BOUNDARY CONDITIONS OF SEVERAL VARIABLES RELATIVE TO THE ROBUSTNESS OF ANALYSIS OF VARIANCE UNDER VIOLATION OF THE ASSUMPTION OF HOMOGENEITY OF VARIANCES

DISSERTATION

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

For the Degree of

DOCTOR OF PHILOSOPHY

By

Grady M. Grizzle, B.S., M. Ed.
Denton, Texas
December, 1977

Grizzle, Grady M., Boundary Conditions of Several Variables Relative to the Robustness of Analysis of Variance Under Violation of the Assumption of Homogeneity of Variances. Doctor of Philosophy (Educational Research), December, 1977, 100 pp., 10 tables, 8 illustrations, bibliography, 28 titles.

The problem with which this investigation is concerned is that of determining boundary conditions of several variables relative to the robustness of analysis of variance under violation of the assumption of homogeneity of variances.

The purpose of this study is to determine boundary conditions associated with the number of treatment groups (K), the common treatment group sample size (n), and an index of the extent to which the assumption of equality of treatment population variances is violated (Q) with regard to user confidence in application of the one-way analysis of variance F-test for determining equality of treatment population means.

A Monte Carlo simulation technique was employed in the generation of data. Through this technique, random samples from K normally distributed treatment populations with equal means were generated in such a manner to produce

K samples with n subjects per treatment group. Throughout the study, K was restricted to values of 3, 5, 7, or 9 while the common number of subjects per treatment group sample (n) ranged from 3 to 19. The variances of the treatment populations were specified in a manner so that the variance of the first K-1 treatment populations was unity and the variance of the K^{th} treatment population was Q where Q was assigned odd integral values between 3 and 21 inclusively.

For each parametric combination of K, n, and Q, 2,000 independent samples were generated and the one-way analysis of variance F-test technique for determining equality of means performed. This analysis produced a distribution containing 2,000 F-ratios for each combination of K, n, and Q and was called the actual F-distribution. In order to compare the actual F-distribution to the theoretical nominal F-distribution the proportion of F-ratios in the actual F-distribution exceeding $_{1-\alpha}F_{(K-N)(N-1)}$ for $\alpha\varepsilon\{.010, .050, .100, .200\}$ was determined and compared to α. For the purpose of the study questionable confidence in using the analysis of variance technique for a given combination of K, n, and Q was defined as resulting in an actual F-distribution in which the proportions of F-ratios exceeding $_{1-\alpha}F_{(K-N)(N-1)}$ were significantly different from α for all $\alpha\varepsilon\{.010, .050, .100, .200\}$.

The study concludes that the analysis of variance F-test is robust when the number of treatment groups is less than seven and when the extreme ratio of variances is less than 1:5, but when the violation of the assumption is more severe or the number of treatment groups is seven or more, serious discrepancies between actual and nominal significance levels occur. It was also concluded that for seven treatment groups confidence in the application of the analysis of variance should be limited to the values of Q and n so that $n \geq 10 \ln \frac{1}{2}Q$. For nine treatment groups, it was concluded that confidence be limited to those values of Q and n so that $n \geq -\frac{2}{3} + 12 \ln \frac{1}{2}Q$. No definitive boundary could be developed for analyses with five treatment groups.

In light of these conclusions, it is recommended that in any experimental design which employs more than five treatment groups, the experimenter should attempt to obtain estimates of the variance within each treatment group. If heterogeneity of variances exists to any real degree, the researcher should consider parametric and nonparametric alternatives to the analysis of variance F-test. It is alsrecommended that additional studies be conducted on various aspects of the robustness of the analysis of variance F-test.

TABLE OF CONTENTS

Page

LIST OF TABLES iv

LIST OF ILLUSTRATIONS. vi

Chapter

I. INTRODUCTION. 1

Statement of the Problem
Purpose of the Study
Rationale
The Model
Definition of Terms
Limitations
Hypotheses

II. SURVEY OF RELATED LITERATURE. 13

III. PROCEDURES. 34

IV. ANALYSIS OF DATA AND FINDINGS 44

Hypothesis I
Hypothesis II
Hypothesis III

V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS . . 88

Summary
Conclusions
Recommendations

BIBLIOGRAPHY 98

LIST OF TABLES

Table		Page
I.	Effect of Inequality of Population Variances on Probability of Type I Error with Two-tailed t Test for Equality of Means at Nominal 5 Per Cent Significance Level.	17
II.	The Range of Type I Errors Associated with Various Sample Sizes for the 5 Per Cent Level of Significance	20
III.	Effect of Inequality of Variances on Probability of Type I Error with F Test for Equality of Means in One Way Lay Out at the 5 Per Cent Level of Significance. .	22
IV.	Chi-square Goodness of Fit Test of Observed and Expected Rejected Null Hypotheses for the F(PPM): Model: N(50,10)30; N(50,5)30;N(50,15)30	28
V.	Chi-square Goodness of Fit Test of Observed and Expected Rejected Null Hypotheses for the F(PPM): Model: N(500,100)30; N(500,5)30;N(500,15)30	29
VI.	Summary Data from Simulations Used to Produce Actual F-distributions Containing 500 F-ratios with Selected Number of Groups, Sample Size, and Homogeneity of Variance Violation Parameters	45
VII.	Summary Data from Simulations Used to Produce Actual F-distributions Containing 500 F-ratios with Five Treatment Groups and Selected Sample Size and Homogeneity of Various Violation Parameters.	57
VIII.	Summary Data from Simulations Used to Produce Actual F-distributions Containing 500 F-ratios with Seven Treatment Groups and Selected Sample Size and Homogeneity of Various Violation Parameters.	68

LIST OF TABLES (Continued)

Table Page

IX. Summary Data from Simulations Used to Produce Actual F-distributions Containing 500 F-ratios with Nine Treatment Groups and Selected Sample Size and Homogeneity of Various Violation Parameters 78

X. Minimum and Maximum $\alpha' - \alpha$ Deviations for Specified α-levels 87

LIST OF ILLUSTRATIONS

Graph Page

1. Distribution of Significant and Non-significant Simulation Models with Five Treatment Groups and Hypothesized Boundary Location, K = 5. . . 64

2. Distribution of Boundary Points Determined Through Simulations with Five Treatment Groups and Hypothesized Boundary Location, K = 5. 66

3. Distribution of Significant and Non-significant Simulation Models with Seven Treatment Groups and Hypothesized Boundary Location, K = 7 74

4. Distribution of Boundary Points Determined Through Simulations with Seven Treatment Groups and Hypothesized Boundary Location, K = 7. 76

5. Distribution of Significant and Non-significant Simulation Models with Nine Treatment Groups and Hypothesized Boundary Location, K = 9 84

6. Distribution of Boundary Points Determined Through Simulations with Nine Treatment Groups and Hypothesized Boundary Location, K = 9. 86

7. Distribution of Significant and Non-significant Simulation Models with Seven Treatment Groups and Recommended Boundary Location, K = 7 92

8. Distribution of Significant and Non-significant Simulation Models with Nine Treatment Groups and Recommended Boundary Location, K = 9 94

CHAPTER I

INTRODUCTION

The major purpose of any experimental research is to describe the relationship between variation in the independent variable and variation in the dependent variable. The analysis of variance is a statistical procedure which provides an extremely flexible method for determining the factors that influence the variation of the dependent variable (10). An F-ratio in analysis of variance provides a test of the hypothesis that all treatment population means are equal.

The proper application of the analysis of variance technique is dependent upon the fulfillment of certain assumptions. It is highly probable that in a practical application, the exact fulfillment of these assumptions is never achieved. The question then arises as to the extent to which these assumptions may be violated without seriously affecting the results obtained through application of the statistical technique.

The derivation of the statistical model upon which analysis of variance is based requires the acceptance of four assumptions. Among the assumptions is that of homogeneity of group variance. This assumption stipulates that

each of the treatment populations criterion measures has the same variance.

The effects of heterogeneous group variance on the analysis of variance F statistic have been probed from both the empirical and mathematical points of view. With respect to the mathematical investigation of robustness, Scheffé states ". . . we realize that standards of rigor possible in deducing a mathematical theory from certain assumptions generally cannot be maintained in deriving the consequences of departures from these assumptions" (11, p. 331).

The later generation computer affords a device which can greatly facilitate the empirical investigations of this nature. Until the advent of the high-speed computer, time and energy limitations prevented this type of investigation.

Throughout the literature, there is evidence of considerable work in the area of heterogeneous group variances. Because of the reason cited by Scheffé, most of the work has been empirically based. The current theory which states that if the number of subjects in each group are equal, then the violation of the assumption of homogeneity of variance does not significantly affect the actual probabilities of the occurrence of a type I error has been based on the work of Hsu (6), Box (2; 3), Scheffé (11), Pratt (9), Norton (8), and others. There is some evidence, however, that this may not always be the case. Under certain

conditions such as extreme heterogeneity, a large number of treatments, and a small number of subjects, it appears that the actual probability of a type I error may be significantly different from the nominal or expected level.

Statement of the Problem

The problem of this study was the establishment of boundary conditions of several variables relative to the robustness of analysis of variance under violation of the assumption of homogeneity of variances.

Purpose of the Study

The purpose of this study was to determine boundary conditions of several variables under which confidence in using the analysis of variance technique is justified.

Rationale

In a 1960 publication, Boneau stated:

> As psychologists who perform in a research capacity are well aware, psychological data too frequently have an exasperating tendency to manifest themselves in a form which violates one or more of the assumptions underlying the usual statistical tests of significance. Faced with the problem of analyzing such data, the researcher usually attempts to transform them in such a way that the assumptions are tenable or he may look elsewhere for a statistical test.
>
> Confronted with this discouraging prospect and a perhaps equally discouraging one of laboriously transforming data, performing related tests, and then perhaps having difficulty in interpreting the results, the researcher is often tempted to ignore such considerations and go ahead and run a t test or an analysis of variance. In most cases he is deterred by the feeling that such a procedure will not solve

> the problem. If a significant result is forthcoming, is it due to difference between means or is it due to the violation of assumptions? The latter possibility is usually sufficient to preclude the use of the t or F test (1, p. 49).

Glass, Peckham, and Sanders stated, after reviewing robustness studies:

> . . . we find it significant to note that subsequent investigations have not extended Box's work in the direction of this curious finding. The conventional conclusion that heterogeneous variances are not important when N's are equal seems to have boundary conditions like all other conclusions in this area, and the boundary conditions may not have been sufficiently probed (5, p. 244).

In a final summary, given as Table 16 in their 1972 review, Glass, Peckham, and Sanders conclude:

> Clearly there are boundary conditions on the conclusions on Table 16. There must surely be some breaking points at which a distribution is so pathologically skewed that nominal levels of significance and power are seriously misleading, for example (5, p. 272).

Bradley stated:

> When an assumption is violated, the discrepancy between true and nominal significance levels is not a simple function of the "degree" to which the assumption was violated. Instead, it is influenced by a multitude of additional factors which are not involved in the statement of the assumption (and which do not appear in the quoted statement claiming robustness) but which interact with the violation when it occurs (4, p. 47).

The Model

The one-way fixed effects analysis of variance requires that the following assumptions be made

1. The criterion variable y_{ij} is expressed as a linear combination of independent components: μ, the common location parameter; α_j, the incremental or decremental effect of treatment j on the dependent variable for all observations in group j; and e_{ij}, the error of the (i,j) observation.
2. The e_{ij}'s are normally distributed with a mean of zero and a variance which is constant across treatment groups.
3. The treatment effects α_j are constants (11, p. 55).

Assumption one leads to the expression of the criterion variable y as $y_{ij} = \mu + \alpha_j + e_{ij}$.

Assumption two imposes rather severe limitations on the ANOVA model. These include normality, equality of group variances and independence. It is a consequence of these that a large number of observations taken under the j treatments should have nearly a normal distribution within each group and the variance of this distribution from one group to the next should be the same.

The validity of the model under violations of these assumptions has been investigated under selected conditions using both mathematical and empirical techniques. A study of such investigation leads to the question under investigation in this study:

Do boundary conditions of several variables exist relative to the robustness of the model under the violation of the assumption of homogeneity of variance and, if so, are they quantifiable and identifiable?

Definition of Terms

For the purposes of this study, the following terms are defined.

Nominal Significance Level--The nominal significance level, α, is the percentage of F ratios which exceed the tabled value of F associated with k-1 and N-k degrees of freedom. This tabled F associated with α is the point on the central F distribution above which 100α per cent of F ratios will occur when all assumptions of analyais of variance are met.

Actual Significance Level--The actual significance level, α', is the percentage of F ratios which exceed the tabled value of F in an empirical distribution.

Robust--A statistical model is said to be robust with respect to the violation of a particular assumption to the degree that the model can tolerate the violation without seriously affecting the results (2).

Significant Difference Between Nominal and Actual Significance Levels--An actual significance level is said to significantly differ from the nominal significance level when it fails to fall within a 95 per cent confidence interval about the nominal level.

Monte Carlo Simulation--A Monte Carlo Simulation is a procedure in which random samples are drawn from populations having specified parameters and then a particular statistic

is computed. After repetition of the process has yielded the empirical sampling distribution, the distribution is compared to the theoretical distribution to determine possible deviations from the theoretical distribution.

Pseudorandom Numbers--Pseudorandom numbers are "pseudo" because once the generating sequence is started, each number is precisely determined by the preceding number. Even though pseudorandom numbers are determined by a recursive formula, they have the basic properties of randomness which make them quite usable in simulation studies (7). Throughout this study, pseudorandom numbers are referred to as random numbers.

Boundary Point--For a specified value of k, (Q,n) is a boundary point if and only if, for a specified value of n, Q is the minimal integral value which produces an actual F distribution which is significantly different from the nominal F distribution at the .05 level for all specified nominal levels of .010, .050, .100, .200.

L-Distance--If (x,y) is a point in the Qn-plane and $\{(Q,n):n=f(Q)\}$ is a curve in the Qn-plane and $(Q_1,f(Q_1))$ is the intersection of the line $\{(Q,n):Q=x\}$ with the curve $\{(Q,n):n=f(Q)\}$ and $(Q_2,f(Q_2))$ is the intersection of the line $\{(Q,n):n=y\}$ and the curve $\{(Q,n):n=f(Q)\}$, then the L-Distance from (x,y) to $\{(Q,n):n=f(Q)\}$ is the minimum of $|f(Q_2) - f(Q_1)|$ and $|Q_2 - Q_1|$.

Limitations

This study was limited to experimental conditions simulated with the following conditions

1. In all simulations, the assumptions of random selection, normality of the distribution of criterion measures for each treatment population, and equality of means of criterion measures for each treatment group were met;

2. The number of treatment groups, sizes of samples of each treatment group, and ratios of variances considered were then deemed necessary to identify boundary conditions.

Hypotheses

To carry out the purpose of this study, the following hypotheses were formulated. (K = Number of treatment groups) (n = Number of subjects per treatment group) (Ratio of variances = 1:1:1:...:Q).

I. The proportion of F ratios in the actual F distribution exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ will not differ significantly from the proportion of F ratios in the nominal F distribution exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ at the .10 level of significance for:

A. K = 3, n = 15, 7, 5, 3, with Q = 3, 7, 9, 17 for all specified nominal levels of .010, .050, .100, .200.

B. $K = 5$, $n = 15, 7, 5$ with $Q = 3, 7, 9, 17$ and $n = 3$ with $Q = 3, 7, 9$ for all specified nominal levels.

C. $K = 7$, $n = 15, 7, 5, 3$ with $Q = 3$, $n = 15, 7$ with $Q = 7, 9, 17$ and $n = 5$ with $Q = 7$ for all specified nominal levels.

D. $K = 9$, $n = 15, 7, 5, 3$ with $Q = 3$ and for $n = 15$ with $Q = 7, 9, 17$ for all specified nominal levels.

II. The proportion of F ratios in the actual F distribution exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ will be significantly greater than the proportion of F ratios in the nominal F distribution exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ at the .10 level of signifiance for:

A. $K = 5$, $n = 3$, with $Q = 17$ for all specified nominal levels.

B. $K = 7$, $n = 3$ with $Q = 7, 9, 17$ and $n = 5$ with $Q = 19, 17$ for all specified nominal levels.

C. $K = 9$, $n = 7, 5, 3$, with $Q = 7, 9, 17$ for all specified nominal levels.

III. A. At least 70 per cent of the boundary points for $K = 5$ will lie within an L-Distance of one unit of the line defined by $\{(Q,n): n=2 \ln(Q-13)\}$.

B. At least 70 per cent of the boundary points for $K = 7$ will lie within an L-Distance of one unit of the line defined by $\{(Q,n): n=15/8 \ln 3(Q-4)\}$.

C. At least 70 per cent of the boundary points for $K = 9$ will lie within an L-Distance of one unit of the line defined by $\{(Q,n):n=5\ \ln(Q-2)\}$.

CHAPTER BIBLIOGRAPHY

1. Boneau, C. A., "The Effects of Violations of Assumptions Underlying the t-test," Psychological Bulletin, LVII (1960), 49-64.

2. Box, G. E. P., "Non-Normality and Tests of Variance," Biometrica, XL (1953), 318-335.

3. _____, "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-way Classification," Annals of Mathematical Statistics, XXV (1954), 290-302.

4. Bradley, J. V., "Studies in Research Methodology VI. The Central Limit Effect for a Variety of Populations and the Robustness of the z, t, and F." USAF Aerospace Medical Division Report AMRL-TR-64-123, 1964.

5. Glass, G. V., Peckham, P. D., and Sanders, J. R., "Consequences of Failure to Meet Assumptions Underlying the Analysis of Variance and Covariance," Review of Educational Research, XLII (Summer, 1972), 237-288.

6. Hsu, P. L., "Contributions to the Theory of Students T Test as Applied to the Problem of Two Samples," Statistical Research Memoirs, II (1938), 1-24.

7. Lehman, R., and Bailey, D. E., Digital Computing, Fortran N and Its Application to the Behavioral Sciences, New York, John Wiley and Sons, 1968.

8. Norton, D. W., "An Empirical Investigation of the Effects of Non-Normality and Heterogeneity Upon the F Test of Analysis of Variance," unpublished doctoral dissertation, State University of Iowa, Iowa City, Iowa, 1952.

9. Pratt, J. W., "Robustness of Some Procedures for the Two-Sample Location Problem," Journal of the American Statistical Association, LIX (1964), 665-680.

10. Roscoe, J. T., Fundamental Research Statistics for the Behavioral Sciences, New York, Holt, Rinehart and Winston, Inc., 1969.

11. Scheffé, H., The Analysis of Variance, New York, John Wiley & Sons, 1959.

CHAPTER II

SURVEY OF RELATED LITERATURE

Analysis of variance is a statistical procedure to simultaneously investigate the differences among means of several populations. This method involves estimating how much of the total variation in a set of data can be attributed to certain identifiable causes of variation and how much can be attributed to chance.

The simplest design in which analysis of variance is employed is the simple randomized design. This design partitions the total variance in a set of samples into the between variance and the within variance. The interest in such a design is the between variation, commonly called the treatment effect (12, p. 1).

The mathematical derivation of the model used for the one-way fixed effects analysis of variance is based upon the fulfillment of several assumptions. These include:

1. The criterion variable y_{ij} is expressed as a linear combination of independent components: μ, the common location parameter; α_j, the incremental or decremental effect of treatment j on the dependent variable for all observations in group j; and e_{ij}, the error of the (i,j) observation.
2. The e_{ij}'s are normally distributed with a mean of zero and a variance which is constant across treatment groups.
3. The treatment effects α_j are constants (20, p. 55).

Assumption one leads to the expression of the criterion variable y as $y_{ij} = \mu + \alpha_j + e_{ij}$.

In application as an inferential technique, the assumptions of the mathematical model impose the following restrictions on the application of the analysis of variance techniques:

1. All treatment groups were originally drawn at random from the same parent population.
2. The variance of the criterion measures is the same for each of the treatment populations.
3. The distribution of the population criterion measures is normal.
4. The means of the criterion measures is the same for each treatment population (13, p. 73).

If any one of the assumptions is not satisfied, the sampling distribution of the mean squares ratio will differ from the theoretical F-distribution. For example, if a significant mean square ratio is found, it could have resulted from a failure to satisfy any one of these assumptions. Since the most frequent usage of analysis of variance is in testing the equality of treatment population means, the interpretation of the F ratio is dependent upon the successful fulfillment of the assumptions of randomness, normality of distribution, and homogeneity of treatment population variances.

The assumptions upon which the mathematical development of the F test for analysis of variances is based have been under scrutiny since Pearson's study in 1929. Although Pearson's study dealt with only the violation of the

assumption of normality, it is noteworthy in that it first suggested the robustness of the F-ratio for analysis of variance.

Pearson used samples of five or ten for the case of two groups having equal variances and equal means. As a result of the study, Pearson found that if the distributions from which the samples are taken are symmetrical, the F-ratio is affected very little by departure from normality. Pearson concluded:

> There are so many ways in which the population form may be modified and so many changes to be rung in the values of N, variance, and mean, that it would be dangerous to draw too sweeping conclusions from a single experiment. Yet, as far as it goes, this study concludes that, where slight variations from normality exist, Fisher's test may be used with confidence (17, p. 360).

Although Pearson's 1929 study did not explicitly deal with the problem of heterogeneity of variances, it is significant to note that the conclusion stated above forms a thread of uncertainty which is present in nearly all such "robustness" studies from 1929 to the present. This uncertainty is based on the very complex inter-relationships between variates which could conceivably affect the outcome of any single experiment. The mathematical permutations of the possible numerical values of only a few such variates dictate that the entire answer of the "robustness" question of analysis of variance techniques may only be answered

through the use of many differing studies performed over a great number of years.

In 1931, Pearson turned his attention toward the robustness of the procedure with reference to the assumption of homogeneity of treatment population variances. He concluded that the F test is basically unaffected by heterogeneity of variance. He emphasized, however, that extreme violation caused the F test to be conservative so that a significant difference of means might be overlooked (16).

Hsu (1938) was one of the first statisticians to obtain precise mathematical results in this area. In his study, Hsu determined the actual probability of a significant result at the .05 level for various values of the ratio of σ_1^2 to σ_2^2 in a two-tailed t-test. Hsu found that as long as the n's were equal, heterogeneity of variance had little effect. As the ratio N_1/N_2 increased, however, serious discrepancies appeared between the nominal significance level of .05 and the actual probability of a type I error (11). Table I summarizes the findings of this study.

David and Johnson in 1951 were the first to use the calculator and electronic devices for selecting samples. Their empirical approach involved selecting 10,000 sets of three groups each having n = 24. In their study they dealt with only the violation of the assumption of normality

(6, pp. 43-57). The empirical procedure they used, however, was employed the very next year by Norton.

TABLE I

EFFECT OF INEQUALITY OF POPULATION VARIANCES ON PROBABILITY OF TYPE I ERROR WITH TWO-TAILED t TEST FOR EQUALITY OF MEANS AT NOMINAL 5 PER CENT SIGNIFICANCE LEVEL

σ_1^2/σ_2^2	(N_1N_2)		
	(15,5)	(5,3)	(7,7)
.0	32	.22	.072
.1	.23	.14	.070
.2	.18	.10	.063
.5	.098	.072	.058
1	.050	.050	.050
2	.025	.38	.051
5	.008	.031	.058
10	.005	.030	.063
	.002	.031	.072

Norton's study involved not only the question of normality and the question of homogeneity of variance but a combination of types of violations of these assumptions. Having noticed that the research on the assumptions underlying the F test seemed to indicate that the sampling distribution of the F statistic is not very sensitive to violations of the theoretical requirements, Norton studied

various combinations of violations (15). The results of the Norton study can be summarized as follows

1. When samples are taken from populations having the same shape, the shape of distribution has little effect on the distribution of the F-ratio;

2. When samples are taken from populations having the same shape but heterogeneous variance or from populations having heterogeneous shapes and homogeneous variances, the discrepancies between observed and expected type I errors are quite small;

3. Serious discrepancies between observed and expected type I errors are likely when the populations from which the samples are taken are heterogeneous with respect to both shape and variance.

Since Norton was particularly concerned with the interaction between non-normal distributions and heterogeneous variances, the number of parametric values employed for normal distributions with unequal variances is limited. However, he did find that when the number of groups was three, the number of subjects per group was also three, and the ratio of variances was 1:4:9, 7.46 per cent of the F-ratios obtained exceeding $_2F_6$ instead of the anticipated 5 per cent. When the number of subjects per group was increased to 10, he determined an actual F distribution where 6.56 per cent of the F ratios exceed $_2F_{27}$. In reporting on this study, Lindquist states:

> It is apparent from these results that marked heterogeneity of variance has a small but real effect on the form of the F-distribution. If one used the probabilities read from the normal theory F-table in interpreting the results of an experiment with this degree of heterogeneity, he might think he was making a test at the 5% level when actually he was making it at the 7% level, or might think he was testing at the 1% level, when he actually was doing so at the 2+% level of significance, etc. Accordingly, where marked (but not extreme) heterogeneity is expected, it is desirable to allow for the discrepancy by setting a slightly higher "apparent" level of significance for this test than one would otherwise employ (13, p. 83).

Although Lindquist made the above statement with reference to Norton's study, he had been concerned with the problem of unequal variances for many years. From as early as 1940, Lindquist had been studying the problem. In a 1940 publication with R. H. Godard, Lindquist investigated the problem using actual data. In this study, he found evidence of the same trend that Norton was to clarify fifteen years later. Lindquist's conclusions in this study, however, must be discounted to some degree because of the questionable application of the Chi-square Goodness-of-Fit test used to investigate similarities between the actual and nominal F distributions (8).

In 1953 Horsnell conducted a study in which only homogeneity of variances was violated. The sample sizes which were investigated were five, ten, fifteen, and twenty. The study included four equal-sized groups in each model. The ratios of the four standard deviations ranged

from 1:1:1:1 to 1:1:2:3. Table II summarizes the results of the Horsnell study.

TABLE II

THE RANGE OF TYPE I ERRORS ASSOCIATED WITH VARIOUS SAMPLE SIZES FOR THE 5 PER CENT LEVEL OF SIGNIFICANCE

Sample Size	Range of
5	.055 - .069
10	.052 - .063
15	.049 - .062
20	.049 - .061

Horsnell's conclusions lend evidence to the conjecture that the F test is affected only slightly by small group differences in standard deviations as long as the assumption of normality is not violated (10).

In light of this conclusion, Box (1953) studied the error involved when tests of equality of variance, such as the Bartlett test, are used to determine whether the F test should be used to test equality of means. It was found that tests of equality of variance are so sensitive that to ignore tests of variance leads to fewer incorrect decisions than does using them. Box states, "To use a test of equality of variance to approve the use of the F test for equality of means is like sending a row boat to find out whether conditions are sufficiently calm for an ocean liner to leave

port" (3, p. 331). Box again reiterates the same concept that Pearson stated in 1929. That is, the entire question of robustness of the F test under violation of the assumption of homogeneous variances is too complex to be answered by only noting the extent of the violation. Implied is the concept that only can one determine the effects when taken in conjunction with several other variables.

Box, in 1954, published some of his early results in a mathematical approach to the problem of the effect of heterogeneous variances on alpha. Box's results evidence the same trend earlier discovered by Hsu. That is, when the n's are equal, the actual and nominal levels of significance agree quite closely. There is, however, one notable exception to this conclusion. When Box used seven groups with a ratio of variances equal to 1:1:1:1:1:1:7 and n = 3 in each group, he found the actual level of significance to be .12 as compared to the nominal level of .05 (4). As only equal sample sizes are to be used in this study, only those results of Box's study using equal samples are reported in Table III. Although these results were not confirmed by Scheffé in 1959 (20) nor by Pratt in 1968 (18, pp. 678-680), Hsu's original work in the area seems to indicate the same tendency. It is significant to note that Hsu's work involved only two groups where Box's study involved more.

TABLE III

EFFECT OF INEQUALITY OF VARIANCES ON PROBABILITY OF TYPE I ERROR WITH F TEST FOR EQUALITY OF MEANS IN ONE WAY LAY OUT AT THE 5 PER CENT LEVEL OF SIGNIFICANCE

Number of Groups	Ratio of Group Variances	Group Sizes	n	Probability of Type I Errors
3	1:2:3	5	15	.056
3	1:1:3	5	15	.059
5	1:1:1:1:3	5	25	.074
7	1:1:1:1:1:1:7	3	21	.120

In looking at a slightly different problem in 1954, Behrens extensively studied the effects of heterogeneous population variances upon the simple analysis of variance F test. He found that the calculated F values on samples taken from populations having unequal variances would themselves have a larger variance than expected. This increase in the variance of the F ratios would cause adherence to the F test assumption. Behrens recommended setting the level of significance more stringently if it is known that the population variances are unequal (1, p. 27).

Gronow (1951) developed theoretical significance levels and some power values for the two-sample t by means of a series expansion. He reported that in the equal sample case m=n=10 and for ratio of variances = 1/3, 1/2, 1, 2, and

3, variance heterogeneity had very little effect on the significance level of t. However, when samples of differing sizes were studied, the effect was more pronounced (9, p. 255).

An extensive study by Boneau (1960) used a Monte Carlo procedure to investigate the robustness of the F test. Due to the utilization of electronic computing equipment, 1,000 cases of each model were used. A model consisted of two groups of $n = 5$ or $n = 15$ taken from populations having variances $\sigma^2 = 1$ or $\sigma^2 = 4$. Three distributions were generated and the samples were taken from various combinations of these distributions. Although Boneau worked exclusively with the t test at the 5 per cent and 1 per cent levels of significance, the fact that the two group analysis of variance F with one and n degrees of freedom is equal to t^2 for n degrees of freedom, many of the results of the Boneau study can be extended to the simple analysis of variance F test. Boneau concluded that the t test and consequently the F test are extremely robust tests. It should be noted, however, that for heterogeneous variance it is very important to have equal sized groups. It also would appear that, for extreme violations of the assumptions for the t test, a sufficiently large sample size is necessary to allow the statistical effects of averaging to be eminent. The results of the Boneau study tend to agree with most

results of the Norton study. Boneau summarized the results of his study with the following statement:

> We may conclude that for a large number of different situations confronting the researcher, the use of the ordinary t test and its associated table will result in probability statements which are accurate to a high degree, even though the assumptions of homogeneity of variance and normality of the underlying distributions are untenable. This large number of situations has the following general characteristics: (a) the two sample sizes are equal or nearly so, (b) the assumed underlying population distributions are of the same shape or nearly so. (If the distributions are skewed they should have nearly the same variance.) . . . If the sample sizes are unequal, one is in no difficulty provided the variances are compensatingly equal. . . . If the two underlying populations are not of the same shape, there seems to be little difficulty if the distributions are both symmetrical (2, p. 64).

These conclusions seem rather strong on the basis of the restricted range of Boneau's investigation and in view of the instances of rather large relative discrepancies between the nominal and the observed α-values.

Bradley (1964), in one of a series of studies on the central limit effect of normal distribution theory statistics, investigated empirically the true significance level of t under conditions of both non-normality and heterogeneity of variance. He sampled from both normal and a compound of normal distributions with equal variances and with unequal variances in a ratio of 4 to 1. Various combinations of sample sizes, ranging from 2 to 204 were also used. For each pair of samples of a given size

selected from a given pair of populations, the value of t was obtained. Thus, empirically determined sampling distributions of 10,000 t-values were developed. After extensively analyzing the data, Bradley made several conclusions.

Bradley began by pointing out that statements such as "t is robust to non-normality and/or variance heterogeneity" are rampant in statistical literature. He continues as follows:

> When an assumption is violated, the discrepancy between true and nominal significance levels is not a simple function of the "degree" to which the assumption was violated. Instead, it is influenced by a multitude of additional factors which are not involved in the statement of the assumption (and which do not appear in the quoted statement claiming robustness) but which interact with the violation when it occurs. Such factors are:
>
> (1) size of nominal significance level
> (2) location of rejection region
> (3) absolute sample sizes
> (4) relative sample sizes
> (5) relative sizes of population variances (when homogeneity of variance is not an assumption)
> (6) relative shapes of the sampled populations
> (7) absolute correlation between sample means and variances for each population
> (8) relative correlations between sample means and variances among the various populations sampled.
>
> The interactions are likely to be exceedingly complex and of high order. That is, the extent to which a given factor influences robustness generally depends not only upon its own value, but also upon the particular combination of factors involved and the value of each, and the interdependency is often quite strong (5, p. 73).

The Bradley quote indicates that the concerns voiced by Pearson (1929) and Box (1954) are still present. The "robustness" of a statistical measure cannot be adequately investigated without considering a number of variables which may interact with the violation of an assumption.

Murphy (1967) did research quite similar to the 1954 Behrens study and found compatible results. Murphy used a number of variance ratios (θ^2) with a variety of sample sizes and compared calculated t values to the critical t value at the 5 per cent level of significance. Only one instance ($\theta^2 = 2$, $n_1 = n_2 = 8$) yielded a smaller empirical alpha than the expected .05 level (14).

The importance of having equal sample sizes when the variances are unequal was very much evidenced from Murphey's results. This tendency has been found in many previous studies in this area. These include the work of Gronow (9), Boneau (2), and Box (3). In a 1972 publication, Glass, Peckham, and Sanders (7) described many of these studies in detail. In reviewing these studies, Glass, et al. (1972) conclude that the following seem justified

1. When n's are unequal and variances are heterogeneous, actual significance level may greatly exceed the nominal significance level when samples with smaller n's come from populations with larger variances.
2. When n's are unequal and variances are heterogeneous, the actual significance level may be greatly exceeded by the nominal significance level when samples with smaller n's come from populations with smaller variances (7, p. 245).

Of studies conducted since 1970, the dissertation by Melvin Roy (1971) appears to have major implication with regard to the subject of this study. Roy used a Monte Carlo simulation technique in order to investigate the effect of heterogeneous variances on the analysis of variance test. In one model where the number of treatment groups equals three, $n_1 = n_2 = n_3 = 30$, and a ratio of variances = 1:2:3, he found a significant chi-square when comparing the actual and nominal F distributions. In a second model with the same parametric values with the exception of a ratio of variances = 1:3:20, he likewise found a significant chi-square value. Tables IV and V contain the results of the Roy study on these two models (19).

The use of the Chi-square Goodness of Fit statistic in this study is questionable. This is due to the basic shape of the F distribution and the fact that most of the differences between the two distributions will occur in a region of the distributions which is usually of little concern in hypothesis testing (7). It can still be seen from this data, however, that with the violation of the assumption, the probability of making a Type I error is elevated at the .01, .02, .05, .10, and .20 levels.

Much research has been performed on the question of the robustness of the F test in the analysis of variance under violation of the assumption of homogeneity of variance.

TABLE IV

CHI-SQUARE GOODNESS OF FIT TEST OF OBSERVED AND EXPECTED REJECTED NULL HYPOTHESES FOR THE F(PPM): MODEL: N(50,10)30;N(50,5)30;N(50,15)30

Alpha Level	Observed	Expected	(O-E)
1.00	2,000	2,000	0
.99	1,975	1,980	5
.98	1,949	1,960	11
.95	1,879	1,900	21
.90	1,764	1,800	36
.80	1,541	1,600	59
.70	1,327	1,400	73
.30	583	600	17
.20	431	400	31
.10	237	200	37
.05	128	100	28
.02	65	40	25
.01	30	20	10

Computed chi-square = 45.202.

TABLE V

CHI-SQUARE GOODNESS OF FIT TEST OF OBSERVED AND EXPECTED REJECTED NULL HYPOTHESES FOR THE F(PPM): MODEL: N(500,100)30;N(500,5)30,N(500,15)30

Alpha Level	Observed	Expected	(O-E)
1.00	2,000	2,000	0
.99	1,924	1,980	56
.98	1,872	1,960	88
.95	1,719	1,900	181
.90	1,547	1,800	253
.80	1,301	1,600	289
.70	1,130	1,400	270
.30	552	600	48
.20	412	400	12
.10	270	200	70
.05	178	100	78
.02	100	40	60
.01	60	20	40

Computed Chi-square = 412.149.

Generally, the research supports the theory that when n's are equal, the F test of analysis of variance is robust, Box's curious results under extreme conditions and trends occurring in Hsu's and Behren's data indicate that the

question has not been fully answered. Glass, Pecham, and Sanders in 1972 stated:

> Whatever the cause, we find it significant to note that subsequent investigators have not extended Box's work in the direction of this curious finding. The conventional conclusion that heterogeneous variances are not important when n's are equal seems to have boundary conditions like all other conclusions in this area, and the boundary conditions may not have been sufficiently probed (7, pp. 244-245).

As reflected in statements by Pearson, Box, and Bradley, one overriding concern in the study of the robustness of the F test of analysis of variance under the violation of the assumption of homogeneity of variances appears to be a continual effort to over-simplify the problem. Bradley indicated that several variables should possibly be included in such a study in order to gain insight into the interaction of the variables with the extent of the violation of the assumption (5).

In explaining their Table 16 in a 1972 article which displayed a summary of consequences of violations of assumptions of the fixed effects ANOVA, Glass, Pecham, and Sanders concluded the following:

> Clearly there are boundary conditions on the conclusions in Table 16. There must surely be some breaking point at which a distribution is so pathologically skewed that nominal levels of significance and power are seriously misleading (7, p. 272).

The purpose of this research project is to further investigate this question and to empirically extend Box's study in order to see if boundary conditions exist.

CHAPTER BIBLIOGRAPHY

1. Behrens, W. U., "The Comparison of Means of Independent Normal Distributions with Different Variances," Biometrica, XX (1964), 16-27.

2. Boneau, C. A., "The Effects of Violation of Assumptions Underlying the t-test," Psychological Bulletin, LVII (1960), 49-64.

3. Box, G. E. P., "Non-Normality and Tests of Variance," Biometrica, XL (1953), 318-335.

4. _____, "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-way Classification," Annals of Mathematical Statistics, XXV (1954), 290-302.

5. Bradley, J. V., "Studies in Research Methodology VI. The Central Limit Effect for a Variety of Populations and the Robustness of the z, t, or F," USAF Aerospace Medical Division Report AMAL-TR-64-123, 1964.

6. David, F. N., and Johnson, N. L., "The Effect of Non-Normality on the Power Function of the F Test in the Analysis of Variance," Biometrica, XXXVIII (1951), 43-57.

7. Glass, G. V., Peckham, P. D., and Sanders, J. R., "Consequences of Failure to Meet Assumptions Underlying the Analysis of Variance and Covariance," Review of Educational Research, XLII (Summer, 1972), 37-288.

8. Godard, R. H., and Lindquist, E. F., "An Empirical Study of the Effect of Heterogeneous Within-groups Variance upon Certain F-tests of Significance in Analysis of Variance," Psychometrika, V (December, 1940), 263-273.

9. Gronow, D. G. C., "Test for the Significance of the Difference Between Means in Two Normal Populations Having Unequal Variance," Biometrica, XXXVIII (1951), 252-256.

10. Horsnell, G., "The Effects of Unequal Group Variances on the F Test for the Homogeneity of Group Means," Biometrica, XL (1953), 128-136.

11. Hsu, P. L., "Contributions to the Theory of Student's T Test as Applied to the Problem of Two Samples," Statistical Research Memoirs, II (1938), 1-24.

12. Lang, M. S., "Empirical Comparisons of the Treatments by Levels in Analysis of Variance and Analysis of Covariances," unpublished doctoral dissertation, College of Education, Department of Research and Statistical Methodology, University of Northern Colorado, Greeley, Colorado, 1972.

13. Lindquist, E. F., Design and Analysis of Experiments in Psychology and Education, Boston, Houghton-Mifflin Company, 1953.

14. Murphey, B. P., "Some Two-Sample Tests When the Variances are Unequal: A Simulation Study," Biometrica, LIV (1967), 676-683.

15. Norton, D. W., "An Empirical Investigation of the Effects of Non-Normality and Heterogeneity Upon the F Test of Analysis of Variance," unpublished doctoral dissertation, State University of Iowa, Iowa City, Iowa, 1952.

16. Pearson, E. S., "Analysis of Variance in Cases of Non-Normal Variation," Biometrica, XXIII (1931), 114-133.

17. _____, "Some Notes on Sampling Tests with Two Variables," Biometrica, XXI (1929), 337-360.

18. Pratt, J. W., "Robustness of Some Procedures for the Two-Sample Location Problem," Journal of the American Statistical Association, LIX (1964), 665-680.

19. Roy, M. R. "An Investigation of the Robustness of the Simple Analysis of Variance F Test," unpublished doctoral dissertation, College of Education, University of Northern Colorado, Greeley, Colorado, 1971.

20. Scheffé, H., The Analysis of Variance, New York, John Wiley & Sons, 1959.

CHAPTER III

PROCEDURES

The one-way fixed effects analysis of variance requires that the following assumptions be made:

1. The criterion variable y_{ij} is expressed as a linear combination of independent components: μ, the common location parameter, α_j, the incremental or decremental effect of treatment j on the dependent variable for all observations in group j; and e_{ij}, the error of the (i,j) the observation.
2. The e_{ij}'s are normally distributed with a mean of zero and a variance which is constant across treatment groups.
3. The treatment effects, α_j, are constants (10, p. 55).

Assumption one leads to the expression of the criterion variable, y, as $y_{ij} = \mu + \alpha_j + e_{ij}$.

In application, the assumption of the mathematical model imposes the following restrictions on the application of the analysis of variance technique.

1. All treatment groups were originally drawn at random from the same parent population.
2. The variance of the criterion measures is the same for each of the treatment population.
3. Distribution of the population criterion measures is normal.
4. The means of the criterion measures are the same for each treatment population (6, p. 73).

For the purpose of this study, the assumption requiring homogeneity of treatment population variances was studied in

order to determine relationships between this assumption and other selected variants. In developing procedures which would allow definitive conclusions, the question became one of how to empirically derive distributions of F-ratios obtained under specific conditions and how to compare the empirically generated distributions with the central F-distribution.

The procedures used in Monte Carlo methods are of utmost importance to this study. The usual procedures of sample selection, data collection, and data analysis must be considered from a slightly different point of view. The data used in the study were generated by a computer in a manner to conform to pre-determined conditions with minimal variation. The variations existed to a limited degree and were resultant from sampling error. The term "random number" used in the context of this study is not truly random and should be understood to be a "pseudorandom number." Pseudorandom number sequences generated internally by a computer are not random in a true sense because they are determined by a finite "starting point" and have limited precision (8).

The random number generators used in this study were IBM subroutines GAUSS and RANDU. The purpose of GAUSS is to compute a normally distributed random number sequence with a prespecified mean and standard deviation (7). The

IBM subroutine RANDU generates a sequence, x_i, where for each i, x_i is a uniformly distributed random number such that $0 \leq x_i \leq 1$. The GAUSS subroutine uses twelve uniform random numbers generated by RANDU to compute a normal random number of the Central Limit Theorem. The conversion of the twelve random numbers was accomplished through the use of the formula (5):

$$y = \frac{\sum_{i=1}^{k} (x_i - \frac{k}{2})}{\sqrt{\frac{k}{12}}}$$

As k increases without bound, y asymptotically approaches a true normal distribution. For the application of these procedures in this study, the value of k chosen was 12. The above equation for k equal to 12 reduces to:

$$y = \sum_{i=1}^{k} (x_i - 6)$$

The resulting normal random number obtained was then adjusted to yield the given mean and standard deviation using the formula:

$$y' = y(s) + AM$$

where y' is the required normally distributed number
s is the required standard deviation
AM is the required mean

A review of the literature in the field revealed directions for this study and, thusly, affected the procedures developed. Bradley (1) indicated that an investigation of the effects of violating the assumption of homogeneous variances must include factors or parameters to be studied other than simple indicators of extent of the violation. Likewise, Glass, Pecham, and Sanders indicated a need to explore "boundary conditions" (3).

The boundary conditions which were investigated in this study were determined by the effects of three unique parameters on the empirically derived distributions of F-ratios. Selected other parameters were held constant in order to fulfill other assumptions underlying the use of analysis of variance. The computer-based derivation of samples produced samples which were randomly selected from normally distributed populations. Likewise, the mean of each treatment population was equal to zero.

The manipulated parameters used in the study to define boundary conditions were the number of treatment groups (K), the sample size of each treatment group (n), and an index of the extent to which the assumption under question had been violated. This index was represented by Q where for K treatment groups, the ratio of variances across the groups was 1:1:...:Q. When Q is equal to 1, all assumptions upon which the use of analysis of variance is predicated are met

and, in theory, the empirically derived distribution of F-ratios should be identical to the nominal, central F-distribution and any variation between the distributions should be a result of sampling error and should not be systematic in nature.

All data used in the study were generated by either an IBM 360/model 50 or an IBM 370/model 155 computer utilizing IBM subroutines GAUSS and RANDU and standard statistical programs used by the North Texas State University Computing Center (4; 9). Starting points or "seed numbers" used were nine-digit odd numbers which were randomly selected. Other statistical tests such as the proportions test were performed through the use of the IBM 370/model 155 computer utilizing the FORTRAN IV and the APL computer languages.

Hypotheses I and II were tested using empirically derived distributions of F-ratios based upon 500 replications. That is, for each of the 64 simulation models of the form (K,n,Q) where $K \varepsilon \{3,5,7,9\}$, $n \varepsilon \{3,5,7,15\}$, and $Q \varepsilon \{3,7,9,17\}$, 500 independent sets of samples were drawn according to specified simulation model parameters and 500 F-ratios were computed. For each of the 64 simulation models, the 500 F-ratios constituted the empirical F-distribution to be compared to the nominal F-distribution.

After the F-distributions were generated, the next question was concerned with how to statistically compare the

empirically derived F-distributions to the appropriate central F-distribution. Goodness-of-Fit tests such as the chi-square were not appropriate. According to Glass, et al.:

> Inferential tests of the hypothesis of exact correspondence between actual and theoretical distributions of tests statistics are unnecessary (because the null hypothesis is a prior almost certainly false) and potentially misleading (because they would tend to reject because of lack of fit in the central regions of the distribution which could be irrelevant to the use of extreme percentiles for example, 95, 97.5, 99.5 in actual inferential applications of the test statistic) (3, p. 282).

The procedure in comparing the actual or derived F-distribution to the central F-distribution which was used in this study parallels the use of theoretical application of the statistical method to actual inferential testing. For a specified K and N, where N-Kn, n is the number of subjects per treatment group and for $\alpha\varepsilon\{.005, .010, .025, .050, .100, .200, .300, .400, .500, .600, .700, .800, .900, .950, .975, .990, .995\}$, the number of F-ratios in the empirically derived F-distribution exceeding $_{1-\alpha}F_{(K-1)(N-K)}$, as determined by a central F-distribution, was determined. When this number was compared to the number of generated F-ratios in the empirical distribution, the proportion of F-ratios in the empirically derived distribution exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ was computed. For hypothesis I and II, the generated proportions corresponding to specific nominal

levels of .010, .050, .100, and .200 were tested to determine possible differences using standard tests of significance of proportions as described by Glass and Stanley (2) as

$$z = \frac{P - a}{\sqrt{a(1-a)/n}}$$

where P is the proportion of N objects possessing trait under question

a is a real number so that $0 \leq a \leq 1$.

To test the hypotheses at the .10 level as specified, the value of z was compared with the $100(\alpha)$ and the $100(1-\alpha)$ percentiles in the unit normal distribution.

The basic purposes of formulating and testing hypotheses I and II were (1) to test programs and methodology, (2) to provide base-line data which could be compared to data found in other studies, and (3) to provide "starting points" for the remainder of the study.

The literature available in this area indicated that, although the analysis of variance procedure was robust with regard to some types of violations of the assumption of homogeneity of variance, it was probable that extreme violations of the assumption in conjunction with other extreme parametric values may produce conditions where confidence in the statistical procedures is questionable. The "starting points" provided through testing hypotheses

I and II provided indicators to guide the extension of the investigation.

For the second part of the study as directed by hypothesis III, K, the number of treatment groups, was limited to parametric values of 5, 7, and 9. However, n, the number of observations per treatment group and Q, the index of the extent of violation of the assumption were basically unrestricted. For testing hypothesis III, the actual data generating mechanism was the same as used in testing the first two hypotheses. The same computer programs utilizing the subroutines GAUSS and RANDU were used. For hypothesis III, however, the number of repititions used for each simulation was increased from 500 as used in hypotheses I and II to 2,000. This resulted in the generation of F-distributions containing 2,000 F-ratios.

One of the major procedural questions posed was of how to describe "boundary conditions." For the purpose of this study, a "boundary" for a specified value of K, was an area in the Qn-plane defined about a curve $f = \{(Q,n): n=F(Q)\}$ so that the "boundary" was the locus of points (r,s) such that the L-distance from (r,s) to f is less than or equal to one unit. This "boundary" divided that Qn-plane into two mutually exclusive regions for each specified value of K.

In order to test hypothesis III, starting points used in hypotheses I and II for values for K of 5, 7, and 9 were selected. For a given value of n, Q was increased or decreased in order for the point (Q,n) to approach the curve identified in the hypothesis. When the minimal value of Q had been reached which provided a proportion of F-ratios in the actual distribution exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which was significantly different from α for all $\alpha\varepsilon\{.010, .050, .100, .200\}$ at the .05 level, the point (Q,n) was designated to be a boundary point. The L-distance from each boundary point to the hypothesized curve was determined.

CHAPTER BIBLIOGRAPHY

1. Bradley, J. V., "Studies in Research Methodology VI. The Central Limit Effect for a Variety of Populations and the Robustness of the z, t, and F," USAF Aerospace Medical Division Report AMRL-TR-64-123, 1964.

2. Glass, Gene V., and Stanley, Julian C., Statistical Methods in Education and Psychology, Englewood Cliffs, N.J., Prentice-Hall, 1971.

3. Glass, G. V., Peckham, P. D., and Sanders, J. R., "Consequences of Failure to Meet Assumptions Underlying the Analysis and Covariance," Review of Educational Research, XLII (Summer, 1972), 237-288.

4. IBM System/360 Scientific Subroutine Package, Programmer's Manual, Version III, Program number 360A-CM-03X, 5th Edition, August, 1970.

5. Lehman, R., and Bailey, D. E., Digital Computing, Fortran IV and Its Application to the Behavioral Sciences, New York, John Wiley and Sons, Inc., 1966.

6. Lindquist, E. F., Design and Analysis of Experiments in Psychology and Education, Boston, Houghton-Mifflin Company, 1953.

7. McClaran, V. R., "An Investigation of the Effect of Violating the Assumption of Homogeneity of Regression Slopes of the Analysis of Covariance upon the F-statistic," unpublished doctoral dissertation, College of Education, North Texas State University, Denton, Texas, 1972.

8. Naylor, Thomas H., and others, Computer Simulation Techniques, New York, John Wiley and Sons, Inc., 1966.

9. NTSU Subroutine Library User's Guide, Version I, September, 1971.

10. Scheffé, H., The Analysis of Variance, New York, John Wiley & Sons, 1959.

CHAPTER IV

ANALYSIS OF DATA AND FINDINGS

A comparison of the actual significance levels as determined by the empirically derived distributions of F-ratios to the nominal significance levels of .010, .050, .100, and .200 is presented in Table VI. The number of groups (K), the common treatment group sample size (n), and the index of the extent of the violation (Q) are also presented in Table VI. The sixty-four simulations displayed in Table VI provided data for testing hypotheses I and II and each simulation was based upon a distribution of F-ratios containing 500 F-ratios.

Inspection of the data presented in Table VI revealed several patterns. First, with few exceptions, a minor violation of the homogeneity of variance assumption such as Q = 3 yielded a significance level closely approximating the nominal significance level. Of the sixteen simulations with Q = 3, only three, simulation 29 with K = 5, n = 15; simulation 53 with K = 9, n = 5; and simulation 61 with K = 9, n = 15, produced actual significance levels corresponding to .010, .050, .100, and .200, which were significantly different from the nominal levels with a 90 per cent confidence interval was employed.

TABLE VI

SUMMARY DATA FROM SIMULATIONS USED TO PRODUCE ACTUAL F-DISTRIBUTIONS CONTAINING 500 F-RATIOS WITH SELECTED NUMBER OF GROUPS, SAMPLE SIZE, AND HOMOGENEITY OF VARIANCE VIOLATION PARAMETERS

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
1	3	3	3	.008	.068*	.118*	.212
2	3	3	7	.024*	.094*	.160*	.252*
3	3	3	9	.026*	.102*	.160*	.268*
4	3	3	17	.042*	.118*	.176*	.266*
5	3	3	3	.008	.066*	.100	.214
6	3	5	7	.034*	.086*	.132*	.198
7	3	5	9	.020*	.074*	.142*	.240*
8	3	5	17	.052*	.108*	.162*	.258*
9	3	7	3	.024*	.064*	.090	.218
10	3	7	7	.030*	.078*	.124*	.208
11	3	7	9	.038*	.084*	.138*	.224*
12	3	7	17	.048*	.112*	.160*	.234*
13	3	15	3	.006	.044	.086	.162*
14	3	15	7	.038*	.096*	.154*	.242*
15	3	15	9	.036*	.100*	.138*	.204
16	3	15	17	.036*	.090*	.154*	.218
17	5	3	3	.014	.058	.108	.208
18	5	3	7	.050*	.104*	.144*	.232*
19	5	3	9	.054*	.120*	.172*	.246*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VI--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
20	5	3	17	.062*	.128*	.194*	.274*
21	5	5	3	.024*	.070*	.126*	.204
22	5	5	7	.036*	.098*	.126*	.218
23	5	5	9	.048*	.100*	.132*	.250*
24	5	5	17	.072*	.132*	.176*	.250*
25	5	7	3	.014	.076*	.124*	.216
26	5	7	7	.030*	.072*	.128*	.218
27	5	7	9	.038*	.096*	.138*	.228*
28	5	7	17	.072*	.120*	.158*	.216
29	5	15	3	.022*	.076*	.140*	.228*
30	5	15	7	.050*	.110*	.166*	.252*
31	5	15	9	.044*	.082*	.130*	.212
32	5	15	17	.040*	.092*	.144*	.198
33	7	3	3	.024*	.068*	.114	.192
34	7	3	7	.032*	.108*	.172*	.250*
35	7	3	9	.044*	.106*	.160*	.248*
36	7	3	17	.080*	.150*	.202*	.274*
37	7	5	3	.026*	.056	.110	.226
38	7	5	7	.034*	.102*	.144*	.230*
39	7	5	9	.054*	.112*	.172*	.232*
40	7	5	17	.078*	.144*	.194*	.262*
41	7	7	3	.012	.052	.094	.204
42	7	7	7	.046*	.126*	.164*	.254*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VI--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
43	7	7	9	.058*	.120*	.172*	.236*
44	7	7	17	.088*	.150*	.198*	.258*
45	7	15	3	.004*	.062	.096	.178
46	7	15	7	.046*	.092*	.138*	.214
47	7	15	9	.044*	.092*	.132*	.210
48	7	15	17	.050*	.108*	.160*	.200
49	9	3	3	.012	.042	.106	.212
50	9	3	7	.030*	.102*	.156*	.264*
51	9	3	9	.062*	.120*	.164*	.264*
52	9	3	17	.060*	.140*	.188*	.262*
53	9	5	3	.022*	.088*	.150*	.240*
54	9	5	7	.024*	.092*	.142*	.216
55	9	5	9	.060*	.106*	.154*	.256*
56	9	5	17	.096*	.146*	.176*	.244*
57	9	7	3	.020*	.058	.114	.214
58	9	7	7	.046*	.104*	.164*	.248*
59	9	7	9	.068*	.112*	.172*	.252*
60	9	7	17	.085*	.140*	.178*	.236*
61	9	15	3	.022*	.070*	.134*	.232*
62	9	15	7	.042*	.092*	.134*	.194
63	9	15	9	.048*	.100*	.126*	.218
64	9	15	17	.088*	.140*	.170*	.232*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

Second, violation of the assumption of homogeneity of variance of degrees with Q greater than or equal to seven produced significance levels which differed from nominal significance levels. For example, of the sixteen simulations with Q = 7, nine simulations produced significance levels which differed from the nominal significance levels for the values of .010, .050, .100, and .200. Likewise for Q = 9, twelve of the sixteen simulations produced actual significance levels which differed from the nominal levels. For Q = 17, twelve of sixteen simulations produced actual significance levels which differed from anticipated nominal levels.

Third, sample size appeared to have a direct bearing upon the effects of violating the assumption of homogeneity of variances. For n = 3, 5, 7, and 15 the corresponding numbers of simulations of the sixteen studied for each parametric value of n which yielded actual levels which differed from nominal levels was 12, 11, 9, and 4, respectively. It appeared that as the common group sample size decreased, the effect of violating the assumption became more pronounced.

Fourth, an anticipated relationship between the number of groups in the simulation and the effect on signifiance levels under violation of the assumption was not clearly evident. For K = 3 and K = 5, eight of the sixteen

simulations for each parametric value produced significance levels which differed from the nominal levels. For K = 7, however, only nine such simulations were found out of the sixteen and for K = 9, only eleven were found. While some evidence of a possible trend existed, the actual differences in the number of simulations producing significance levels which differed from the nominal levels could have been a function of sampling error.

The last pattern which emerged dealt with the significance levels themselves. For the purposes set forth in this portion of the study, only nominal significance levels of .010, .050, .100, and .200 were studied. Data, however, for other levels was produced. For the nominal significance level of .010, of the sixty-four unique simulations, fifty-seven simulations produced actual significance levels which were significantly different from .010. In all but one of these, the proportion of F-ratios in the actual F-distributions significantly exceeded $_{.010}F_{(K-1)(N-K)}$. This singular exception was simulation 45 with K = 7, n = 15, and n = 3 in which proportion of F-ratios exceeding $_{.010}F_{(6,98)}$ was .004. For the significance level of .050, again fifty-seven of the sixty-four simulations resulted in actual significance levels which differed from .050. Fifty-four of the sixty-four simulations produced a proportion of F-ratios which significantly exceeded .100.

For the nominal level of .200, however, only thirty-eight of the sixty-four simulations produced actual significance levels which significantly departed from .200.

From the above mentioned patterns and other possible trends exhibited in Table VI, it became apparent that the study of the violation of this assumption in analysis of variance is quite complex and must be studied in a context which includes the interrelationships between other parameters.

Since the purpose of this study was to determine boundary conditions of selected variables which may affect the application of the analysis of variance technique, hypotheses I and II were set forth in order to identify points of departure from which the search for boundary conditions might proceed. For this reason hypotheses I and II were tested at the .10 level of significance and were restricted to specified values for the parameters of sample size, number of groups, departure from the assumption and nominal significance levels.

Hypothesis I

As originally set forth, hypothesis I asserted that the proportion of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ will not differ significantly from the proportion of F-ratios in the nominal F-distribution

exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ at the .10 level of significance for:

A. K = 3, n = 15, 7, 5, and 3 with Q = 3, 7, 9, and 17 for all specified nominal levels of .010, .050, .100, and .200.

B. K = 5, n = 15, 7, and 5 with Q = 3, 7, 9, and 17 and n = 3 with Q = 3, 7, and 9 for all specified nominal levels.

C. K = 7, n = 15, 7, 5, and 3 with Q = 3, n = 15, and 7 with Q = 7, 9, and 17 and n = 5 with Q = 7 for all specified nominal levels.

D. K = 9, n = 15, 7, 5, and 3 with Q = 3 and for n = 15 with Q = 7, 9, and 17 for all specified nominal levels.

Part A of the hypothesis I asserted that all simulations with three treatment groups would result in actual F-distributions in which all actual significance levels corresponding to .010, .050, .100, and .200 would not be significantly different from nominal significance levels. Part A of the hypothesis I was rejected. Out of the sixteen simulations included in Part A, eight simulations produced actual distributions of F-ratios in which the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ was significantly different from α for all $\alpha\varepsilon\{.010, .050, .100, .200\}$. Employing notation of the form Z(K,n,Q) where Z is the

simulation number, K is the number of treatment group, n is the common number of subjects per treatment group sample, and Q is the index of the violation of the assumption where the ratio of variances is 1:1:...:Q, the eight simulations were 2(3,3,7); 3(3,3,9); 4(3,3,17); 7(3,5,9); 8(3,5,17); 11(3,7,9); 12(3,7,17); 14(3,15,7)

Part B of hypothesis I was also rejected. Of the fifteen simulations included in part B, seven produced actual distributions of F-ratios in which the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ was significantly different from α for all $\alpha\varepsilon\{.010, .050, .100, .200\}$. These seven simulations were identified as 18(5,3,7); 19(5,3,9); 23(5,5,9); 24(5,5,17); 27(5,7,9); 29(5,15,3); and 30(5,15,7).

Part C of hypothesis I was also rejected. Of the eleven simulations included in part C, four produced actual distributions of F-ratios in which the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ was significantly different from α for all $\alpha\varepsilon\{.010, .050, .100, .200\}$. These four simulations were identified as 38(7,5,7); 42(7,7,7); 43(7,7,9), and 44(7,7,17).

Part D of hypothesis I was rejected. Of the seven simulations included in part D, three simulations produced actual distributions of F-ratios in which the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ was significantly different from α for all $\alpha\varepsilon\{.010, .050, .100, .200\}$.

These three simulations were 53(9,5,3); 61(9,15,3), and 64(9,15,17).

Hypothesis II

Hypothesis II asserted the following: The proportion of F-ratios in the actual F-distribution exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ will be significantly greater than the proportion of F-ratios in the nominal F-distribution exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ at the .10 level of significance for:

A. K = 5, n = 3 with Q = 17 for all specified nominal levels.

B. K = 7, n = 3 with Q = 9, 7, and 17 and n = 5 with Q = 19 and 17 for all specified nominal levels.

C. K = 9, n = 7, 5, and 3 with Q = 7, 9, and 17 for all specified nominal levels.

Part A of hypothesis II was accepted. The simulation used in testing part A of hypothesis II was 20(5,3,17). For all $\alpha\varepsilon\{.010, .050, .100, .200\}$ the proportion of F-ratios in the F-distribution with (5,3,17) parameters which exceeded $_{1-\alpha}F_{(4,11)}$ was significantly greater than α.

Part B of hypothesis II was accepted. Of the five simulations in part B, 34(7,3,7); 35(7,3,9); 36(7,3,17); 39(7,5,9), and 40(7,5,17), each yielded a F-distribution in which the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$

was significantly greater than α for all $\alpha \varepsilon \{.010, .050, .100, .200\}$.

Part C of hypothesis II was rejected. Of the nine simulations used in testing part C of this hypothesis, one simulation, 54(9,5,7) yielded a F-distribution in which the proportion of F-ratios exceeding $_{.80}F_{(8,37)}$ was not significantly greater than .20.

Hypothesis III

As originally asserted, hypothesis III with three subparts was stated as:

A. At least 70 per cent of the boundary points for K = 5 will lie within an L-Distance of one unit of the line defined by $\{(Q,n):n = 2 \ln(Q - 13)\}$.

B. At least 70 per cent of the boundary points for K = 7 will lie within an L-Distance of one unit of the line defined by $\{(Q,n):n = 15/8 \ln 3(Q - 4\}$.

C. At least 70 per cent of the boundary points for K = 9 will lie within an L-Distance of one unit of the line defined by $\{(Q,n):n = 5 \ln (Q - 2)\}$.

The formulation of this hypothesis and its subsequent testing required the explicit definitions for two distinct concepts, boundary point and L-distance. As previously set forth, definitions for these concepts are as follows:

Boundary Point--For a specified value of K, (Q,n) is a boundary point if and only if, for a specified value of

n, Q is the minimal integral value which produces an actual F-distribution which is significantly different from the nominal F-distribution at the .05 level for all specified nominal levels of .010, .050, .100, and .200.

L-Distance--If (x,y) is a point in the Qn-plane and $\{(Q,n):n = f(Q)\}$ is a curve in the Qn-plane and $(Q_1,f(Q_1))$ is the intersection of the line $\{(Q,n):Q = x\}$ with the curve $\{(Q,n): n = f(Q)\}$ and $(Q_2, f(Q_2))$ is the intersection of the line $\{(Q,n):n = y)\}$ and the curve $\{(Q,n):n = f(Q)\}$, then the L-Distance from (x,y) to $\{(Q,n):n = f(Q)\}$ is the minimum of $|f(Q_2) - f(Q_1)|$ and $|Q_2 - Q_1|$.

The statement of this hypothesis in this manner not only allowed the flexibility required by this type of investigation but facilitated the remainder of the investigation by dictating a sense of directionality to be followed. Unfortunately, actual testing of this hypothesis was impaired by the nature of the data. This is particularly true with Hypothesis III, part A, which asserts a boundary region for the case K = 5. The sixteen original simulation models served as starting points for the remainder of the investigation. After the starting points were determined, values for the parameters Q and n were selected for the generation of F distributions. The testing of actual proportions with nominal proportions produced data upon which the decision of fulfillment of boundary point criteria was based.

For the purpose of testing hypothesis III, part A, the original sixteen simulations with K = 5 were replicated in order to produce actual F-distributions containing 2,000 F-ratios. In addition, eighty-nine other combinations of parametric values of the homogeneity of variance violation index (Q) and sample size (n) were used in order to produce a total of 105 simulation models with K = 5. The proportion of F-ratios in the actual F-distribution corresponding to nominal significance levels of .010, .050, .100, and .200 for each of the 105 simulation models for K = 5 are found in Table VII.

Several characteristics of the data contained in Table VII appeared to be noteworthy. First, extreme parametric values of Q such as Q > 19 produced, with only one exception, model 151(5,12,21), proportions of F-ratios which were significantly greater than $_{1-\alpha}F_{(K-1)(N-K)}$ for $\alpha\varepsilon\{.010, .050, .100, .200\}$. Likewise, for parametric values of Q < 6, of the twenty simulation models tested, only five or 20 per cent produced actual F-distributions such that the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ was significantly different from α for all specified nominal levels.

Second, the deviation of the actual significance level (α') or the proportion of F-ratios in the empirically derived F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ from the

TABLE VII

SUMMARY DATA FROM SIMULATIONS USED TO PRODUCE ACTUAL F-DISTRIBUTIONS CONTAINING 500 F-RATIOS WITH FIVE TREATMENT GROUPS AND SELECTED SAMPLE SIZE AND HOMOGENEITY OF VARIOUS VIOLATION PARAMETERS

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
65	5	3	3	.018*	.065*	.114*	.190
66	5	3	4	.018*	.074*	.129*	.224*
67	5	3	5	.027*	.088*	.134*	.235*
68	5	3	6	.036*	.100*	.151*	.251*
69	5	3	7	.036*	.093*	.151*	.239*
70	5	3	8	.042*	.103*	.155*	.241*
71	5	3	9	.040*	.116*	.172*	.259*
72	5	3	17	.068*	.135*	.189*	.261*
73	5	3	21	.071*	.134*	.180*	.256*
74	5	5	3	.017*	.060*	.107	.198
75	5	5	4	.024*	.071*	.123*	.217
76	5	5	5	.022*	.072*	.124*	.212
77	5	5	6	.034*	.083*	.131*	.210
78	5	5	7	.039*	.098*	.144*	.223*
79	5	5	8	.032*	.092*	.140*	.223*
80	5	5	9	.044*	.107*	.158*	.234*
81	5	5	11	.052*	.120*	.165*	.239*
82	5	5	13	.054*	.122*	.170*	.246*
83	5	5	15	.063*	.134*	.174*	.234*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VII--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
84	5	5	17	.070*	.138*	.186*	.249*
85	5	5	21	.069*	.125*	.176*	.245*
86	5	7	3	.016*	.066*	.121*	.222*
87	5	7	4	.025*	.073*	.126*	.212
88	5	7	5	.030*	.074*	.123*	.197
89	5	7	6	.033*	.078*	.135*	.229*
90	5	7	7	.047*	.101*	.149*	.229*
91	5	7	8	.045*	.097*	.142*	.233*
92	5	7	9	.044*	.100*	.153*	.223
93	5	7	10	.053*	.106*	.151*	.213
94	5	7	11	.050*	.113*	.157*	.228*
95	5	7	12	.044*	.095*	.136*	.209
96	5	7	13	.057*	.115*	.161*	.228*
97	5	7	15	.049*	.103*	.148*	.207
98	5	7	17	.060*	.119*	.169*	.231*
99	5	7	19	.066*	.117*	.162*	.228*
100	5	7	21	.064*	.123*	.171*	.249*
101	5	9	3	.039*	.065*	.106	.192
102	5	9	4	.022*	.065*	.121*	.214
103	5	9	5	.035*	.086*	.135*	.215
104	5	9	6	.037*	.093*	.145*	.222*
105	5	9	7	.041*	.100*	.146*	.220*
106	5	9	8	.038*	.087*	.133*	.211

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VII--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
107	5	9	9	.045*	.105*	.153*	.232*
108	5	9	10	.039*	.105*	.149*	.215
109	5	9	11	.050*	.099*	.141*	.203
110	5	9	12	.039*	.105*	.149*	.215
111	5	9	13	.057*	.115*	.159*	.237*
112	5	9	14	.039*	.105*	.140*	.215
113	5	9	15	.054*	.112*	.155*	.233*
114	5	9	16	.055*	.105*	.148*	.215
115	5	9	17	.063*	.126*	.172*	.240*
116	5	9	19	.067*	.116*	.155*	.218*
117	5	9	21	.063*	.117*	.170*	.229*
118	5	11	3	.021*	.064*	.110	.207
119	5	11	4	.017*	.068*	.118*	.201
120	5	11	5	.029*	.077*	.129*	.221*
121	5	11	6	.021*	.072*	.123*	.214
122	5	11	7	.026*	.077*	.126*	.203
123	5	11	8	.044*	.096*	.138*	.217
124	5	11	9	.036*	.097*	.139*	.217
125	5	11	10	.043*	.098*	.124*	.213
126	5	11	11	.053*	.105*	.146*	.215
127	5	11	12	.051*	.106*	.155*	.219*
128	5	11	13	.048*	.110*	.152*	.215
129	5	11	14	.051*	.108*	.153*	.214

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VII--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
130	5	11	15	.052*	.118*	.163*	.233*
131	5	11	16	.056*	.104*	.146*	.207
132	5	11	17	.050*	.108*	.150*	.220*
133	5	11	19	.062*	.115*	.154*	.219*
134	5	11	21	.066*	.117*	.163*	.221*
135	5	13	3	.028*	.070*	.116*	.211
136	5	13	4	.023*	.072*	.122*	.219*
137	5	13	5	.028*	.076*	.124*	.201
138	5	13	6	.037*	.061*	.131	.198
139	5	13	7	.034*	.060*	.140*	.218*
140	5	13	8	.040*	.099*	.144*	.217
141	5	13	9	.043*	.097*	.142*	.223*
142	5	13	10	.052*	.110*	.154*	.229*
143	5	13	11	.047*	.097*	.150*	.221*
144	5	13	12	.044*	.091*	.127*	.202
145	5	13	13	.048*	.106*	.150*	.226*
106	5	13	14	.052*	.108*	.145*	.207
147	5	13	15	.050*	.107*	.141*	.201
148	5	13	16	.051*	.104*	.135*	.194
149	5	13	17	.051*	.099*	.140*	.210
150	5	13	19	.074*	.134*	.181*	.232*
151	5	13	21	.056*	.113*	.155*	.217
152	5	15	3	.018*	.063*	.110	.194

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VII--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
153	5	15	5	.030*	.082*	.134*	.204
154	5	15	6	.028*	.081*	.127*	.212
155	5	15	7	.032*	.084*	.134*	.212
156	5	15	8	.036*	.086*	.133*	.211
157	5	15	9	.041*	.088*	.131*	.211
158	5	15	10	.042*	.096*	.133*	.203
159	5	15	11	.044*	.098*	.139*	.209
160	5	15	12	.051*	.107*	.142*	.217
161	5	15	13	.050*	.109*	.147*	.214
162	5	15	14	.051*	.106*	.140*	.202
163	5	15	15	.054*	.116*	.158*	.221*
164	5	15	16	.057*	.117*	.153*	.221*
165	5	15	17	.053*	.098*	.146*	.214
166	5	15	18	.058*	.112*	.145*	.205
167	5	15	19	.055*	.111*	.154*	.220*
168	5	15	20	.058*	.118*	.172*	.236*
169	5	15	21	.068*	.123*	.161*	.218*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

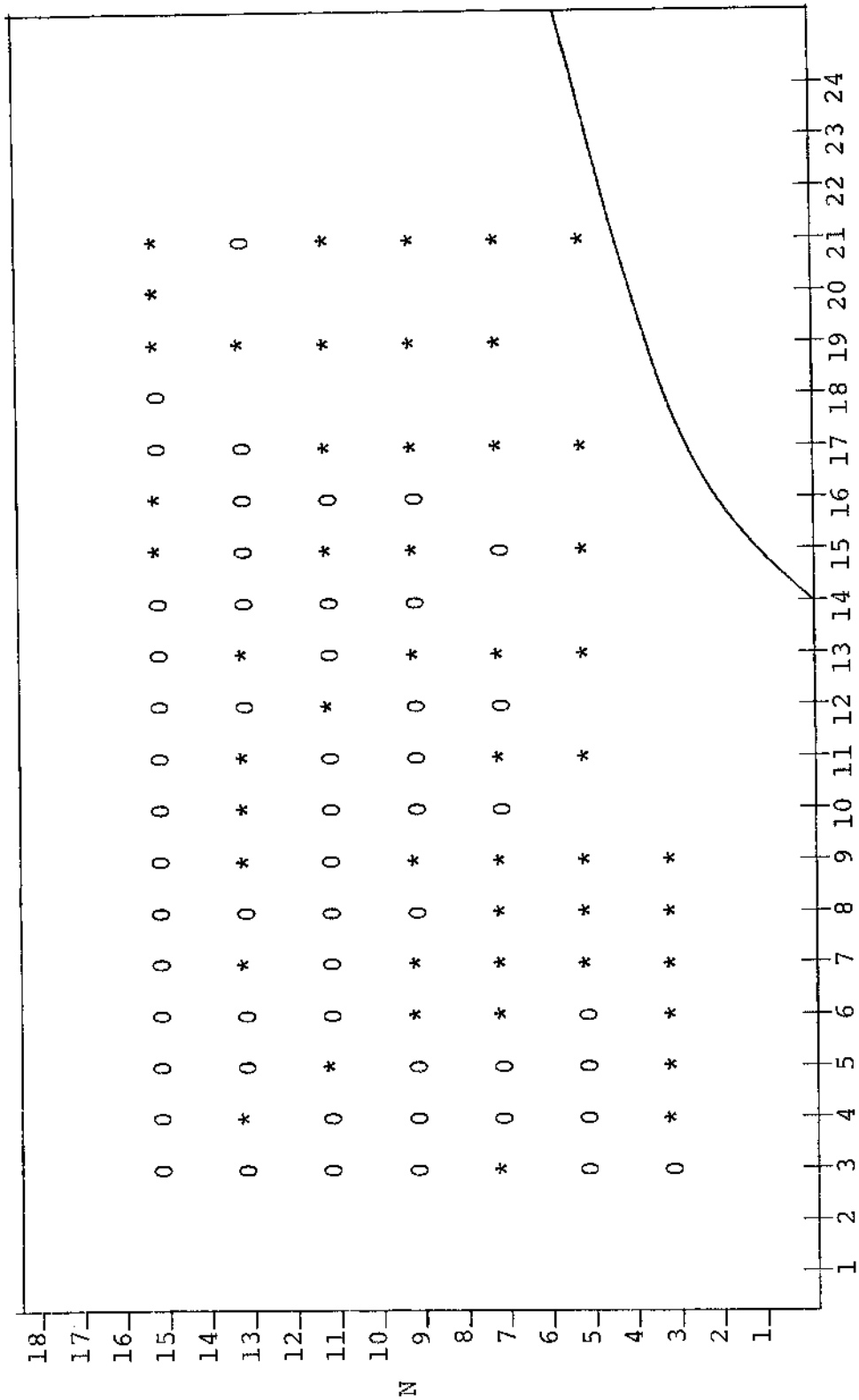
nominal significance level (α) should be recognized. For $\alpha = .010$, α' was greater than α for all α' generated in the 105 simulations studied for K = 5. These deviations ranged from $\alpha' - \alpha = .006$ for 86(5,7,3) to .059 for 73(5,3,21). For $\alpha = .050$ the deviation ranged from .01 for simulations 74(5,5,3) and 139(5,13,7) to .085 for simulation 72(5,3,17) when $\alpha' = .135$. For $\alpha = .100$ the smallest value of α' generated was .106 for simulation 101(5,9,3), while the largest value of α' generated was .189 for simulation 72(5,3,17). These extreme values of α' determined a range of deviations of α' from α to be between .006 to .089.

For all 105 simulation models analyzed, only in comparison to $\alpha = .200$, did the achieved value of α' underestimate α. This occurred in models 65(5,3,3); 74(5,5,3); 88(5,7,5); 101(5,9,3); 138(5,13,6); 148(5,13,16), and 152(5,15,3). With the singular exception of simulation 148 with Q = 16, every incidence of $\alpha' < \alpha$ was found to be in simulation models with smaller parametric values of Q. The largest values of α' corresponding to $\alpha = .200$ were observed in model 72(5,3,15) with $\alpha' = .261$ and model 71(5,3,9) with $\alpha' = .259$. On a comparison of $\alpha' - \alpha$ deviations, the range of values extended from -.010 to .061. In summary, it should be noted that for all α levels of .010, .050, .100, and .200 investigated, the ranges of deviations of $\alpha' - \alpha$ were

generally of the same length with slightly longer ranges occurring when values of α of .100 and .200 were employed.

As can be seen in comparing Table VII to Table VI, which displayed summary data on the original sixty-four simulation models employing F-distributions containing 500 F-ratios, the same general patterns in data displayed in Table VI and discussed at that point are observable in Table VII also.

As originally stated, hypothesis III, part A, asserted the existence and location of a boundary region for selected analysis of variance models with K = 5 to be determined by the positions of boundary points. Graph I displays data relative to the investigation of this assertion. In this graphic, simulation models which produced F-distributions such that the proportion of F-ratios in the distribution exceeded $_{1-\alpha}F_{(K-1)(N-K)}$ for all $\alpha\varepsilon\{.010, .050, .100, .200\}$ at the .05 level of significance are indicated by asterisks (*) while all other simulation models which do not meet this criterion are indicated by zeros (0). The hypothesis asserted that the boundary would be defined by a region about $\{(Q,n): n = 2 \ln(Q - 13)\}$. The graph of this defined set of ordered pairs (Q,n) are displayed on Graph I as a continuous solid line. For the sake of clarity and brevity, the function $\{(Q,n): n = 2 \ln(Q - 13)\}$ will be denoted as G_5.

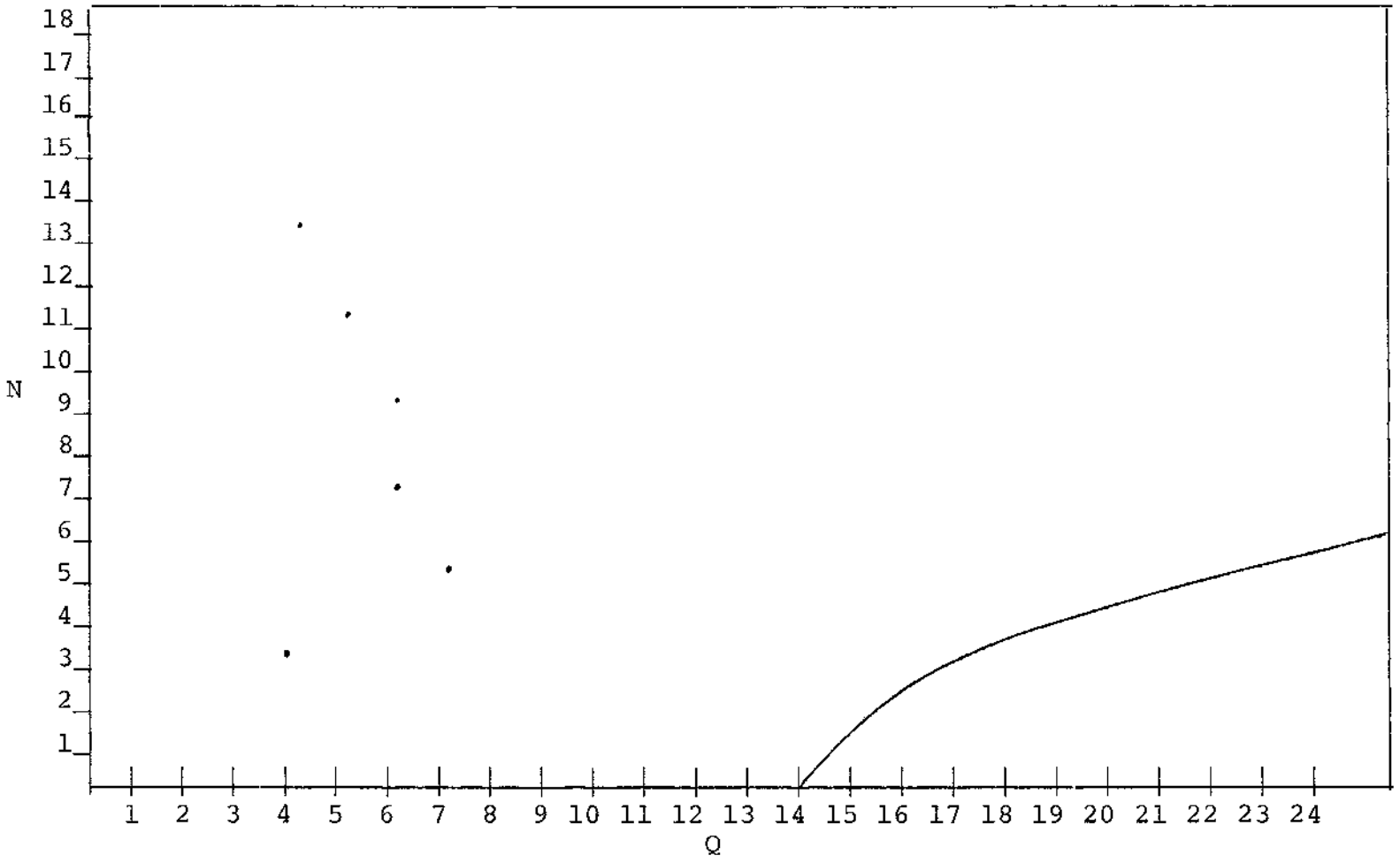


Graph 1--Distribution of significant and non-significant simulation models with five treatment groups and hypothesized boundary location, K = 5.

As defined, the boundary points for simulation models with K = 5 are representations of simulation models 66(5,3,4); 78(5,5,7); 104(5,9,6); 120(5,11,5); 136(5,13,4), and 163(5,15,15). Since no simulations were attempted with a value of Q less than three, there did not exist a boundary point for the set of simulation models with n = 7. A comparison of the location of these defined boundary points and G_5 is shown in Graph 2. As indicated in Graph 2, no boundary point fell within a L-distance of one unit of G_5. Hence, the assertion of G_5 as defining a boundary area for simulation models with K = 5 was not supported. Research hypothesis III, part A was, therefore, rejected.

As seen in Graph 1, the diffused pattern of simulations producing F-distributions in which α' was significantly different from α for all corresponding specified values of α was revealed. This diffused pattern restricted the definition of a boundary area for Analysis of Variance simulation models with K = 5.

For the purpose of testing hypothesis III, part B, the original sixteen simulation models with K = 7 which were employed in testing hypotheses I and II were replicated in order to product actual F-distributions containing 2,000 F-ratios. A total of 98 unique simulation models was investigated using K = 7 and differing combinations of parametric values for n and Q. The proportion of F-ratios in the actual F-distribution corresponding to nominal



Graph 2--Distribution of boundary points determined through simulations with five treatment groups and hypothesized boundary locations, K = 5.

significance levels of .010, .050, .100, and .200 for each of the ninety-eight simulation models for K = 7 are found in Table VIII.

Of the eighteen simulation models presented in Table VIII with K = 7 and $n < 7$, only two, 170(7,3,3) and 178(7,5,3), failed to produce an F-distribution in which each proportion, α', was significantly different from α for $\alpha \varepsilon \{.010, .050, .100, .200\}$.

For $\alpha = .010$, significant deviations of α' from α were found at all sample size levels and across all values of Q with the exception of Q = 3, the minimal violation of the homogeneity of variance assumption studied. For all ninety-eight simulation models with K = 7, the smallest value of α' obtained was .012 in simulation model 170(7,3,3) while the greatest value was found to be .083 in simulation model 186(7,5,17); therefore, the deviation range when $\alpha = .010$ extended from .002 to .073. It was found that for $\alpha = .010$, no combination of parametric values of n and Q selected produced an F-distribution in which α' under-estimated α.

For the α-level of .050, the least deviation, $\alpha' - \alpha$, occurred with α' derived from simulation model 188(7,7,3) where the achieved α' of .056 yielded a deviation from .050 of .006. The largest deviation of .103 was a result of $\alpha' = .153$ derived from simulation model 242(7,13,19). As

TABLE VIII

SUMMARY DATA FROM SIMULATIONS USED TO PRODUCE ACTUAL F-DISTRIBUTIONS CONTAINING 500 F-RATIOS WITH SEVEN TREATMENT GROUPS AND SELECTED SAMPLE SIZE AND HOMOGENEITY OF VARIOUS VIOLATION PARAMETERS

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
170	7	3	3	.012	.061*	.112	.211
171	7	3	4	.025*	.074*	.141*	.245*
172	7	3	5	.027*	.082*	.138*	.233*
173	7	3	6	.028*	.091*	.147*	.243*
174	7	3	7	.035*	.103*	.161*	.245*
175	7	3	9	.055*	.122*	.176*	.263*
176	7	3	11	.055*	.128*	.185*	.266*
177	7	3	17	.077*	.136*	.185*	.252*
178	7	5	3	.014	.060	.103	.189
179	7	5	4	.026*	.079*	.134*	.225*
180	7	5	5	.029*	.081*	.140*	.229*
181	7	5	7	.037*	.092*	.135*	.221*
182	7	5	9	.049*	.110*	.150*	.230*
183	7	5	11	.059*	.132*	.178*	.243*
184	7	5	13	.068*	.127*	.174*	.247*
185	7	5	15	.062*	.129*	.182*	.250*
186	7	5	17	.083*	.139*	.194*	.258*
187	7	5	21	.079*	.142*	.190*	.249*
188	7	7	3	.017*	.056	.105	.193

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VIII--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
189	7	7	4	.026*	.080*	.138*	.234*
190	7	7	3	.034*	.083*	.138*	.229*
191	7	7	7	.046*	.097*	.140*	.215
192	7	7	8	.036*	.097*	.140*	.212
193	7	7	9	.047*	.108*	.149*	.229*
194	7	7	10	.050*	.097*	.140*	.214
195	7	7	11	.057*	.104*	.151*	.220*
196	7	7	12	.067*	.118*	.167*	.236*
197	7	7	13	.070*	.123*	.166*	.233*
198	7	7	14	.060*	.121*	.165*	.246*
199	7	7	16	.067*	.124*	.173*	.238*
200	7	7	17	.072*	.127*	.180*	.244*
201	7	7	21	.073*	.134*	.177*	.238*
202	7	9	4	.025*	.077*	.121*	.202
203	7	9	6	.031*	.078*	.125*	.205
204	7	9	7	.046*	.104*	.144*	.210
205	7	9	8	.056*	.108*	.154*	.229*
206	7	9	9	.054*	.113*	.153*	.220*
207	7	9	10	.043*	.089*	.128*	.198
208	7	9	11	.060*	.125*	.166*	.235*
209	7	9	12	.057*	.120*	.158*	.219*
210	7	9	13	.064*	.126*	.163*	.231*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VIII--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
211	7	9	14	.066*	.128*	.174*	.236*
212	7	9	15	.068*	.117*	.169*	.232*
213	7	9	16	.063*	.114*	.158*	.221*
214	7	9	17	.072*	.123*	.170*	.230*
215	7	11	5	.021*	.078*	.125*	.216
216	7	11	6	.031*	.086*	.130*	.209
217	7	11	7	.044*	.097*	.149*	.229*
218	7	11	8	.044*	.098*	.141*	.207
219	7	11	9	.057*	.104*	.145*	.215
220	7	11	10	.051*	.106*	.149*	.220*
221	7	11	11	.047*	.105*	.144*	.211
222	7	11	12	.067*	.118*	.158*	.224*
223	7	11	13	.067*	.119*	.163*	.215
224	7	11	14	.068*	.125*	.164*	.225*
225	7	11	15	.072*	.122*	.164*	.222*
226	7	11	16	.078*	.132*	.175*	.239*
227	7	11	17	.074*	.136*	.172*	.231*
228	7	11	21	.079*	.133*	.169*	.221*
229	7	13	6	.031*	.089*	.136*	.208
230	7	13	7	.048*	.101*	.143*	.211
231	7	13	8	.048*	.103*	.142*	.219*
232	7	13	9	.052*	.110*	.140*	.217*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VIII--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
233	7	13	10	.057*	.110*	.154*	.218*
234	7	13	11	.063*	.107*	.158*	.216
235	7	13	12	.069*	.124*	.166*	.232*
236	7	13	13	.068*	.120*	.165*	.219*
237	7	13	14	.078*	.129*	.162*	.232*
238	7	13	15	.067*	.127*	.168*	.225*
239	7	13	16	.070*	.124*	.162*	.222*
240	7	13	17	.076*	.130*	.169*	.228*
241	7	13	18	.071*	.134*	.177*	.235*
242	7	13	19	.082*	.153*	.190*	.248*
243	7	13	21	.081*	.138*	.185*	.242*
244	7	15	3	.023*	.063*	.120*	.214
245	7	15	7	.038*	.086*	.127*	.198
246	7	15	9	.052*	.097*	.134*	.203
247	7	15	10	.053*	.112*	.144*	.207
248	7	15	11	.065*	.119*	.167*	.236*
249	7	15	12	.058*	.108*	.151*	.209
250	7	15	13	.059*	.115*	.150*	.219*
251	7	15	14	.064*	.125*	.160*	.223*
252	7	15	15	.064*	.115*	.159*	.212
253	7	15	17	.073*	.140*	.171*	.225*
254	7	15	18	.068*	.147*	.159*	.221*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE VIII--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
255	7	15	15	.071*	.122*	.159*	.224*
256	7	17	9	.042*	.091*	.134*	.192
257	7	17	10	.044*	.089*	.119*	.183
258	7	17	11	.051*	.101*	.146*	.208
259	7	17	12	.054*	.111*	.149*	.228*
260	7	17	13	.074*	.125*	.164*	.224*
261	7	17	14	.068*	.120*	.163*	.218*
262	7	17	15	.066*	.118*	.154*	.203
263	7	17	16	.071*	.124*	.167*	.221*
264	7	17	17	.062*	.121*	.168*	.228*
265	7	17	18	.060*	.121*	.160*	.217*
266	7	17	19	.071*	.127*	.169*	.231*
267	7	17	21	.080*	.137*	.168*	.233*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

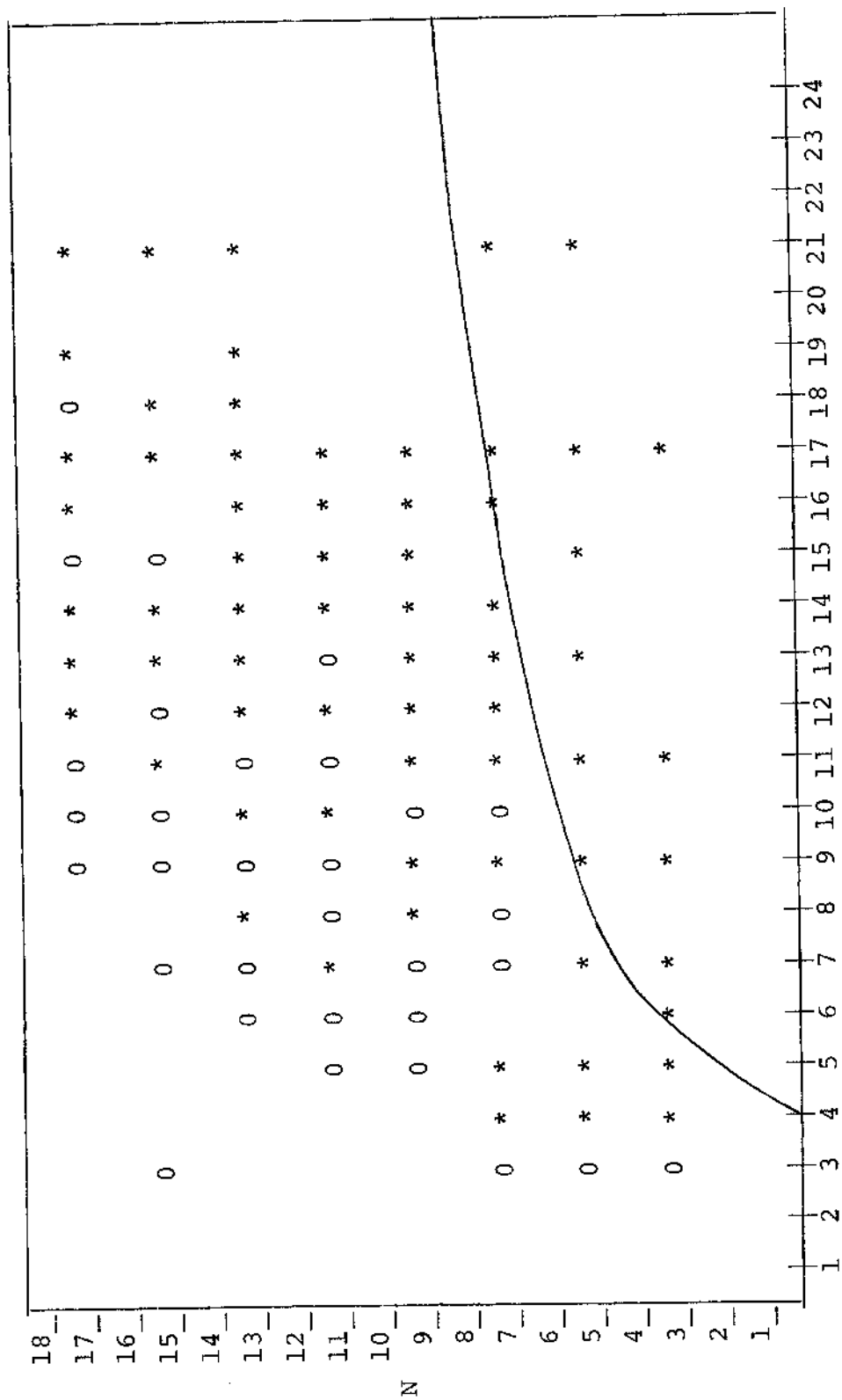
previously noted for the α-level of .010, no α' obtained corresponding to α = .050 underestimated α.

For the α-level of .100, the minimum α' achieved was .103 obtained from simulation model 178(7,5,3). The minimum $\alpha' - \alpha$ deviation for α = .100 was, therefore, .003. The largest α' corresponding to α = .100 achieved was .194 from simulation model 186(7,5,17), and thus the maximum $\alpha' - \alpha$ deviation for α = .100 was .094.

Simulation models producing an α' corresponding to α = .200 did, on some occasions, underestimate α. In Table VIII these simulation models can be identified as 178(7,5,3,); 188(7,7,3); 207(7,9,10); 245(7,15,7); 256(7,17,9), and 257(7,17,10). Simulation model 257(7,17,10) produced an α' of .183 so that $\alpha' - \alpha$ was equal to -.017. The largest α' corresponding to α = .200 generated was .266 from simulation model 176(7,3,11). The maximum $\alpha' - \alpha$ deviation was, therefore, .066.

When deviation ranges for $\alpha \varepsilon \{.010, .050, .100, .200\}$ were compared, it was found that the range length for values of α of .050, .100, and .200 were similar and the range length for .010 was less than that of the other α-levels. The range lengths for the K = 7 simulation models were generally longer than corresponding α-levels range lengths for the K = 5 simulation models.

Graph 3 illustrates summary data relative to testing hypothesis III, part B. The ninety-eight simulation models

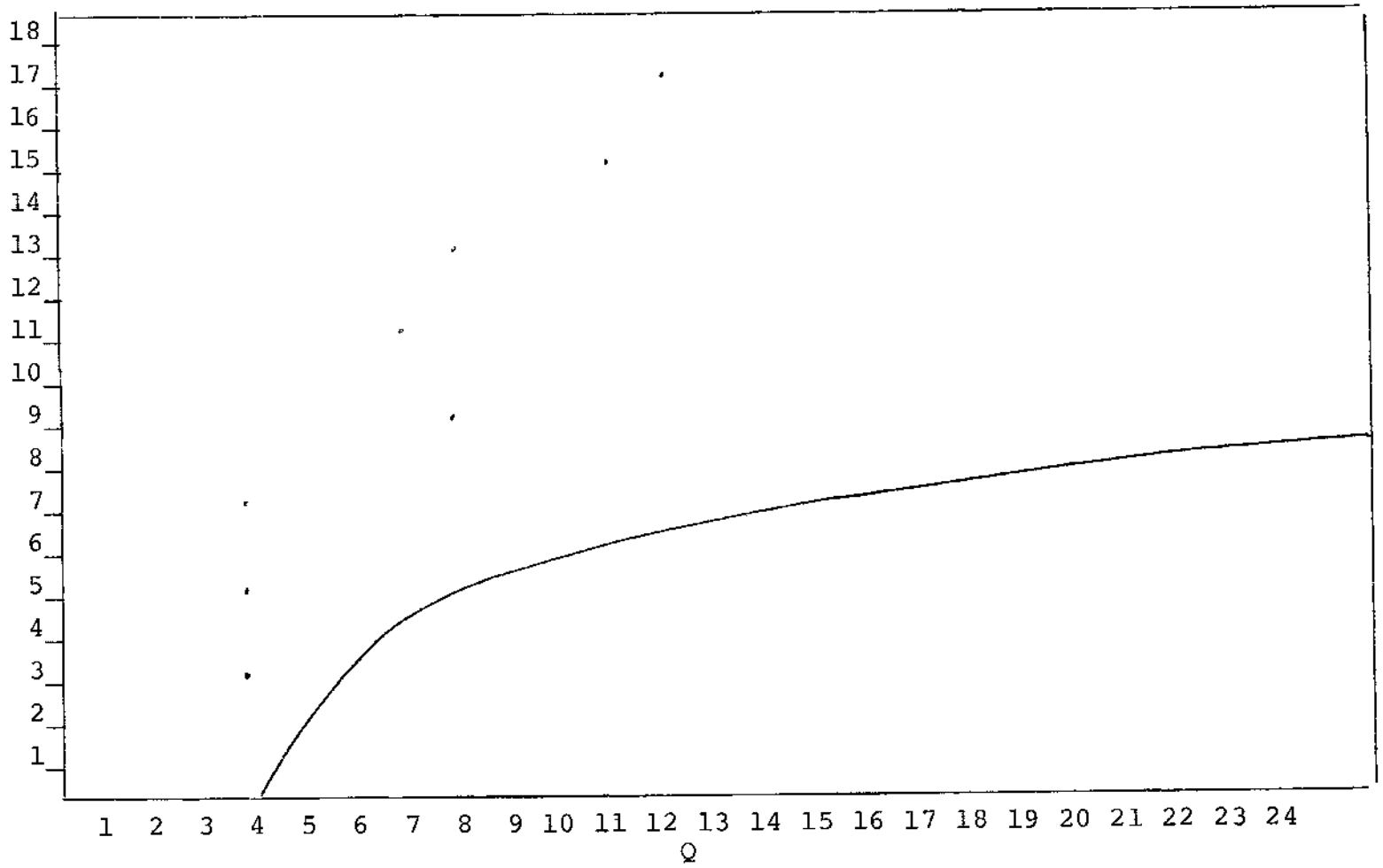


Graph 3--Distribution of significant and non-significant simulation models with seven treatment groups and hypothesized boundary location, K = 7.

are represented on this graphic, as is the location of the hypothesized boundary line. Hypothesis III, part B, asserted that the graph of $G_7 = \{(Q,n):n=15/8(\ln 3(Q - 4))\}$ was the basis for defining a boundary region for K = 7. Simulation models which produced F-distributions in which the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ for all $\alpha\varepsilon\{.010, .050, .100, .200\}$ are indicated by asterisks (*). Simulation models in which this criterion was not met are indicated by zeros (0). The graph of $G_7 = \{(Q,n) : N = 15/8 (\ln 3(Q - 4))\}$ is represented by a continuous solid line.

The boundary points for hypothesis III, part B, were determined to be attributable to the following simulation models: 171(7,3,4); 179(7,5,4); 189(7,7,4); 205(7,9,8); 217(7,11,7); 231(7,13,8); 248(7,15,11), and 259(7,17,12). Graph 4 displays the locations of the defined boundary points and the graph of G_7. As can be seen in Graph 4, no boundary points were found to be within an L-distance of one unit of G_7. Hence, hypothesis III, part B, was rejected.

As was previously found to be the case in the set of simulation models with K = 5, the pattern of models in which the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ was significantly different from for all $\alpha\varepsilon\{.010, .050, .100, .200\}$ was diffused throughout ranges of values for parameters Q and n.



Graph 4--Distribution of boundary points determined through simulations with seven treatment groups and hypothesized boundary location, K = 7.

Hypothesis III, part C asserted the existence and location of a boundary region in the Qn-plane for the case of K = 9. In order to test this hypothesis a total of ninety simulation models were studied. Sixteen of these models were replications of models used in hypothesis I and II with the number of F-ratios contained in the actual F-distributions increased from 500 to 2,000. Proportions of F-ratios in the empirically derived distributions corresponding to nominal significance levels of .010, .050, .100, and .200 are found in Table IX.

For an α-level of .010, the smallest α' generated was .014 as a result of simulation model 268(9,3,3) while the largest α' generated was .097 as a result of simulation model 291(9,7,21). When $\alpha' - \alpha$ was used as a measure of deviation, the range of deviation determined was .004 through .087.

The range of deviation determined for the ninety simulation models in the .050 α-level was .008 through .111 as a result of the same simulation models found to produce maximum and minimum α' values at the .010 level. Simulation model 268(9,3,3) generated an α' of .058 while model 291(9,7,21) generated an α' of .161.

For the α-level of .100, the minimum α' obtained was .104 in simulation model 292(9,9,3) and the maximum α' obtained was .207 in simulation model 291(9,7,21). The

TABLE IX

SUMMARY DATA FROM SIMULATIONS USED TO PRODUCE ACTUAL F-DISTRIBUTIONS CONTAINING 500 F-RATIOS WITH NINE TREATMENT GROUPS AND SELECTED SAMPLE SIZE AND HOMOGENEITY OF VARIOUS VIOLATION PARAMETERS

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
268	9	3	3	.014	.058	.113	.222*
269	9	3	4	.023*	.071*	.121*	.223*
270	9	3	5	.025*	.083*	.133*	.226*
271	9	3	7	.037*	.100*	.156*	.250*
272	9	3	9	.056*	.123*	.178*	.272*
273	9	3	17	.083*	.155*	.205*	.277*
274	9	5	3	.018*	.063*	.113	.209
275	9	5	4	.018*	.065*	.108	.202
276	9	5	5	.029*	.084*	.144*	.235*
277	9	5	6	.039*	.092*	.133*	.222*
278	9	5	7	.044*	.110*	.156*	.247*
279	9	5	9	.054*	.109*	.166*	.236*
280	9	5	13	.075*	.134*	.174*	.239*
281	9	5	17	.082*	.146*	.187*	.241*
282	9	7	3	.021*	.071*	.125*	.214
283	9	7	4	.027*	.076*	.124*	.211
284	9	7	5	.036*	.088*	.138*	.220*
285	9	7	6	.041*	.094*	.147*	.232*
286	9	7	7	.045*	.098*	.146*	.231*
287	9	7	9	.054*	.113*	.159*	.236*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE IX--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
288	9	7	11	.061*	.116*	.187*	.226*
289	9	7	13	.070*	.134*	.171*	.230*
290	9	7	17	.092*	.146*	.190*	.244*
291	9	7	21	.097*	.161*	.207*	.262*
292	9	9	3	.015	.062*	.104	.188
293	9	9	4	.023*	.074*	.114*	.199
294	9	9	5	.030*	.083*	.135*	.229*
295	9	9	6	.036*	.090*	.140*	.223*
296	9	9	7	.047*	.101*	.147*	.221*
297	9	9	8	.057*	.109*	.158*	.224*
298	9	9	9	.053*	.112*	.156*	.225*
299	9	9	11	.066*	.118*	.166*	.229*
300	9	9	13	.066*	.126*	.172*	.245*
301	9	9	17	.080*	.130*	.169*	.226*
302	9	9	21	.092*	.144*	.187*	.241*
303	9	11	3	.023*	.068*	.111	.211
304	9	11	4	.024*	.072*	.119*	.213
305	9	11	5	.032*	.084*	.136*	.206
306	9	11	6	.037*	.083*	.131*	.211
307	9	11	7	.050*	.091*	.148*	.220*
308	9	11	8	.051*	.107*	.149*	.218*
309	9	11	9	.059*	.112*	.157*	.226*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE IX--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
310	9	11	11	.058*	.114*	.150*	.217
311	9	11	13	.073*	.121*	.164*	.225*
312	9	11	15	.072*	.134*	.171*	.226*
313	9	11	17	.078*	.130*	.170*	.231*
314	9	11	21	.086*	.144*	.179*	.225*
315	9	13	4	.024*	.076*	.124*	.210
316	9	13	5	.031*	.078*	.124*	.199
317	9	13	6	.046*	.105*	.146*	.213
318	9	13	7	.052*	.105*	.158*	.217
319	9	13	8	.038*	.089*	.140*	.201
320	9	13	9	.058*	.118*	.167*	.243*
321	9	13	10	.068*	.118*	.159*	.224*
322	9	13	11	.063*	.107*	.150*	.220*
323	9	13	13	.069*	.128*	.167*	.232*
324	9	13	15	.071*	.122*	.168*	.234*
325	9	13	17	.077*	.140*	.179*	.238*
326	9	13	21	.087*	.140*	.174*	.228*
327	9	15	3	.016*	.060*	.112	.191
328	9	15	6	.042*	.086*	.129*	.207
329	9	15	7	.053*	.104*	.150*	.217
330	9	15	8	.041*	.094*	.137*	.206
331	9	15	9	.061*	.113*	.158*	.223*
332	9	15	10	.058*	.110*	.150*	.213

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE IX--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
333	9	15	11	.066*	.125*	.167*	.225*
334	9	15	12	.065*	.115*	.151*	.225*
335	9	15	13	.073*	.120*	.157*	.219*
336	9	15	15	.079*	.123*	.163*	.224*
337	9	15	17	.081*	.134*	.168*	.232*
338	9	15	21	.089*	.145*	.185*	.243*
339	9	17	7	.052*	.107*	.148*	.211
340	9	17	8	.043*	.115*	.155*	.213
341	9	17	9	.062*	.111*	.149*	.216
342	9	17	10	.069*	.113*	.161*	.221*
343	9	17	11	.073*	.128*	.168*	.225*
344	9	17	12	.070*	.128*	.171*	.237*
345	9	17	13	.078*	.127*	.158*	.223*
346	9	17	15	.081*	.139*	.182*	.231*
347	9	17	17	.081*	.142*	.182*	.232*
348	9	17	21	.090*	.149*	.185*	.233*
349	9	19	9	.054*	.111*	.151*	.214
350	9	19	10	.059*	.113*	.161*	.235*
351	9	19	11	.058*	.114*	.155*	.220*
352	9	19	12	.057*	.113*	.151*	.220*
353	9	19	13	.067*	.119*	.162*	.221*

*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

TABLE IX--Continued

Simulation Number	Number of Groups (K)	Sample Size (n)	Homogeneity of Variance Violation Index (Q)	Proportion of F-Ratios in the Actual F-Distribution Corresponding to Nominal Significance Levels of			
				.010	.050	.100	.200
354	9	19	14	.075*	.126*	.158*	.226*
355	9	19	15	.072*	.127*	.163*	.220*
356	9	19	17	.069*	.126*	.163*	.230*
357	9	19	21	.085*	.138*	.176*	.231*

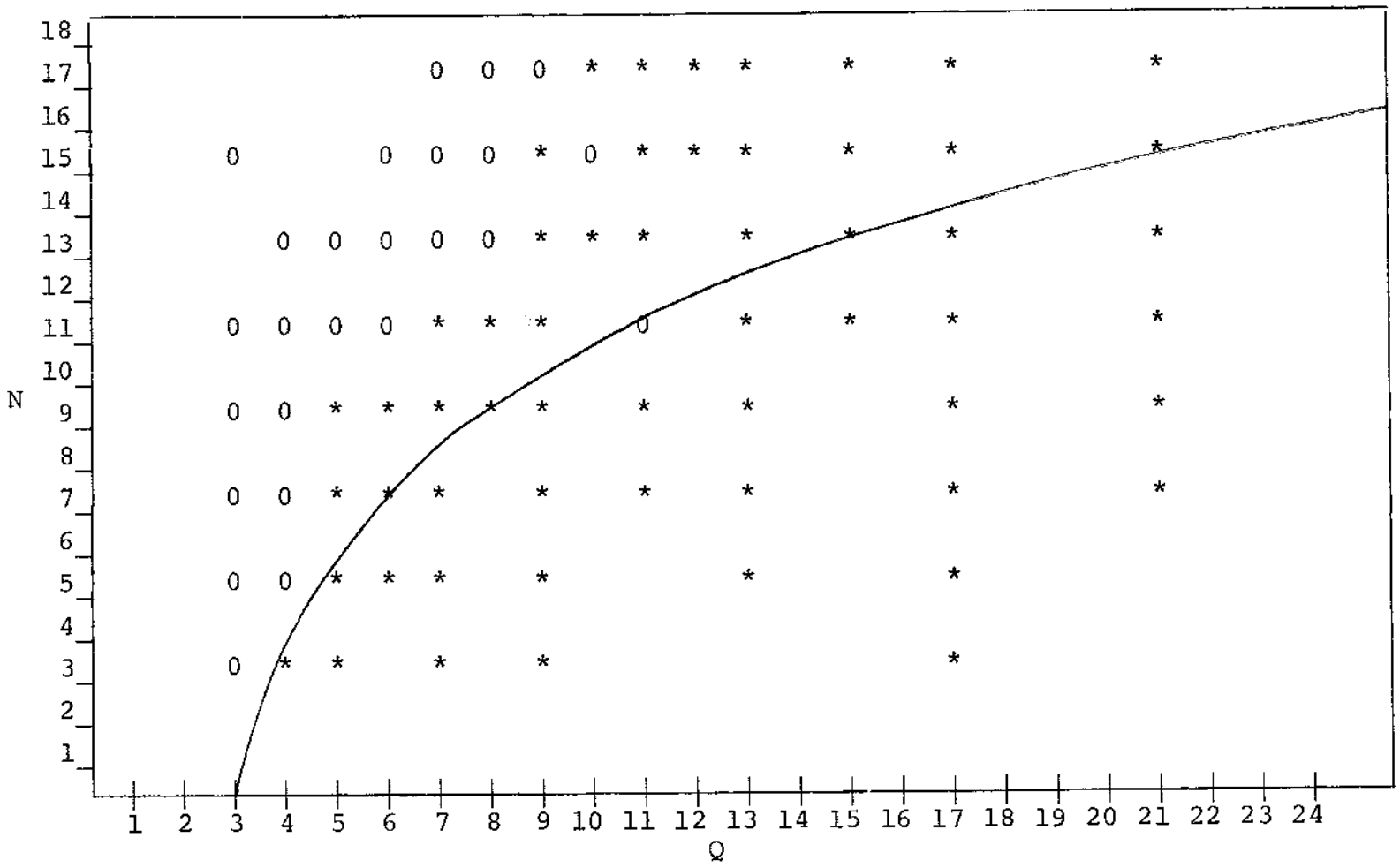
*Proportions of F-ratios in the actual F-distributions exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ which are significantly different from corresponding proportions of .010, .050, .100, and .200 at the .10 level of significance.

$\alpha' - \alpha$ deviation analysis resulted in a range of deviations from .004 through .107.

For the .200 α-level, the minimum α' obtained was .188. Again, the only instances where α' underestimated α was at the .200 α-level. This occurred in simulation models 292(9,9,2); 293(,9,94); 316(9,13,5), and 327 (9,15,3). The minimum α' of .188 was achieved in the 292(9,9,3) model. The largest α' corresponding to .200 was .277 from simulation model 273(9,3,17). The deviation range $\alpha' - \alpha$ was determined to be from -.012 through .077.

Graph 5 illustrates summary data relative to testing hypothesis III, part C. Simulation models with K = 9 which produced F-distributions in which the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ was significantly different from α for all $\alpha\varepsilon\{.010, .050, .100, .200\}$ are denoted by asterisks (*). Simulation models with K = 9 which did not meet the above mentioned criterion are denoted by zeros (0). The graph of $G_9 = \{(Q,n) : n = 5 \ln(Q - 2)\}$ is represented by a continuous solid line.

As can be seen in Graph 5, every simulation model with $Q > 11$ produced an F-distribution such that the proportion of F-ratios exceeding $_{1-\alpha}F_{(K-1)(N-K)}$ was significantly different from α for all $\alpha\varepsilon\{.010, .050, .100, .200\}$. Likewise, most simulation models not meeting this criterion



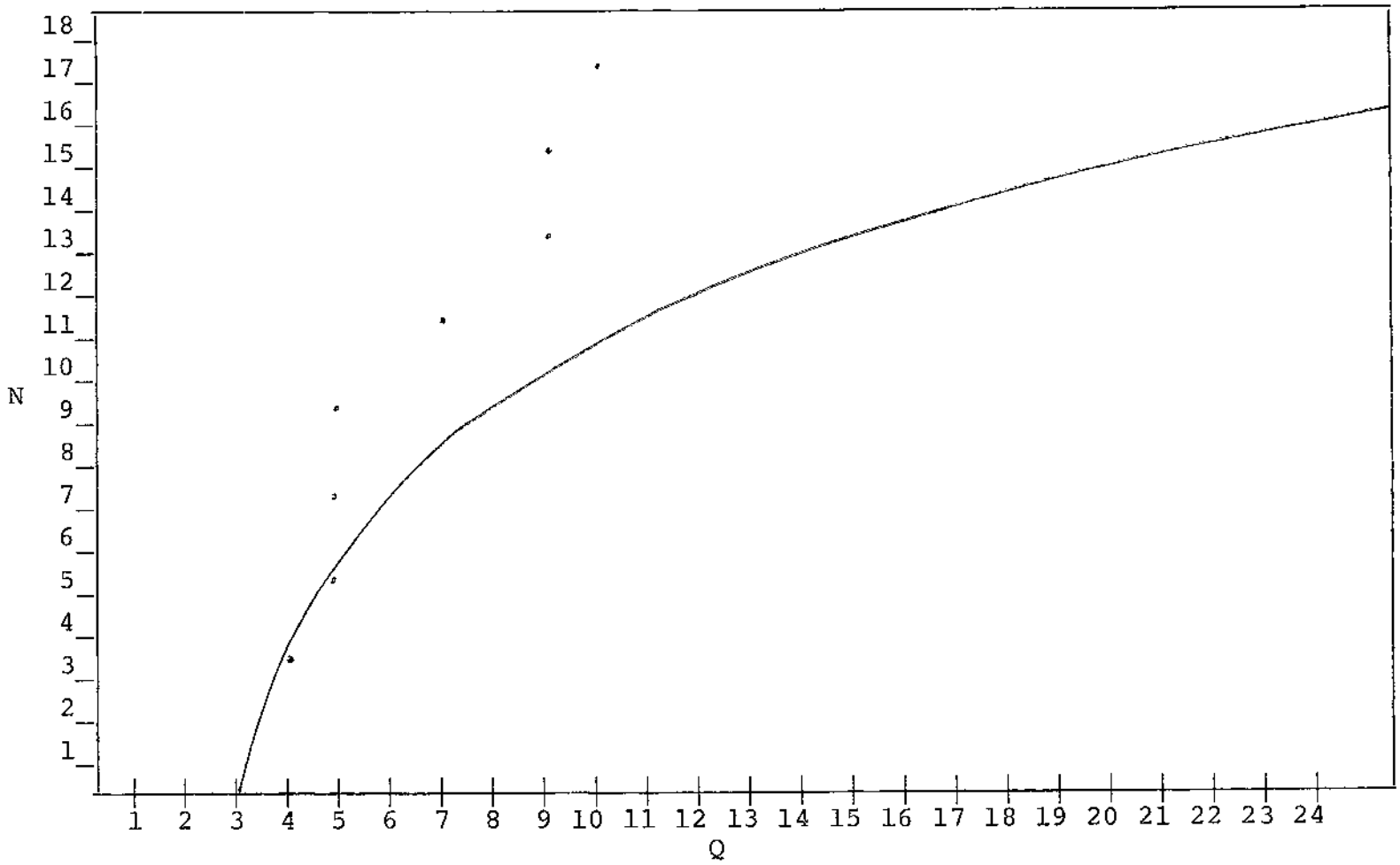
Graph 5--Distribution of significant and non-significant simulation models with nine treatment groups and hypothesized boundary locations, K = 9.

had less severe violation of the assumption of homogeneity of variance.

In order to test hypothesis III, part C, the following boundary points were determined: 269(9,3,4); 276(9,5,5); 284(9,7,5); 294(9,9,5); 307(9,11,7); 320(9,13,9); 331(9,15,9); 342(9,17,10), and 350(9,19,10). Graph 6 illustrates the location of each determined boundary point and the graph of the function G_9. When the L-distance from each boundary point to G_9 was determined, only boundary point 284(9,7,5) fell within an L-distance of one unit from G_9. Therefore, the assertion of a boundary region existing with reference to function G_9 was not supported. Hypothesis III, part C was rejected.

Table X summarizes deviation range data determined for all simulation models with $K \varepsilon \{5,7,9\}$. Inspection of the data revealed the following points. First, regardless of the number of groups, sample size, or extent of violation the proportion of F-ratios in the actual F-distribution exceeding $_{1-\alpha}F_{(K-1)(K-N)}$ for $\alpha \varepsilon \{.010, .050, .100\}$ exceeded α in every simulation. When compared to $\alpha = .200$, however, α' did, in some cases, underestimate α.

Second, the largest $\alpha' - \alpha$ maximum deviation occurred at the .050 α-level for $K = 7$ and $K = 9$ and at the .100 α-level for $K = 9$. When the length of the deviation range



Graph 6--Distribution of boundary points determined through simulations with nine treatment groups and hypothesized boundary location, K = 9.

was determined by assessing the numerical difference between each corresponding minimum and maximum, it was determined that the length of the deviation appeared to be, in part, a function of the K value. As the value of K progressed from 5 to 9, deviation range lengths increased all α'-levels accordingly.

TABLE X

MINIMUM AND MAXIMUM $\alpha' - \alpha$ DEVIATIONS FOR SPECIFIED α-LEVELS

	Number of Groups (K)					
	5		7		9	
α-levels of	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
.010	.006	.059	.002	.073	.004	.087
.050	.010	.085	.006	.103	.008	.111
.100	.006	.089	.003	.094	.004	.107
.200	-.010	.061	-.017	.066	-.012	.077

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The purpose of this study was to determine boundary conditions of several variables under which confidence in using the analysis of variance F-test is justified. In order to fulfill this purpose, a Monte Carlo simulation technique was employed to generate distributions of F-ratios under controlled violation of the homogeneity of variance assumption. The relationships among the variates sample size, number of treatment groups, significance levels, and the extent of violation of the assumption were investigated.

A total of 357 simulation models were studied. In each of the simulation models an empirical distribution of F-ratios was compared with the theoretical F-distribution. This was performed by a comparison of actual significance levels to the corresponding nominal significance levels of .01, .05, .10, and .20. In the initial sixty-four simulations the comparison between actual and nominal proportions was analyzed at the .10 level of significance for proportions based upon distributions of 500 F-ratios. In the remaining simulations, the comparisons were analyzed

at the .05 level of significance for proportions based upon distributions of 2,000 F-ratios.

In all simulation models generating an F-distribution of 2,000 F-ratios, parametric values for sample size, number of treatment groups, and extent of violation of the homogeneity of variance assumption were selected in a patterned manner in order to determine boundary conditions of these variables with respect to the robustness of the analysis of variance F-test.

Conclusions

On the basis of the synthesis and analysis of the data collected, the following conclusions were drawn. These conclusions were reached through both the original hypotheses directing the study and through trends which emerged in the data.

First, the analysis of variance F-test was found to be generally robust when the violation of the homogeneity of variances was of small magnitude, for example, when the extreme ratios of variances is less than 1:5, and the number of treatment groups is less than seven. This conclusion is consistent with much of the previous work in this area.

Second, when either the violation of the homogeneity of variances assumption was more pronounced or the number of treatment groups was greater than or equal to seven, serious discrepancies between actual and nominal significance levels

occurred. While Roy (2, p. 76) in the 1971 dissertation study did not specifically deal with large numbers of treatment groups, he did conclude that when the ratio of any pair of population variance exceeds 1:9, nonparametric statistics should be employed in lieu of the analysis of variance F-statistic. This conclusion is likewise supported by Spreckelmeyer (3, p. 99) who, in reporting his 1970 study, concluded that where moderate to extreme heterogeneity exists, other alternate tests were more stable than F in terms of type I error control. While few previous studies have investigated the effects of the number of treatment groups on robustness when the number of groups is greater than five, this conclusion is also consistent with the results of Box's 1954 study (1, p. 300).

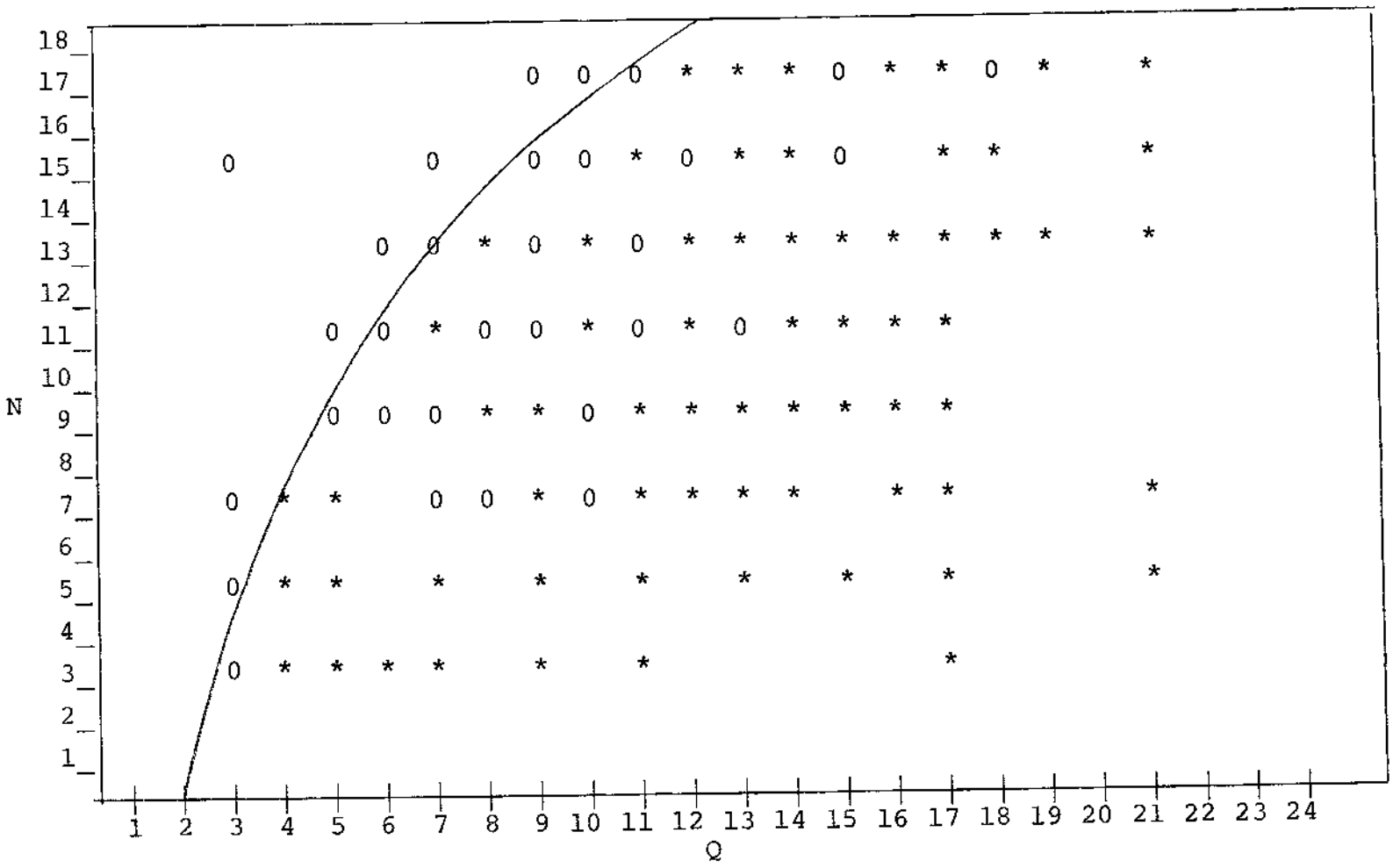
Third, for the simulation models with five treatment groups, it was concluded that the relative inconsistencies in the data would not allow definitive establishment of boundary conditions for sample size and extent of violation of the assumption under question. As was previously noted, no boundary patterns emerged for simulations employing five treatment groups. The data indicate that significant discrepancies between actual and nominal significance levels consistently occurred only when the extent of violation was severe. Moderate violation of the assumption produced a

pattern of results upon which no definitive conclusions could be drawn.

The fourth conclusion was based upon the analysis of data collected through simulations in which the number of treatment groups was seven. It was concluded that confidence in the application of the analysis of variance F-test with seven treatment groups was limited to values of Q (the extent of violation indicator and n (the common treatment group sample size) such that $n \geq 10 \ln \frac{1}{2}Q$.

In Graph 7 simulation models which produced F-distributions in which the proportion of F-ratios exceeding $1-\alpha^{F}(K-1)(N-K)$ for all $\alpha\varepsilon\{.010, .050, .100, .200\}$ are indicated by asterisks (*). Simulation models in which this criterion was not met are indicated by zeros (0). The graph of $\{(Q,n):n = 10 \ln \frac{1}{2}Q\}$ is represented by a continuous solid line. The region $\{(Q,n): n \geq 10 \ln \frac{1}{2}Q\}$ is that portion of the Qn-plane found above and to the left of the line in Graph 7.

When n was less than eight to ten subjects per treatment group, minor violation of the assumption of homogeneity of variances produced actual significance levels which markedly deviated from nominal levels. When n was greater than 10 or 12, the violation of the assumption had to be more pronounced in order to produce serious discrepancies between actual and nominal significance levels.



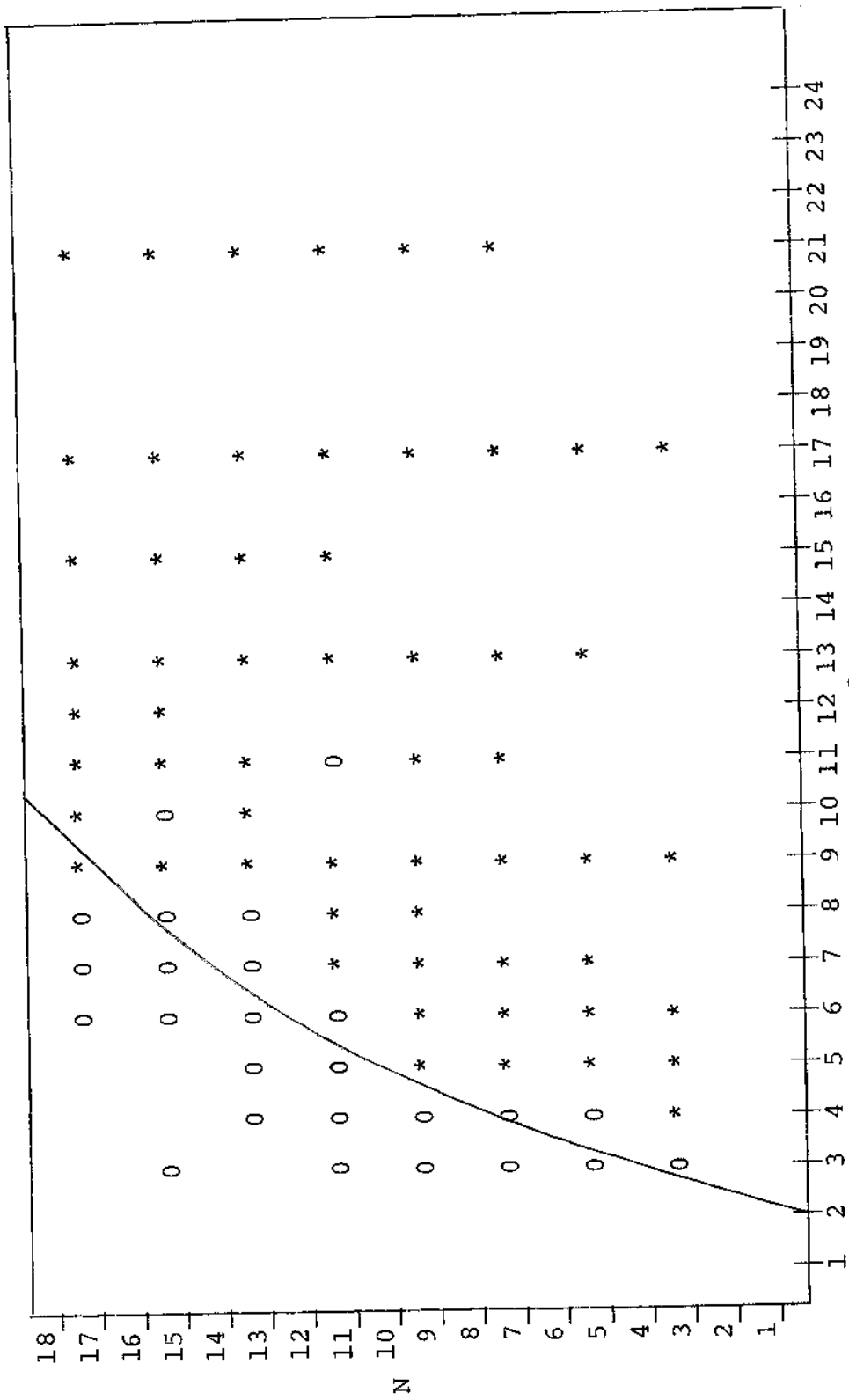
Graph 7--Distribution of significant and non-significant simulation models with seven treatment groups and recommended boundary location, K = 7.

The fifth conclusion was based upon the analysis of data collected through simulation in which the number of treatment groups was nine. In a pattern similar to the case with seven treatment groups, it was concluded that confidence in the application of the analysis of variance F-test with nine treatment groups was limited to values of Q and n such that $n \geq -\frac{3}{2} + 12 \ln \frac{1}{2}Q$.

In Graph 8 simulation models which produced F-distributions in which the proportion of F-ratios exceeding $1-\alpha^{F(K-1)(N-K)}$ for all $\alpha \varepsilon \{.010, .050, .100, .200\}$ are indicated by asterisks (*). Simulation models in which this criterion was not met are indicated by zeros (0). The graph of $\{(Q,n): n = -\frac{3}{2} + 12 \ln \frac{1}{2}Q\}$ is represented by a continuous solid line. The region $\{(Q,n): n \geq -\frac{3}{2} + 12 \ln \frac{1}{2}Q\}$ is that portion of the Qn-plane found above and to the left of the line in Graph 8. When n was less than eleven subjects per treatment group, even minor violations of the assumption of homogeneity of variances produced actual significance levels which markedly deviated from nominal levels.

Recommendations

The following recommendation was based on these conclusions. It was recommended that when the researcher is confronted with a research design with more than five treatment groups, the treatment group variances should be estimated and, if heterogeneity of variances exists to any



Graph 8--Distribution of significant and non-significant simulation models with nine treatment groups and recommended boundary location, K = 9.

real degree, the researcher should consider parametric and nonparametric alternatives to the analysis of variance F-test. Spreckelmeyer (3) in his 1970 study compared both control of type I error and power of various analysis of variance F-test alternatives.

Recommendations for further study are many, as throughout the investigation, numerous unanswered questions were posed. First, this study should be replicated and should be expanded to include wider ranges of parametric values for sample size, number of treatment groups, and extent of violation of the assumption of homogeneity of variances indicators. Also, any consideration of replication should include the selection of multi-variate analytical methods to be used in analyzing the data. The complex interactions between the variables require this type of analysis.

As this study only used violation of the homogeneity of variances assumption of the form 1:1:1:...:Q, it is recommended that the effect of differing patterns of unequal variance ratios on the robustness of the F-test be studied.

It was observed in this study that for each value of K, the number of treatment groups, there was a set of parametric combinations for n and Q which was relatively compact when plotted on the Qn-plane and, over which decisions

concerning confidence in the analysis of variance F-test could not be drawn. It is recommended that further study should be conducted in such a way that probability distributions for combinations of parameters in the region can be developed. It appeared quite likely that in cases with marginal true effects of violation of the assumption, the effects of sampling error restrict definitive boundary determination and, thus, require the development of probability distributions to be used in ascertaining confidence estimates in using the analysis of variance F-test when applied in the actual research environment.

CHAPTER BIBLIOGRAPHY

1. Box, G. E. P., "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-way Classification," Annals of Mathematical Statistics, XXV (1954), 290-302.

2. Roy, M. R., "An Investigation of the Robustness of the Simple Analysis of Variance F Test," unpublished doctoral dissertation, College of Education, University of Northern Colorado, Greeley, Colorado, 1971.

3. Spreckelmeyer, R. L., "The Effect of Unequal Sample Sizes and Variance Heterogeneity on Some K-sample Tests: An Empirical Investigation," unpublished doctoral dissertation, College of Education, University of Iowa, Iowa City, Iowa, 1970.

BIBLIOGRAPHY

Books

Glass, Gene V., and Stanley, Julian C., Statistical Methods in Education and Psychology, Englewood Cliffs, N.J., Prentice-Hall, 1971.

IBM System/360 Scientific Subroutine Package, Programmer's Manual, Version III, Program number 360A-CM-03X, 5th Edition, August, 1970.

Lehman, R., and Bailey, D. E., Digital Computing, Fortran IV and Its Application to the Behavioral Sciences, New York, John Wiley and Sons, Inc., 1966.

Lindquist, E. F., Design and Analysis of Experiments in Psychology and Education, Boston, Houghton-Mifflin Company, 1953.

Naylor, Thomas H., and others, Computer Simulation Techniques, New York, John Wiley and Sons, Inc., 1966.

NTSU Subroutine Library User's Guide, Version I, September, 1971.

Roscoe, J. T., Fundamental Research Statistics for the Behavioral Sciences, New York, Holt, Rinehart and Winston, Inc., 1969.

Scheffé, H., The Analysis of Variance, New York, John Wiley & Sons, 1959.

Articles

Behrens, W. U., "The Comparison of Means of Independent Normal Distributions with Different Variances," Biometrica, XX (1964), 16-27.

Boneau, C. A., "The Effects of Violation of Assumptions Underlying the t-test," Psychological Bulletin, LVII (1960), 49-64.

Box, G. E. P., "Non-Normality and Tests of Variance," Biometrica, XL (1953), 318-335.

_____, "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-way Classification," Annals of Mathematical Statistics, XXV (1954), 290-302.

David, F. N., and Johnson, N. L., "The Effect of Non-Normality on the Power Function of the F Test in the Analysis of Variance," Biometrica, XXXVIII (1951), 43-57.

Glass, G. V., Peckham, P. D., and Sanders, J. R., "Consequences of Failure to Meet Assumptions Underlying the Analysis of Variance and Covariance," Review of Educational Research, XLII (Summer, 1972), 37-288.

Godard, R. H., and Lindquist, E. F., "An Empirical Study of the Effect of Heterogeneous Within-groups Variance upon Certain F-tests of Significance in Analysis of Variance," Psychometrika, V (December, 1940), 263-273.

Gronow, D. G. C., "Test for the Significance of the Difference Between Means in Two Normal Populations Having Unequal Variance," Biometrica, XXXVIII (1951), 252-256.

Horsnell, G., "The Effects of Unequal Group Variances on the F Test for the Homogeneity of Group Means," Biometrica, XL (1953), 128-136.

Hsu, P. L., "Contributions to the Theory of Student's T Test as Applied to the Problem of Two Samples," Statistical Research Memoirs, II (1938), 1-24.

Murphey, B. P., "Some Two-Sample Tests When the Variances Are Unequal: A Simulation Study," Biometrica, LIV (1967), 676-683.

Pearson, E. S., "Analysis of Variance in Cases of Non-Normal Variation," Biometrica, XXIIII (1931), 114-133.

_____, "Some Notes on Sampling Tests with Two Variables," Biometrica, XXI (1929), 337-360.

Pratt, J. W., "Robustness of Some Procedures for the Two-Sample Location Problem," Journal of the American Statistical Association, LIX (1964), 665-680.

Public Documents

Bradley, J. V., "Studies in Research Methodology VI. The Central Limit Effect for a Variety of Populations and the Robustness of the z, t, or F," USAF Aerospace Medical Division Report AMAL-TR-64-123, 1964.

Unpublished Materials

Lang, M. S., "Empirical Comparisons of the Treatments by Levels in Analysis of Variance and Analysis of Covariances," unpublished doctoral dissertation, College of Education, Department of Research and Statistical Methodology, University of Northern Colorado, Greeley, Colorado, 1972.

McClaran, V. R., "An Investigation of the Effect of Violating the Assumption of Homogeneity of Regression Slopes of the Analysis of Covariance upon the F-statistic," unpublished doctoral dissertation, College of Education, North Texas State University, Denton, Texas, 1972.

Norton, D. W., "An Empirical Investigation of the Effects of Non-Normality and Heterogeneity Upon the F Test of Analysis of Variance," unpublished doctoral dissertation, State University of Iowa, Iowa City, Iowa, 1952.

Roy, M. R., "An Investigation of the Robustness of the Simple Analysis of Variance F Test," unpublished doctoral dissertation, College of Education, University of Northern Colorado, Greeley, Colorado, 1971.

Spreckelmeyer, R. L., "The Effect of Unequal Sample Sizes and Variance Heterogeneity on Some K-Sample Tests: An Empirical Investigation," unpublished doctoral dissertation, College of Education, University of Iowa, Iowa City, Iowa, 1970.