,3,4,5,6} is seventy. Consequently $\alpha^* = .05/70 = .000714$. Let $1-\alpha^*$ be represented by L(z) in Table 14 in the APPENDIX. In this case, $L(z) = .999286$, which corresponds to $z = 1.992$. $\hat{F}_A$ and $\hat{F}_B$ are different

if

$$\max_x \left| \frac{\hat{F}_A(x) - \hat{F}_B(x)}{\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \right| > 1.992.$$

$\hat{F}_A(x)$ and $\hat{F}_B(x)$ are given in Table 7 and for the values A and B of interest, Table 8 contains the values

$$D_{AB}/\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

where D_{AB} is the maximum of $|\hat{F}_A(x) - \hat{F}_B(x)|$ and n_A and n_B are the sample size for A and B. By consulting this table, and comparing the values with the value 1.992 five of the seventy contrasts are found to be significant. The age groups 20-40, 30-50, 10-50, 20-50, and 10-40 are all different from the age group 50+. Note that no pair of single age groups had significantly different sampling distributions.

The second sample to be presented is taken from a published paper by Weigand and Gaylor [24]. In this study the theory was tested that Negro skin resists irritants better than caucasian skin. Using concentrations of dinitrochlorobenzene (DNCB) as the variable, the minimal dose producing perceptible erythema was determined on the skin of caucasians, light- and dark-skinned Negroes. This was done twice, first on normal skin, then on skin stripped of stratum corneum. Table 9 gives the sampling distribution for the groups, caucasians (1), light-skinned Negroes (2), dark-skinned Negroes (3), all blacks (23), and light-skinned Negroes and whites combined (12), for normal skin only.

The Caucasian group contained twenty-three subjects, the light Negro group contained twenty-six subjects and the dark Negro group contained twenty-nine.

TABLE 7

THE CUMULATIVE SAMPLING DISTRIBUTIONS OF BLOOD PHOSPHORUS LEVEL OF THE AGE GROUPINGS EMPLOYED IN THE CONTRASTS

Phosphorus (mg %)	GROUPS (1)	(2)	(3)	(4)	(5)	(6)	(12)	(23)	(34)	(45)
1.000000	.008264	.004444	.004310	.000000	.000000	.000000	.005253	.004398	.001736	.000000
1.733333	.008264	.004444	.004310	.000000	.000000	.000000	.005253	.004398	.001736	.000000
2.066666	.016528	.013333	.008620	.005813	.000000	.000000	.014010	.011730	.006944	.002898
2.399999	.057851	.033333	.012931	.017441	.000000	.000000	.038528	.026392	.015625	.008695
2.733333	.107438	.084444	.077586	.093023	.031791	.013888	.089316	.082111	.086805	.062318
3.066666	.173553	.197777	.193965	.165697	.101156	.083333	.192644	.196480	.177083	.133333
3.399999	.305785	.373333	.396551	.357558	.260115	.138888	.359019	.381231	.373263	.308695
3.733333	.619834	.640000	.676724	.645348	.531791	.569444	.635726	.652492	.657986	.588405
4.066666	.768595	.833333	.862068	.808139	.748554	.763888	.819614	.843108	.829861	.778260
4.399999	.826446	.933333	.922413	.909883	.898843	.986111	.910683	.929618	.914930	.904347
4.733333	.925619	.980000	.974137	.982558	.976878	1.000000	.968476	.978005	.979166	.979710
5.066666	.942148	1.000000	.982758	.991279	.994219	1.000000	.987740	.994134	.987847	.992753
5.399999	.991735	1.000000	.991379	1.000000	1.000000	1.000000	.998248	.997067	.996527	1.000000
>5.399999	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	.999996	1.000000	1.000000	1.000000

TABLE 7--Continued

Phosphorus (mg %)	GROUPS (56)	(123)	(234)	(345)	(456)	(1234)	(2345)	(3456)	(12345)	(23456)
1.000000	.000000	.004981	.002923	.001084	.000000	.003487	.002186	.001006	.002679	.000938
1.733333	.000000	.004981	.002923	.001084	.000000	.003487	.002186	.001006	.002679	.000938
2.066666	.000000	.012453	.009746	.004338	.002624	.010462	.007288	.004024	.008037	.003752
2.399999	.000000	.031133	.023391	.009761	.007874	.027027	.017492	.009054	.020763	.008442
2.733333	.028708	.085927	.085769	.066160	.057742	.088055	.072157	.062374	.075016	.059099
3.066666	.098086	.193026	.186159	.148590	.128608	.184829	.164723	.143863	.165438	.139774
3.399999	.239234	.369863	.373294	.330802	.292650	.366172	.344752	.316901	.341594	.304878
3.733333	.538277	.647571	.650097	.610629	.586614	.646904	.620262	.607645	.620227	.605065
4.066666	.751196	.831880	.831384	.799349	.776902	.824760	.810495	.796780	.807099	.794559
4.399999	.913875	.914072	.923001	.908893	.912073	.912816	.916909	.914486	.909578	.919324
4.733333	.980861	.970112	.979532	.978308	.981627	.973844	.978862	.979879	.974547	.981238
5.066666	.995215	.986301	.993177	.990238	.993438	.987794	.993440	.990945	.989283	.991557
5.399999	1.000000	.996264	.998050	.997830	1.000000	.997384	.998542	.997987	.997990	.998123
>5.399999	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

TABLE 8

THE MAXIMUM OBSERVED DIFFERENCES OF THE CUMULATIVE SAMPLING DISTRIBUTIONS FOR VARIOUS SPECIFIC CONTRASTS CORRECTED FOR THE SAMPLE SIZE

C(A,B)	n_A	n_B	D_{AB}	$D_{AB}\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}$
C(1,2)	121	450	.1069	1.0439
C(1,3)	121	232	.0960	.8561
C(1,4)	121	344	.0834	.7891
C(1,5)	121	346	.0881	.8341
C(1,6)	121	72	.1597	1.0729
C(2,3)	450	232	.0367	.4541
C(2,4)	450	344	.0321	.4482
C(2,5)	450	346	.1132	1.5831
C(2,6)	450	72	.2345	1.8470
C(3,4)	232	344	.0539	.6344
C(3,5)	232	346	.1450	1.7088
C(3,6)	232	72	.2577	1.9102
C(4,5)	344	346	.1136	1.4920
C(4,6)	344	72	.2187	1.6875
C(5,6)	346	72	.1213	.9364
C(1,23)	121	682	.1032	1.0462
C(1,34)	121	576	.0885	.8849
C(1,45)	121	690	.1801	1.8273
C(1,56)	121	418	.0874	.8466
C(1,234)	121	1026	.0966	1.0050
C(1,345)	121	922	.0824	.8522
C(1,456)	121	762	.0856	.8747
C(1,2345)	121	1372	.0905	.9543
C(1,3456)	121	994	.0880	.9139
C(1,23456)	121	1066	.0929	.9684
C(2,34)	450	576	.0207	.3290
C(2,45)	450	690	.0647	1.0677
C(2,56)	450	418	.1341	1.9740
C(2,345)	450	922	.0492	.8555
C(2,456)	450	762	.0807	1.3570
C(2,3456)	450	994	.0564	.9856
C(3,12)	232	571	.0424	.5445
C(3,45)	232	690	.0883	1.1634
C(3,56)	232	418	.1573	1.9213
C(3,456)	232	762	.1004	1.3389

TABLE 8--Continued

C(A,B)	n_A	n_B	D_{AB}	$D_{AB}\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}$
C(4,12)	344	571	.0270	.3955
C(4,23)	344	682	.0350	.5292
C(4,56)	344	418	.1183	1.6250
C(4,123)	344	803	.0274	.4252
C(5,12)	346	571	.1040	1.5265
C(5,23)	346	682	.1211	1.8347
C(5,34)	346	576	.1262	1.8554
C(5,123)	346	803	.1158	1.8007
C(5,234)	346	1026	.1183	1.9023
C(5,1234)	346	1147	.1152	1.8782
C(6,12)	72	571	.2202	1.7607
C(6,23)	72	682	.2424	1.9561
C(6,34)	72	576	.2344	1.8752
C(6,45)	72	690	.1698	1.3710
C(6,123)	72	803	.2310	1.8777
C(6,234)	72	1026	.2344	1.9226
C(6,345)	72	922	.1920	1.1605
C(6,1234)	72	1147	.2273	1.8708
C(6,2345)	72	1372	.2059	1.7030
C(6,12345)	72	1493	.2027	1.6800
C(12,34)	571	576	.0222	.3759
C(12,45)	571	690	.0593	1.0428
C(12,56)	571	418	.1198	1.8610
C(12,345)	571	922	.0440	.8624
C(12,456)	571	672	.0640	1.1244
C(12,3456)	571	994	.0488	.9293
C(23,45)	682	690	.7260	1.3445
C(23,56)	682	418	.1420	2.2859
C(23,456)	682	762	.0886	1.6808
C(34,56)	576	418	.1340	2.0855
C(45,123)	690	803	.0612	1.1790
C(56,123)	418	803	.1306	2.1653
C(56,234)	418	1026	.1340	2.3090
C(56,1234)	418	1147	.1269	2.2211
C(123,456)	803	762	.0772	1.5265

TABLE 9

THE CUMULATIVE SAMPLING DISTRIBUTIONS OF DNCB CONCENTRATION OF THE GROUPS COMPOSING THE CONTRASTS

DNCB Concentration	n=23 Caucasians	n=26 Light Skin	n=29 Dark Skin	n=55 All Blacks	n=49 White and Light Skin
0.025	.05	.08	0	.04	.07
0.1	.30	.16	.03	.09	.21
0.25	.88	.48	.42	.45	.65
1.0	1.00	.96	.77	.86	.98
2.5	1.00	.96	.97	.96	.98
+	1.00	1.00	1.00	1.00	1.00

There are several different hypotheses which a researcher might wish to test. Two sets of hypotheses are discussed here.

The first set of hypotheses contains all the contrasts which are relevant to the researcher. These are C(1,2), C(1,3), C(2,3), C(1,23) and C(12,3). For $\alpha = .05$, $\alpha^* = .01 = \alpha/5$, and the critical value from Table 14 in the APPENDIX is 1.6277.

In Table 10 the values of $D_{AB}/\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$ where D_{AB} is the maximum of $|\hat{F}_A(x) - \hat{F}_B(x)|$ is given for all above contrasts.

Two contrasts were significant, C(1,3) and C(1,23). Hence the groups which could be said to have different sampling distributions are Caucasians and dark-skinned Negroes, and Caucasians and the Negro groups as a whole.

The problem could be viewed differently in the sense that the researcher could have tested this hypothesis: Using Caucasians as a control and the two groups of Negroes as experimental groups, one would wish to ask about C(1,2) and C(1,3). For $\alpha = .05$, $\alpha^* = .05/2 = .025$. From

TABLE 10

THE MAXIMUM OBSERVED DIFFERENCE OF THE CUMULATIVE SAMPLING DISTRIBUTIONS FOR VARIOUS SPECIFIC CONTRASTS CORRECTED FOR THE SAMPLE SIZES

$C_{A,B}$	n_A	n_B	D_{AB}	$D_{AB}/\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$
(1,2)	23	26	.40	1.3974
(1,3)	23	29	.46	1.6475
(1,23)	23	55	.43	1.7317
(2,3)	26	29	.19	.7405
(12,3)	48	29	.23	.9779

Table 14 in the APPENDIX, the critical value is 1.4812. In Table 10 the appropriate values are given, since these do not change. This test does not obtain significance for Caucasians versus light-skinned Negroes.

We now turn to the second part, that for which stratum corneum was first stripped from the skin. In this section the Caucasian group contained sixteen subjects, the light-Negro and dark Negro groups contained twelve each. Again two sets of hypotheses will be tested. The first set to be chosen is C(1,2), C(1,3), C(2,3) and C(1,23).

Table 11 gives the sampling distribution of the groups, Caucasians (1), light-skinned Negroes (2), dark-skinned Negroes (3) and all blacks (23).

For $\alpha = .05$, $\alpha^* = .0125$. The critical value from Table 14 in the APPENDIX is 1.593. In Table 12 the values of $D_{AB}/\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$ are given for all above contrasts.

TABLE 11

THE CUMULATIVE SAMPLING DISTRIBUTIONS OF DNCB CONCENTRATION OF THE SPECIAL GROUPS COMPOSING THE CONTRASTS

DNCB Concentration	n=16 Caucasian	n=12 Light Skin	n=12 Dark Skin	n=24 All Blacks
.01	.06	.08	.25	.165
.025	.06	.50	.33	.415
.1	1.00	.84	.92	.88
.25	1.00	1.00	1.00	1.00

TABLE 12

MAXIMUM OBSERVED DIFFERENCES OF THE CUMULATIVE SAMPLING DISTRIBUTIONS FOR VARIOUS SPECIFIC CONTRASTS CORRECTED FOR THE SAMPLE SIZES

$C_{A,B}$	n_A	n_B	D_{AB}	$D_{AB}/\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}$
(1,2)	16	12	.44	1.1522
(1,3)	16	12	.27	.7070
(2,3)	12	12	.17	.4164
(1,23)	16	24	.355	1.1000

None of the contrasts was significant. Hence we cannot conclude that any of the pairs chosen have different distributions. Thus removal of stratum corneum produced more uniform results among the groups.

If one wishes again to consider the Caucasian group as a control group, then contrasts C(1,2) and C(1,3) should be chosen. That is, one is comparing both the light-skinned and dark-skinned Negro groups with the Caucasian group. In this case for $\alpha = .05$, $\alpha^* = .025$. From Table 14 in the APPENDIX the critical value is 1.4812. The values of $D_{AB}/\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$ for C(1,2) and C(1,3) are the same as those in Table 12. By comparison with the critical value 1.4812, there is not enough evidence to say that the light-skinned or dark-skinned group differ from the Caucasian control group.

The third example is hypothetical, but is a common type of problem occurring in epidemiologic studies. Let us assume that the researcher wishes to analyze the incidence of some rare disease in Oklahoma and has available the incidence rate for each of the seventy-seven Oklahoma counties. The customary usage is followed here of calling a rate the numerator of a standardized rate. For example if the incidence is 3.7 per hundred thousand, the number 3.7 is called the rate rather than the actual rate 3.7/100,000. It is the random variable whose values are the numerators which has expected value λ in the Poisson distribution of parameter λ. In this case the researcher wishes to ask if there exists any statistically significant difference between any pair of counties. Let x_i be the rate for the i^{th} county. It is not unusual to assume that these rates each have a Poisson distribution. The null hypothesis is $H_o: \lambda_i = \lambda_j = \lambda$ for every i,j. Assuming that the overall α for

$\lambda_1 = \lambda_2 = \ldots = \lambda_{77}$ is .05, then by the methods discussed in CHAPTER IV,

$$\alpha^* = \frac{\frac{1}{20}}{\frac{77 \cdot 76}{2}} = \frac{1}{58520} = .000017.$$

$$P(x_i - x_j = K) = e^{-(\lambda_i + \lambda_j)} \left(\sqrt{\frac{\lambda_i}{\lambda_j}} \right)^K I_K(2\sqrt{\lambda_i \lambda_j}),$$

where $I_K(Z)$ is the modified Bessel function of the first kind of order K.

Under the null hypothesis $\lambda_i = \lambda_j = \lambda$ the above equation reduces to

$$P(x_k = x_j = K) = e^{-2\lambda} I_K(2\lambda)$$

Therefore $\sum_{K=-a}^{a} e^{-2\lambda} I_K(2\lambda) = 1 - \alpha^* \sim .999983$. For different values of λ, the required value of (a) has been computed in Table 13 representing the difference that must be observed between any pair of rates in order to conclude that a difference exists with $\alpha = .05$.

TABLE 13

VALUES REQUIRED TO SHOW SIGNIFICANT DIFFERENCES OF TWO RATES FOR GIVEN λ AT $\alpha = .05$

2λ	a
1	6
2	7
5	10
10	14
50	>21

CHAPTER VI

SUMMARY

There are several aims in this dissertation. The first is to illustrate some of the difficulties in obtaining the individual significance levels associated with a multiple comparisons test for a given significance level. It is shown that obtaining these individual significance levels is not simple unless the individual comparisons are independent. In this case the relationship between the overall significance level and the individual significance level is $\alpha = 1 - (1-\alpha^*)^p$, with α the overall significance level, α^* the individual significance level and p the number of contrasts. The second example gives another means of showing the relationship between α and α^*. However, for illustrative purposes it is assumed that the observations are uniformly distributed. This is done simply because in this case it is possible to exhibit the exact relationships and demonstrate, more easily, the nature of the difficulties. Since it is clear that the sample means cannot be distributed uniformly, this is not a practical method. In the third example, under the more practical assumption that the observations are independently and identically distributed normally with mean μ and variance σ^2, the relationship between α and α^* is, again, obtained. Under these assumptions the problems have been treated, rather thoroughly, by Scheffé, Tukey, Duncan, and others.

In CHAPTER II a description is given of most of the existing methods for multiple comparisons. Most of the emphasis is on nonparametric methods, since that is a principal interest of this dissertation.

CHAPTER III gives a complete discussion in mathematical terms of some general aspects of the multiple comparisons problem. Included in this chapter is a discussion of the important concept of error rate. It is assumed in this dissertation that errors refer to the situation in which all hypotheses are considered to hold. In this context each rejection becomes an error relative to the hypothesis, and the error rate statistic is the ratio of the number of rejections to the total number of statements. Another important concept considered is the means of defining a contrast in a nonparametric situation in such a way as to make statistical sense. It is reasonable to ask the following question: If the random variables $\{X_i\}_{i\in A}$ are considered to have a common distribution function, and if $\{X_j\}_{j\in B}$ have a common distribution function, are the two the same? Let $I_j = \{1,\ldots,g\}$ where g is the number of groups. The sampling distribution for the pooled data is

$$\hat{F}_A(x) = \sum_{i\in A} n_i/n_A \; 1/n_i \sum_{j=1}^{n_i} U(x - x_{ij}),$$

where $n_A = \Sigma\, n_i$. A statistic to be used in comparing the subgroups determined by (A,B), a pair of nonempty disjoint subsets of I_g, is $\hat{F}_A(x) - \hat{F}_B(x)$. If $C_{A,B}$ is a vector whose i^{th} component is n_i/n_A if $i\in A$, $-n_i/n_B$ if $i\in B$, 0 if $i\notin A\cup B$, then $C'_{A,B}\,\Theta(x) =$

$$\sum_{i\in A}(n_i/n_A)\hat{F}_i(x) - \sum_{i\in B} n_i/n_B\, \hat{F}_i(x) = \hat{F}_A(x) - \hat{F}_B(x).$$

The vector $C_{A,B}$ is a special contrast.

Lastly, a model is established for the most general problem in multiple comparisons. The problem may be stated as follows: Let $\{X_{ij}\}$ $j = 1, 2, \ldots, n_i$; $i = 1, 2, \ldots, g$ be independent random variables, distributed F_i, $i = 1, 2, \ldots, g$, H is a subset of the set of all pairs (A,B) of non-empty disjoint subsets of $I_g = 1, 2, \ldots, g$, with $\{C_{(A,B)} | (A,B) \in H\}$ the contrasts of interest to the researcher.

In CHAPTER IV two different methods are discussed. The first method is the F.S.D. method first proposed by Fisher, in which if one is interested in m comparisons, α/m is used for the individual significance level, α being the overall significance level.

The second technique discussed here centers around problems arising from having univariate observations of members from several groups of possibly unequal size. Because of the extremely complex intrinsic nature of the problem, the analysis is brought only to the point where it is clear how to proceed with the use of large-scale computers.

The fifth chapter is devoted to applications and examples utilizing the method developed in CHAPTER IV. Three examples are given, two illustrating the use of multiple comparisons in Kolmogorov-Smirnov statistics, and the third illustrating a parametric problem involving Poisson processes for which existing methods are not appropriate.

LIST OF REFERENCES

1. Duncan, David R. (1952): On the properties of the multiple comparisons test, Virginia J. Sci., 2:171-189.

2. Duncan, David R. (1955): Multiple Range and Multiple F tests, Biometrics, 11:1-42.

3. Dunnett, C. W. (1955): Multiple comparisons with a standard, Proc. 9th Ann. Conv. Amer. Soc. Qual. Control, May, 1955, 485-492.

4. Federer, W. T. (1955): "Experimental Design, Theory and Applications," The Macmillan Company, New York.

5. Fisher, R. A. (1935): "The Design of Experiments," Oliver and Boyd, Ltd., Edinburgh and London.

6. Fraser, D. A. S. (1957): "Nonparametric Methods in Statistics," John Wiley and Sons, Inc., New York.

7. Greenberg, B. G., and Sarhan, A. E.(1959): Matrix Inversion, Its Interest and Application in Analysis of Data, J. Amer. Stat. Assoc., 54:755-766.

8. Keuls, M. (1952): The use of "studentized range" in connection with an analysis of variance, Euphytica, 1:112-122.

9. Kolmogorov, A. N. (1941): Confidence limits for an unknown distribution function, Ann. Math. Stat., 12:461-463.

10. Kurtz, T. E., Link, R. F., Tukey, J. W., and Wallace, D. L. (1965): Short-cut multiple comparisons for balanced single and double classifications: Part I, Results, Technometrics, 7:95-165.

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

12. Nemenyi, P. (1963): Distribution-free multiple comparisons. Unpublished doctoral thesis, Princeton University, Princeton, N. J.

13. Newman, D. (1939): The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of standard deviation, Biometrika, 31:20-30.

14. O'Neill, R., and Wetherill, G. B. (1971): The Present State of Multiple Comparison Methods, J. Royal Stat. Soc., 2:1971

15. Rao, C. R. (1965): "Linear Statistical Inference and Its Applications," John Wiley and Sons, New York.

16. Scheffé, H. (1953): A method for judging all contrasts in the analysis of variance, Biometrika, 40:87-104.

17. Scheffé, H. (1959): "The Analysis of Variance," John Wiley and Sons, New York.

18. Seeger, P. (1966): "Variance Analysis of Complete Designs. Some Practical Aspects," Almqvist and Wicksell, Stockholm.

19. Smirnov, N. V. (1948): Table for estimating the goodness of fit of empirical distributions, Annals Math. Stat., 19:279-281.

20. Steel, R. G. D. (1959): A multiple comparison sign test: treatments versus control, J. Am. Stat. Assoc., 54:767-775.

21. Steel, R. G. D. (1959): A multiple comparison rank sum test: treatment versus control, Biometrics, 15:560-572.

22. Steel, R. G. D. (1960): A rank sum test for comparing all pairs of treatments, Technometrics, 2:197-207.

23. Tukey, J. W. (1953): The problem of multiple comparisons. Unpublished manuscript.

24. Weigand, D. A., and Gaylor, J. R. (1973): Irritant Reactions in Negro and Caucasian Skin, Unpublished manuscript.

25. Wilks, S. S. (1962): "Mathematical Statistics," John Wiley and Sons, New York.

APPENDIX

TABLE 14

TABLE FOR ESTIMATING THE GOODNESS OF FIT OF EMPIRICAL DISTRIBUTIONS BY SMIRNOV [19]

Table of L(z)								
z	L(z)		z	L(z)		z	L(z)	
.29	.000	004	.69	.272	189	1.09	.814	342
.30	.000	009	.70	.288	765	1.10	.822	282
.31	.000	021	.71	.305	471	1.11	.829	950
.32	.000	046	.72	.322	265	1.12	.837	356
.33	.000	091	.73	.339	113	1.13	.844	502
.34	.000	171	.74	.355	981	1.14	.851	394
.35	.000	303	.75	.372	833	1.15	.858	038
.36	.000	511	.76	.389	640	1.16	.864	442
.37	.000	826	.77	.406	372	1.17	.870	612
.38	.001	285	.78	.423	002	1.18	.876	548
.39	.001	929	.79	.439	505	1.19	.882	258
.40	.002	808	.80	.455	857	1.20	.887	750
.41	.003	972	.81	.472	041	1.21	.893	030
.42	.005	476	.82	.488	030	1.22	.898	104
.43	.007	377	.83	.503	808	1.23	.902	972
.44	.009	730	.84	.519	366	1.24	.907	648
.45	.012	590	.85	.534	682	1.25	.912	132
.46	.016	005	.86	.549	744	1.26	.916	432
.47	.020	022	.87	.564	546	1.27	.920	556
.48	.024	682	.88	.579	070	1.28	.924	505
.49	.030	017	.89	.593	316	1.29	.928	288
.50	.036	055	.90	.607	270	1.30	.931	908
.51	.042	814	.91	.620	928	1.31	.935	370
.52	.050	306	.92	.634	286	1.32	.938	682
.53	.058	534	.93	.647	338	1.33	.941	848
.54	.067	497	.94	.660	082	1.34	.944	872
.55	.077	183	.95	.672	516	1.35	.947	756
.56	.087	577	.96	.684	636	1.36	.950	512
.57	.098	656	.97	.696	444	1.37	.953	142
.58	.110	395	.98	.707	940	1.38	.955	650
.59	.122	760	.99	.719	126	1.39	.958	040
.60	.135	718	1.00	.730	000	1.40	.960	318
.61	.149	229	1.01	.740	566	1.41	.962	486
.62	.163	225	1.02	.750	826	1.42	.964	552
.63	.177	753	1.03	.760	780	1.43	.966	516
.64	.192	677	1.04	.770	434	1.44	.968	382
.65	.207	987	1.05	.779	794	1.45	.970	158
.66	.223	637	1.06	.788	860	1.46	.971	846
.67	.239	582	1.07	.797	636	1.47	.973	448
.68	.255	780	1.08	.806	128	1.48	.974	970

TABLE 14--Continued

Table of L(z)								
z	L(z)		z	L(z)		z	L(z)	
1.49	.976	412	1.89	.998	421	2.29	.999	944
1.50	.977	782	1.90	.998	536	2.30	.999	949
1.51	.979	080	1.91	.998	644	2.31	.999	954
1.52	.980	310	1.92	.998	744	2.32	.999	958
1.53	.981	476	1.93	.998	837	2.33	.999	962
1.54	.982	578	1.94	.998	924	2.34	.999	965
1.55	.983	622	1.95	.999	004	2.35	.999	968
1.56	.984	610	1.96	.999	079	2.36	.999	970
1.57	.985	544	1.97	.999	149	2.37	.999	973
1.58	.986	426	1.98	.999	213	2.38	.999	976
1.59	.987	260	1.99	.999	273	2.39	.999	978
1.60	.988	048	2.00	.999	329	2.40	.999	980
1.61	.988	791	2.01	.999	380	2.41	.999	982
1.62	.989	492	2.02	.999	428	2.42	.999	984
1.63	.990	154	2.03	.999	474	2.43	.999	986
1.64	.990	777	2.04	.999	516	2.44	.999	987
1.65	.991	364	2.05	.999	552	2.45	.999	988
1.66	.991	917	2.06	.999	588	2.46	.999	989
1.67	.992	438	2.07	.999	620	2.47	.999	990
1.68	.992	928	2.08	.999	650	2.48	.999	991
1.69	.993	389	2.09	.999	680	2.49	.999	992
1.70	.993	823	2.10	.999	705	2.50	.999	9925
1.71	.994	230	2.11	.999	728	2.55	.999	9956
1.72	.994	612	2.12	.999	750	2.60	.999	9974
1.73	.994	972	2.13	.999	770	2.65	.999	9984
1.74	.995	309	2.14	.999	790	2.70	.999	9990
1.75	.995	625	2.15	.999	806	2.75	.999	9994
1.76	.995	922	2.16	.999	822	2.80	.999	9997
1.77	.996	200	2.17	.999	838	2.85	.999	99982
1.78	.996	560	2.18	.999	852	2.90	.999	99990
1.79	.996	704	2.19	.999	864	2.95	.999	99994
1.80	.996	932	2.20	.999	874	3.00	.999	99997
1.81	.997	146	2.21	.999	886			
1.82	.997	346	2.22	.999	896			
1.83	.997	533	2.23	.999	904			
1.84	.997	707	2.24	.999	912			
1.85	.997	870	2.25	.999	920			
1.86	.998	023	2.26	.999	926			
1.87	.998	145	2.27	.999	934			
1.88	.998	297	2.28	.999	940			