THE ANALYSIS OF THE ACCUMULATION OF TYPE II ERROR IN MULTIPLE COMPARISONS FOR SPECIFIED LEVELS OF POWER TO VIOLATION OF NORMALITY WITH THE DUNN-BONFERRONI PROCEDURE, A MONTE CARLO STUDY

DISSERTATION

Presented to the Graduate Council of the
University of North Texas in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Bonnie Ann Powers-Prather, B.S.N., M.S.N.

Denton, Texas

August, 1989

wkb

Powers-Prather, Bonnie Ann, The Analysis of the Accumulation of Type II Error in Multiple Comparisons for Specified Levels of Power to Violation of Normality with the Dunn-Bonferroni Procedure, A Monte Carlo Study. Doctor of Philosophy (Educational Research), August, 1989, 96 pp., 22 tables, 21 illustrations, bibliography, 25 titles.

The study seeks to determine the degree of accumulation of Type II error rates, while violating the assumptions of normality, for different specified levels of power among sample means. The study employs a Monte Carlo simulation procedure with three different specified levels of power, methodologies, and population distributions.

On the basis of the comparisons of actual and observed error rates, the following conclusions appear to be appropriate.

1. Under the strict criteria for evaluation of the hypotheses, Type II experimentwise error does accumulate at a rate that the probability of accepting at least one null hypothesis in a family of tests, when in theory all of the alternate hypotheses are true, is high, precluding valid tests at the beginning of the study.

2. The Dunn-Bonferroni procedure of setting the critical value based on the beta value per contrast did not significantly reduce the probability of committing a Type II error in a family of tests.

3. The use of an adequate sample size and orthogonal contrasts, or limiting the number of pairwise comparisons to the number of means, is the best method to control for the accumulation of Type II errors.

4. The accumulation of Type II error is irrespective of distributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

The use of statistical packages in the current computers allows the researcher to test multiple hypotheses from the data collected. The researcher is faced, however, with the problem of accumulation of errors of each hypothesis tested. Even though the individual test is the conceptual unit for defining error probabilities, the family of tests is often considered as a unit. While many manuscripts have been published on how to protect against the accumulation of Type I error, it is a rare article that addresses the accumulation of Type II error. Westermann and Hager (1986), for example, noted that even when the problem has been addressed in the literature, it has been dealt with in an oblique manner.

Westermann and Hager (1986) have suggested that three general hypotheses be used in any research question: (a) the substantive (or educational) hypothesis (EH), the hypothesis of interest, which is a linear relationship that can not be directly tested; (b) the statistical hypothesis that can be tested directly; and (c) the derived statistical hypothesis (DSH) which is the research hypothesis stated either in the null or alternate form and which has optimal agreement between the substantive hypothesis under study and a statistical hypothesis. The connection between the EH and the DSH can be symbolized by EH --> DSH. An example of the three hypotheses would be: the EH postulates that no correlation exists between two variables and is examined by a Pearson product-moment correlation: EH --> (DSH: $\rho = 0$) = H_0. The advantages of examining the research under the three hypotheses are in the evaluation of the error terms.

Westermann and Hager (1986) also suggest two nontraditional error probabilities:

1. ε (epsilon)-- the probability of accepting the DSH when it is not valid; and

2. Φ (phi)-- the probability of rejecting the DSH when it is valid.

The relationship between the probabilities of correct and incorrect decisions concerning the Derived Statistical Hypothesis (DSH) that most adequately represents the Substantive Hypothesis (EH) and the error terms of alpha and beta can be seen in Table 1.

In the general usage of educational research, the alternate hypothesis ($u_1=u_2=...=u_k$) is true in relationship to the substantive hypothesis, in which $\underline{C}$+1 means are tested by $\underline{C}$ orthogonal contrasts. Since the relationship of alpha and beta errors is symmetrical, Westermann and Hager have suggested that the Dunn-Bonferroni inequality, a procedure utilized to protect against the accumulation of Type I error, can also be applied to Type II errors. In Table 1, when the alternate hypothesis is true, epsilon will be as low as the maximum value of all alphas associated with the tests. Therefore, any control of Type I errors, in this case, is to control the wrong error term at the expense of the right error term.

Table 1

Relationship of Error Probability in the DSH Decision

	DSH = H_1		DSH = H_0	
	Valid	Not Valid	Valid	Not Valid
DSH Accepted	$1-\Phi=1-\beta$ (Power)	$\varepsilon=\alpha$ (Type I)	$1-\Phi=1-\alpha$	$\varepsilon=\beta$ (Type II)
DSH Rejected	$\Phi=\beta$ (Type II)	$1-\varepsilon=1-\alpha$	$\Phi=\alpha$ (Type I)	$\varepsilon=1-\beta$ (Power)

Source: Westermann & Hager, 1986.

Publications concerning Type II errors have in general focused on how to limit the accumulation of Type I errors, while maintaining adequate power (Rodger, 1974; Rosenthal & Rubin, 1984; Silverstein, 1986). Inasmuch as power is defined as 1 - β, maintaining adequate power is actually a result of limiting the Type II error. A Type I error can only occur if the null hypothesis is true, and a Type II error can only occur if the alternate hypothesis is true. So, logically, one cannot limit both Type I errors _and_ Type II errors simultaneously. Only one error term can be of concern at a time. If the null hypothesis is true, then it is appropriate to limit the accumulation of Type I errors. Likewise, if the alternate hypothesis is true, then it is appropriate to limit the accumulation of Type II errors. In practice, the researcher does not know which situation is true. In the case that the actual value of beta (.20) is much larger than alpha (.05), the accumulation of beta and the subsequent loss of power are more dramatic than the accumulation of alpha. Thus, the probability of falsely accepting at least one null hypothesis in a family of tests, when the alternate hypothesis is true, is high, precluding valid tests at the beginning of the study.

Statement of the Problem

The problem of this study is to determine the degree of accumulation of Type II error rates, while violating the assumptions of normality, for different specified levels of power among sample means.

Purpose of the Study

The purpose of this study is to analyze the accumulation of Type II error rates in a Helmert contrast and all possible pairwise comparisons at specified levels of power, and to analyze the effect of violating the assumption of normality in data generated by Monte Carlo methods where the alternate hypothesis is true.

Questions

The following questions were formulated to carry out the purpose of this study:

1. What is the difference between the expected Type II error rate and the observed error rate for Helmert orthogonal contrasts and all possible pairwise comparisons over different levels of power and shape of the distributions?

2. What is the difference between the expected and the observed experimentwise Type II error rates of the Helmert orthogonal contrasts and all possible pairwise comparisons for the different levels of power and shape of the distributions?

3. How do the following procedures compare in the number of Type II errors for the levels of power and distribution:

(a) Fisher Least Significant Difference with $\alpha_c = .05$ and the sample size determined from the beta error per family (β_F);

(b) Dunn-Bonferroni inequality procedure with $\alpha_c = .05$ and the sample size determined from the beta error per contrast (β/c); and

(c) Dunn-Bonferroni inequality procedure with $\alpha_c = \beta/c$ and the sample size based on beta error per family (β_F)?

Significance of the Problem

In the application of research, decisions are made based on the results of statistical findings. The researcher protects the findings by taking a conservative approach because a wrong decision could affect lives and/or money. Since the determination of a correct decision is so critical, the traditional approach is to protect the null hypothesis against the probability of a Type I error. Westermann and Hager (1986) argue that the stated null hypothesis may not be the hypothesis to protect, but that reality should be protected. If the alternate hypothesis is in reality true, then it also should be protected. Since beta and alpha errors are reciprocal in their relationship, the study of the accumulation of beta errors is equally important as the accumulation of alpha errors.

Definition of Terms

The following definitions are specifically related to this study.

<u>Monte Carlo simulation methods</u> were invented at Los Alamos, New Mexico, to deal with the difficult calculations for nuclear research. Random samples from populations of specific parameters are generated, and then a statistic is computed. This is the technique by which the analysis of the accumulation of beta will be studied (Tietjen, 1986).

A Type I error concerns the decision to reject the null hypothesis when it is true. The probability of committing a Type I error, called a level of significance, is determined by the researcher and is designated by the Greek letter α (alpha) (Tietjen,1986).

A Type II error occurs when the researcher fails to reject the null hypothesis when it is false. The probability of making a Type II error is designated by the Greek letter β (beta). This error term is determined by the variables of: (a) the level of significance and whether a one-or two-tailed test is used, (b) sample size, (c) size of the population standard deviation, and (d) the magnitude of the differences between the means (Kirk, 1982).

The power of the test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. The probability of making a correct rejection is equal to $1 - \beta$ (Kirk, 1982).

The effect size index is the degree of departure from the null hypothesis that is detectable. It is a standardized raw effect size expressed as the difference of the population means divided by the standard deviation of either population (since they are assumed equal). A priori, the effect size can be estimated by the following: small effect size $\underline{d}$ = .2, medium effect size $\underline{d}$ = .5, and large effect size $\underline{d}$ = .8 (Cohen, 1977).

A priori orthogonal contrast is a linear relationship of the combination of means with coefficients such that (a) at least two coefficients are not equal to zero, (b) the sum of the coefficients is equal to zero, and (c) the sum of the crossproducts is zero. The number of orthogonal contrasts in any set of $\underline{c}$ + 1 means is equal to $\underline{c}$. Thus, the contrasts are mutually nonredundant and uncorrelated (Ferguson, 1981; Kirk, 1982).

Pairwise comparisons involve a linear relationship of the combination of means with coefficients such that two of the

coefficients are equal to 1 and -1, and all others are equal to zero. The number of pairwise comparisons for $\underline{c}$ + 1 means is equal to $[(\underline{c}+1)-1](\underline{c}+1)/2$ (Kirk, 1982).

Helmert contrast of four means is an orthogonal set of contrasts with the coefficients of:

(1)	1	-1/3	-1/3	-1/3
(2)	0	1	-1/2	-1/2
(3)	0	0	1	-1 .

The substantive (or educational) hypothesis (EH) refers to a linear or other functional relationship in which the researcher is interested (Westermann & Hager, 1986).

The derived statistical hypothesis (DSH) usually concerns the correlation or variance and corresponds to the substantive hypothesis in an optimal manner (Westermann & Hager, 1986).

The statistical hypothesis is the actual statistical test applied to the null or the alternate hypothesis (Westermann & Hager, 1986).

Strict criteria for evaluation of the hypotheses would specify that the DSH would be accepted only if all means are significantly different.

Lenient criteria for evaluation of the hypotheses would allow the DSH to be accepted if at least one contrast or comparison is significant (Westermann & Hager, 1986).

Uniform or rectangular distribution is defined on an interval where the probability of a random variable is equal anywhere in the interval (Tietjen, 1986).

Exponential or J-shaped distribution is a continuous probability distribution whose density function can be derived as: $f(x) = e^{-x}$, for $x > 0$ (Robinson, 1985).

Error rate per contrast (α_{PC}) is equal to the number of contrasts falsely declared significant divided by the total number of contrasts (Kirk, 1982). For this study, the error rate per contrast will be based on the number of contrasts falsely declared nonsignificant.

Error rate per experiment (α_{PE}) is the number of contrasts falsely declared significant, or for this study nonsignificant, divided by the number of experiments. This is the error rate most utilized and is an expected number of errors per experiment (Kirk, 1982).

Error rate experimentwise (α_{EW}) is the number of experiments with at least one contrast falsely declared significant, or for this study nonsignificant, divided by the number of experiments. This is a more conservative approach than the per experiment error rate and is a probability. For orthogonal contrasts, the relationship is defined as: $\beta_{EW} = 1 - (1 - \beta_{PC})^{C}$ and the error rate experimentwise cannot exceed the error rate per experiment (Kirk, 1982).

Error rate familywise is utilized in ANOVA as the unit of concern for the family of contrasts, with the error rate per family and the error rate familywise determined similarly to the per experiment and experimentwise (Kirk, 1982).

Delimitations

The present study is limited to the following experimental, simulated conditions:

1. All samples are from one of the three population distributions: normal, uniform, and exponential.

2. Only the case of four equally spaced treatment groups of equal variance are considered.

3. The set of Helmert contrast and the set of all possible pairwise comparisons will be utilized.

4. The levels of power to be considered will be .70, .80, and .90.

5. The sample size of each group will be equal and will be determined by the specified level of power for the overall *F*-test and the beta error per contrast.

CHAPTER II

REVIEW OF RELATED LITERATURE

History of Power

The literature concerning Type II errors revolves around the concept of power analysis. Before 1925, the field of statistical analysis was dominated by two men: Karl Pearson, most known for his product-moment correlation coefficient, and R. A. Fisher, probably the most widely known statistician of all time and the founder of many statistical techniques, including the analysis of variance. These two men were challenged by the appearance in the field of J. Neyman and E. S. Pearson. A controversy soon developed between the two factions regarding the general area of hypothesis testing and the interpretation of statistical tests.

The philosophies of the Fisherian school and the Neyman-Pearson school, respectively, can be compared, as noted by Hogben (1957), as the "Backward Look" and the "Forward Look." Upon the detection of a significant difference, for example, the Fisher approach would note that the null hypothesis was invalid. The Neyman-Pearson approach, on the other hand, would only conclude this upon the completion of a series of tests that had repeatedly rejected the null. Further, the two schools differed on whether the significance level should be stated a priori. Contrary to popular belief, while Fisher favored the .05 level, he did not believe it to be firm nor that it had to be stated a priori. In contrast, the Neyman-Pearson approach was to state a priori the significance level and adhere to it for all statistical decisions.

The third disagreement between the two approaches involves the interpretation of the results of research. The Fisherian approach is asymmetrical in its process. If the null is rejected, then it can be stated that the effect size is not zero, while if the null is retained, then it cannot be stated that the effect size <u>is</u> zero. The effect size of the alternate hypothesis is stated as not equal to zero, but to what extent this is true is never specified, while the Neyman-Pearson approach is to state an exact value for the alternate hypothesis.

The issue of the asymmetrical approach of the Fisher school prompted Neyman and Pearson to introduce the concept of power and Type II error. Type I error was recognized by the Fisher school, but only in context of the level of probability a researcher was willing to accept for a false rejection of the null. The Neyman-Pearson approach determined the Type I error probability, over a series of tests, by the ratio of incorrect decisions (Chase & Tucker, 1976).

Despite the theoretical work and controversies on the subject, the concept of power did not find common usage in statistical application until Cohen's article on power in 1962. Cohen later developed his concept more fully in his book, <u>Statistical Power Analysis for the Behavioral Sciences</u>, in 1969. After the appearance of Cohen's article and book, surveys of current research journals and later publications by other authors attempted to determine and define the evolving role of the power concept in statistical application, a process which continued into the 1980's (Brewer, 1972; Sawyer & Ball, 1981; Woolley & Dawson, 1983). The second area of publication in the field concerned the containing of Type I errors while minimizing the loss of Type II errors (de Cani, 1984; Games, 1971; Petrinovich & Hardyck, 1969; Rodger, 1967; Rosenthal & Rubin, 1984).

Cohen is the expert in power analysis most often cited by researchers, and the tables in his book are the source of reference for nearly all power analysis. Since power is defined in its relationship to beta, one can not discuss the accumulation of beta without discussing power. The focus of power analysis is to determine if a research design has the power to detect a significant difference between the null and alternate population distribution.

Cohen envisioned power analysis as a situation in which, a priori, a researcher could ascertain the probability of having significant research results by consulting the tables in his book to determine the power of a statistical test. Cohen's calculation of power is a direct result of the effect size, sample size, and the level of alpha. The researcher could fix the values of two of the three variables for a certain level of power and choose the third variable (usually sample size). The researcher could then conclude if, a priori, an adequate sample size was available for a certain research design.

However, excessively high power would increase the likelihood of detecting a trivial effect due to the decrease in the standard error as a result of an increase in the sample size. With an increase in power, a trivial difference could be found significant if the sample size were large enough. Therefore, Cohen (1977) recommended that an adequate level of power be .80. A power level of .80 is a beta level of .20. The ratio of an alpha level of .05 to a beta level of .20 indicates that the probability of committing a Type I error is more serious than the probability of committing a Type II error.

Multiple Comparisons

Simultaneous multiple comparison procedures are performed on (a) a limited number of a priori comparisons based on specific hypotheses; (b) all or most of the pairwise comparisons of means; and (c) exploratory analysis of combinations of means (Klockars & Saxs, 1986). Debate exists among authors as to the optimal method or procedure of dealing with the accumulation of errors (usually Type I) in these multiple comparison procedures. Games (1971) offers ten procedures, based in part on the *F* test and the multiple *t* statistic. Seven of the procedures use the multiple *t* statistic, with the differences found in the determination of the critical values. Four of the procedures, as outlined by Games, include:

1. Specify the per comparison rate for each orthogonal contrast and allow the experimentwise error rate to increase to $1 - (1 - \alpha)^c$ (Games, 1971).

The variability among means is divided into exclusive parts and is equal to the variability found in the between-group sum of squares. The main advantage of using the orthogonal contrasts is in the error rate, since the probability of a Type I (or Type II) error in one comparison is more likely to be isolated and not repeated on the other contrasts of the set. In view of this property, some authorities debate the need to have a significant overall *F* test before the testing of the individual contrasts (Games, 1971; Kirk, 1982; Klockars & Sax, 1986).

According to the multiplicative rule of independent events, the probability of not making a Type I error for *C* contrasts is $(1 - \alpha)^c$. Therefore, the probability of making one or more Type I errors will be equal to $1 - (1 - \alpha)^c$. As the number of contrasts increases, the probability of Type I errors also increases (Games, 1971; Kirk, 1982).

An example of the comparisons of alpha and beta levels with error rate per contrasts can be seen in Table 2.

Table 2

Expected Type I and Type II Error Rates Per C Number of Orthogonal Contrasts

Number of: Means	Contrasts	Alpha .05	Beta .10	Beta .20	Beta .30
3	2	.0975	.1900	.3600	.5100
4	3	.1426	.2710	.4880	.6570
5	4	.1855	.3439	.5904	.7599
6	5	.2262	.4095	.6723	.8319
10	9	.3697	.6513	.8658	.9718

Prob. of Type I error (alpha) = $1 - (1 - \alpha)^c$

Prob. of Type II error (beta) = $1 - (1 - \beta)^c$

2. Specify the per comparison rate for all possible comparisons and allow the experimentwise error rate to increase as the number of means increases (Games, 1971).

The major difference in orthogonal and nonorthogonal comparisons is in the number of errors in an experiment. The errors are isolated in orthogonal tests. However, in nonorthogonal tests, errors are likely to be replicated in the experiment (Klockars & Sax, 1986). As a result, in multiple comparisons, the emphasis is on the accumulation of Type I errors.

3. Specify the experimentwise error rate by the use of an overall *F* test (Games, 1971).

From the upper limit of the error rate per family of tests, the Type I error rate per contrast can be derived accordingly by the expression $\alpha = 1 - (1 - \alpha_F)^c$. Similarly, according to Westermann and Hager (1983), the upper limit for Type II errors per contrast in a family of tests is equal to $1 - (1 - \beta_F)^c$.

An alternative could be to utilize the Fisher's Least Significant Difference, which offers the least protection of a Type I error and, therefore, the most protection for a Type II error. The Fisher's LSD is often not recommended, however, since the probability of a Type I error is likely to be larger than the specified level of significance. The Fisher's LSD is also seen in the idea of "protected" and "unprotected." If the Fisher LSD is performed after a significant F-ratio, then the test is referred to as a "Fisher-protected" LSD. Otherwise it is referred to as the "unprotected LSD." The implication is that the error rate per contrast will be higher than expected when the *F* test is unprotected (Roscoe, 1975).

A significant overall *F* test means that the researcher rejects H_0: $\mu_1 = \mu_2 = \mu_3$, but does not indicate which means are different or how the means differ. For a larger numbers of means it is possible to obtain a significant F with all nonsignificant t's. When the boundaries of the areas of rejection of the F test and the multiple t tests are diagrammed, the difference between regions can be examined to explain how the tests differ in results. Since the significant F test does not indicate which contrast is significant, a Type II error will occur in the LSD test (Games, 1971).

4. Limit the upper boundary of the experimentwise error rate by the Dunn-Bonferroni inequality (Games, 1971).

One of the statistical methods used to contain the accumulation of Type I errors is the Dunn-Bonferroni procedure. Dunn (1961) originally examined the properties of the Bonferroni inequality, which shows that the error rate experimentwise could not exceed the sum of the per contrast error rate, that is, $\alpha_{EW} \leq \Sigma\ \alpha_{PC}$. If each C contrast is tested at the α/C level of significance, then the total error rate experimentwise will not exceed α (Games, 1971).

The Dunn-Bonferroni procedure is additive in definition. The error rate for any family of tests (α_{EW}) is not exceeded by the sum of the C per contrast error rate (Kirk, 1982). The procedure can be based on orthogonal or nonorthogonal comparisons, but in general the procedure is more powerful as the number of contrasts tested approaches the number of means (Games, 1971).

In the ordered Bonferroni procedure, power is saved for contrasts of more interest, while sacrificing power for other contrasts of less interest. An example of subdividing the error term based on the Dunn-Bonferroni procedure can be found in the study by de Cani (1984), which found that in ordered or weighted Bonferroni procedures: (a) the loss of power is larger at the overall alpha level at .05 than at .10; (b) main effects have less loss of power, while interaction suffers the most loss of power; (c) loss of power is larger at low levels of power rather than at higher levels of power; and (d) the average level of power suffers less relative to the change of overall alpha level.

Rosenthal and Rubin (1984), in a related study of unordered, partially ordered, and completely ordered contrasts, found that the power of the Dunn-Bonferroni procedure increases as one progresses from unordered to completely ordered contrasts. In this study as well as the de Cani article, the emphasis is still on undergoing procedures to limit the accumulation of Type I error, while maintaining power.

Silverstein (1986) also conducted a study of the Dunn-Bonferroni procedure, Type I error rate, and power. This study found that although the Dunn-Bonferroni procedure controls the risk of Type I errors, when the number of tests performed increases, the risk of failing to reject a false null also increases. Silverstein also found that increasing sample size was more effective in increasing the power of a test than was reducing the level of significance. Indeed, Games (1971) noted that there is little logic in resorting to sophisticated, statistical methods to control the probability of accumulation of Type I error for all contrasts when no attempt is made to control the probability of Type II error accumulating in the contrasts. While Westermann and Hager (1986) suggest the use of the Dunn-Bonferroni inequality procedure to control for the accumulation of either Type I or Type II errors, the procedure may have difficulties that preclude its use.

Westermann and Hager (1986) noted in their article summary that the researcher should "always adjust the error probability α and/or β that is connected with the false rejection of the statistical metahypothesis" (p. 38). In order to accomplish these adjustments, the Westermann and Hager article posed two perpectives.

First, in the case where the value of the Type II error probability is determined for the family of tests, with $\beta_c = \beta_{Fc}/C = \Phi_c/C$, the critical value is then based on β_c. The article suggests that the researcher is free to choose "comparatively large values of α_c" (p. 132). However, the example demonstrated utilizes $\alpha_c = \varepsilon_c = \Phi_c = .05$ and $\beta_c = \Phi_c/3 = .01667$, with the sample size based on this critical value.

The problems with this procedure are two-fold in the determination of the critical value and sample size. From the example, the power level of the test is set at .95, which according to Cohen (1977) is an excessive level of power that will require a much larger sample size than that required for power set at the adequate level of .80. The procedure to determine the sample size based on the beta rate per familywise (β_{Fc}) is found in Cohen's tables, but Cohen's tables are not available for unusual critical values that would result from the beta rate per comparison.

The article could also be viewed from the second perspective of setting the critical values of $\alpha_c = \beta_c$ and determining the sample size with the overall beta rate per familywise (β_F). If the β_{Fc} were equal to .20, for example, then the α_c would equal .20/$\underline{c}$, which could be considered an excessively large critical level of significance in the publishing circles. Although the critical value of this procedure is nontraditional, the procedure for determining similar sample size is found in current research. For example, in the Keselman (1976) study, the sample size for the Tukey procedure was based on the beta rate per familywise.

In comparing the two procedures, the power of the test can be examined through the area of the region of rejection. For example, if one starts with a sample size based on the overall stated $\underline{F}$-test power, the change in the region of rejection for the individual sets of contrasts/comparisons will increase or decrease the originally stated $\underline{F}$ test power, as seen in the Tables 3 and 4.

Table 3

Comparisons of the Proportion of Area under the Curve for Two-Tailed Test of Orthogonal Contrasts with the Least Significant Difference and the Dunn-Bonferroni Procedure with Four Groups

Power Level	.70	.80	.90
LSD	.05/2=.025	.05/2=.025	.05/2=.025
DB(.70)	(.30/2)/3=.05		
DB(.80)		(.20/2)/3=.033	
DB(.90)			(10/2)/3=.0167
Difference:	DB(.70) > LSD .025	DB(.80) > LSD .0083	LSD > DB(.90) .0083

For the orthogonal contrasts, when the critical value is subdivided by the Dunn-Bonferroni procedure, a larger area of rejection (hence more power in the test) occurs in the power levels of .70 and .80, but not in the .90 level of power. For the pairwise comparisons, the LSD provides a larger or equal area of rejection than does the subdivided Dunn-Bonferroni, thus confirming that the power of the test is maximized as the number of contrasts approaches the number of means. Indeed, when a researcher utilizes orthogonal contrasts, the probability of both a Type I and

Type II error can be specified. However, if a researcher tests all possible pairwise means, then the probability of Type I error or Type II error, depending on the situation, increases as the number of means increases (Games, 1971).

Table 4.

Comparisons of the Proportion of Area under the Curve for Two-Tailed Test of Pairwise Contrasts with the Least Significant Difference and the Dunn-Bonferroni Procedure with Four Groups

Power	.70	.80	.90
LSD	.05/2=.025	.05/2=.025	.05/2=.025
DB(.70)	(.30/2)/6=.025		
DB(.80)		(.20/2)/6=.0167	
DB(.90)			(.10/2)/6=.0083
Difference:	DB(.70) = LSD 0.0	LSD > DB(.80) .0083	LSD > DB(.90) .0167

In summary, Type I and Type II errors are reciprocal in nature, but the researcher is usually faced with the problem of which error term to protect. Traditionally, research methodology has protected against the occurrence of the Type I error, while ignoring the probability of the occurrence of the Type II error. The researcher could utilize (a) the traditional approaches as outlined by Games (1971) for both Type I and Type II errors, or (b) the nontraditional approaches such as the ones outlined by Westermann and Hager (1986).

CHAPTER III

PROCEDURES

The comparisons of contrasts and the analysis of variance in this Monte Carlo simulation require that the following assumptions be made:

1. The simulated observations will be samples from three population distributions: normal, uniform, and exponential.

2. Observations are random samples from the populations.

3. In the null case, the numerator and denominator of the F ratio are estimates of the same population variance.

4. The numerator and denominator of the F ratio are independent of each other.

5. The model equation, $Y_{ij} = u + a_j + e_{i(j)}$, reflects the sum of all the sources of variation that affect each Y_{ij}.

6. The experiment contains all the treatment levels, a_j's, of interest.

7. The error term, $e_{i(j)}$, (a) is independent of all other error terms, and (b) is normally distributed within each treatment population, with (c) mean equal to zero and (d) variance equal to one (Kirk, 1982).

The Models

The model of interest is the fixed-effects for a completely randomized design model:

$$\underline{Y}_{ij} = \underline{u} + \underline{a}_j + \underline{e}_{ij} \quad (i = 1,\ldots,\underline{n};\ \underline{j} = 1,\ldots,\underline{c} + 1).$$

In the fixed effects model, the treatment effect, $\underline{a}_{ij} = \underline{u}_j - \underline{u}$, is a constant for all observations within a group but may vary for each $\underline{j} = 1,\ldots, \underline{c} + 1$ group and the sum of all $\underline{a}_j$'s is equal to zero. Since $\underline{u}$ and $\underline{a}_j$ are constant for all observations within the population $\underline{j}$, the only source of variation is due to the error effect, $\underline{e}_{ij}$, which can be shown to equal to $\underline{Y}_{ij} - \underline{u} - \underline{a}_j$ (Kirk, 1982).

The hypotheses of interest included the overall null hypothesis:

$$H_o\colon \underline{u}_1 = \underline{u}_2 = \underline{u}_3 = \underline{u}_4$$

and the two sets of alternate hypotheses, which included:

Set 1: Helmert Contrasts

$$H_{a1}\colon \underline{u}_1 - (\underline{u}_2 + \underline{u}_3 + \underline{u}_4)/3 = 0,$$
$$H_{a2}\colon \underline{u}_2 - (\underline{u}_3 + \underline{u}_4)/2 = 0,$$
$$H_{a3}\colon \underline{u}_3 - \underline{u}_4 = 0.$$

Set 2: All Pairwise Comparisons

$$H_{a1}\colon \underline{u}_1 - \underline{u}_2 = 0, \qquad H_{a4}\colon \underline{u}_2 - \underline{u}_3 = 0,$$
$$H_{a2}\colon \underline{u}_1 - \underline{u}_3 = 0, \qquad H_{a5}\colon \underline{u}_2 - \underline{u}_4 = 0,$$
$$H_{a3}\colon \underline{u}_1 - \underline{u}_4 = 0, \qquad H_{a6}\colon \underline{u}_3 - \underline{u}_4 = 0.$$

The test statistics utilized for $\underline{c} + 1 = 4$ included:

1. To test the overall $\underline{F}$ significance, the SWEEP function of SAS(IML) was utilized to calculate the sums of squares (Goodnight, 1979):

$$\underline{F} = \frac{(\underline{SSE}_1 - \underline{SSE}_2)/\underline{c}}{(\underline{SSE}_2)/(\underline{c} + 1)(\underline{n} - 1)},$$

where $\underline{SSE}_1$ is the total sum of squares and $\underline{SSE}_2$ is the sum of squares error calculated from the regression analysis on the model equation.

2. The multiple $\underline{t}$ statistic, Fisher's Least Significant Difference, will test (a) the orthogonal contrasts and pairwise comparisons and (b) the orthogonal contrasts where the sample size is based on the power of $1 - \beta/c$:

$$\underline{t} = \frac{\Sigma \underline{c}_j \underline{Y}_{.j}}{\sqrt{\underline{MS}_{error} (\Sigma(\underline{c}^2{}_j/\underline{n}_j))}}$$

in which $\underline{c}_j$ is the coefficient term and $\underline{MS}_{error} = (\underline{SSE}2)/[\underline{N}-(\underline{c}+1)]$, and the critical value of $\underline{t}_{.05/2}$ with the degrees of freedom equal to that of the $\underline{MS}_{error}$ (Kirk, 1982).

3. The Dunn-Bonferroni procedure ($\underline{t}$D) utilizes the $\underline{t}$ statistic, but uses a different critical value,

$$\underline{t}D\ \alpha_{/2,C,v} = \underline{t}_{(\alpha/2)/C,v},$$

where $\underline{C}$ is the number of planned contrasts, and $\underline{v}$ is the number of degrees of freedom for the $\underline{MS}_{error}$. As suggested by Westermann and Hager (1986), this procedure was to be utilized with the respective β values of .10, .20, and .30.

The Simulation Plan

In order to generate data for the study, a plan was employed of applying the *F*-test and the two specified multiple comparison procedures, and presenting the summary statistic.

This study was conducted by means of a computer simulation using the SAS Matrix (IML) procedure. For the normal distribution, the error terms for each observation were produced by the RANNOR random number generator which generates numbers with a mean of zero and variance of one. For the uniform distribution, the error term was generated by the RANUNI procedure on the interval of 0 to 1. The exponential distribution was produced by the RANEXP, which generates uniform random numbers with a parameter of one (SAS Institute, Inc., 1988).

The following procedure was used to obtain the sample size, means, and treatment effects for each group:

1. The medium effect size (f = .25) was utilized from the Table 8.3.14 (Cohen, 1969, p. 308,) to determine the sample size for each group, where:

Power = .90, n_j = 58,
Power = .80, n_j = 44, and
Power = .70, n_j = 36.

The sample size was based on β/c values, where:

Power = .97, n_j = 76,
Power = .93, n_j = 64, and
Power = .90, n_j = 58.

2. The range ($\underline{d}_j = \underline{b}_j\underline{f}$) is the distance between the largest and smallest of the $\underline{c}$ + 1 means. For intermediate variability with equally spaced means, the range was determined from Table 8.2.1 (Cohen, 1969) where $\underline{b}_2 = 2.68$, and $\underline{f}$ is computed by the formula:

$$\underline{f} = \frac{\underline{d}}{2}\sqrt{\frac{(\underline{c} + 1) + 1}{3(\underline{c})}},$$

and where ($\underline{c}$ + 1) = 4. For $\underline{f}$ = .25, the range would be $\underline{d}_2 = 2.68(.25) = .67$ (Cohen, 1969, pp. 270-272).

3. The means were equally spaced over the range of .67 of a within-population standard deviation, at an interval of $\underline{d}/(\underline{c})$ to give intermediate variability. Since the error term has a variance of 1, the within-population standard deviation is 1, and, therefore, the range will be .67 , with intervals of .2233.

4. The grand mean was arbitrarily set at 10; then each observation was the sum of a randomly generated error term, the grand mean, and the treatment effect of that group. Therefore, the scores for each treatment group would be:

$$\underline{Y}_{i1} = 10 + (-.3350) + \underline{e}_{ij},$$
$$\underline{Y}_{i2} = 10 + (-.1117) + \underline{e}_{ij},$$
$$\underline{Y}_{i3} = 10 + .1117 + \underline{e}_{ij}, \text{ and}$$
$$\underline{Y}_{i4} = 10 + .3350 + \underline{e}_{ij}.$$

Analysis of Data

The simulations in the study involved two methods. The first method involved the computation of an overall F test, and the unprotected Least Significance Difference multiple t statistic, with the critical value of .05 for the orthogonal and all pairwise comparisons, and the Dunn-Bonferroni t statistic, with the orthogonal contrasts and critical values based on β/c. The sample size for each group was based on the power of the overall F test. This procedure was replicated 1,000 times per each level of power for each distribution.

Method two involved the computations of an overall F test and the unprotected Least Significant Difference multiple t statistic with the critical value of .05 for the orthogonal and all pairwise comparisons, but the sample size would be determined from the power of β/c instead of the overall F test. This procedure was also replicated 1000 times per level of power for each distribution.

On the data from the contrasts/comparisons the following statistical analysis was performed on each proposed distribution. For Method One with sample size based on overall F Test:

1. The number of significant results for the Dunn-Bonferroni (tD) procedure was tabulated for each of the 3 contrasts over the 1,000 simulations and divided by 1000 to determine the level of observed power. The results were

subtracted from 1.00 to determine the Type II error rate for the orthogonal contrasts. For each level of power, the resulting Type II observed error rate per contrast was compared to the expected error rate.

2. The number of significant results for the Fisher's LSD statistic was tabulated for each simulation, then divided by 1000 to determine the level of observed power. The results were subtracted from 1.00 to determine the Type II error rate for the orthogonal contrasts and pairwise comparisons. For each level of power, the actual Type II error rate computed was compared to the expected error rate.

3. For the strict evaluation criteria of each Fisher LSD and Dunn-Bonferroni procedure, the number of significant contrasts/comparisons per experiment was tabulated. The resulting number of experiments in which all the contrasts/comparisons were significant was divided by 1000 and subtracted from 1.00 to give the experimentwise error rate. For the lenient criteria evaluation, the experiments in which at least one of the contrasts/comparisons was significant were also tabulated and divided by 1000 to indicate the percent of experiments in which at least one of the hypotheses was significant.

4. The number of significant results for the $\underline{F}$ test was tabulated for each simulation (as would be found if the researcher were utilizing the PROC GLM command of SAS) and

then divided by 1000 to determine the observed level of power. The results were subtracted from 1.00 to determine the Type II error rate familywise. The resulting error rate familywise was then compared to the expected Type II error rate for each level of power.

5. The Kolmogorov-Smirnov test of goodness-of-fit was used to test whether the observed frequency distribution departed significantly from the hypothesized frequency distribution of the noncentral $\underline{F}$. A noncentral $\underline{F}$ probability was determined and counted according to the cumulative frequency. The cumulative frequency distribution was divided into 17 intervals of .005, .01, .025, .05, .1, .2, .3, .4, .5, .6, .7, .8, .9, .95, .975, .990, and .995. The level of significance for the goodness-of-fit tests was set at .05 and by the formula $1.36/\sqrt{N}$, resulting in a critical value of Dmax = .043 for 1,000 replications (Roscoe, 1975).

For Method Two with sample size based on beta per contrast:

1. The number of significant results for the Fisher's LSD statistic was tabulated for each simulation, then divided by 1000 to determine the level of observed power. The results were subtracted from 1.00 to determine the Type II error rate for the orthogonal contrasts and pairwise comparisons. For each level of power, the actual Type II observed error rate was compared to the expected error rate.

2. For the strict evaluation criteria of each Fisher LSD and Dunn-Bonferroni procedure, the number of significant contrasts/comparisons per experiment was tabulated. The resulting number of experiments in which all the contrasts/comparisons were significant was divided by 1000 and subtracted from 1.00 to give the experimentwise error rate. For the lenient criteria evaluation, the experiments in which at least one of the contrasts/comparisons were significant were tabulated and divided by 1000 to indicate the percent of experiments in which at least one of the hypotheses was significant.

3. The number of significant results for the *F* test was tabulated for each simulation (as would be found if the researcher was utilizing the PROC GLM command of SAS) and then divided by 1000 to determine the observed level of power. The results were then subtracted from 1.00 to determine the Type II error rate familywise. The resulting error rate familywise was then compared to the expected Type II error rate for each level of power.

4. The Kolmogorov-Smirnov test of goodness-of-fit is used to test whether the observed frequency distribution departed significantly from the hypothesized frequency distribution of the noncentral *F*. A noncentral *F* probability was determined and counted according to the cumulative frequency. The cumulative frequency distribution

was divided into 17 intervals of .005, .01, .025, .05, .1, .2, .3, .4, .5, .6, .7, .8, .9, .95, .975, .990, and .995. The level of significance for the goodness-of-fit tests was set at .05 and by the formula $1.36/\sqrt{\underline{N}}$, the critical value of Dmax = .043 for 1,000 replications (Roscoe, 1975).

CHAPTER IV

RESULTS OF THE SIMULATIONS

Introduction

Before the simulations of this study were executed, a preliminary study was initiated to verify the mathematical formulas and coding of the eighteen computer programs, consequently verifying the three different random number generators, the coding, and the resulting findings. The simulations were then executed according to the two methodologies under study.

Method I examines the procedure in which the sample size is based on the power of the overall *F* test, and then compares the results of (a) the Fisher LSD (T1 to T3) with α = .05, to the Dunn-Bonferroni procedure (TD1 to TD3) with $\alpha/c = \beta/c$, with the Helmert orthogonal contrasts, (b) all pairwise comparisons (C1 to C6), and (c) the *F* tests for the specified levels of power and distributions to violations of normality. Method II examines the procedure in which the sample size is based on the power of the Dunn-Bonferroni inequality β/c, and then compares the results of (a) the Fisher LSD with α= .05 for the Helmert orthogonal contrasts (T1 to T3) to all pairwise comparisons (C1 to C6) with α_c = .05, and (b) the *F* tests for the specified levels of power and distributions to violations of normality.

Preliminary Study

The final results of the preliminary study are presented in Tables 5 through 7. From the initial results, the standard deviation of the uniform distribution was found not to be in accordance to the specifications of the study. A variety of approaches were examined, and it was found that a standard deviation of approximately 1.00 would result if each error term was multiplied by the constant four. For the overall $\underline{F}$ test, the sweep operations were tested to determine which type of coding would result in the correct sums of squares as found in the PROC GLM procedure of SAS. For theoretical reasons and accuracy of results, the Helmert orthogonal coding was utilized in the sweep operations, which did produce the required sums of squares for the computations of the $\underline{F}$ test.

Table 5

Verification of the Normal Random Number Generator

Group	Mean	Std. Dev.	Min. Value	Max. Value	Range
I	9.74	1.1387	7.155	11.667	4.51
II	9.93	.8606	8.126	12.309	4.18
III	10.14	1.0389	8.253	12.745	4.49
IV	10.39	1.0885	7.811	12.960	5.14
AVG.	10.05	1.0316	7.836	12.420	4.58

Table 6

Verification of the Exponential Random Number Generator

Group	Mean	St. Dev.	Min. Value	Max. Value	Range
I	10.75	1.1196	9.699	15.621	5.92
II	10.63	.7100	9.916	13.596	3.67
III	11.17	1.1553	10.117	15.468	5.35
IV	11.15	.7622	10.341	14.040	3.69
AVG.	10.93	.9367	10.018	14.681	4.65

Table 7

Verification of the Uniform Random Number Generator

Group	Mean	St. Dev.	Min. Value	Max. Value	Range
I	11.91	1.1711	9.794	13.657	3.86
II	11.97	1.2163	9.997	13.885	3.89
III	12.16	1.1102	10.409	14.059	3.65
IV	12.43	1.0446	10.410	14.163	3.75
AVG.	12.11	1.1355	10.153	13.941	3.03

To better understand the tables and discussion, the coding of the orthogonal contrasts and pairwise comparisons will be repeated:

T1/TD1 (3 -1 -1 -1),
T2/TD2 (0 2 -1 -1),
T3/TD3 (0 0 1 -1),
C1 (1 -1 0 0), C4 (0 1 -1 0),
C2 (1 0 -1 0), C5 (0 1 0 -1),
C3 (1 0 0 -1), AND C6 (0 0 1 -1).

As previously stated, the purpose of Method I of the study was to evaluate the accumulation of the Type II error for: (a) the Helmert contrasts and all pairwise comparisons, as measured by the number of significant results per contrast and comparison; (b) the significant results experimentwise of the contrasts and comparisons; and (c) the number of significant results familywise of the overall F test for each level of power and for the different distributions when the sample size is based on the overall F test. In Table 8, the results of Method I are presented in the form of the observed level of power for the statistical procedures utilized.

In Table 8, the levels of power for the orthogonal contrasts were, in general, less than the expected levels of power. The Fisher LSD procedure was found to have a smaller level of power than the Dunn-Bonferroni procedure. The levels of power for the pairwise comparisons were considerably less than the expected levels, except for the C3 comparison. The observed levels of power for the F test were as expected except for the uniform distribution simulations, which were lower than expected, especially for the lower levels of power.

The purpose of Method II of this study was to evaluate the accumulation of the Type II error for the (a) Helmert contrasts and all pairwise comparisons, as measured by the

Table 8

Observed Levels of Power of Method I for the Distributions and Specified Levels of Power

	Normal			Exponential			Uniform		
Power	70	80	90	70	80	90	70	80	90
				Orthogonal Contrasts					
Critical Values									
α = .05									
T1	.642	.700	.821	.651	.728	.838	.529	.712	.722
T2	.383	.418	.518	.398	.461	.576	.283	.413	.445
T3	.178	.183	.213	.159	.190	.244	.124	.196	.166
α = β/C									
TD1	.729	.727	.768	.742	.761	.794	.633	.649	.660
TD2	.514	.471	.468	.520	.515	.511	.389	.369	.367
TD3	.273	.228	.167	.249	.229	.194	.210	.143	.127
				Pairwise Comparisons					
α = .05									
C1	.159	.154	.213	.151	.187	.234	.123	.214	.185
C2	.465	.526	.648	.491	.566	.679	.384	.539	.575
C3	.799	.870	.945	.795	.875	.944	.678	.879	.880
C4	.164	.184	.216	.146	.190	.238	.127	.165	.192
C5	.483	.570	.646	.475	.572	.688	.364	.533	.546
C6	.178	.183	.213	.159	.190	.244	.124	.196	.166
				F Tests					
α = .05									
	.721	.765	.899	.690	.800	.898	.551	.791	.798

number of significant results per contrast and comparison; (b) the number of significant results experimentwise of the contrasts and comparisons; and (c) the number of significant results familywise of the overall *F* test for each level of power and for the different distributions when the sample size is based on $\alpha/\underline{c}$. The results of Method II, in Table 9, are presented in the form of the observed levels of power for the various statistical procedures utilized.

Table 9

Observed Levels of Power of Method II for the Distributions and Specified Levels of Power

	Normal			Exponential			Uniform		
Power	90	93	97	90	93	97	90	93	97
				Orthogonal Contrasts					
α = .05									
T1	.831	.870	.932	.843	.853	.901	.721	.749	.835
T2	.524	.599	.652	.563	.609	.656	.411	.472	.531
T3	.219	.246	.271	.236	.245	.284	.186	.182	.205
				Pairwise Comparisons					
α = .05									
C1	.247	.255	.311	.213	.236	.271	.182	.178	.232
C2	.680	.738	.797	.662	.780	.768	.533	.560	.693
C3	.949	.960	.990	.952	.959	.979	.861	.896	.943
C4	.216	.254	.263	.238	.258	.277	.166	.171	.224
C5	.645	.704	.770	.674	.733	.778	.547	.608	.662
C6	.219	.246	.271	.236	.245	.284	.186	.182	.205
				F Tests					
α = .05									
	.904	.932	.973	.902	.927	.956	.766	.832	.906

The power levels of Method II, as seen in Table 9, are similar to those in the Method I in findings, with the levels of power being slightly higher in all areas than they were in Method I. However, it can be noted that the power levels of Method II are still lower than the expected levels of power.

Findings

The data will be presented in accordance with the research questions and procedural questions examined.

Research Questions

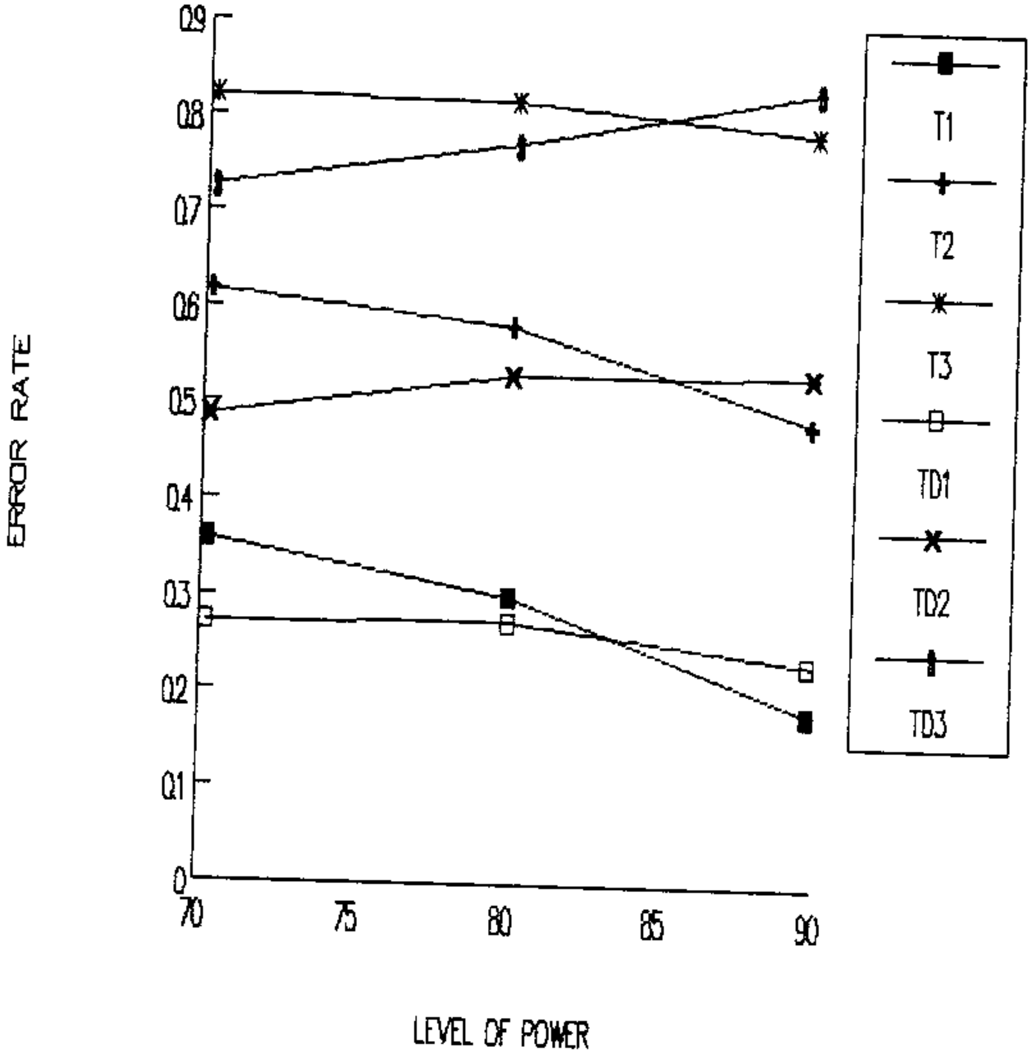
1. What is the difference between the expected Type II error rate and the observed error rate between Helmert orthogonal contrasts and all possible pairwise comparisons for different levels of power and distributions?
(See Tables 10 and 11 in Appendix A.)

The observed error rates per contrast and per comparison were in general larger than the expected error rates. In Figures 1, 2, 3 and 5, it can be noted that the lowest error occurred with the Helmert orthogonal contrasts T1/TD1 (3 -1 -1 -1), followed by T2/TD2 (0 2 -1 -1), and with T3/TD3 (0 0 1 -1) for both Methods I and II. In Method I, the error rates of the LSD and the Dunn-Bonferroni procedures behaved as expected. With the normal and exponential distributions, the Dunn-Bonferroni procedure produced lower error with the lower beta levels, but higher error for the beta level of .10. With the uniform distribution, the LSD procedure produced lower error rates for the beta value of .30, then crossed to higher error rates for the lower beta values.

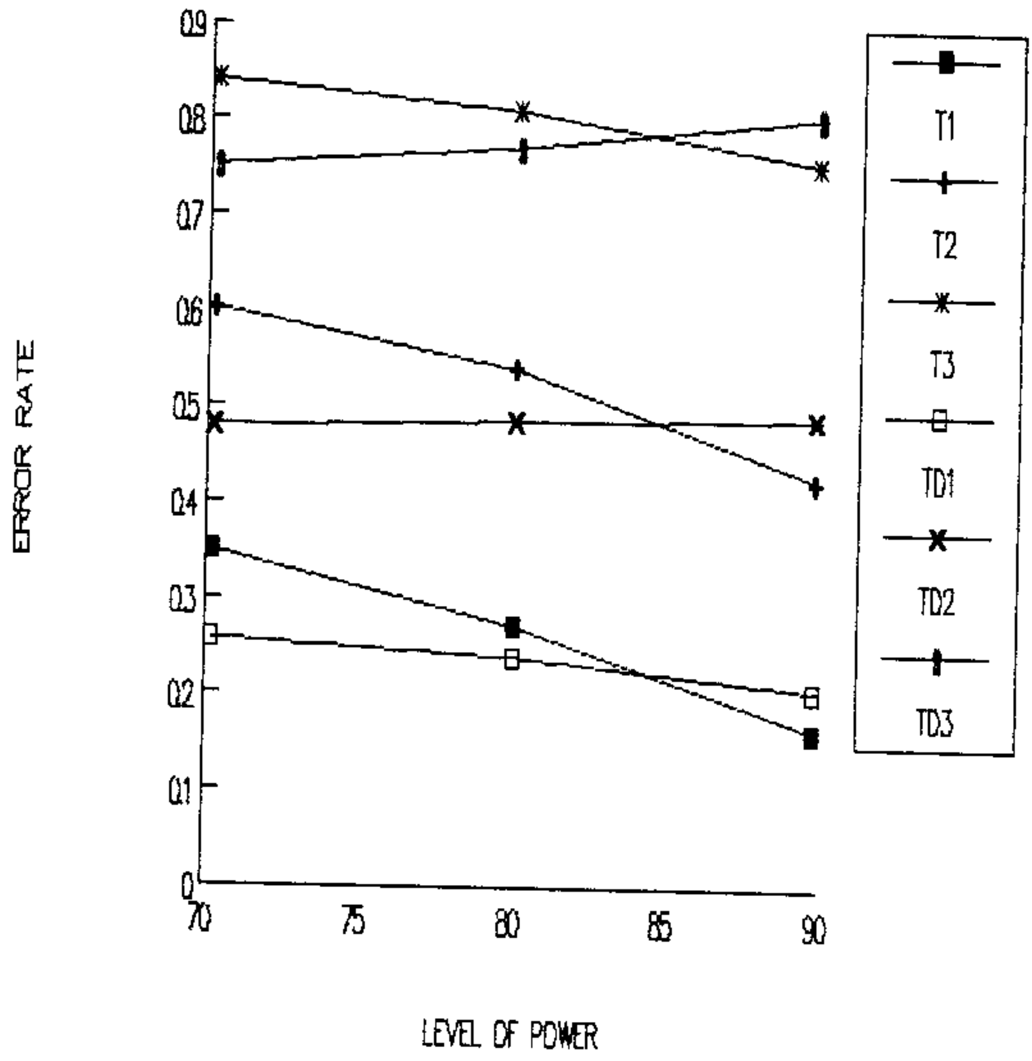
For the pairwise comparisons, as seen in Figures 4 and 6, the LSD of the contrast C3 (1 0 0 -1) for both methods and distributions produced lower than expected error rates. Both methods are testing, at least for this study, the extreme linear range of the group means. Of all the contrasts and comparisons tested, these methods <u>should</u> have tested significantly different. The intermediate results were produced by the comparisons of C2 (1 0 -1 0) and C5 (0 1 0 -1). The largest error was produced by the comparisons of C1 (1 -1 0 0), C4 (0 1 -1 0), and C6 (0 0 1 -1).

The contrasts or comparisons that tested groups in which the means were next to each other had the highest error rate. For example, the orthogonal contrast T3 or TD3, and the pairwise comparisons C1, C4, and C6 produced the largest error rate. The orthogonal contrasts and pairwise comparisons in which, in theory, the means tested were separated by at least one group did produce lower error rates. Therefore, the linear relationship of the means did affect the observed error rates. This phenomenon occurred across the levels of power, method, and distributions.

The familywise error was as expected with the exception of a higher than expected error rate for the uniform distributions. As seen in Figures 7 and 8, the familywise error rate was more greatly affected in the power level of .70 of Method I and the power level of .90 for Method II.



<u>Figure 1</u>. The error rate per contrast for Fisher LSD and the Dunn-Bonferroni procedure of the Helmert contrasts of Method I: Normal distribution.



<u>Figure 2</u>. The error rate per contrast for Fisher LSD and the Dunn-Bonferroni procedure of the Helmert contrasts of Method I: Exponential distribution.

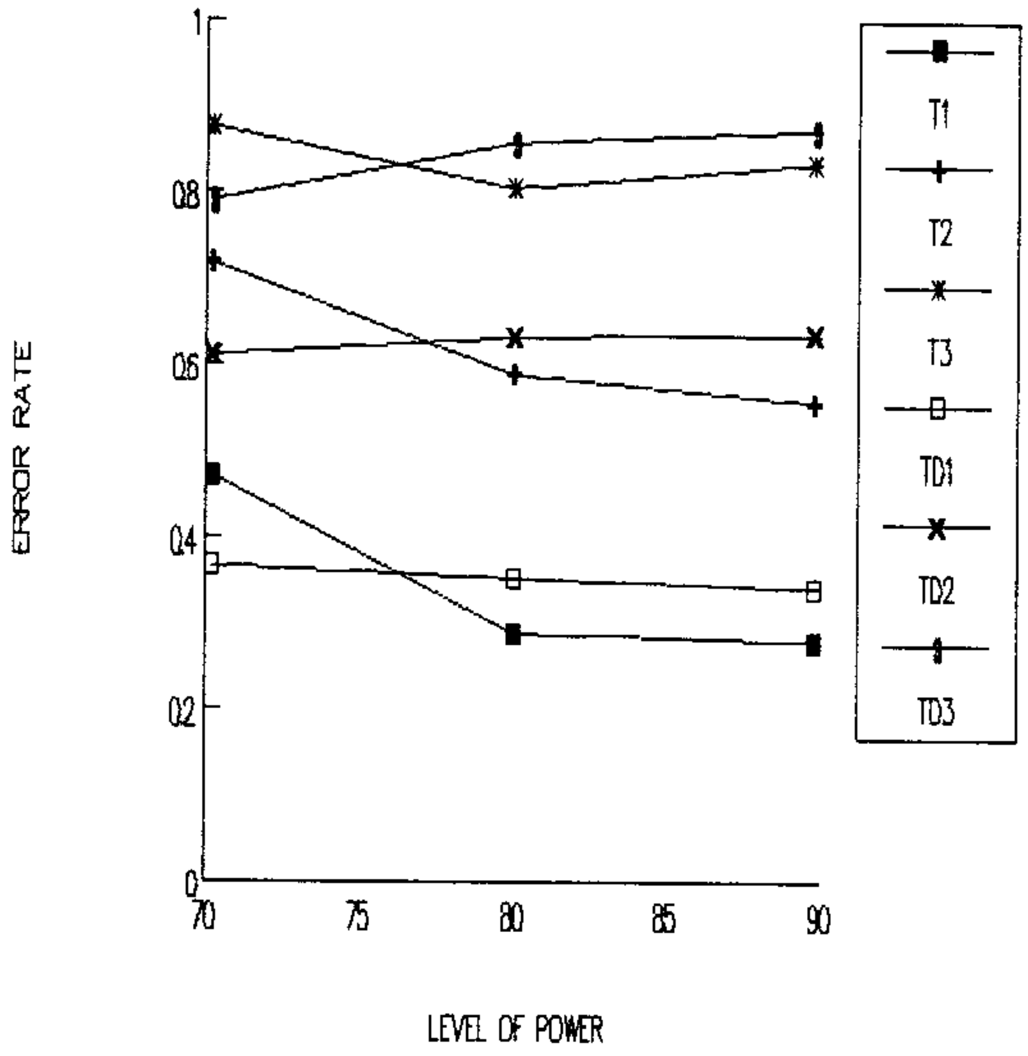
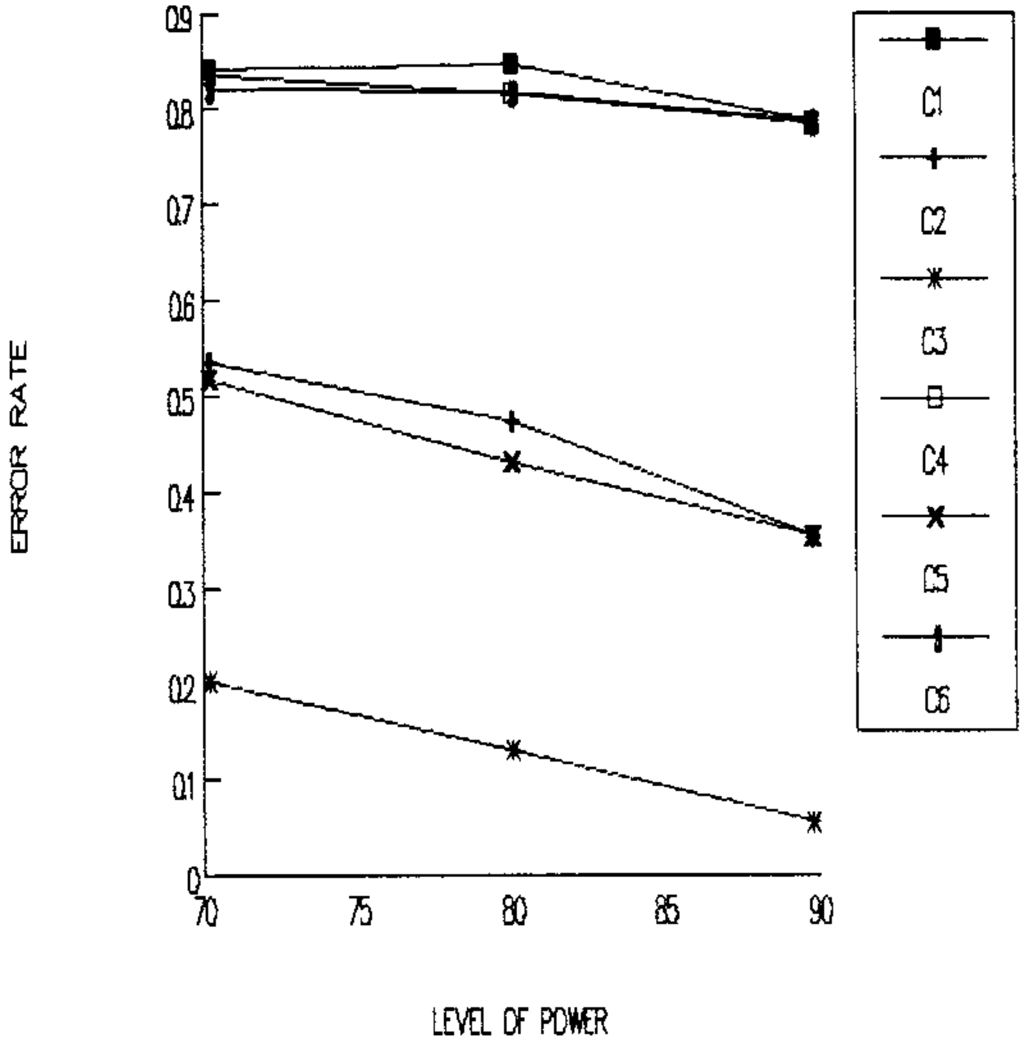
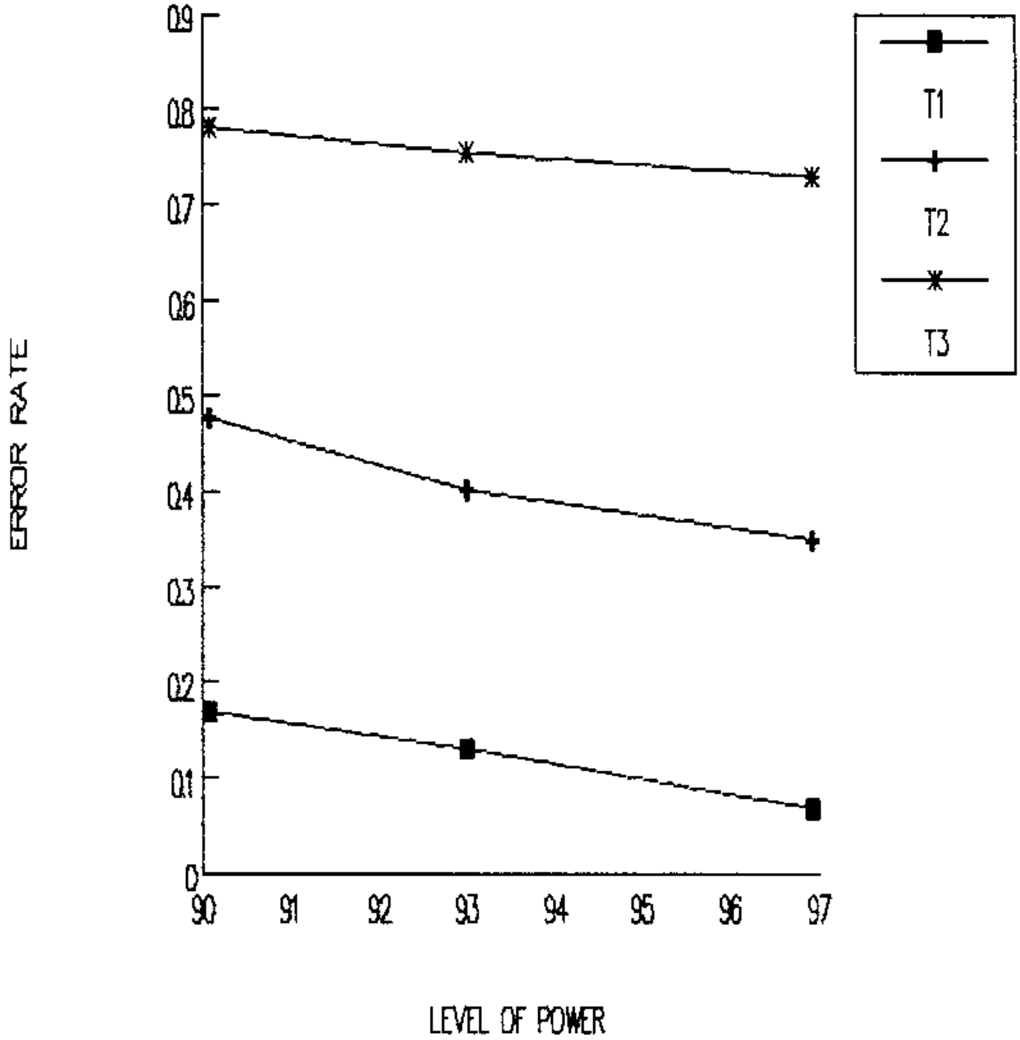


Figure 3. The error rate per contrast for Fisher LSD and the Dunn-Bonferroni procedure of the Helmert contrasts of Method I: Uniform distribution.



<u>Figure 4</u>. The error rate per pairwise comparison for Fisher LSD of Method I: Normal distribution.



<u>Figure 5</u>. The error rate per contrast for Fisher LSD of the Helmert contrasts of Method II: Normal distribution.

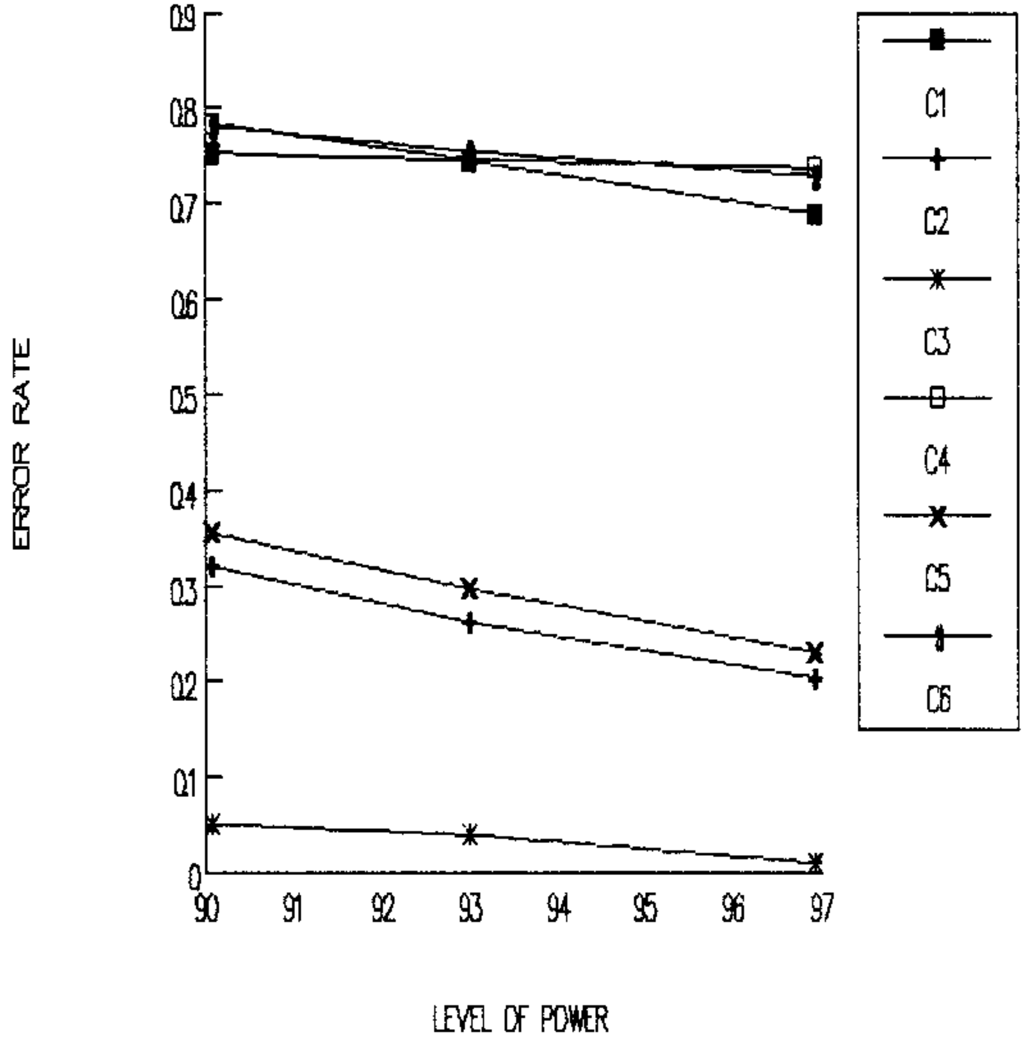


Figure 6. The error rate per pairwise comparison for Fisher LSD of Method II: Normal distribution.

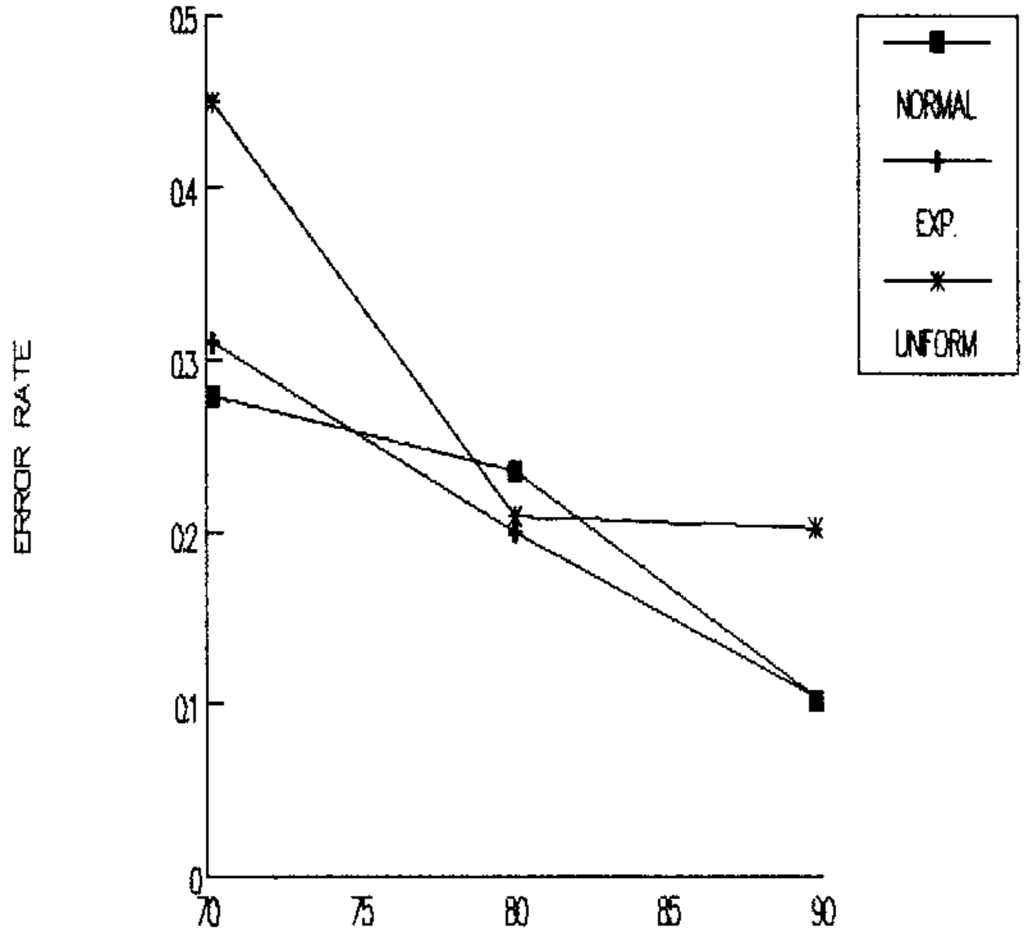


Figure 7. The familywise error rate for the different distributions and levels of power: Method I.

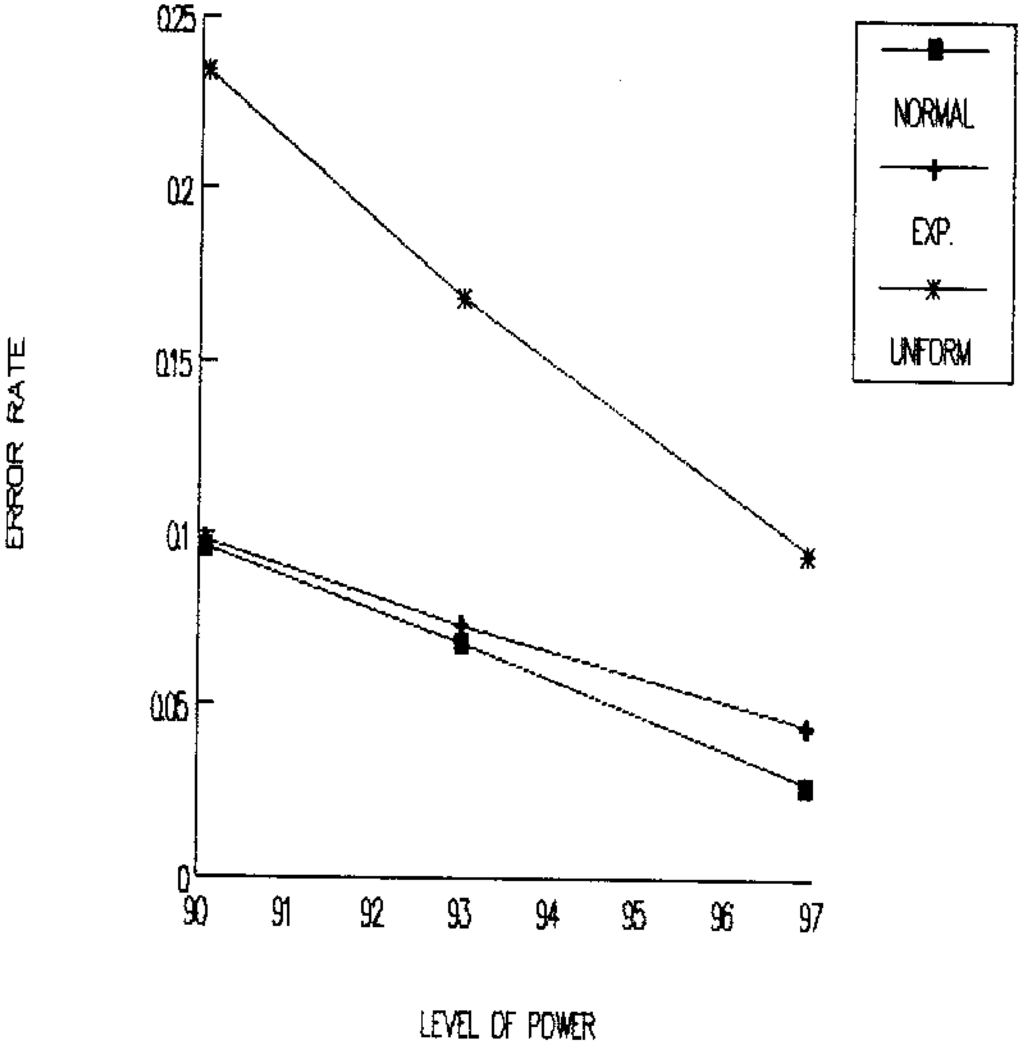


Figure 8. The familywise error rate for the different distributions and levels of power: Method II.

2. What is the difference between the expected and observed experimentwise Type II error rates of the Helmert orthogonal contrasts and all possible pairwise comparisons when using the different levels of power and the distributions? (See Tables 12 to 16, Appendix A.)

If the experiments are evaluated from a strict criteria, the results indicate that all of the observed experimentwise error rates were much larger than the expected experimentwise error rates for all levels of power and the distributions (Tables 12 and 13, Appendix A). For the orthogonal contrasts, the number of complete significant experiments was less than 18 percent at the most for all levels of power, distribution, and method. The largest experimentwise error did occur with the largest beta values and the least error for the smallest beta value, as was expected. The error values did not vary to any extent from distribution to distribution (Figure 9). The pairwise comparisons contained only 3 experiments in the 18 simulation situations in which all of the comparisons were significant (Table 13, Appendix A). Thus only 3 experiments in 18,000 were significant for all of the pairwise comparisons. All 3 experiments were from Method II, which did have larger sample sizes than Method I.

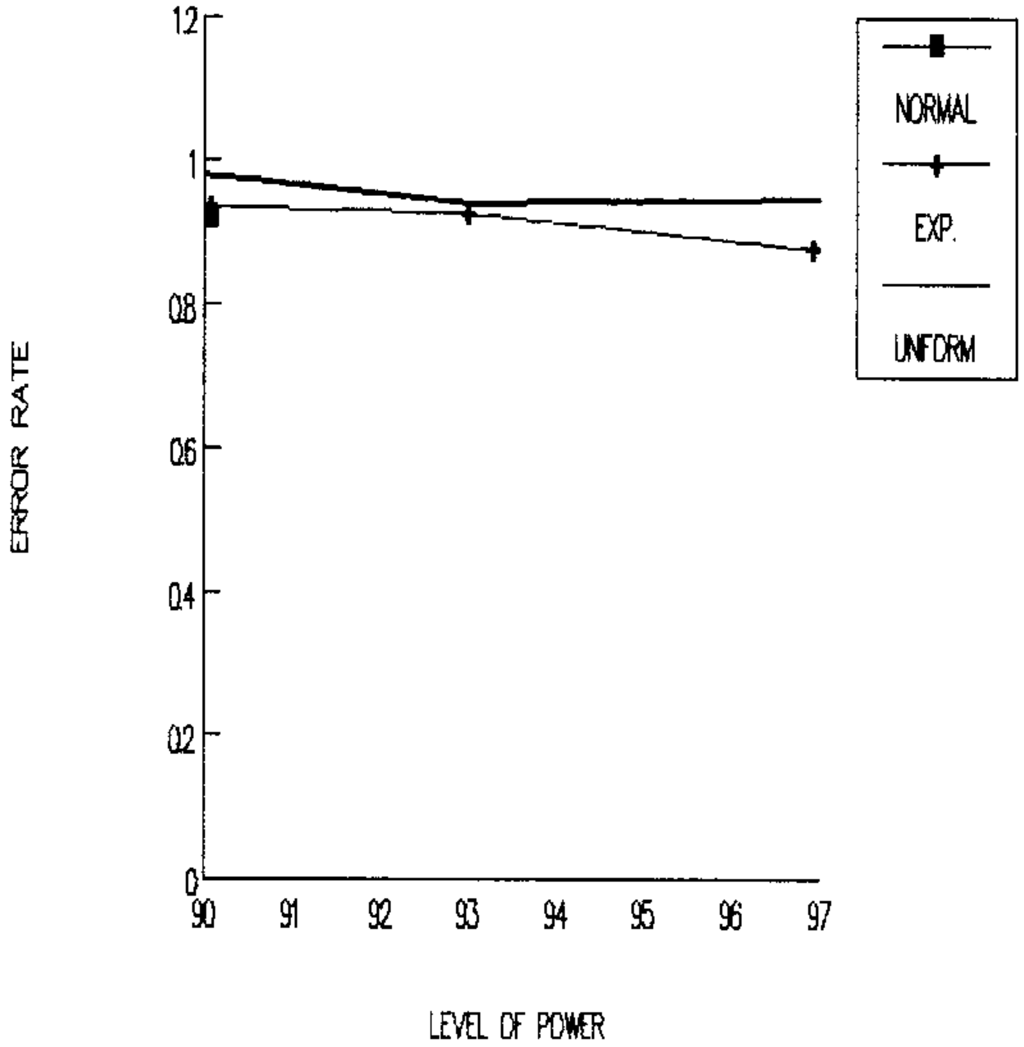


Figure 9. The experimentwise error rate for the Fisher LSD procedure of Method II for all levels of power and the distributions: Strict criteria.

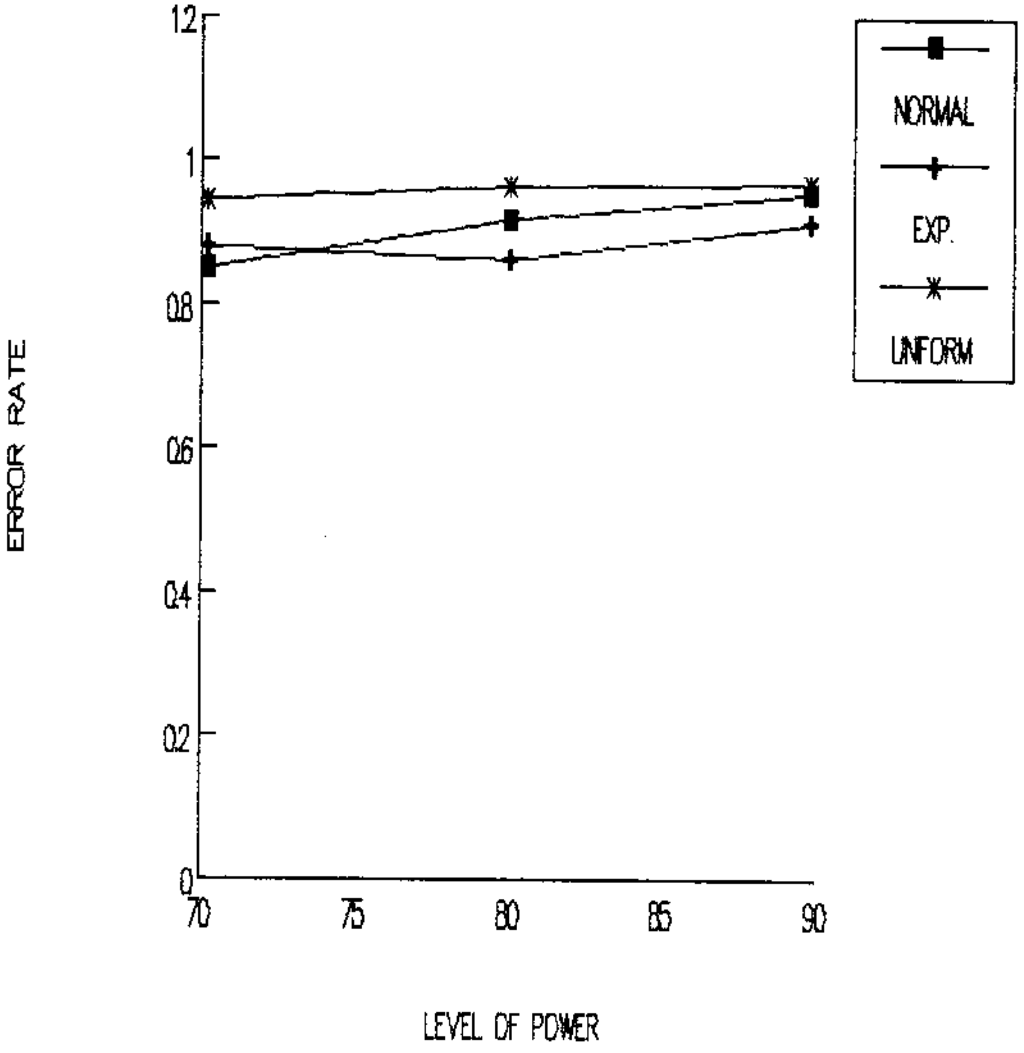


Figure 10. The experimentwise error rate for the Dunn-Bonferroni procedure of Method I for all levels of power and the distributions.

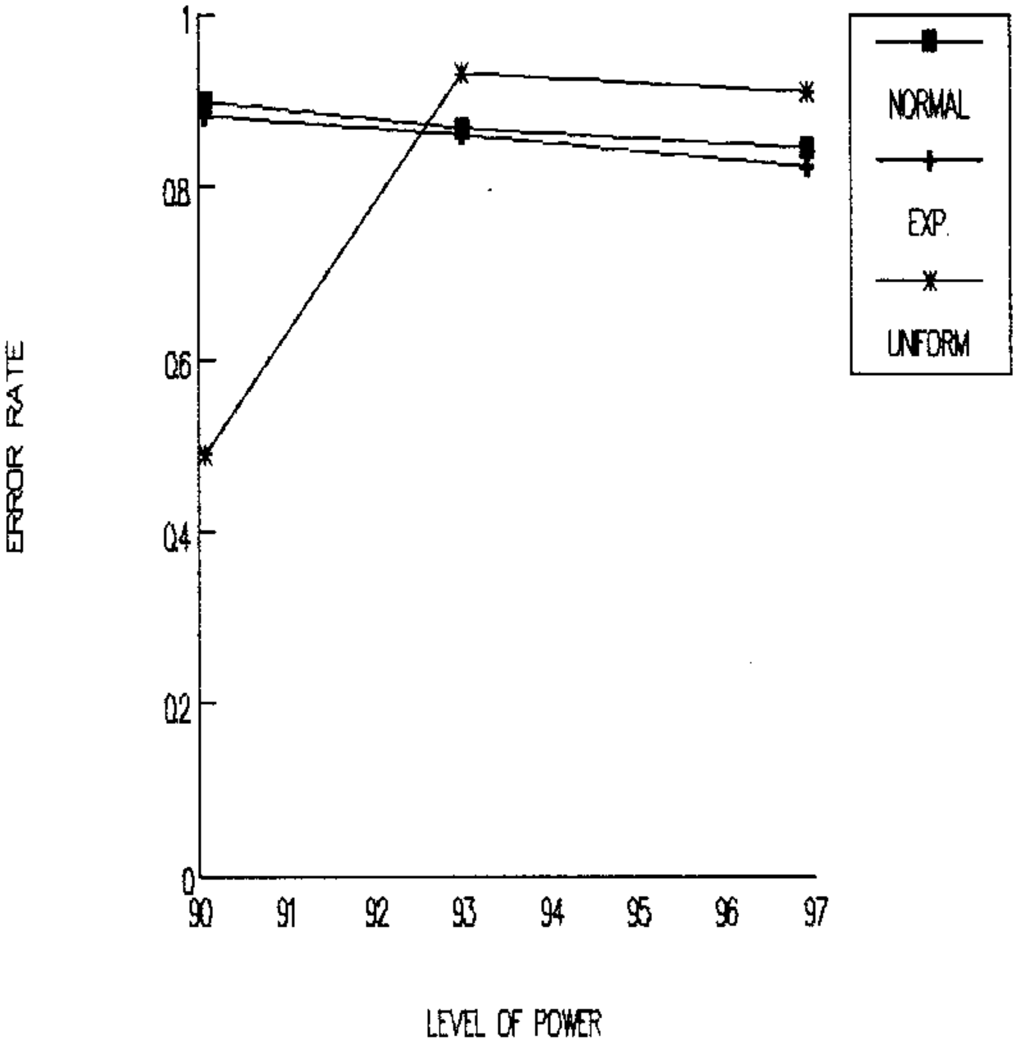


Figure 11. The experimentwise error rate for the Fisher LSD procedure of Method II for all levels of power and the distributions.

Table 17

Experimentwise Error Rate for Method I: Strict Criteria

	Normal			Exponential			Uniform		
Power	70	80	90	70	80	90	70	80	90
				Orthogonal Contrasts					
α = .05	.950	.941	.920	.936	.925	.876	.980	.940	.946
$\alpha = \beta/C$	.850	.916	.953	.880	.860	.913	.945	.963	.966
				Pairwise Comparisons					
α = .05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 18

Experimentwise Error Rate for Method II: Strict Criteria

	Normal			Exponential			Uniform		
Power	90	93	97	90	93	97	90	93	97
				Orthogonal Contrasts					
α = .05	.898	.868	.845	.882	.860	.822	.490	.932	.909
				Pairwise Comparisons					
α = .05	1.00	.999	.999	1.00	.999	1.00	1.00	1.00	1.00

If, however, the contrasts/comparisons are examined in the lenient criteria evaluation, in which at least one of the planned contrasts/comparisons are significant, then some interesting trends emerge. First, by the definition of lenient criteria, the percentage of such experiments either meets or exceeds the initial level of power for all levels of power, method, and distribution.

Second, when examining the individual contrasts or comparisons, as the level of power increases, the number of significant experiments increases and the number of nonsignificant experiments decreases. Method II had the least number of nonsignificant experiments across levels of power and to the violations of normality. The number of nonsignificant experiments for pairwise comparisons was less than the orthogonal contrasts in most of the simulations.

For the Helmert contrasts in Method I, the largest number of significant experiments occurred when only one of the contrasts was significant. For Method II, the largest number of significant Helmert contrasts occurred with at least two of the three contrasts significant. For the pairwise comparison, the combination of at least three of the six comparisons being significant produces the largest number of results for all levels, distributions, and methods. While zero results were found under the strict criteria, the lenient criteria producing the least number of nonsignificant results were two experiments for the power of .97 under Method II (normal distribution).

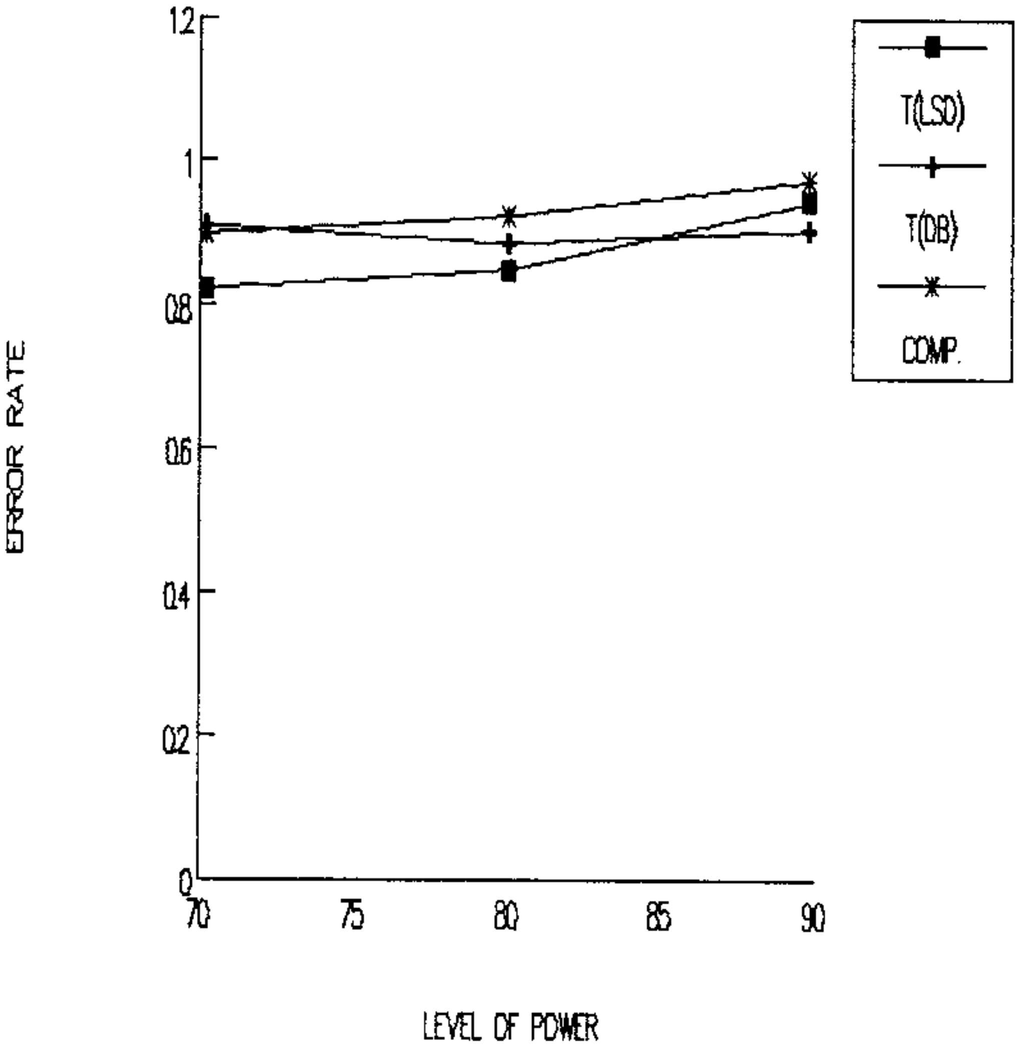


Figure 12. The percent of significant contrasts/comparisons under the lenient criteria for Method I: Normal distribution.

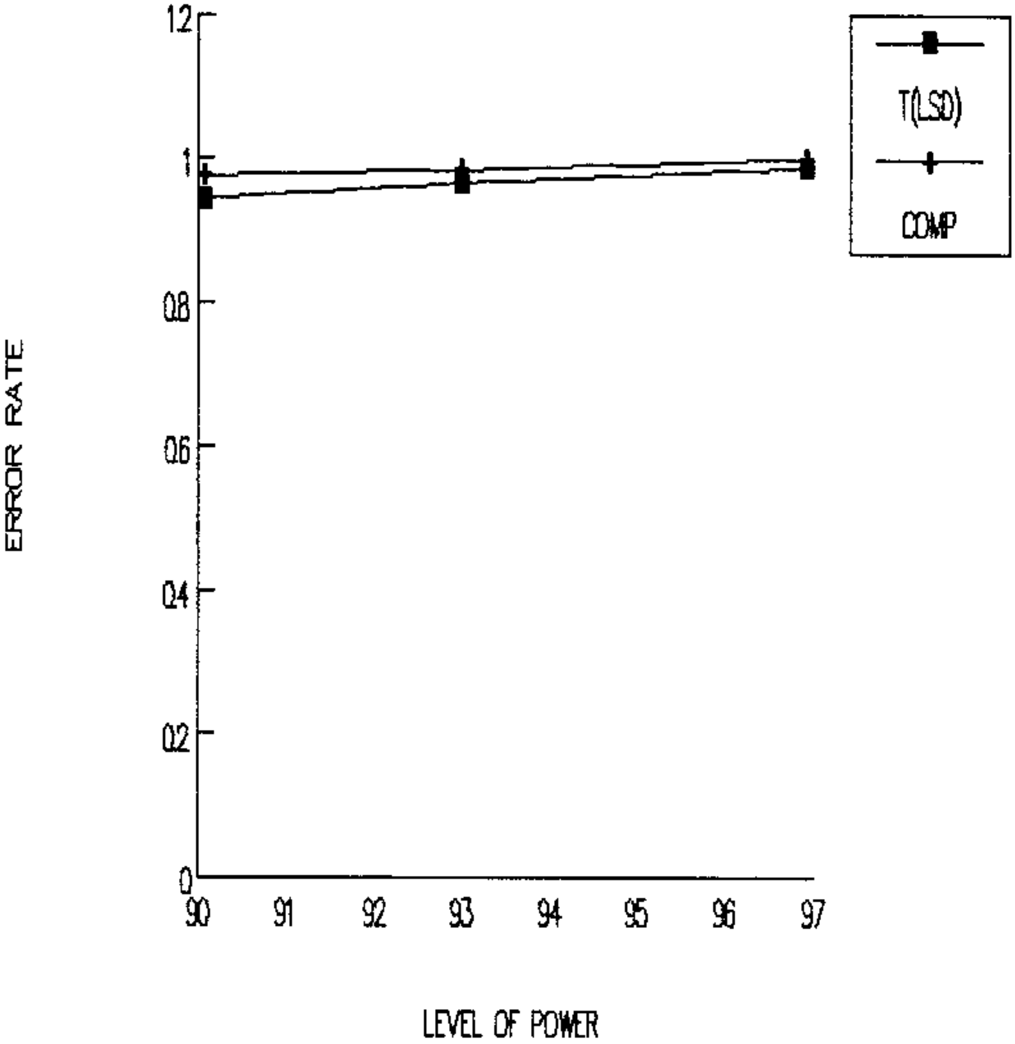
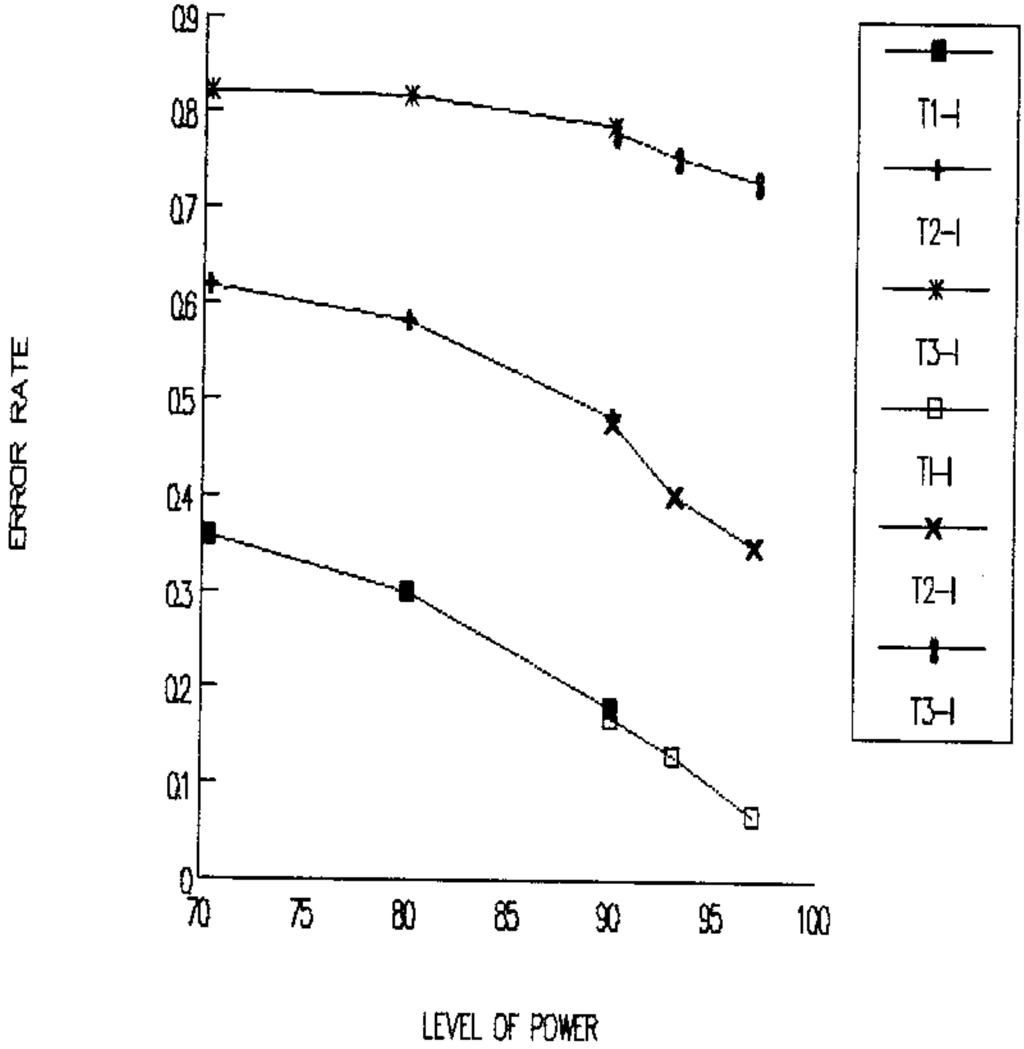


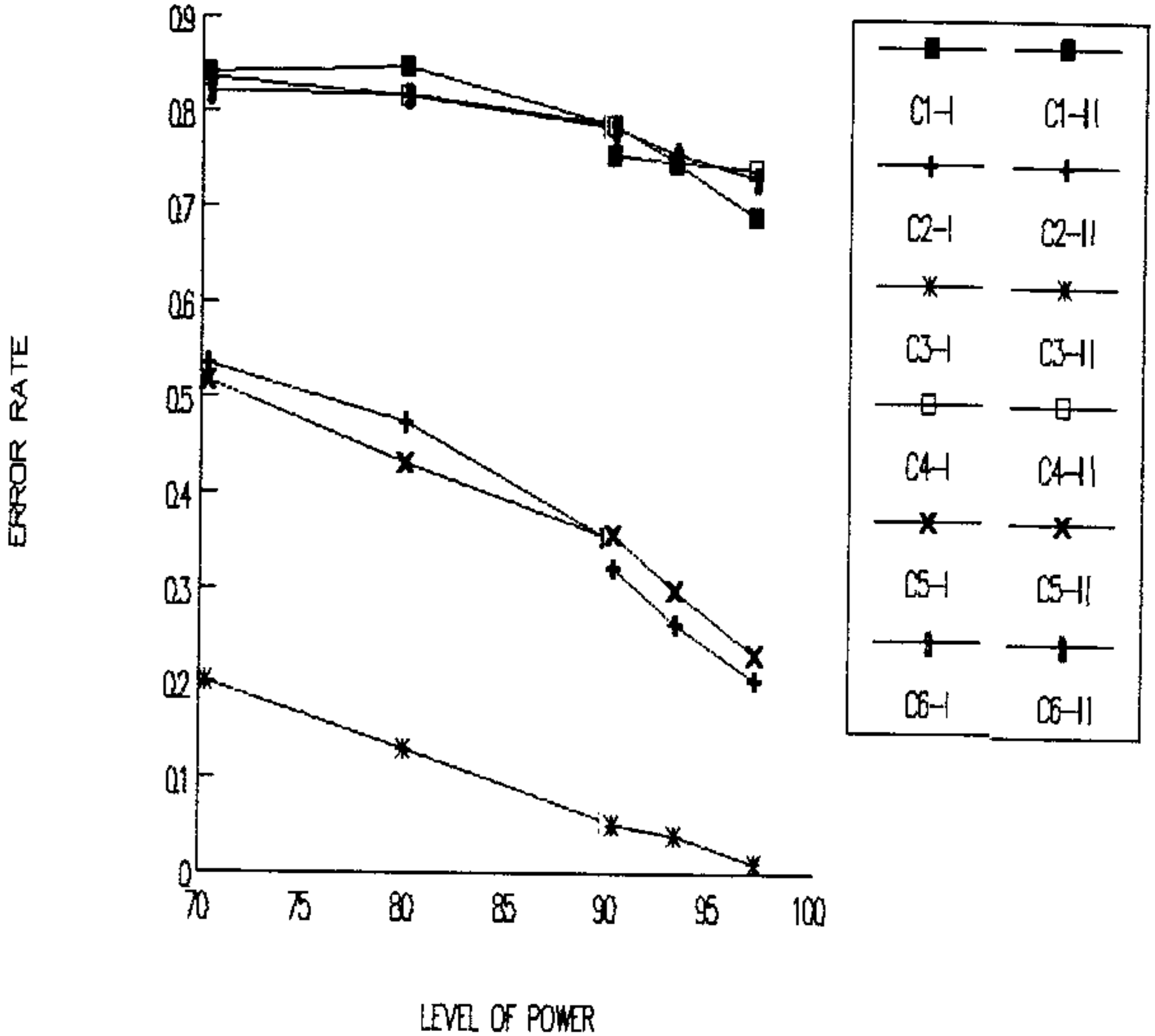
Figure 13. The percent of significant contrasts/comparisons under the lenient criteria for Method I: Normal distribution.

3. Regarding the three procedures--(a) the Fisher Least Significant Difference with α = .05 and the sample size determined from the beta error per family (β_F); (b) the Dunn-Bonferroni inequality procedure with α = .05 and the sample size determined from the beta error per contrast (/c); and (c) the Dunn-Bonferroni inequality procedure with α = β/c and the sample size based on the beta error per family (β_F)--how do the procedures compare in the number of Type II error rates for the levels of power and distribution? (See Tables 19 to 21, Appendix A.)

Of the three procedures, both of the Dunn-Bonferroni procedures produced, in general, fewer Type II error rates than did the Fisher LSD. The exception is the Beta value of .10 in Method I, which, when subdivided by the number of contrasts, has a smaller area of rejection than the Fisher LSD. The Dunn-Bonferroni procedure of Method II did produce a lower Type II error rate than did the Dunn-Bonferroni procedure of Method I. As seen in Figure 14, when comparing the orthogonal contrasts of T1/TD1 and T2/TD2 in Method I and of T1 and T2 in Method II, Method II clearly produced lower error rates. For the pairwise comparisons, Method II produced smaller error rates than did Method I (Figure 15). Method II was able to produce this reduced error rate, but did so with a larger sample size requirement. Indeed,



<u>Figure 14</u>. The error rates per comparison of the Helmert orthogonal contrast for Methods I and II: Normal distribution.



<u>Figure 15</u>. The error rates per comparison of the pairwise comparisons for Methods I and II: Normal distribution.

across the methods, levels, and distributions, a larger sample size produced fewer Type II error rates.

<u>Procedural Question</u>

I. Are the *F* test probabilities distributed according to the noncentral *F* distribution?

The Kolmogorov-Smirnov test of goodness of fit determines if an observed distribution departs from a theoretical distribution. For this, the observed distributions were the *F* test probabilities of the noncentral *F* distribution for the three distributions utilized in the study. (See Table 22, Appendix A.)

As seen in Figures 16 to 21, the normal and exponential distributions for the level of power .70 are not significantly different from those expected, indicating that the observed noncentral *F* distributions closely approximate the theoretical distribution. The simulations of the exponential distribution of Method I with a power of .90 had four significant intervals but the differences were small. All of the observed distributions of the uniform distributions are statistically different from the theoretical distribution. For Method I, 9 of 18 levels are significantly different. For Method II, 11 of the levels are significantly different. The distribution probability levels from .10 to .975 are less than the theoretical distribution, producing a pronounced curve instead of a straight line as expected. This confirms the general

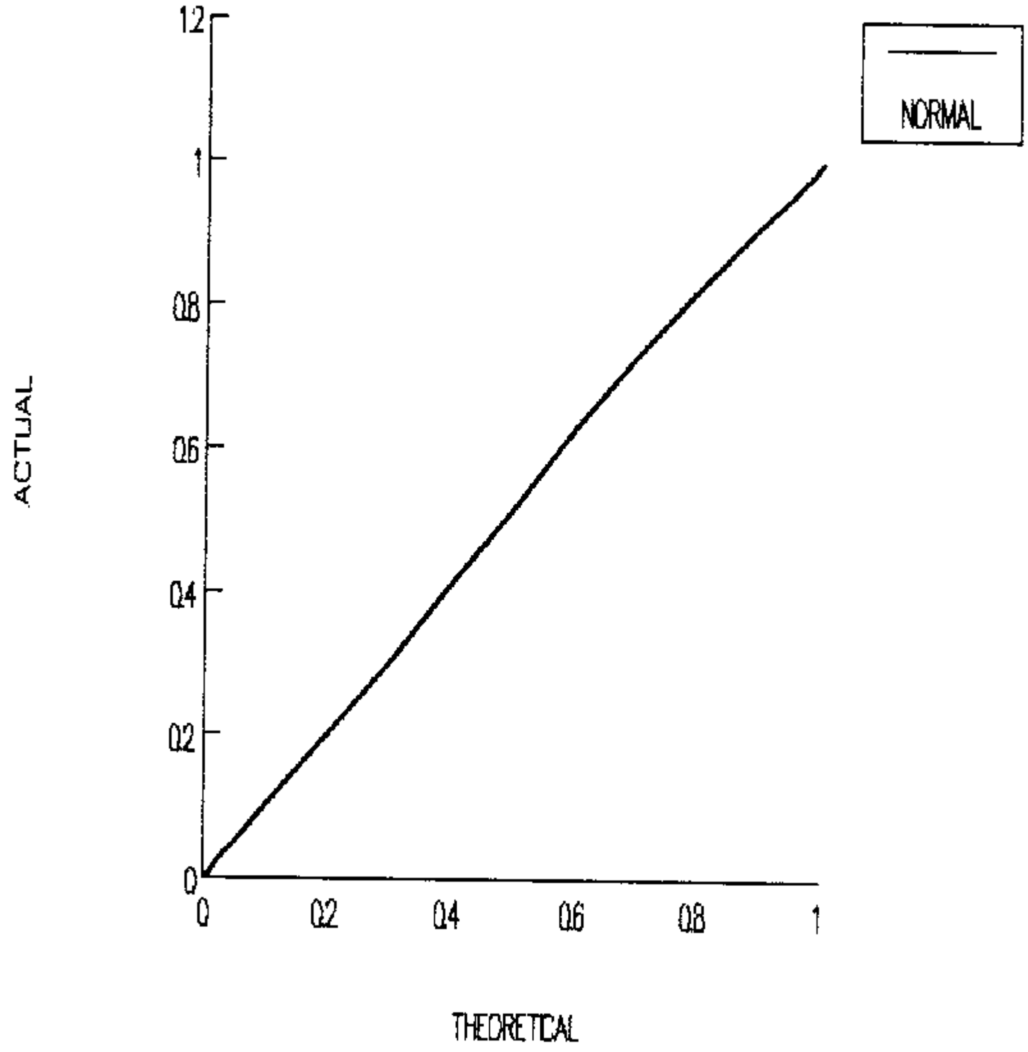


Figure 16. Kolmogorov-Smirnov Goodness of Fit Test comparison of actual and theoretical levels of significance of the distributions of the F test probabilities theoretical distributions levels: Normal distribution, Method I, Power level .70.

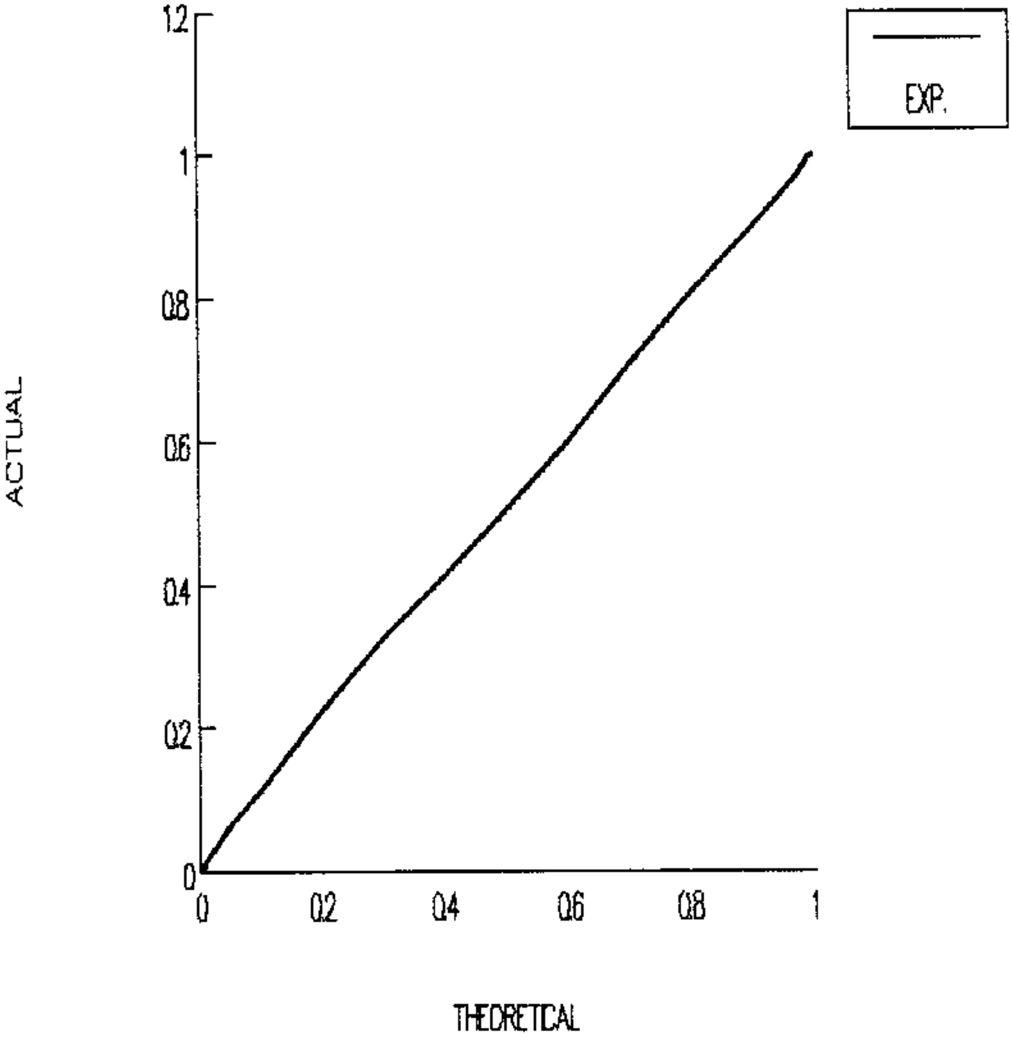


Figure 17. Kolmogorov-Smirnov Goodness of Fit Test comparison of actual and theoretical levels of significance of the distributions of the F test probabilities theoretical distribution Levels: Exponential distribution, Method I, Power level .70.

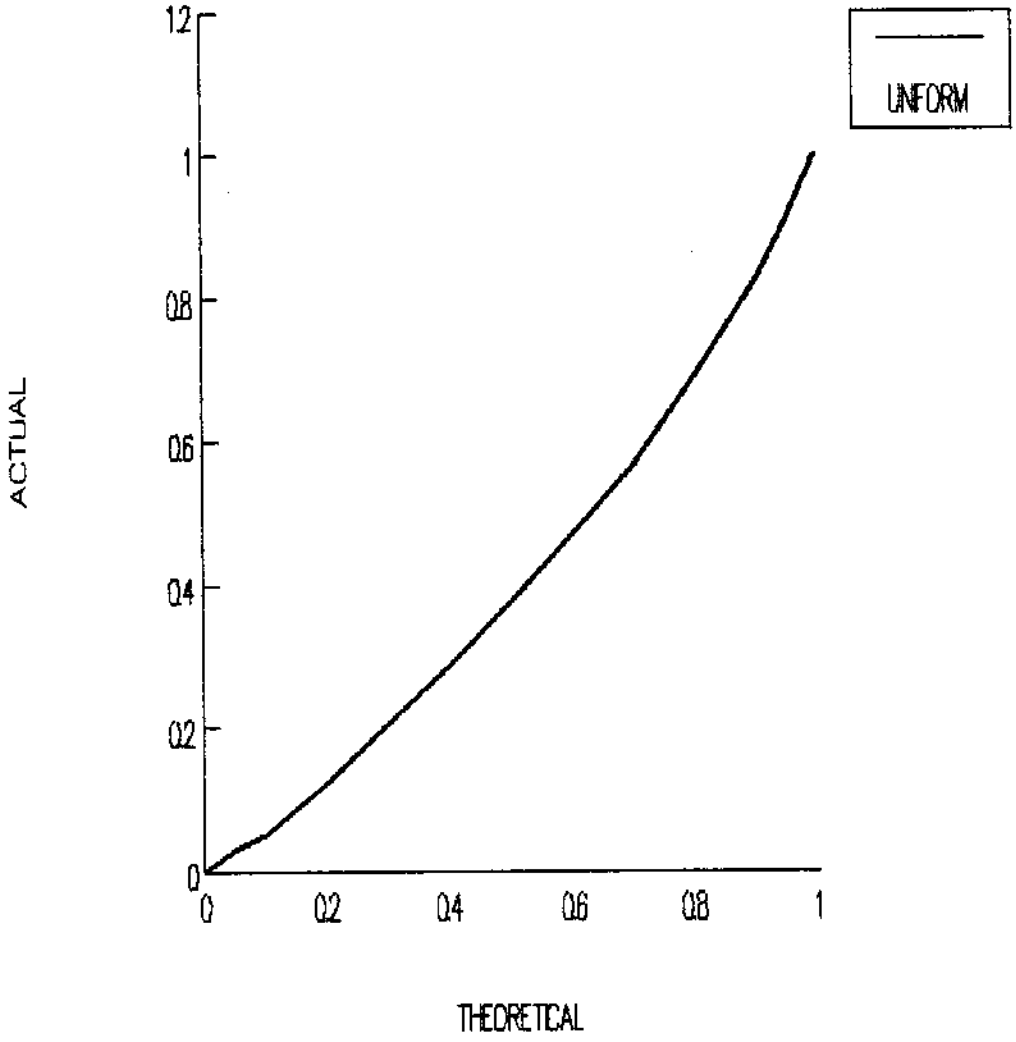


Figure 18. Kolmogorov-Smirnov Goodness of Fit Test comparison of actual and theoretical levels of significance of the distributions of the *F* test probabilities theoretical distributions levels: Uniform distribution, Method I, Power level .70.

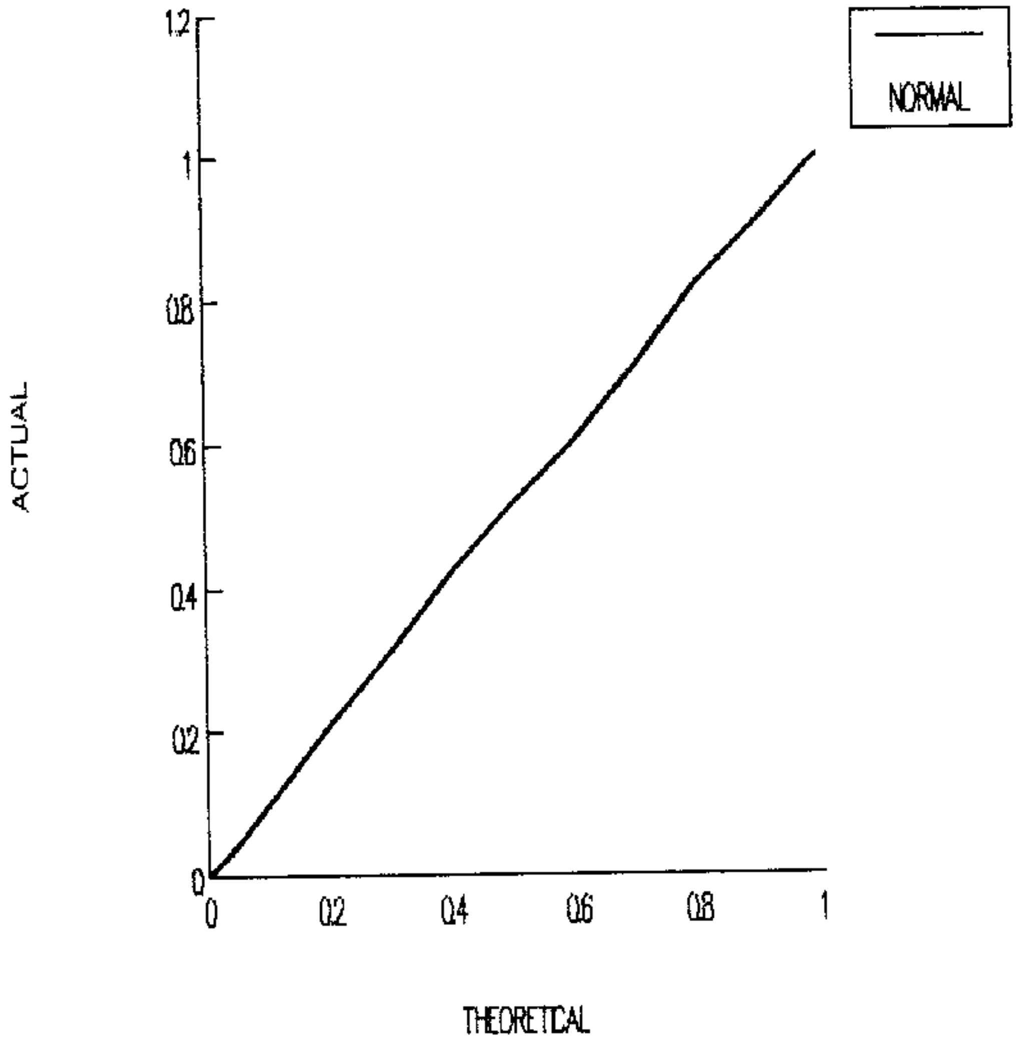


Figure 19. Kolmogorov-Smirnov Goodness of Fit Test comparison of actual and theoretical levels of significance of the distributions of the *F* test probabilities theoretical distributions levels: Normal distribution, Method II, Power level .70.

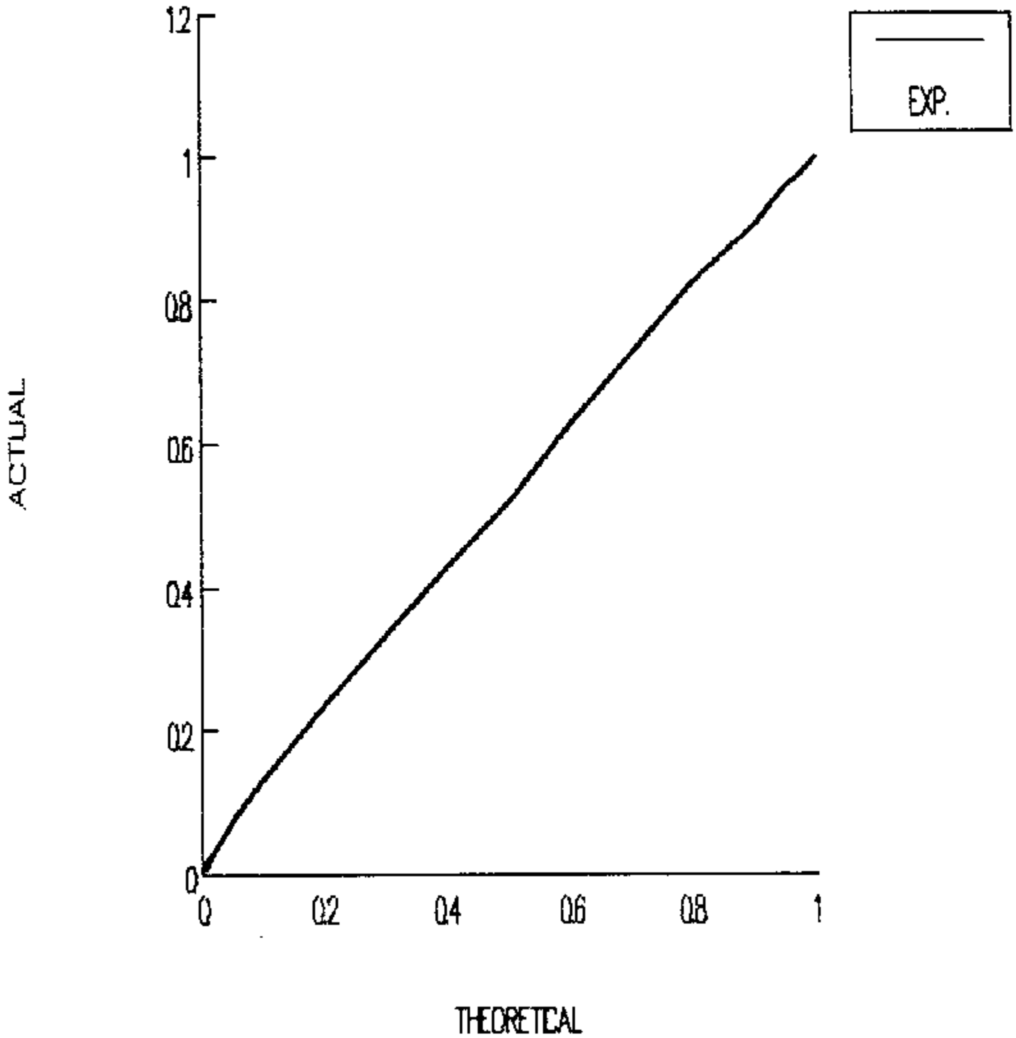


Figure 20. Kolmogorov-Smirnov Goodness of Fit Test comparison of actual and theoretical levels of significance of the distributions of the F test probabilities theoretical distribution levels: Exponential distribution, Method II, Power level .70.

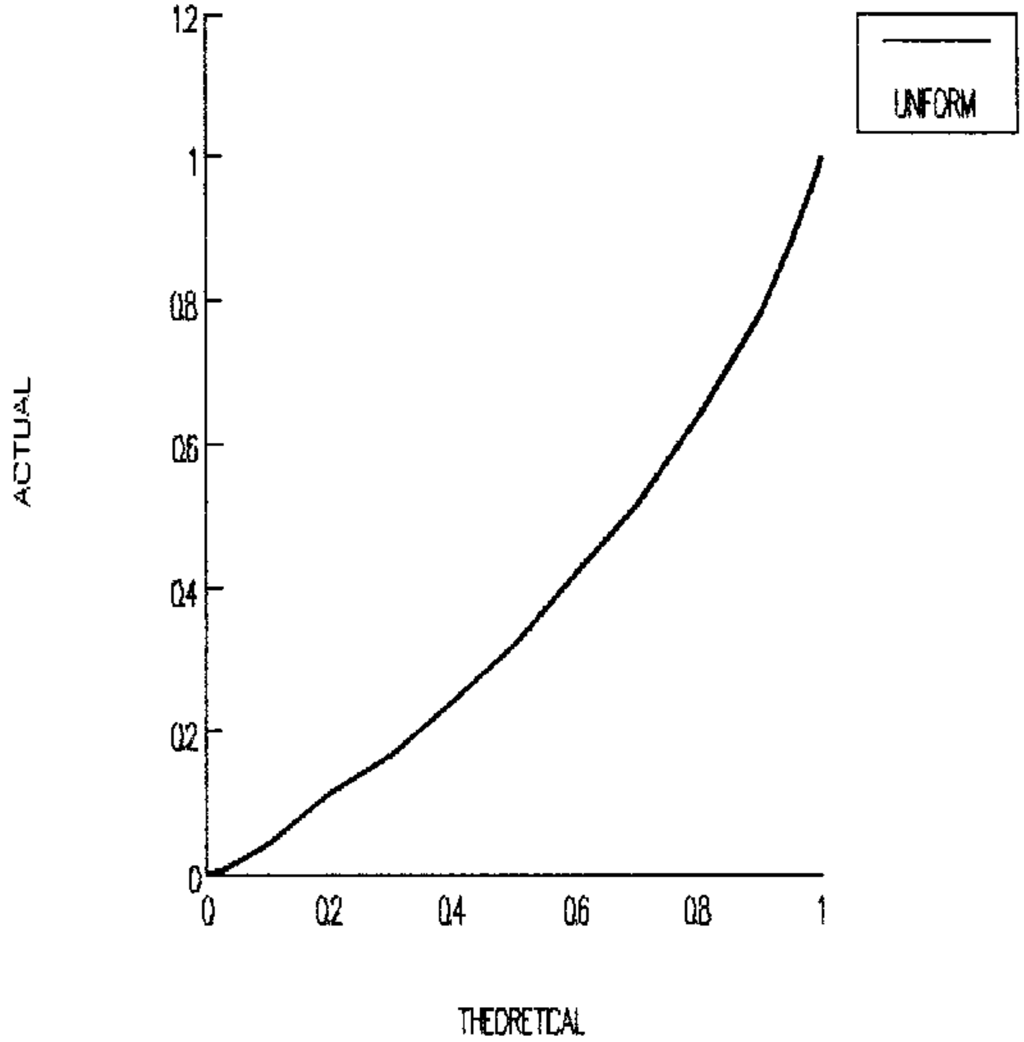


Figure 21. Kolmogorov-Smirnov Goodness of Fit Test comparison of actual and theoretical levels of significance of the distributions of the F test probabilities theoretical distribution levels: Uniform distribution, Method II, Power level .70.

findings of the study, in that overall the uniform distribution experiments produced larger error terms and some unusual behavior of the contrasts.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The main purpose of this study was to investigate, through the use of Monte Carlo simulations, the accumulation of Type II errors to the varying levels of power, population distributions, and methodology. The Monte Carlo computer simulations controlled the various parameters of the study within each experiment. The simulation conditions included in this study involved equal sample size, equally spaced means, and three specified levels of power. The population distributions, from which the error terms were sampled, were normal, exponential, or uniform, with a mean of zero and standard deviation of one.

The observed levels of power produced were obtained by computing the proportion of times each experiment yielded a rejection of the hypothesis at the specified levels of significance. The various error terms were computed by the proportion of times the experiment did not reject the hypothesis at the specified levels of power.

The study was reported by the methodology procedure examined and the various research questions. Several points emerge from the study:

1. Under the strict criteria, the accumulation of Type II experimentwise error does occur at a significant rate, especially for pairwise comparisons, at all specified levels of power. The high rate of accumulation can be expected since the value of beta is always much higher than alpha. The study does indicate that the probability of falsely accepting at least one null hypothesis in a family of tests, when in theory all of the alternate hypotheses are true, is high, precluding valid tests at the beginning of the study.

2. Neither of the Dunn-Bonferroni procedures utilized in this study resulted in significant improvement over the Fisher LSD for the Helmert contrasts. Although the Dunn-Bonferroni procedure with the sample size based on the beta per contrast did produce an improvement in the number of significant results, it did so at the cost of requiring a much larger sample size. The Dunn-Bonferroni procedure basing the critical value on the beta per contrast is too nontraditional of an approach to statistical methodology to be widely accepted unless dramatic changes could be verified. Therefore, the usage of either procedure could not be justified for a change in methodology.

3. From the lenient criteria evaluation of the means, the optimum approach to limiting the accumulation of Type II errors is two-pronged. First, orthogonal contrasts are more favorable than pairwise comparisons in limiting the number of falsely accepted null hypotheses. The problem lies in

that some researchers view the use of contrasts as not being as easy as the pairwise approach, although this study does indicate that the extra effort and forethought of planning orthogonal contrasts will pay off in the increased number of significant findings. If pairwise comparisons are to be utilized, then the number of comparisons should be close to the number of means.

Second, adequate sample size is indicated as a method of limiting the accumulation of Type II error. The use of Cohen's tables for the determination of sample size as indicated from this study only assures that the extreme differences in means will be detected, not the differences in adjacent means. Therefore, the probability of Type II errors remains high even if the proper sample size is utilized. Cohen's tables for sample size determination are fairly simple to use for the average researcher. Again, adequate sample size does require planning and forethought by the researcher.

The use of orthogonal contrasts or of a limited number of comparisons <u>and</u> adequate sample size is not a panacea for the accumulation of Type II error. As a part of proper methodology, however, it will improve the probability of rejecting the null hypothesis when the alternate hypothesis is true.

4. The accumulation of Type II error is irrespective of distribution. Although some variation among distributions was noted, no significant differences could be found.

Conclusions

On the basis of the results of the various data presented, the following conclusions are appropriate.

1. The accumulation of Type II exerimentwise error is as substantial as Westermann and Hager (1986) theorized for both orthogonal contrasts and pairwise comparisons. Using the strict criteria of evaluation, one could argue that with an experimentwise error rate of 1.00 (as was found in 15 of the 18 simulation situations), the accumulaltion of error is larger than one would imagine.

2. The accumulation of Type II error is not affected by the violation of normality. Research that would support or contradict this finding is not found in the literature.

3. The two proposed Dunn-Bonferroni methods under investigation to limit the accumulation of Type II error were not as effective as expected. Westermann and Hager's (1986) proposal that the Dunn-Bonferroni method be used to deal with this issue is interesting but is not applicable for the researcher.

4. The procedures as outlined by Games (1971) to deal with Type I error are applicable for Type II errors, as well, namely: (a) isolate the error by the use of orthogonal contrasts; or (b) limit the number of pairwise comparisons to no more than the number of means.

Recommendations

The purpose of this study was to examine the issue of the accumulation of Type II error and possible methodologies to limit that accumulation when the assumptions of normality are violated. The study did not deal with all of the issues of the accumulation of Type II error, since only the medium effect size and medium variability were explored. Additional situations for research are indicated to explore fully the implications of the accumulation of Type II error.

It would be of interest to see if similar results would occur in the accumulation of Type II error:

1. when the combination of the assumptions of normality and equal variance are violated;

2. when the effect size and range variability are other than medium; and

3. when the number of group means are less than four.

APPENDIX A

TABLES 10 TO 16

AND 19 TO 22

RESULTS OF THE SIMULATIONS

Table 10

Observed Error Rate per Contrast and Comparison with Method I for the Distributions and Specified Levels of Power

	Normal			Exponential			Uniform		
Power	70	80	90	70	80	90	70	80	90
				Orthogonal Contrasts					
α = .05									
T1	.358	.300	.179	.349	.272	.162	.471	.288	.278
T2	.617	.582	.482	.602	.539	.424	.717	.587	.555
T3	.822	.817	.787	.841	.810	.756	.876	.804	.834
$\alpha = \beta/C$									
TD1	.271	.273	.232	.258	.239	.206	.376	.351	.340
TD2	.486	.529	.532	.480	.485	.489	.611	.631	.633
TD3	.727	.772	.833	.751	.771	.806	.790	.857	.873
				Pairwise Comparisons					
α = .05									
C1	.841	.846	.787	.849	.813	.766	.877	.786	.815
C2	.535	.474	.352	.509	.434	.321	.616	.461	.425
C3	.201	.130	.055	.205	.125	.056	.322	.121	.120
C4	.836	.816	.784	.854	.810	.762	.873	.835	.808
C5	.517	.430	.354	.525	.428	.312	.636	.467	.454
C6	.822	.817	.787	.841	.810	.756	.876	.804	.834

Table 11

Observed Error Rate per Contrast and Comparison with Method II for the Distributions and Specified Levels of Power

	Normal			Exponential			Uniform		
Power	90	93	97	90	93	97	90	93	97
				Orthogonal Contrasts					
α = .05									
T1	.169	.130	.068	.157	.147	.099	.279	.251	.165
T2	.476	.401	.348	.437	.391	.344	.589	.528	.469
T3	.781	.754	.729	.764	.755	.716	.814	.818	.795
				Pairwise Comparisons					
α = .05									
C1	.753	.745	.689	.787	.764	.729	.818	.822	.768
C2	.320	.262	.203	.383	.220	.232	.467	.440	.307
C3	.051	.040	.010	.048	.041	.021	.139	.104	.057
C4	.784	.746	.737	.762	.742	.723	.834	.829	.776
C5	.355	.296	.230	.326	.267	.222	.453	.392	.338
C6	.781	.754	.729	.764	.755	.716	.814	.818	.795

Table 12

The Percent of Significant Contrasts/Comparisons per Experiment with Method I for the Strict and Lenient Criteria

	Normal			Exponential			Uniform		
Power	70	80	90	70	80	90	70	80	90
				Orthogonal Contrasts					
α = .05									
Strict Criteria									
3 of 3	.050	.059	.080	.064	.075	.124	.020	.060	.054
Lenient Criteria									
2 of 3	.281	.336	.449	.274	.366	.475	.203	.339	.352
1 of 3	.491	.452	.414	.468	.422	.336	.470	.463	.467
Sum	.822	.848	.943	.806	.863	.930	.693	.862	.873
0 of 3	.178	.153	.057	.194	.137	.065	.307	.138	.127
$\alpha = \beta/C$									
Strict Criteria									
3 of 3	.105	.084	.047	.120	.140	.087	.055	.037	.034
Lenient Criteria									
2 of 3	.397	.374	.405	.375	.410	.420	.299	.277	.267
1 of 3	.407	.726	.452	.401	.373	.398	.469	.493	.518
Sum	.909	.884	.904	.895	.887	.905	.803	.807	.819
0 of 3	.091	.116	.096	.105	.113	.095	.117	.793	.181
				Pairwise Comparisons					
α = .05									
Strict Criteria									
6 of 6	.000	.000	.000	.000	.000	.000	.000	.000	.000
Lenient Criteria									
5 of 6	.001	.025	.036	.011	.031	.061	.006	.023	.021
6 of 6	.123	.172	.241	.141	.205	.293	.084	.192	.175
3 of 6	.337	.347	.418	.313	.351	.386	.236	.337	.383
2 of 6	.263	.255	.205	.251	.221	.163	.361	.349	.233
1 of 6	.164	.123	.073	.157	.110	.066	.204	.134	.124
Sum	.898	.922	.973	.873	.918	.969	.791	.935	.936
0 of 6	.102	.078	.027	.127	.082	.031	.209	.065	.064

Table 13

The Percent of Significant Contrasts/Comparisons per Experiment With Method II for the Strict and Lenient Criteria

	Normal			Exponential			Uniform		
Power	90	93	97	90	93	97	90	93	97
				Orthogonal Contrasts					
α = .05									
Strict Criteria									
3 of 3	.102	.132	.155	.118	.140	.178	.510	.068	.091
Lenient Criteria									
2 of 3	.425	.487	.557	.459	.474	.511	.346	.378	.441
1 of 3	.418	.345	.276	.370	.339	.285	.473	.443	.416
Sum	.945	.964	.988	.947	.953	.974	.870	.889	.948
0 of 3	.055	.036	.012	.053	.047	.026	.130	.111	.052
				Pairwise Comparisons					
α = .05									
Strict Criteria									
6 of 6	.000	.001	.001	.000	.001	.000	.000	.000	.000
Lenient Criteria									
5 of 6	.050	.079	.106	.063	.065	.105	.021	.026	.050
4 of 6	.247	.294	.106	.063	.065	.105	.021	.026	.050
3 of 6	.420	.407	.385	.384	.359	.378	.342	.379	.404
2 of 6	.198	.157	.107	.179	.166	.109	.256	.251	.194
1 of 6	.062	.045	.033	.078	.055	.036	.128	.102	.065
Sum	.977	.983	.998	.977	.982	.989	.923	.939	.974
0 of 6	.023	.017	.002	.028	.018	.011	.077	.061	.026

Table 14

Percent of Significant Experiments for Beta Value of .30

	Method I			Method II		
	Normal	Exp.	Uniform	Normal	Exp.	Uniform
Power	70	70	70	90	90	90
			Orthogonal Contrasts			
α = .05						
T1	.642	.651	.529	.831	.843	.721
T2	.383	.398	.283	.524	.563	.411
T3	.178	.159	.124	.219	.236	.186
$\alpha = \beta/C$						
TD1	.729	.742	.633			
TD2	.514	.520	.389			
TD3	.273	.249	.210			
			Pairwise Comparisons			
α = .05						
C1	.159	.151	.123	.247	.213	.182
C2	.465	.491	.384	.680	.662	.533
C3	.799	.795	.678	.947	.952	.861
C4	.164	.146	.127	.216	.236	.166
C5	.483	.475	.364	.645	.674	.547
C6	.178	.159	.124	.219	.236	.186
			Observed Level of Power			
α = .05						
F Test	.721	.690	.551	.904	.920	.766
			Experimentwise Orthogonal Contrasts			
α = .05						
3 of 3	.050	.064	.020	.102	.118	.510
2 of 3	.281	.274	.203	.425	.459	.346
1 of 3	.491	.468	.470	.418	.370	.473
0 of 3	.178	.194	.307	.055	.053	.130
$\alpha = \beta/C$						
3 of 3	.150	.120	.055			
2 of 3	.397	.375	.299			
1 of 3	.407	.401	.469			
			Pairwise Comparisons			
α = .05						
6 of 6	.000	.000	.000	.000	.000	.000
5 of 6	.011	.011	.006	.050	.063	.176
4 of 6	.123	.141	.084	.247	.268	.176
3 of 6	.337	.313	.236	.420	.384	.342
2 of 6	.263	.251	.261	.198	.179	.246
1 of 6	.164	.157	.204	.062	.078	.128
0 of 6	.102	.127	.209	.023	.280	.077

Table 15

Percent of Significant Experiments for Beta Value of .20

	Method I			Method II		
	Normal	Exp.	Uniform	Normal	Exp.	Uniform
Power	80	80	80	93	93	93
			Orthogonal Contrasts			
α = .05						
T1	.700	.728	.712	.870	.853	.749
T2	.418	.461	.413	.599	.609	.472
T3	.183	.190	.196	.246	.245	.182
$\alpha = \beta/C$						
TD1	.727	.761	.649			
TD2	.471	.515	.369			
TD3	.228	.229	.143			
			Pairwise Comparisons			
α = .05						
C1	.154	.187	.214	.255	.236	.178
C2	.526	.566	.539	.738	.780	.560
C3	.870	.875	.879	.960	.959	.896
C4	.184	.190	.165	.254	.258	.171
C5	.570	.572	.533	.704	.733	.608
C6	.183	.190	.196	.246	.245	.182
			Observed Level of Power α = .05			
F test	.765	.800	.791	.932	.927	.832
			Experimentwise Orthogonal Contrasts			
α = .05						
3 of 3	.059	.075	.060	.132	.140	.068
2 of 3	.336	.366	.339	.487	.474	.378
1 of 3	.452	.422	.463	.345	.339	.443
0 of 3	.153	.137	.138	.036	.047	.111
$\alpha = \beta/C$						
3 of 3	.084	.140	.037			
2 of 3	.374	.410	.277			
1 of 3	.426	.373	.493			
0 of 3	.116	.113	.193			
			Pairwise Comparisons			
α = .05						
6 of 6	.000	.000	.000	.001	.001	.000
5 of 6	.025	.031	.023	.079	.065	.026
4 of 6	.172	.205	.192	.294	.336	.181
3 of 6	.347	.351	.337	.407	.359	.379
2 of 6	.255	.221	.249	.157	.166	.251
1 of 6	.123	.110	.134	.045	.055	.102
0 of 6	.078	.082	.065	.017	.018	.061

Table 16

Percent of Significant Experiments for Beta Value of .10

Power	Method I Normal 90	Method I Exp. 90	Method I Uniform 90	Method II Normal 97	Method II Exp. 97	Method II Uniform 97
			Orthogonal Contrasts			
α = .05						
T1	.821	.838	.722	.932	.901	.835
T2	.518	.576	.445	.652	.656	.531
T3	.213	.244	.166	.271	.284	.205
$\alpha = \beta/C = .03$						
TD1	.768	.794	.660			
TD2	.468	.511	.367			
TD3	.167	.194	.127			
			Pairwise Comparisons			
α = .05						
C1	.213	.234	.185	.311	.271	.232
C2	.449	.475	.352	.557	.511	.441
C3	.945	.944	.880	.990	.979	.943
C4	.216	.238	.192	.263	.277	.224
C5	.646	.688	.546	.770	.778	.662
C6	.213	.244	.166	.271	.284	.205
			Observed Level of Power α = .05			
_F_ Test	.899	.898	.798	.973	.956	.906
			Experimentwise Orthogonal Contrasts			
α = .05						
3 of 3	.080	.124	.054	.155	.178	.091
2 of 3	.449	.475	.352	.557	.511	.441
1 of 3	.414	.336	.467	.276	.285	.416
0 of 3	.057	.065	.127	.120	.026	.052
$\alpha = \beta/C = .03$						
3 of 3	.047	.087	.034			
2 of 3	.405	.420	.267			
1 of 3	.452	.398	.518			
0 of 3	.096	.095	.181			
			Pairwise Comparisons			
α = .05						
	1.000	1.000	1.000	.999	.999	1.000
			Familywise Error Rate			
_F_ Test	.235	.200	.209	.068	.073	.168

Table 19

Error Rates for Beta Value of .30

	Method I			Method II		
	Normal	Exp.	Uniform	Normal	Exp.	Uniform
Power	70	70	70	90	90	90
			Error Rates per Orthogonal Contrasts			
α = .05						
T1	.358	.351	.471	.169	.157	.279
T2	.617	.602	.717	.476	.437	.589
T3	.822	.841	.876	.781	.764	.814
$\alpha = \beta/C$						
TD1	.271	.258	.367			
TD2	.486	.480	.611			
TD3	.727	.751	.790			
			Error Rates per Comparisons			
α = .05						
C1	.841	.849	.877	.753	.787	.818
C2	.535	.509	.616	.320	.338	.467
C3	.201	.205	.322	.051	.048	.139
C4	.836	.854	.873	.784	.762	.834
C5	.517	.525	.636	.355	.326	.453
C6	.822	.841	.876	.781	.764	.814
			Experimentwise Error Rate			
			Orthogonal Contrasts			
α = .05						
	.950	.936	.980	.898	.882	.490
$\alpha = \beta/C$						
	.850	.880	.945			
			Pairwise Comparisons			
α = .05						
	1.000	1.000	1.000	1.000	1.000	1.000
			Familywise Error Rate			
F Test	.279	.310	.449	.096	.080	.234

Table 20

Error Rates for Beta Value of .20

	Method I			Method II		
	Normal	Exp.	Uniform	Normal	Exp.	Uniform
Power	80	80	80	93	93	93
			Error Rates per Orthogonal Contrasts			
α = .05						
T1	.300	.272	.288	.130	.147	.251
T2	.582	.539	.587	.401	.391	.528
T3	.817	.810	.804	.754	.755	.818
$\alpha = \beta / C =$ .07						
TD1	.273	.239	.351			
TD2	.529	.485	.631			
TD3	.772	.771	.857			
			Error rates per Comparisons			
α = .05						
C1	.846	.813	.786	.745	.764	.822
C2	.474	.434	.461	.262	.220	.440
C3	.130	.125	.121	.040	.041	.104
C4	.816	.810	.835	.746	.742	.829
C5	.430	.428	.467	.296	.267	.392
C6	.817	.810	.804	.754	.755	.818
			Experimentwise Error Rate			
			Orthogonal Contrasts			
α = .05	.941	.925	.940	.868	.860	.932
$\alpha = \beta / C$	.916	.860	.963			
			Pairwise Comparisons			
α = .05	1.000	1.000	1.000	.999	.999	1.000
			Familywise Error Rate			
F test	.235	.200	.209	.068	.073	.168

Table 21

Error Rates for Beta Value of .10

	Method I			Method II		
	Normal	Exp.	Uniform	Normal	Exp.	Uniform
Power	80	80	80	97	97	97
	Error Rates per Orthogonal Contrasts					
α = .05						
T1	.179	.162	.278	.068	.099	.165
T2	.482	.424	.555	.348	.344	.469
T3	.787	.756	.834	.729	.716	.795
$\alpha = \beta/C$ = .03						
TD1	.232	.206	.340			
TD2	.532	.489	.633			
TD3	.833	.806	.873			
	Error Rates per Comparisons					
α = .05						
C1	.789	.766	.815	.689	.729	.768
C2	.352	.321	.425	.203	.232	.307
C3	.055	.056	.120	.010	.021	.057
C4	.784	.762	.808	.737	.723	.776
C5	.354	.312	.454	.230	.222	.338
C6	.787	.756	.834	.729	.716	.795
	Experimentwise Error Rate					
	Orthogonal Contrasts					
α = .05	.920	.876	.946	.845	.822	.909
$\alpha = \beta/C$ = .03	.953	.913	.966			
	Pairwise Comparisons					
α = .05	1.000	1.000	1.000	.999	1.000	1.000
	Familywise Error Rate					
F test	.101	.102	.202	.027	.044	.094

Table 22

Kolmogorov-Smirnov Goodness of Fit Test Comparison of Actual and Theoretical Levels of Significance of the Distributions of the F Test Probabilities

	Theoretical Distributions Levels															
	0.005	0.010	0.025	0.050	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	0.950	0.975	0.990
METHOD I NORMAL																
70	0.003	0.009	0.028	0.049	0.104	0.202	0.301	0.410	0.517	0.628	0.729	0.822	0.913	0.951	0.974	0.987
80	0.008	0.011	0.027	0.053	0.108	0.211	0.308	0.397	0.483	0.568	0.680	0.777	0.887	0.943	0.977	0.994
90	0.003	0.007	0.019	0.041	0.093	0.187	0.282	0.396	0.482	0.584	0.687	0.781	0.903	0.952	0.974	0.989
EXPONENTAL																
70	0.004	0.015	0.034	0.064	0.116	0.226	0.326	0.413	0.505	0.600	0.706	0.805	0.898	0.947	0.969	0.993
80	0.006	0.012	0.025	0.052	0.107	0.226	0.326	0.414	0.504	0.619	0.706	0.870	0.898	0.940	0.964	0.991
90	0.008	0.021	0.043	0.068	0.124	0.247	0.351	0.457	0.542	0.644	0.735	0.825	0.901	0.944	0.971	0.992
UNIFORM																
70	0.002	0.006	0.015	0.030	0.053	0.124	0.205	0.285	0.376	0.470	0.564	0.687	0.824	0.907	0.956	0.984
80	0.002	0.004	0.006	0.020	0.041	0.102	0.173	0.254	0.334	0.418	0.535	0.658	0.796	0.871	0.930	0.971
90	0.001	0.001	0.004	0.020	0.047	0.107	0.169	0.247	0.341	0.438	0.544	0.668	0.810	0.890	0.934	0.972
METHOD II NORMAL																
70	0.003	0.009	0.022	0.043	0.101	0.212	0.311	0.421	0.519	0.604	0.707	0.820	0.909	0.956	0.979	0.993
80	0.006	0.008	0.021	0.046	0.099	0.196	0.293	0.430	0.518	0.610	0.706	0.806	0.960	0.953	0.974	0.998
90	0.004	0.009	0.022	0.051	0.084	0.187	0.276	0.367	0.472	0.572	0.677	0.782	0.898	0.948	0.977	0.991
EXPONENTAL																
70	0.009	0.021	0.040	0.075	0.135	0.237	0.334	0.428	0.520	0.628	0.724	0.824	0.903	0.958	0.975	0.990
80	0.007	0.017	0.036	0.060	0.119	0.233	0.328	0.422	0.514	0.670	0.708	0.788	0.885	0.953	0.982	0.992
90	0.006	0.010	0.024	0.041	0.087	0.182	0.284	0.378	0.471	0.563	0.652	0.740	0.862	0.933	0.962	0.986
UNIFORM																
70	0.001	0.003	0.007	0.017	0.044	0.115	0.166	0.240	0.320	0.420	0.515	0.635	0.778	0.878	0.931	0.969
80	0.002	0.005	0.012	0.022	0.049	0.096	0.155	0.216	0.291	0.375	0.493	0.620	0.773	0.872	0.919	0.962
90	0.000	0.001	0.002	0.009	0.022	0.005	0.115	0.188	0.265	0.337	0.465	0.568	0.737	0.851	0.913	0.950

APPENDIX B

LISTING OF THE COMPUTER PROGRAM

USED IN THE SIMULATION STUDY

```
*********************************************;
*                                           *;
*    SET UP AND INITIALIZE VARIABLES        *;
*                                           *;
*********************************************;
*;
*;
*;
PROC MATRIX FUZZ ;
*;
*;
*;
*;
*;
*;
*****************************************************;
*

*                    INITIALIZED              VARIABLES

*

*****************************************************;
*;
*;
*;
*;
N = 36;
CNT=J(18,1,0);
CFKS=J(18,1,0);
CFKS(,1)=.005/.01/.025/.05/.1/.2/.3/.4/.5/.6/.7/.8/
          .9/.95/.990/.995/1/0;
TCNT=J(4,1,0);
TDNT=J(4,1,0);
SUMT=J(1,1,0);
SUMTD=J(1,1,0);
SUMC=J(1,1,0);
SUMTCNT=J(4,1,0);
SUMTDNT=J(4,1,0);
FCNT=J(1,1,0);
CPCNT=J(7,1,0);
SUMCP=J(7,1,0);
******************;
O=J(N,1,1);
S1=J(N,1,9.665);
S2=J(N,1,9.8883);
S3=J(N,1,10.1117);
S4=J(N,1,10.335);
******************;
X2G1=J(N,1,3);
X3G1=J(N,1,0);
X4G1=J(N,1,0);
GR1=O||X2G1||X3G1||X4G1;
X2G2=J(N,1,-1);
```

```
X3G2=J(N,1,2);
X4G2=J(N,1,0);
GR2=O||X2G2||X3G2||X4G2;
X2G3=J(N,1,-1);
X3G3=J(N,1,-1);
X4G3=J(N,1,1);
GR3=O||X2G3||X3G3||X4G3;
X2G4=J(N,1,-1);
X3G4=J(N,1,-1);
X4G4=J(N,1,-1);
GR4=O||X2G4||X3G4||X4G4;
*******************************************************
*

*   RANDOM   NUMBER  GENERATOR   NORMAL   POWER  70

*

*******************************************************
*;
DO I = 1 TO 1000;
*;
ERR1=RANNOR(J(N,1,0));
Y1=S1+ERR1;
MAT1=GR1||Y1;
ERR2=RANNOR(J(N,1,0));
Y2=S2+ERR2;
MAT2=GR2||Y2;
ERR3=RANNOR(J(N,1,0));
Y3=S3+ERR3;
ERR4=RANNOR(J(N,1,0));
MAT3=GR3||Y3;
Y4=S4+ERR4;
MAT4=GR4||Y4;
************************;
MN1=MAT1(,5);
MEAN1=MN1(.,);
MN2=MAT2(,5);
MEAN2=MN2(.,);
MN3=MAT3(,5);
MEAN3=MN3(.,);
MN4=MAT4(,5);
MEAN4=MN4(.,);
MEAN=MEAN1//MEAN2//MEAN3//MEAN4;
*******************************************************;
*

*        USING      SWEEP      OPERATOR      TO

*     COMPUTE     ERROR     SUM     OF     SQUARES

*

*******************************************************;
```

```
M=MAT1//MAT2//MAT3//MAT4;
A=M'*M;
SSE1=SWEEP(A,1);
SSE2=SWEEP(A,1:4);
*;
****************************************************;
*

* COMPUTE F RATIO AND  COUNTING NUMBER OF REJECTIONS

*

****************************************************;
*;
*;
*;
FTEST=((SSE1(5,5)-SSE2(5,5))#/3)#/(SSE2(5,5)#/140);
PROB=1-PROBF(FTEST,3,140,8.79);
FPB=1-PROBF(FTEST,3,140);
IF FPB < .05 THEN  FCNT(,1)=FCNT(,1)+1;
*;
*;
*;
IF PROB < .005 THEN GOTO CNT005;
ELSE IF PROB< .01 THEN GOTO CNT01;
ELSE IF PROB < .025 THEN GOTO CNT025;
ELSE IF PROB < .05 THEN GOTO CNT05;
ELSE IF PROB < .1 THEN GOTO CNT1;
ELSE IF PROB < .2 THEN GOTO CNT2;
ELSE IF PROB < .3 THEN GOTO CNT3;
ELSE IF PROB < .4 THEN GOTO CNT4;
ELSE IF PROB < .5 THEN GOTO CNT5;
ELSE IF PROB < .6 THEN GOTO CNT6;
ELSE IF PROB < .7 THEN GOTO CNT7;
ELSE IF PROB < .8 THEN GOTO CNT8;
ELSE IF PROB < .9 THEN GOTO CNT9;
ELSE IF PROB < .95 THEN GOTO CNT95;
ELSE IF PROB < .975 THEN GOTO CNT975;
ELSE IF PROB < .990 THEN GOTO CNT990;
ELSE IF PROB < .995 THEN GOTO CNT995;
ELSE GOTO CNTONE;
CNT005: CNT(1,1)=CNT(1,1)+1;
CNT01: CNT(2,1)=CNT(2,1)+1;
CNT025:CNT(3,1)=CNT(3,1)+1;
CNT05: CNT(4,1)=CNT(4,1)+1;
CNT1: CNT(5,1)=CNT(5,1)+1;
CNT2: CNT(6,1)=CNT(6,1)+1;
CNT3: CNT(7,1)=CNT(7,1)+1;
CNT4: CNT(8,1)=CNT(8,1)+1;
CNT5: CNT(9,1)=CNT(9,1)+1;
CNT6: CNT(10,1)=CNT(10,1)+1;
CNT7: CNT(11,1)=CNT(11,1)+1;
```

```
CNT8: CNT(12,1)=CNT(12,1)+1;
CNT9: CNT(13,1)=CNT(13,1)+1;
CNT95: CNT(14,1)=CNT(14,1)+1;
CNT975: CNT(15,1)=CNT(15,1)+1;
CNT990: CNT(16,1)=CNT(16,1)+1;
CNT995: CNT(17,1)=CNT(17,1)+1;
CNTONE: CNT(18,1)=CNT(18,1)+1;
*;
*;
*******************************************;
*;
*       ORTHOGONAL CONTRASTS                 *;
*;
*******************************************;
*;
**CODING************;
C1=3 -1 -1 -1;
C2=0 2 -1 -1;
C3=0 0 1 -1;
*******************;
CMAT=C1//C2//C3;
CSQ=CMAT*CMAT';
CMEAN=CMAT*MEAN;
MSE=SSE2(5,5)#/140;
*******************;
T1=CMEAN(1,1)#/(MSE#(CSQ(1,1)#/36))##.5;
T2=CMEAN(2,1)#/(MSE#(CSQ(2,2)#/36))##.5;
T3=CMEAN(3,1)#/(MSE#(CSQ(3,3)#/36))##.5;
********************************************;
T1PROB=1-PROBT(T1,140);
IF T1PROB < .025 OR T1PROB > .975 THEN T1SIG=1;
ELSE IF PROB > .025 OR T1PROB < .975 THEN T1SIG=0;
T2PROB=1-PROBT(T2,140);
IF T2PROB < .025 OR T2PROB > .975 THEN T2SIG=1;
ELSE IF T2PROB > .025 OR T2PROB < .975 THEN T2SIG=0;
T3PROB=1-PROBT(T3,140);
IF T3PROB < .025 OR T3PROB > .975 THEN T3SIG=1;
ELSE IF T3PROB > .025 OR T3PROB < .975 THEN T3SIG=0;
*;
*;
*;
***DUNN-BONFERRONI  PROCEDURE WITH  BETA  ERRORS***;

*;
*;
IF T1PROB < .05 OR T1PROB > .95 THEN TD1SIG=1;
ELSE IF T1PROB > .05 OR T1PROB < .95 THEN TD1SIG=0;
IF T2PROB < .05 OR T2PROB > .95 THEN TD2SIG=1;
ELSE IF T2PROB > .05 OR T2PROB < .95 THEN TD2SIG=0;
IF T3PROB < .05 OR T3PROB > .95 THEN TD3SIG=1;
ELSE IF T3PROB > .05 OR T3PROB < .95 THEN TD3SIG=0;
```

```
IF T1SIG=1 THEN TCNT(1,1)=TCNT(1,1)+1;
IF T2SIG=1 THEN TCNT(2,1)=TCNT(2,1)+1;
IF T3SIG=1 THEN TCNT(3,1)=TCNT(3,1)+1;
IF T1SIG=1 AND T2SIG=1 AND T3SIG=1 THEN TSUM=1;
ELSE TSUM=0;
IF TSUM=1 THEN TCNT(4,1)=TCNT(4,1)+1;
*;
IF T1SIG=1 THEN SUMT(1,1)=SUMT(1,1)+1;
IF T2SIG=1 THEN SUMT(1,1)=SUMT(1,1)+1;
IF T3SIG=1 THEN SUMT(1,1)=SUMT(1,1)+1;
*;
IF SUMT(1,1)=3 THEN SUMTCNT(1,1)=SUMTCNT(1,1)+1;
IF SUMT(1,1)=2 THEN SUMTCNT(2,1)=SUMTCNT(2,1)+1;
IF SUMT(1,1)=1 THEN SUMTCNT(3,1)=SUMTCNT(3,1)+1;
IF SUMT(1,1)=0 THEN SUMTCNT(4,1)=SUMTCNT(4,1)+1;
***************************************************;

IF TD1SIG=1 THEN TDNT(1,1)=TDNT(1,1)+1;
IF TD2SIG=1 THEN TDNT(2,1)=TDNT(2,1)+1;
IF TD3SIG=1 THEN TDNT(3,1)=TDNT(3,1)+1;
*;
IF TD1SIG=1 AND TD2SIG=1 AND TD3SIG=1 THEN TDSUM=1;
ELSE TDSUM=0;
IF TDSUM=1 THEN TDNT(4,1)=TDNT(4,1)+1;
*;
IF TD1SIG=1 THEN SUMTD(1,1)=SUMTD(1,1)+1;
IF TD2SIG=1 THEN SUMTD(1,1)=SUMTD(1,1)+1;
IF TD3SIG=1 THEN SUMTD(1,1)=SUMTD(1,1)+1;
*;
IF SUMTD(1,1)=3 THEN SUMTDNT(1,1)=SUMTDNT(1,1)+1;
IF SUMTD(1,1)=2 THEN SUMTDNT(2,1)=SUMTDNT(2,1)+1;
IF SUMTD(1,1)=1 THEN SUMTDNT(3,1)=SUMTDNT(3,1)+1;
IF SUMTD(1,1)=0 THEN SUMTDNT(4,1)=SUMTDNT(4,1)+1;
*;
*;
*********************************************************
*

*                                          COMPARISIONS

*

*********************************************************
*;
*;
CP1=1 -1 0 0;
CP2=1 0 -1 0;
CP3=1 0 0 -1;
CP4=0 1 -1 0;
CP5=0 1 0 -1;
CP6=0 0 1 -1;
**************************************;
CPMAT=CP1//CP2//CP3// CP4// CP5// CP6;
CPSQ=CPMAT*CPMAT';
```

```
CPMEAN=CPMAT*MEAN;
***************************************;
P1=CPMEAN(1,1)#/(MSE#(CPSQ(1,1)#/36))##.5;
P2=CPMEAN(2,1)#/(MSE#(CPSQ(2,2)#/36))##.5;
P3=CPMEAN(3,1)#/(MSE#(CPSQ(3,3)#/36))##.5;
P4=CPMEAN(4,1)#/(MSE#(CPSQ(4,4)#/36))##.5;
P5=CPMEAN(5,1)#/(MSE#(CPSQ(5,5)#/36))##.5;
P6=CPMEAN(6,1)#/(MSE#(CPSQ(6,6)#/36))##.5;
*;
CP1PROB=1-PROBT(P1,140);
IF CP1PROB <.025 OR CP1PROB >.975 THEN C1SIG=1;
ELSE IF CP1PROB >.025 OR CP1PROB <.975 THEN C1SIG=0;
CP2PROB=1-PROBT(P2,140);
IF CP2PROB <.025 OR CP2PROB >.975 THEN C2SIG=1;
ELSE IF CP2PROB >.025 OR CP2PROB <.975 THEN C2SIG=0;
CP3PROB=1-PROBT(P3,140);
IF CP3PROB <.025 OR CP3PROB >.975 THEN C3SIG=1;
ELSE IF CP3PROB >.025 OR CP3PROB <.975 THEN C3SIG=0;
CP4PROB=1-PROBT(P4,140);
IF CP4PROB < .025 OR CP4PROB > .975 THEN C4SIG=1;
ELSE IF CP4PROB >.025 OR CP4PROB <.975 THEN C4SIG=0;
CP5PROB=1-PROBT(P5,140);
IF CP5PROB <.025 OR CP5PROB >.975 THEN C5SIG=1;
ELSE IF CP5PROB >.025 OR CP5PROB <.975 THEN C5SIG=0;
CP6PROB=1-PROBT(P6,140);
IF CP6PROB <.025 OR CP6PROB >.975 THEN C6SIG=1;
ELSE IF CP6PROB >.025 OR CP6PROB <.975 THEN C6SIG=0;

*******************************************************;

IF C1SIG=1 THEN CPCNT(1,1)=CPCNT(1,1)+1;
IF C2SIG=1 THEN CPCNT(2,1)=CPCNT(2,1)+1;
IF C3SIG=1 THEN CPCNT(3,1)=CPCNT(3,1)+1;
IF C4SIG=1 THEN CPCNT(4,1)=CPCNT(4,1)+1;
IF C5SIG=1 THEN CPCNT(5,1)=CPCNT(5,1)+1;
IF C6SIG=1 THEN CPCNT(6,1)=CPCNT(6,1)+1;
IF C1SIG=1 AND C2SIG=1 AND C3SIG=1 AND C4SIG=1 AND
C5SIG=1 AND C6SIG=1 THEN CSUM=1;
ELSE CSUM=0;
IF CSUM=1 THEN CPCNT(7,1)=CPCNT(7,1)+1;
*;
*;
****************************************************;
*;
IF C1SIG=1 THEN SUMC(1,1)=SUMC(1,1)+1;
IF C2SIG=1 THEN SUMC(1,1)=SUMC(1,1)+1;
IF C3SIG=1 THEN SUMC(1,1)=SUMC(1,1)+1;
IF C4SIG=1 THEN SUMC(1,1)=SUMC(1,1)+1;
IF C5SIG=1 THEN SUMC(1,1)=SUMC(1,1)+1;
IF C6SIG=1 THEN SUMC(1,1)=SUMC(1,1)+1;
*;
IF SUMC(1,1)=6 THEN SUMCP(1,1)=SUMCP(1,1)+1;
IF SUMC(1,1)=5 THEN SUMCP(2,1)=SUMCP(2,1)+1;
IF SUMC(1,1)=4 THEN SUMCP(3,1)=SUMCP(3,1)+1;
```

```
IF SUMC(1,1)=3 THEN SUMCP(4,1)=SUMCP(4,1)+1;
IF SUMC(1,1)=2 THEN SUMCP(5,1)=SUMCP(5,1)+1;
IF SUMC(1,1)=1 THEN SUMCP(6,1)=SUMCP(6,1)+1;
IF SUMC(1,1)=0 THEN SUMCP(7,1)=SUMCP(7,1)+1;
*;
*;
*;
SUMT=J(1,1,0);
SUMTD=J(1,1,0);
SUMC=J(1,1,0);
END;
******* EXPERIMENT RATE COMPUTATIONS***************;
*;
*;
TERPC1=TCNT(1,1)#/1000;
TERPC2=TCNT(2,1)#/1000;
TERPC3=TCNT(3,1)#/1000;
TDERPC4=TDNT(1,1)#/1000;
TDERPC5=TDNT(2,1)#/1000;
TDERPC6=TDNT(3,1)#/1000;
TERPE=TCNT(4,1)#/1000;
TDERPE=TDNT(4,1)#/1000;
ERPF=FCNT#/1000;
*;
CPERPC1=CPCNT(1,1)#/1000;
CPERPC2=CPCNT(2,1)#/1000;
CPERPC3=CPCNT(3,1)#/1000;
CPERPC4=CPCNT(4,1)#/1000;
CPERPC5=CPCNT(5,1)#/1000;
CPERPC6=CPCNT(6,1)#/1000;
CPERPE=CPCNT(7,1)#/1000;
*;
*;
*;
***********KOLMOGOROV-SMIRNOV*******************;
*;
*;
CFCNT=CNT#/1000;
MAXDIFF=CFKS-CFCNT;
*;
*;
************PRINT*******************************;
*;
TITLE ' NORMAL DISTRIBUTION POWER70 ';
*;
PRINT
TERPC1 TERPC2 TERPC3 TDERPC4 TDERPC5 TDERPC6
TERPE TDERPE ERPF CPERPC1 CPERPC2 CPERPC3 CPERPC4
CPERPC5 CPERPC6 CPERPE CFCNT MAXDIFF SUMTCNT SUMTDNT
SUMCP;
RUN;
```

REFERENCES

Brewer, J. K. (1972). On the power of statistical tests in the American Educational Research. American Educational Research, 9, 391-401.

Chase L. J., & Tucker, R. K. (1976). Statistical power: Derivation development, and data-analytic implications. The Psychological Record, 26, 473-486.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. Journal of Abnormal and Social Psychology, 65, 145-153.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev. ed.). New York: Academic Press.

de Cani, J. S. (1984). Balancing Type I risk and loss of power in ordered Bonferroni procedures. Journal of Educational Psychology, 76(6), 1035-1037.

Dunn, O. J. (1961). Multiple comparisons among means. Journal of the American Statistical Association, 56, 52-64.

Ferguson, G. A. (1981). Statistical analysis in psychology and education. New York: McGraw.

Games, P. A. (1971). Multiple comparisons of means. American Educational Research Journal, 8(3), 531-565.

Goodnight, J. H. (1979). The SWEEP operator: Its importance in statistical computing (SAS Technical Report Series, R-106). Cary, NC:SAS Institute Inc.

Hogben, L. (1957). Statistical theory: The relationship of probability, credibility, and error. An examination of the contemporary crisis in statistical theory from a behaviorist viewpoint. New York: W. W. Norton.

Keselman, H. J. (1976). A power investigation of the Tukey multiple comparison statistic. Educational and Psychological Measurement, 36, 97-104.

Kirk, R. E. (1982). Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole.

Klockars, A. J. & Sax, G. (1986). Multiple comparisons. In J. L. Sullivan & R. G. Niemi (Eds.), Series: Quantitative applications in the social sciences (No. 61). Beverly Hills: Sage.

Petrinovich, L. F., & Hardyck, C. D. (1969). Error rates for multiple comparison methods: Some evidence concerning the frequency of erroneous conclusions. Psychological Bulletin, 71, 43-54.

Robinson, E. A. (1985). Probability theory and applications. Boston: International Human Resources Development Corporation.

Rodger, R. S. (1967a). Type I errors and their decision basis. The British Journal of Mathematical and Statistical Psychology, 20(Part 1), 51-62.

Rodger, R. S. (1967b). Type II errors and their decision basis. The British Journal of Mathematical and Statistical Psychology, 20(Part 2), 187-204.

Rodger, R. S. (1974). Multiple contrasts, factors, error rate and power. British Journal of Mathematical and Statistical Psychology, 27, 179-198.

Roscoe, J. T.(1975). Fundamental research statistics for the behavioral sciences. New York: Holt, Rinehart and Winston, Inc..

Rosenthal, R., & Rubin, D. B. (1984). Multiple contrasts and ordered Bonferroni procedures. Journal of Educational Psychology, 76, 1028-1034.

SAS Institute Inc.(1988). SAS/IML user's guide, Release 6.03 Editions. Cary, NC:SAS Institute Inc..

Silverstein, A. B. (1986). Statistical power lost and statistical power regained: The Bonferroni procedure in exploratory research. Educational and Psychological Measurement, 46, 303-307.

Tietjen, G. L. (1986). A Topical dictionary of statistics. New York: Chapman and Hall.

Westermann, R. & Hager, W. (1986). Error probabilities in education and psychological research. Journal of Educational Statistics, 11(2), 117-146.