SHORT TITLE

JOINT EFFECTS OF UNEQUAL SLOPES AND VARIANCES

IN ANCOVA

THE JOINT EFFECTS OF SIMULTANEOUSLY VIOLATING THE HOMOGENEITY OF REGRESSION AND HOMOGENEITY OF VARIANCE ASSUMPTIONS ON THE F-TEST IN THE ANALYSIS OF COVARIANCE - A MONTE CARLO SIMULATION

by

R. Lorcan Scanlon

A Thesis Submitted to the Faculty of Graduate Studies and Research in Partial Fulfilment of the requirements for the Degree of Doctor of Philosophy

Department of Psychology
McGill University
Montreal

October 1972

Ph.D. Psychology

THE JOINT EFFECTS OF SIMULTANEOUSLY VIOLATING THE HOMOGENEITY OF REGRESSION AND HOMOGENEITY OF VARIANCE ASSUMPTIONS ON THE F-TEST IN THE ANALYSIS OF COVARIANCE - A MONTE CARLO SIMULATION

R. Lorcan Scanlon

ABSTRACT

The present study was designed as an attempt to evaluate the effects on the Analysis of Covariance F-test of varying combinations of degrees of violation of the homogeneity of variance and the homogeneity of regression assumptions. Measures of violation, invariant under stated constraints on the covariate, were derived for each of the two assumptions. Sampling distributions of simulated ANCOVA experiments embodying combinations of values of these measures were generated. The effect of each combination on the F-test was determined by a comparison of obtained and theoretical percentage points. Results suggested that the two violations tend to neutralize each other in their effects, leaving the F-test remarkably robust with respect to their joint presence. An attempt was made to establish a predictive relationship between levels of violation of the two assumptions on the one hand, and their effect on probability of Type 1 error on the other.

PREFACE

This investigation was supported by a grant from the National Research Council of Canada (Grant No. 246-07) to Dr. James O. Ramsay.

The contribution of the study is to show how the violation of the homogeneity of regression and the homogeneity of variance assumptions in the analysis of covariance, tend to cancel each other out in their effects on the F-test, leaving the test remarkably robust with respect to their joint presence.

I wish to thank my supervisor Dr. James O. Ramsay for his continued encouragement and helpful direction throughout the study.

I am also indebted to Mr. Nick Paul and Dr. Kasi Ananthanarayanan both of the School of Computer Science, McGill University, for their very willing advice on matters related to the random number generator.

Finally I wish to express my thanks to my wife Susan for help in preparing the text.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The application of statistical tests to the analysis and interpretation of data implies that the assumed conditions of the test are met in the experimental situation. In practice, however, these assumptions are never fully justified. Thus, an important area of investigation has necessarily arisen to answer questions concerning the extent to which actual experimental conditions can depart from those assumed in any test before its application is rendered invalid. In other words, this field of research attempts to determine how 'robust' any given test is. If a statistical test is not robust, and if, in actual experimental practice, its underlying assumptions are not upheld, then it becomes extremely difficult for the researcher to isolate that component of his results which is attributable to the violation of the assumptions, from that part which purports to answer his experimental question.

Two ways of approaching questions concerning the robustness of statistical tests have been developed. The first employs traditional mathematical analysis and attempts to determine the behaviour of the test under less than ideal conditions. A difficulty with this analytic approach is that once the assumptions of the test are relaxed the statistics frequently become difficult to handle, and in many cases the statistician has to resort to distributional approximations (often asymptotic) or other simplifying procedures, which yield results that are more indicative than definite.

The second approach is that of Monte Carlo simulation in which the robustness of a statistical test is evaluated by generating an empirical sampling distribution of the statistic, derived from the simulated sets of data embodying some specified degree of violation of one or more of the test's underlying assumptions . This distribution can then be compared with its corresponding theoretical,violation-free distribution and the effect of that degree of violation is determined from the comparison.

This kind of work is much more costly than the first approach, and harbours some intrinsic limitations; whereas there is only one way in which the assumptions can be fully satisfied, there is no practical limit to the number of ways in which, singly or in combination, they can be violated. A Monte Carlo investigator must consequently choose a more or less arbitrarily circumscribed set of possibilities to study, without any guarantee that his findings are generalizable to other contingencies of violation.

The present study was designed as an attempt to ascertain, through the use of the Monte Carlo procedure, the effects of the simultaneous violation of two assumptions underlying a statistical test commonly applied to experimental data in Educational and Psychological research - the Analysis of Covariance. The two subjects of this investigation were the homogeneity of regression and the homogeneity of variance assumptions.

THE ANALYSIS OF COVARIANCE:

The Model and Its Application

Covariance analysis (ANCOVA) incorporates two systems of data analysis, Analysis of Variance and Regression Analysis, which have come to be thought of by many psychologists as distinct, but which in fact are both subsumed under the general linear model.

The mathematical model in the Analysis of Covariance (one-way, linear, fixed effects) is as follows:

$$Y_{ij} = \mu + \tau_j + \beta (X_{ij} - \overline{X}..) + e_{ij} \quad .....(1.1)$$

where Y_{ij} is the measure of the dependent variable and X_{ij} is a concomitant variable or covariatee with grand mean $\overline{X}..$, on which Y_{ij} has a linear regression with regression coefficient β. The constants μ and τ_j are the grand mean of treatment populations and the effect of the jth treatment level, respectively, with $\sum_j \tau_j = 0$. The variable e_{ij} is the random error term assumed to be normally and independently distributed with zero mean and constant variance.

From the point of view of Analysis of Variance, the model can be expressed as:

$$Y_{ij} - \beta (X_{ij} - \overline{X}..) = \mu + \tau_j + e_{ij}, \quad .....(1.2)$$

where a modification (due to the estimation of β) of the usual F-test is applied to the Y-scores now adjusted for X.

On the other hand, considered from the point of view of Regression Analysis, the treatment effects can be conceptualized as dummy variables, so that the design takes on the form of an F-test for multiple regression. These two F-tests are equivalent.

The Covariance model is designed for the following experimental situation: different levels of a treatment, or independent variable, are being applied to randomly selected groups of experimental units. The purpose of the experiment is to determine whether or not these treatment levels differentially affect some dependent response Y. Before the treatments are applied, however, a measure of another variable X is taken on the experimental units. The treatment levels are then effected, and in each instance the measure of variable Y is recorded. In the usual Analysis of Variance situation, these data are then subjected to an F-test for differences in treatment means. In the Analysis of Covariance situation, however, before the Y data are subjected to the F-test, they are first adjusted to remove the influence of the variable X.

This procedure has the advantage of increasing the precision of the treatment comparisons by reducing extraneous variability in the experiment. Cochran (1957) reports that the gain in precision resulting from covariance adjustment is a function of the size of the correlation coefficient ρ between Y and X on experimental units that receive the same treatment. He states that "... if σ_y^2 is the experimental error variance when no covariance is employed, the adjustments reduce this variance to a value which is effectively about

$$\sigma_y^2 (1 - \rho^2)(1 + \frac{1}{f_e - 2}), \qquad (1.3)$$

where f_e is the error number of degrees of freedom." (Cochran, 1957, p. 262).

Assumptions Underlying the Analysis of Covariance

The F-test for the Covariance Analysis, like all parametric tests, assumes for its valid application that the data to which the model is applied behave in a certain way. The following are the basic assumptions of Covariance Analysis.

1. The experimental units are randomly assigned to treatment groups.
2. The dependent or criterion scores have a linear regression on the covariate and the regression coefficient is constant across treatment levels.
3. The covariate is measured without error.
4. The dependent or criterion scores are a linear combination of independent components - an overall mean, a treatment effect, a linear regression on X, and an error term.

5. For each treatment/covariate combination, the error term e_{ij} is independently and normally distributed with a mean of zero and constant variance.

Finally, it is usual in applying the test to a set of data to make the assumption that $\tau_j = 0$ for all treatment groups (the null hypothesis).

If any of these six conditions is not satisfied, the sampling distribution of the F-ratio may differ from the central F-distribution. This means that a significant, or for that matter, a non-significant F-ratio could result from a failure to fulfill any one of these assumptions. Consequently, before concluding from a significant or non-significant F that the sixth assumption is or is not sustained, one must be able to satisfy oneself that failure to meet the other five assumptions is not seriously affecting the behaviour of the F-ratio distribution.

Effects of Departures from the Underlying Assumptions in ANCOVA - A Selected Review

Some of the studies to be cited in this section pertain directly to the ANCOVA design; others focus primarily on ANOVA, but their results are in most cases directly generalizable to the ANCOVA situation. This review consists of a brief summary of the effects of departures from each of several underlying assumptions not directly connected to the subject of the present study, and of a more elaborate documentation of the results of studies dealing with those directly related assumptions - homogeneity of regression and homogeneity of variance.

The consequences of violating the assumption of random assignment of experimental units to treatment groups have been clearly stated by Lord. He points out that "If the individuals are not assigned to treatments at random, then it is not helpful to demonstrate statistically that the groups after treatment show more difference than would be expected by random assignment." (Lord, 1967, p. 305). The ANCOVA test is particularly open to violation of this assumption since in many empirical situations in psychology to which the test is applied, randomization is not feasible.

Evans and Anastasio (1968) distinguish in this context, three different uses of ANCOVA:

1. Random assignment of experimental units to groups and random assignment of treatments to groups;
2. Already-existing groups used as treatment groups, but treatments randomly assigned to them;
3. Already-existing groups used as treatment groups, with some intrinsic attribute considered as "treatment".

They conclude that only usage 1 leaves interpretation of results unequivocal, and that interpretation becomes less and less meaningful as one goes from usage 2 to usage 3.

As to the assumptions related to properties of the concomcomitant variable, Lord, (1960) examined the situation where the covariate X contains errors of measurement, and concluded that under these circumstances, "... the usual ANCOVA fails to adjust adequately for initial differences between groups." (Lord, 1960, p. 307). He constructed a large-sample significance test in an attempt to deal with the problem, but limited its usefulness in many empirical situations by requiring two sets of measures for the covariate.

Cochran (1968) shows how the situation of X measured with error decreases the precision of the experiment by increasing the error variance by a factor determined by the reliability of X. Porter (1967) developed a covariance design for the situation where X is measured with error and incorporates the reliability of X into his model.

Atiqullah (1964) has examined the effect of non-normality on the ANCOVA F-test. He demonstrates how in the balanced lay-out the sensitivity of the test to non-normality depends on the behaviour of the covariate X. He concludes that the test is robust to non-normality when the distribution of the concomitant variable is normal.

Atiqullah measures non-normality by kurtosis. He does not deal with skewness, which as Peckham (1968) remarks, is also relevant to non-normality in the behavioural sciences.

Scheffé (1959) has examined the effect of serial correlation (violating the assumption of independence of errors) for large samples in ANOVA. He derives probabilities corresponding to a 95% confidence interval for various values of ρ and concludes that "... the effect of serial correlation on inferences about means can be serious" (Scheffé, 1959, p. 338).

Atiqullah (1964) has considered the situation where the usual ANCOVA (linear) test is applied to data which contain a quadratic component of regression. In the case of two treatments, he finds that the expected value of the adjusted difference between the two groups is unbiased only if the covariates in both groups are members of the same normal population. Even this does not hold, however, for the case of more than two groups. He concludes that the presence of a quadratic component, if large, may have serious effects on the ANCOVA F-test (Atiqullah, 1964, p. 372).

Homogeneity of Regression

Until quite recently, very little information was available on the effects of violating the assumption of homogeneity of regression in ANCOVA. The assumption implies that the relationship between Y and X is a linear one, and that the regression coefficient of Y on X is constant across treatment groups.

A preliminary check on the assumption of homogeneity of regression is given by Winer (1971). He provides a sampling F-distribution based on the assumption that $\beta_1 = \beta_2 \cdots\cdot = \beta_K$. This is a useful though indirect precaution, but as Evans and Anastasio point out, "... the decision that assumptions have been met rests on the acceptance of the null hypothesis. Thus the user has only a roundabout procedure (using a large value for α) to guard against the relevant class of error, Type II. He cannot even determine, much less control, the probability of detecting violations which are serious enough to affect his conclusions" (Evans and Anastasio, 1968, p. 226).

The two sources of information on the robustness of the test with respect to this assumption are those by Atiqullah (1964) and Peckham (1968). The former is a theoretical paper while the latter consists of a Monte Carlo empirical investigation.

Atiqullah sets up the following two models:

$$M_1 \quad Y_{ij} = \mu + \tau_j + \beta(X_{ij} - \bar{X}_{..}) + e_{ij} \qquad (1.4a)$$

$$M_2 \quad Y_{ij} = \mu + \tau_j + \beta_j(X_{ij} - \bar{X}_{..}) + e_{ij} \qquad (1.4b)$$

and examines the effect of employing the F-test based upon model M_1, when in fact model M_2 is the appropriate one. He makes use of the following notation:

$$W_{jj} = \Sigma (X_{ij} - \overline{X}_{\cdot j})^2; \quad W_2 = \Sigma W_{jj}$$

His contribution may be divided into three parts.

1. In the case of a comparison of two treatments in a two-group experiment, he found that when $\beta_1 \neq \beta_2$, that is when model M_2 is the appropriate one, but model M_1 is applied, the expected value of the difference in adjusted treatment means is biased unless either $\overline{X}_1 = \overline{X}_2$, or $W_{11} = W_{22}$. He suggests that "... in the absence of a prior presumption that B_1 and B_2 are nearly equal, the model M_2 should be used, separate regressions fitted, and the treatment effects estimated as a function of X.
2. Atiqullah next considers a comparison of two treatments in an experiment involving more than two groups. Again, in the case of model M_2 being the appropriate one, but model M_1 being applied, he finds that the expected value of the adjusted differences between pairs of treatments is biased unless both $\overline{X}_1 = \overline{X}_2$ <u>and</u> $W_{11} = W_{22}$.
3. In the case of experiments involving more than two treatment groups, Atiqullah derived an asymptotic approximation to the ANCOVA F-distribution using Fisher's Z-transformation, and then examined the consequences of applying model M_1 when model M_2 obtains. He concluded that the application of the standard covariance F-test, under these circumstances, may yield misleading results

unless the X's are normally distributed, $\sigma_x^2/\sigma_e^2 > 1$, and the variance of the inhomogeneous regression coefficients $\sigma_{\beta_j}^2$ is in the order of $\frac{1}{K^2}$ where K denotes number of treatment groups.

Peckham (1968) conducted a Monte Carlo investigation into the effects of violating the homogeneity of regression assumption in ANCOVA. His method consisted of generating a sampling distribution of the ANCOVA F-statistic, where each sample consisted of an ANCOVA experiment embodying a specified degree of heterogeneity of regression, all other assumptions of the test being satisfied. (To avoid bias due to non-normality in the concomitant variable, the values of X were chosen to approximate a normal distribution). He then compared the empirical sampling distribution with its corresponding central F-distribution.

The study proceeded in two phases. In phase one, Peckham combined varying degrees of heterogeneity of regression with different sample sizes and numbers of treatment groups. In this phase, he fixed $\overline{X}_j = 0$ and $MS_{x_j} = 1$ for all groups. In phase two, he used only the two-group case and arranged the data so that the relationships among the Y, X and β were such that the expectation of the adjusted mean for each group was equal to the same constant.

In both phases, Peckham found for the degree of violation studied that the analysis was not seriously affected by departures from the assumption, and that as the degree of heterogeneity increased, the test became more conservative with respect to Type I error (see Table 1.1).

TABLE 1.1

Effects of Violating the Assumption of Homogeneity of Regression on the ANCOVA (1-way, fixed-effects) F-test (Abridged from Peckham, 1968, Phase 1).

No. of Groups	Group Size	Regression Coefficient Values	Probability of exceeding 5% pt.	1% pt.
2	10	.4 .6	.050	.010
2	10	.1 .9	.029	.004
3	10	.4 .5 .6	.050	.009
3	10	.1 .5 .9	.035	.006
5	10	.4 .4 .5 .6 .6	.048	.007
5	10	.1 .3 .5 .7 .9	.041	.010

Homogeneity of Variance

ANCOVA also relies for its validity on the assumption that the population variance σ^2 of the error term e_{ij} for each treatment-covariate combination is a constant. As Elashoff (1969) points out, there are two main ways in which this assumption is likely to be violated:

1. The variance σ^2 is dependent on the covariate X, but for a given X is constant across treatments,
2. σ^2 is constant within each treatment group, but differs across treatments.

For the purposes of the present study, attention will be focussed exclusively on the second contingency, since it better represents the empirical situation to which ANCOVA is usually applied.

To date, no theoretical or empirical work has been specifically devoted to studying the effects on the ANCOVA F-test of departures from the assumption of equal variances. However, a good deal of both kinds of work has been carried out on the ANOVA design, and since the effect of unequal variances on the two models is almost identical under certain conditions of the covariate (see below page 26), these ANOVA results will be reviewed as relevant in the context of the present study.

Since the 1930's, statisticians have considered the effects of departures from the assumption of homogeneous variances in ANOVA (Welch, 1937; Daniels, 1938; Horsnell, 1953). The results of these studies suggested that the test is robust with respect to the violation of this assumption. Several tests for homogeneity of variance have been developed (Bartlett, 1937; Cochran 1941; Hartley, 1950), but these tests suffer from the same drawbacks as those outlined above in reference to the test for homogeneity of regression (see section immediately preceding).

The first important Monte Carlo study on the issue was carried out by Norton (1952). As reported in Lindquist (1953, p. 78), Norton obtained samples of ANOVA F's by randomly sampling from 3 card populations, which were normally distributed with constant mean, but whose variances were 25, 100 and 225 respectively. He obtained the percentage of sample F's that exceeded the theoretical 5% and 1% values (expected under the null hypothesis), and used the discrepancy between the expected and obtained percentages as a measure of the effects of the violation. His results indicated that the ANOVA F-test is remarkably insensitive to violation of the assumption of equal variances, when equal numbers are assigned to treatment groups (see Table 1.2).

TABLE 1.2

Norton Study - Percentage Counts of Mean-Square Ratios in Empirical Distributions Exceeding Theoretical Percentage Levels in Normal-Theory F-distribution. (Abridged from Lindquist 1953, p. 84).

		Percentages	
No. of Groups	Group Size	Expected	Obtained
3	3	5%	7.26%
3	10	5%	6.56%
3	3	1%	2.13%
3	10	1%	2.00%

Box (1954), using theorems on quadratic forms in multinormally distributed variables, developed a very close approximation to the distribution of ANOVA F (one-way), when the assumption of equal variances is relaxed. Utilizing this approximation he calculated the probability of Type 1 error corresponding to the .05 level of normal theory, for various levels of violation. His results, in the case of equal groups, support those of Norton. In the case of unequal groups however, he showed that the violation of the homogeneity of variance assumption is drastically disturbing to the distribution of F.(see Table 1.3). Both the results of Norton and Box were later empirically supported by those of Boneau (1960).

TABLE 1.3

Effects of Violating the Assumption of Homogeneity of Variance on the ANOVA (1-way) F-test (Abridged from Box 1954, p.299).

No. of Groups	Variances	Group Sizes	Prob. (%) of exceeding 5% point
3	1:2:3	5:5:5	5.78
3	1:2:3	7:5:3	9.57
3	1:1:3	5:5:5	5.82
3	1:1:3	7:5:3	9.78
5	1:1:1:1:3	5:5:5:5:5	6.86
5	1:1:1:1:3	9:5:5:5:1	15.56

ORIGIN OF THE PRESENT STUDY

The preceding review has considered only the effects on ANCOVA and ANOVA of violating each single underlying assumption in turn. This limitation reflects the paucity of existing work evaluating the joint effects of violating two or more of the basic assumptions in ANCOVA. The use of these results as a guide to data analysis would seem to depend upon the validity of the assumption that, if in any set of data to which ANCOVA is being applied more than one assumption is being violated (as is most likely to be the case), the effects of the violations are independent. That is, the violation of one assumption does not ameliorate or exacerbate the effect of another assumption's being violated.

The present study arose out of indications that such a state of affairs does not hold for at least two of the assumptions underlying the ANCOVA F-test. Peckham (1968) found that, as the violation of the homogeneity of regression assumption increased in severity, fewer and fewer simulated ANCOVA F-values exceeded the theoretical percentage levels; that is, the test became more and more conservative with regard to Type 1 error (see Table 1.1). On the other hand, generalizing from the results of Norton and Box (see Tables 1.2 and 1.3), it seems likely that in ANCOVA, the number of F-values exceeding nominal probability levels increases as the degree of violation of the homogeneity of variance assumption increases.

The present study was therefore an attempt to evaluate the effects on the ANCOVA F-test (1-way, linear, fixed effects, equal group sizes) of simultaneously violating the homogeneity of variance and the homogeneity of regression assumptions: to ascertain in general, the extent to which these violations in their varying degrees of severity, tend to cancel each other out, and to determine the general relationship between cumulative percentage counts and violation levels of the two assumptions.

CHAPTER 2

DESIGN OF THE STUDY

The procedure was first to choose the combinations of degrees of violation of the two assumptions to be investigated; then for each combination, to construct a pseudo-random sampling distribution of F-ratios, derived from a series of simulated ANCOVA experiments embodying the degrees of violation of that combination. For each sampling distribution a count was then taken of the proportion of those F-values exceeding the critical tabled F-values for the corresponding central F-distribution (df_1 = K-1, df_2 = N-K-1) at the 0.05 and 0.01 levels of significance. An estimate of the joint effects of each combination of violation-levels of the two assumptions on the probability of Type 1 error was thus determined.

MEASURES OF VIOLATION OF THE TWO ASSUMPTIONS

In order to proceed systematically, it was necessary at the outset to define an index of violation for each of the two assumptions; that is, to determine a quantitative relationship among the K parametric values (beta's, variances) such that as it increased in magnitude, so also did the effect on the probability of Type 1 error.

I - For the purpose of deriving a measure of violation for the homogeneity of regression assumption, the following models of ANCOVA were compared. (The models are simplified so as to ensure full rank in matrices that are to be inverted).

$$L_1:\ Y_{ij} = \tau_j + \beta X_{ij} + e_{ij} \qquad \begin{matrix} j = 1,2\ldots .k \\ i = 1,2\ldots .n \end{matrix} \qquad (2.1a)$$

$$L_2:\ Y_{ij} = \tau_j + \bar{\beta} X_{ij} + (\beta_j - \bar{\beta}) X_{ij} + e_{ij}, \qquad (2.1b)$$

where L_1 is the usual model for ANCOVA and L_2 incorporates a set of parameters $(\beta_j - \bar{\beta})$ to accommodate the presence of inhomogeneous regression coefficients.

Under L_2, the design matrix X, the coefficient vector β, and the error vector E are represented as follows:

$$X = \begin{bmatrix} 1 & 0 & .. & 0 & X_{11} & X_{11} & 0 & .. & 0 \\ 1 & 0 & .. & 0 & X_{21} & X_{21} & 0 & .. & 0 \\ .. & .. & .. & .. & .. & .. & .. & .. & .. \\ 1 & 0 & .. & 0 & X_{n1} & X_{n1} & 0 & .. & 0 \\ - & - & - & - & - & - & - & - & - \\ 0 & 1 & .. & 0 & X_{12} & 0 & X_{12} & .. & 0 \\ 0 & 1 & .. & 0 & X_{22} & 0 & X_{22} & .. & 0 \\ .. & .. & .. & .. & .. & .. & .. & .. & .. \\ 0 & 1 & .. & 0 & X_{n2} & 0 & X_{n2} & .. & 0 \\ - & - & - & - & - & - & - & - & - \\ .. & .. & .. & .. & .. & .. & .. & .. & .. \\ - & - & - & - & - & - & - & - & - \\ 0 & 0 & .. & 1 & X_{1k} & 0 & 0 & .. & X_{1k} \\ 0 & 0 & .. & 1 & X_{2k} & 0 & 0 & .. & X_{2k} \\ .. & .. & .. & .. & .. & .. & .. & .. & .. \\ 0 & 0 & .. & 1 & X_{nk} & 0 & 0 & .. & X_{nk} \end{bmatrix} \qquad \beta = \begin{bmatrix} \tau_1 \\ \tau_2 \\ .. \\ \tau_k \\ \bar{\beta} \\ \beta_1 - \bar{\beta} \\ \beta_2 - \bar{\beta} \\ .. \\ \beta_k - \bar{\beta} \end{bmatrix} \qquad E = \begin{bmatrix} e_{11} \\ e_{21} \\ .. \\ e_{n1} \\ --- \\ e_{12} \\ e_{22} \\ \ldots \\ e_{n2} \\ --- \\ .. \\ --- \\ e_{1k} \\ e_{2k} \\ .. \\ e_{nk} \end{bmatrix},$$

where

$$Y = X\beta + E \quad \text{and} \quad E(Y) = X\beta \tag{2.2}$$

Now, let Model L_1 be (incorrectly) applied and the following quantities calculated:

1. error sum of squares
2. sum of squares due to treatments

Let S denote the matrix containing the first $(k + 1)$ columns of X

c denote the vector containing the first $(k + 1)$ elements of β

T denote the vector containing the $(k + 1)$th. column of X

d denote the scalar containing the $(k + 1)$th. element of β

$$\begin{aligned} \text{Let } U &= Y - S\hat{c} \text{ be a vector of residuals.} \\ &= Y - S\left[(S'S)^{-1} S'Y\right] \\ &= \left[I - S(S'S)^{-1}S'\right] Y \end{aligned}$$

$$\text{Let } Q = \left[I - S(S'S)^{-1}S'\right] \tag{2.3}$$

Q is a symmetric idempotent matrix since

$$Q' = I - S(S'S)^{-1}S' \tag{2.4a}$$

$$\begin{aligned} Q^2 &= \left[I - S(S'S)^{-1}S'\right]\left[I - S(S'S)^{-1}S'\right] \\ &= I - 2S(S'S)^{-1}S' + S(S'S)^{-1}S'S(S'S)^{-1}S' \\ &= I - S(S'S)^{-1}S' \\ &= Q \end{aligned} \tag{2.4b}$$

Thus the error sum of squares is:

$$U'U = (QY)'\, QY = Y'QY \tag{2.5}$$

Let $W = Y - T\hat{d}$ be a vector of residuals

$$= \left[I - T(T'T)^{-1}\, T'\right] Y$$

where $R = I - T(T'T)^{-1}T'$ is a symmetric idempotent matrix (using 2.4a, and 2.4b). (2.6)

Thus treatment sum of squares $= Y'RY - Y'QY$ (2.7a)

$$= Y'(R - Q)Y \tag{2.7b}$$

where $(R - Q)$ is also a symmetric idempotent matrix, (2.8) (for proof see Appendix A).

Further $\left[(R - Q)Y\right]'QY = Y'(R - Q)'QY = 0$ (2.9)

(for proof see Appendix A).

Thus the two vectors are orthogonal and hence any two quantities based on them respectively will be independently distributed.

A necessary and sufficient condition that $Y'AY$ is distributed as a chi-square is that A is idempotent; the degrees of freedom of such a chi-square are equal to the rank of A, and its non-centrality parameter, λ, has the value of the quadratic form $Y'AY$ when the variables have been substituted by their expected values. (Rao, 1965, p. 150).

Thus, using (2.4a; 2.4b; 2.2), $Y'QY$ the error sum of squares, is distributed as a non-central chi-square with degrees of freedom equal to the rank of Q, and non-centrality parameter = $\beta'X'QX\beta$. So, too, $Y'(R-Q)Y$ the treatment sum of squares, is also distributed as a chi-square with degrees of freedom equal to the rank of (R-Q), and non-centrality parameter = $\beta'X'(R-Q)X\beta$, (using 2.8; 2.2). Furthermore, the two chi-square distributions are independent (using 2.9).

The ratio of these two independent chi-square distributions $\frac{Y'(R-Q)Y}{Y'QY}$ yields a doubly non-central F-distribution whose probability density function is given by:

$$dH(\mu) = e^{-\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \frac{(\frac{1}{2}\lambda_2)^s}{s!} \mu^{\frac{1}{2}V_1+r-1} \left(\frac{1}{1+\mu}\right)^{\frac{1}{2}V+r+s} \times \frac{du}{B(\frac{1}{2}V_1 + r, \frac{1}{2}V_2 + s)}, \qquad (2.10)$$

where
$\mu = Y'(R-Q)Y / Y'QY$

$\lambda_1 = \beta'X'(R-Q)X\beta$

$\lambda_2 = \beta'X'QX\beta$

$\lambda = \lambda_1 + \lambda_2$

V_1 = Rank of (R-Q)

V_2 = Rank of Q

$V = V_1 + V_2$

(Kendall and Stuart, 1961, Vol. II, pp. 252)

From (2.10) it can be seen that the shape of the F-distribution is uniquely determined by the values of the two non-centrality parameters. Since the particular values of the covariate X_{ij} as well as those of the β_j are involved in the determination of the values of λ_1 and λ_2, an invariant measure of the violation of the homogeneity of regression assumption may be expressed in terms of the β_j alone only if the values of X_{ij} are known. For the purposes of convenience in this investigation therefore, $\overline{X}_j$ was set to zero and ΣX_{ij}^2 was given a constant value n - 1 for all j, thereby reducing the value of $\lambda_1 = 0$ and $\lambda_2 = n - 1 \left[\sum_j \beta_j^2 - \kappa \overline{\beta}^2 \right]$ (for proofs see Appendix A).

Thus under the present constraints on the covariate X_{ij}, $\sigma^2_{\beta_j}$, the variance of the inhomogeneous β_j becomes an invariant measure of the violation of the homogeneity of regression assumption.

II - The measure of violation of the homogeneity of variance assumption used in this study is the squared coefficient of variation of the κ population variances:

$$c^2_{\sigma_j^2} = \sigma^2_{\sigma_j^2} / \mu^2_{\sigma_j^2} \tag{2.11}$$

where the numerator is the variance of the κ variances and the denominator is the squared mean of the κ variances.

Although this measure has not been directly derived for the ANCOVA situation, Box (1954) found in the case of ANOVA, with equal group sizes, under the condition of unequal variances, the ratio of mean squares does not follow the distribution $F(K-1, N-K)$ but is distributed approximately as:

$$F\left[(K-1)\varepsilon', (N-K)\varepsilon\right], \tag{2.12}$$

where $\varepsilon' = \left(1 + \frac{K-2}{K-1} c^2\right)^{-1}$; $\varepsilon = (1+c^2)^{-1}$ and c^2 is the squared coefficient of variation of the K variances. Thus when the variances are unequal ε' and ε are less than unity and the "significance of effects is somewhat overestimated" (p. 300); the larger the value of c^2, the larger the overestimation.

Box's approximate measure of violation of the homogeneity of variance assumption in ANOVA extends itself to the ANCOVA situation provided that the behaviour of the covariate does not exacerbate the situation. Potthoff (1965) as reported by Elashoff (1969) suggests that the effect of inequality of variances in the Y scores is minimized in the ANCOVA situation when $\sigma^2_{x_j}$ and n_j are constant across groups. For the purposes of the present study, therefore these restrictions on the X_{ij} obtained.

Thus the measures of violation of the homogeneity of regression and the homogeneity of variance assumptions used in this study are $\sigma^2_{\beta_j}$ and $c^2_{\sigma^2_j}$ respectively.

DETERMINATION OF INPUT VALUES OF β_j AND σ_j^2 CORRESPONDING TO VALUES OF $\sigma_{\beta_j}^2$ and $c_{\sigma_j^2}^2$

For any given value of $\sigma_{\beta_j}^2$, the variance of the regression slopes, or of $c_{\sigma_j^2}^2$, the squared coefficient of variation of the variances, a correspondingly unique set of β_j or σ_j^2 does not exist. Consequently the following non-arbitrary methods of determining values of β_j and σ_j^2 corresponding to given values of $\sigma_{\beta_j}^2$ and $c_{\sigma_j^2}^2$ respectively were derived and applied uniformly throughout the present investigation.

The set of κ β_j's corresponding to a given value of $\sigma_{\beta_j}^2$ were selected so that they were centred on unity and were separated from one another by equal intervals.

Since variances are proportional to a chi-square distribution, an attempt was made to select κ variances corresponding to κ points on the abscissa of such a distribution so that the areas bounded by their vertical projections were equal. To do this, κ points on the abscissa of the normal distribution were chosen so that the areas bounded by their vertical projections were equal and centred on the mean. The corresponding χ^2 values were derived by the following approximation due to Wilson and Hilferty (1931):

$$(\chi^2/\nu)^{1/3} = 1 - 2/9\nu + Z(2/9\nu)^{1/2} \tag{2.13}$$

where the value of ν was chosen to yield a desired value of $c_{\sigma_j^2}^2$ between the κ values. The values were then transformed so as to have a mean of unity. The pair-wise sets of β_j and σ_j^2 were then combined in the same ascending order of magnitude. (The actual values of β_j and σ_j^2 derived in the above manner corresponding to given values of $\sigma^2_{\beta_j}$ and $c^2_{\sigma_j^2}$ for different values of κ are recorded in Appendix B).

THE RANDOM NUMBER GENERATOR

Necessary to the success of any Monte Carlo study is an adequate procedure for generating random numbers described by specific probability density functions. Computer simulation of stochastic processes usually depends on the internal generation of 'pseudo-random' sequences of numbers. The most common procedure is to generate sequences of numbers which are randomly 'sampled' from a uniform distribution; samples from other distributions are then produced by some transformation of these uniform variates.

The McGill Random Number Package 'Super-Duper' developed by Marsaglia, Ananthanarayanan and Paul (1972) was used to generate random normal variates in this study. 'Super-Duper' capitalizes on the idiosyncrasies of the I.B.M./360 hardware on which the present simulations were run. The Package contains a fast uniform random number generator which combines a multiplicative congruential generator and a 'shift-register' generator; both generators were chosen so as to have the greatest possible period and to minimize any regularity patterns. The random number generator, which is very fast relative to other available generators (15,000 normal variates per second) uses the uniform generator to generate the normal variate 95% of the time from a rectangular distribution which is close to the normal curve; for the remaining 5% of the time, the variate is generated so as to compensate for the discrepancy. The result is a sequence of variates whose distribution is described by the normal curve. (A listing of 'Super-Duper' appears in Appendix C).

CONTROL OF CONCOMITANT VARIABLES

In order to evaluate the effects on the size of probability of Type 1 error of any given violation combination, each sampling distribution of simulated ANCOVA experiments was constructed so that any deviation from theoretical percentage levels could be attributed as exclusively as possible to the violation of the two assumptions. This implied that:

1. in each simulated ANCOVA experiment, all assumptions of the test except those under study were upheld; and
2. a large enough series of ANCOVA experiments was simulated; that is, the sampling distribution was sufficiently large to keep the standard error of the probability of Type 1 error acceptably low.

Upholding the Remaining Assumptions of the Test

Throughout the entire study the following specifications for each ANCOVA experiment obtained:

a) the κ error term populations, from which the κ treatment groups were randomly drawn, were independently and normally distributed with zero mean.

b) the fixed (for a given sampling distribution) values of the covariate X were transformed such that they were approximately normally distributed with $\overline{X}_j = 0$ and $S_{x_j} = 1$ for all treatment groups.

c) τ_j was set to zero for all treatment groups (the null hypothesis situation obtained).

d) individual values on the dependent variable were determined by 'predicting' from the X values and then adding a random error term (the additivity assumption).

Size of the Sampling Distribution

3000 ANCOVA F-values constituted each sampling distribution. This ensured that under normal Central-F theory conditions, the standard error of the probability of Type 1 error $\approx$ 0.004 for α = .05 and $\approx$ 0.002 for α = .01.

THE MONTE CARLO SIMULATION PROCESS

In addition to the above specifications, values were also given to the following parameters in the construction of each sampling distribution:

a) κ, the number of treatment groups per ANCOVA experiment.

b) n, the number of experimental units per treatment group.

c) the κ population variances as determined by the value of $c^2_{\sigma^2_j}$ and

d) the κ population regression slopes as determined by the value of $\sigma^2_{\beta_j}$.

With these specifications given as input, each sampling distribution is constructed by the following steps:

1. The simulation program causes κ groups of random normal variates each of size n to be generated. These values are transformed in such a way as to conform to the specifications of the covariate (see above). The values are then used as the covariate values for each ANCOVA experiment in the sampling distribution.

2. Next, κ groups of random normal variates each of size n are generated. The elements of the κ groups are then transformed so that the κ groups constitute κ random samples from κ error populations with respective variances as specified.

3. The values for the dependent variable are determined by $Y_{ij} = X_{ij} \cdot \beta_j + \varepsilon_{ij}$. The data are now ready for ANCOVA computation, and the calculation of F is carried out using a specially written ANCOVA program.

4. Steps 2 and 3 are repeated 3000 times, yielding an empirical sampling distribution of ANCOVA F-values for a particular combination of degrees of violation of the two assumptions. (A listing of the simulation program appears in Appendix C).

CHAPTER 3

THE EXPERIMENTS

THE SCOPE OF THE STUDY

Considerations of time and money kept the number of combinations of degrees of violations of the two assumptions investigated necessarily low, and since the primary purpose of the study was to establish a functional relationship between combinations of degrees of violation on the one hand, their effects on the probability of Type 1 error (α) on the other, it seemed desirable to sample from over a fairly wide range of violation of each assumption. A preliminary study was carried out to ascertain roughly the range of effects on α level of varying degrees of violation of each separate assumption. The results of this pilot suggested in the case of the homogeneity of regression assumption, a maximum level of $\sigma^2_\beta = 0.36$ which yielded empirical percentages of the order of 2% and 0.02% corresponding to 5% and 1% nominal levels respectively; and in the case of the homogeneity of variance assumption a maximum level of $c^2_{\sigma^2_j} = 1.80$ which yielded empirical percentages of the order of 10% and 5% corresponding to the nominal 5% and 1% levels respectively.

PARAMETRIC SPECIFICATIONS

The effects of simultaneously violating the two assumptions under study were examined using the following levels of violation:

$\sigma^2_{\beta_j}$ 0.00, 0.04, 0.08, 0.12, 0.16, 0.20, 0.24, 0.28, 0.32, 0.36

$c^2_{\sigma^2_j}$ 0.00, 0.18, 0.36, 0.54, 0.72, 0.90, 1.08, 1.26, 1.44, 1.62, 1.80.

The investigation proceeded by constructing six blocks of sampling distributions according to the procedure described in Chapter 2; each of the first four blocks contained 110 sampling distributions each of which embodied one of the 110 combinations of $\sigma^2_{\beta_j}$ and $c^2_{\sigma^2_j}$ above. Each of the last two blocks contained 48 sampling distributions each of which embodied one of the 48 combinations of the first eight levels of $\sigma^2_{\beta_j}$ and the first six levels of $c^2_{\sigma^2_j}$. (The reason for the reduced number of sampling distributions in the last two blocks was due to inherent limitations on the procedure for deriving values of σ^2_j corresponding to higher levels of $c^2_{\sigma^2_j}$ in the $k = 2$ case).

The size of the simulated ANCOVA experiment in each block was as follows (k is the number of treatment groups, and n is group size)

BLOCK 1	$k = 5$,	$n = 15$
BLOCK 2	$k = 5$,	$n = 5$
BLOCK 3	$k = 3$,	$n = 15$
BLOCK 4	$k = 3$,	$n = 5$
BLOCK 5	$k = 2$,	$n = 15$
BLOCK 6	$k = 2$,	$n = 5$

A violation-free sampling distribution ($\sigma^2_{\beta_j} = 0.00$, $c^2_{\sigma^2_j} = 0.00$) was included in each block as a check on the accuracy of the simulation procedure. For each sampling distribution in each block, the number of ANCOVA F-values (referred to as Fcounts) exceeding the F-values corresponding to nominal 5% and 1% levels of significance respectively, were recorded. An attempt was then made to establish a predictive relationship between Fcounts and levels of $\sigma^2_{\beta_j}$ and $c^2_{\sigma^2_j}$ and to examine any changes in this relationship which occurred with changes in

a) the ANCOVA design size

b) the α level of which Fcounts were taken.

RESULTS

Tables 3.2 and 3.3 contain the matrices of Fcounts (expressed in proportions) of the six blocks of sampling distributions corresponding to the 0.05 and 0.01 levels of significance respectively. Table 3.1 below shows the 95% confidence regions for $P = 0.5$ and $P = 0.1$; they were calculated using the following angular transformation to normality for binomial variates:

$$A = \frac{1}{2}\left[\sin^{-1}\left(\frac{X}{N+1}\right)^{\frac{1}{2}} + \sin^{-1}\left(\frac{X+1}{N+1}\right)^{\frac{1}{2}}\right], \qquad (3.1)$$

where $X/N = P$ and A is asymptotically normally distributed with variance within $\pm 6\%$ of $821/(N+\frac{1}{2})$. (Keeping, 1962, p. 72).

TABLE 3.1

95% Confidence Limits for Estimating P = 0.05 and P = 0.01 for Sample Size N = 3000

Lower Limit	Upper Limit	P
0.043	0.058	0.05
0.005	0.014	0.01

The first entry of each matrix in Tables 3.2 and 3.3 is an Fcount on a violation-free sampling distribution and as can be seen from Table 3.1 each one falls inside the 95% confidence intervals for P = 0.05 and P = 0.01 as the case may be, thereby validating the accuracy of the simulation process; these intervals are also useful in interpreting the seriousness of violation effects.

Table 3.4 contains the results of multiple regression analysis applied to each block with Fcounts as the dependent variable and levels of violation of the two assumptions as predictors. A subsequent examination of plotted residuals suggested the inclusion of the product of the two measures of violation as a third predictor. The regression equations for the three-predictor case are displayed in Table 3.5; with the exceptions of $\hat{\beta}_3$ for Block 1 and for Block 5, in Table 3.2, all parameter estimates provedto be significantly different from zero at least at the 0.05 level of significance (For details of statistical tests see Appendix D).

As can be seen from inspection of Table 3.5, the values of the regression weights in the six equations corresponding to α = 0.05 differ substantially from those of the six equations corresponding to α = 0.01. In order to ascertain whether the regression equations depended on the size of k (number of treatment groups), tests for differences between the three sets of regression equations, (collapsed across n = 5 and n = 15) at each level of α were carried out, and yielded significant results in each case. Further, pairwise tests between regression equations with the same value of k and the same α level, but differing in the size of n were carried out, and yielded significant differences in all twelve cases. (For details of statistical tests, see Appendix D).

TABLE 3.2(a): Matrix of Fcounts for Block 1. $k = 5$, $n = 15$, $\alpha = .05$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.051	.066	.068	.072	.073	.085	.084	.088	.098	.104	.095
0.04	.045	.052	.059	.068	.067	.077	.075	.077	.082	.093	.091
0.08	.038	.046	.050	.054	.062	.069	.077	.068	.075	.082	.095
0.12	.029	.032	.045	.055	.055	.064	.069	.064	.076	.076	.079
0.16	.033	.036	.035	.052	.053	.048	.052	.068	.066	.083	.070
0.20	.026	.027	.039	.042	.048	.051	.056	.069	.065	.064	.070
0.24	.022	.027	.033	.033	.039	.049	.050	.051	.051	.064	.067
0.28	.016	.020	.027	.039	.030	.037	.045	.052	.048	.060	.061
0.32	.014	.023	.026	.032	.032	.033	.042	.051	.054	.055	.048
0.36	.013	.010	.022	.029	.031	.034	.038	.038	.041	.048	.053

TABLE 3.2(b): Matrix of Fcounts for Block 2. $k = 5$, $n = 15$, $\alpha = .05$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.047	.061	.075	.073	.100	.091	.105	.099	.107	.113	.119
0.04	.052	.055	.057	.066	.082	.078	.081	.091	.094	.099	.091
0.08	.037	.045	.055	.052	.062	.065	.075	.082	.090	.097	.097
0.12	.034	.042	.052	.054	.060	.055	.063	.079	.081	.089	.089
0.16	.026	.032	.046	.051	.055	.056	.069	.066	.075	.075	.080
0.20	.027	.027	.038	.046	.054	.055	.057	.059	.061	.074	.068
0.24	.020	.028	.034	.041	.045	.056	.049	.050	.064	.065	.068
0.28	.026	.025	.030	.039	.035	.048	.049	.050	.056	.064	.065
0.32	.019	.023	.031	.027	.039	.040	.054	.050	.053	.049	.063
0.36	.018	.018	.024	.027	.034	.039	.041	.044	.044	0.50	.050

TABLE 3.2(c): Matrix of Fcounts for Block 3. $k = 3$, $n = 15$, $\alpha = .05$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.056	.057	.061	.066	.070	.076	.072	.081	.084	.077	.090
0.04	.045	.047	.047	.055	.059	.060	.063	.078	.081	.078	.089
0.08	.040	.044	.053	.049	.060	.065	.067	.067	.076	.083	.088
0.12	.037	.041	.043	.044	.053	.059	.052	.064	.063	.069	.065
0.16	.030	.034	.046	.039	.048	.047	.052	.051	.058	.056	.061
0.20	.024	.029	.030	.041	.042	.044	.049	.051	.055	.067	.054
0.24	.025	.021	.030	.034	.032	.046	.047	.044	.056	.060	.056
0.28	.023	.025	.025	.030	.037	.037	.042	.048	.050	.051	.050
0.32	.019	.020	.026	.031	.033	.032	.045	.045	.048	.045	.051
0.36	.019	.021	.031	.028	.031	.037	.036	.036	.040	.047	.046

TABLE 3.2(d): Matrix of Fcounts for Block 4. $k = 3$, $n = 5$, $\alpha = .05$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.053	.061	.066	.070	.073	.073	.090	.095	.105	.110	.105
0.04	.039	.049	.057	.057	.070	.067	.084	.080	.094	.086	.110
0.08	.042	.046	.055	.063	.061	.070	.080	.084	.089	.097	.097
0.12	.034	.045	.049	.049	.052	.061	.073	.072	.075	.093	.086
0.16	.034	.038	.039	.044	.049	.054	.065	.067	.074	.076	.086
0.20	.033	.032	.048	.048	.041	.045	.059	.065	.066	.076	.071
0.24	.031	.027	.040	.033	.050	.051	.047	.050	.058	.064	.067
0.28	.028	.032	.037	.040	.042	.045	.044	.055	.059	.058	.056
0.32	.026	.027	.029	.040	.044	.035	.048	.050	.056	.054	.063
0.36	.020	.024	.025	.038	.034	.038	.043	.037	.051	.055	.054

TABLE 3.2(e): Matrix of Fcounts for Block 5. $\kappa = 2$, $n = 15$, $\alpha = .05$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.048	.054	.052	.053	.051	.068					
0.04	.041	.050	.046	.046	.059	.049					
0.08	.045	.042	.042	.046	.052	.052					
0.12	.033	.043	.040	.043	.045	.047					
0.16	.038	.041	.035	.043	.036	.040					
0.20	.030	.037	.042	.035	.036	.039					
0.24	.025	.022	.031	.028	.039	.037					
0.28	.030	.025	.030	.034	.032	.027					
0.32											
0.36											

TABLE 3.2(f): Matrix of Fcounts for Block 6. $\kappa = 2$, $n = 5$, $\alpha = .05$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.048	.055	.061	.063	.073	.071					
0.04	.056	.043	.057	.059	.060	.070					
0.08	.036	.048	.047	.058	.056	.064					
0.12	.037	.044	.044	.048	.049	.058					
0.16	.033	.038	.044	.048	.047	.053					
0.20	.037	.029	.038	.040	.043	.051					
0.24	.032	.032	.033	.037	.041	.049					
0.28	.029	.026	.033	.034	.036	.041					
0.32											
0.36											

TABLE 3.3(a): Matrix of Fcounts for Block 1. $k = 5$, $n = 15$, $\alpha = .01$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.011	.017	.016	.024	.025	.028	.032	.039	.042	.050	.043
0.04	.006	.011	.019	.020	.025	.025	.028	.032	.034	.044	.037
0.08	.006	.011	.017	.019	.020	.022	.028	.030	.034	.033	.042
0.12	.003	.009	.010	.020	.020	.021	.029	.025	.028	.031	.031
0.16	.005	.008	.010	.018	.006	.013	.015	.023	.026	.031	.030
0.20	.004	.004	.009	.009	.014	.016	.017	.025	.024	.027	.023
0.24	.002	.004	.007	.010	.010	.014	.015	.019	.019	.024	.027
0.28	.001	.003	.004	.010	.006	.012	.014	.015	.017	.019	.024
0.32	.002	.003	.003	.007	.011	.008	.013	.016	.019	.022	.017
0.36	.001	.001	.004	.006	.009	.011	.011	.011	.014	.016	.020

TABLE 3.3(b): Matrix of Fcounts for Block 2. $k = 5$, $n = 5$, $\alpha = .01$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.007	.012	.023	.023	.032	.030	.035	.040	.048	.048	.059
0.04	.011	.011	.016	.020	.031	.024	.031	.036	.038	.042	.043
0.08	.007	.013	.017	.012	.020	.024	.029	.032	.039	.039	.046
0.12	.004	.009	.013	.011	.017	.020	.018	.032	.034	.034	.041
0.16	.004	.006	.010	.014	.016	.018	.026	.020	.027	.028	.033
0.20	.007	.006	.010	.010	.015	.014	.022	.020	.024	.028	.026
0.24	.003	.006	.008	.010	.012	.016	.016	.017	.024	.025	.023
0.28	.004	.003	.005	.010	.010	.014	.016	.019	.018	.019	.023
0.32	.002	.004	.005	.008	.008	.012	.017	.014	.019	.018	.022
0.36	.002	.003	.005	.005	.011	.011	.013	.014	.015	.015	.020

TABLE 3.3(c): Matrix of Fcounts for Block 3. k = 3, n = 15, α = .01

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.011	.011	.015	.016	.021	.026	.023	.030	.034	.029	.034
0.04	.010	.010	.011	.013	.017	.018	.022	.026	.034	.028	.033
0.08	.008	.008	.013	.014	.021	.022	.026	.025	.031	.026	.033
0.12	.007	.011	.013	.010	.016	.017	.015	.017	.025	.024	.023
0.16	.004	.004	.007	.007	.013	.015	.015	.016	.014	.022	.022
0.20	.005	.004	.005	.013	.009	.015	.013	.015	.017	.023	.017
0.24	.005	.003	.004	.009	.007	.013	.014	.018	.018	.018	.019
0.28	.003	.006	.005	.007	.009	.008	.011	.013	.016	.015	.012
0.32	.002	.003	.006	.005	.008	.008	.012	.012	.015	.013	.017
0.36	.003	.002	.007	.007	.008	.010	.012	.010	.012	.014	.014

TABLE 3.3(d): Matrix of Fcounts for Block 4. $k = 3$, $n = 5$, $\alpha = .01$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.009	.015	.016	.020	.020	.025	.031	.037	.039	.046	.043
0.04	.007	.012	.014	.014	.019	.019	.027	.030	.037	.034	.049
0.08	.006	.010	.010	.015	.018	.025	.025	.027	.034	.036	.041
0.12	.007	.007	.011	.014	.012	.017	.022	.026	.034	.038	.034
0.16	.004	.007	.008	.010	.014	.017	.021	.022	.025	.027	.032
0.20	.006	.006	.011	.011	.012	.010	.019	.021	.024	.029	.022
0.24	.005	.005	.009	.009	.010	.013	.012	.018	.019	.024	.026
0.28	.005	.004	.010	.011	.008	.013	.012	.015	.018	.020	.018
0.32	.005	.003	.006	.011	.009	.006	.012	.012	.019	.015	.019
0.36	.004	.004	.004	.009	.007	.011	.013	.009	.014	.017	.020

TABLE 3.3(e): Matrix of Fcounts for Block 5. $\kappa = 2$, $n = 15$, $\alpha = .01$

$\sigma^2_{\beta_j}$ \ $c^2\sigma^2_j$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.009	.008	.011	.011	.014	.015					
0.04	.008	.014	.009	.008	.015	.014					
0.08	.009	.009	.007	.009	.010	.013					
0.12	.008	.006	.010	.012	.008	.011					
0.16	.005	.008	.007	.007	.008	.011					
0.20	.004	.007	.007	.007	.005	.007					
0.24	.003	.002	.006	.006	.004	.008					
0.28	.005	.005	.007	.005	.005	.003					
0.32											
0.36											

TABLE 3.3(f): Matrix of Fcounts for Block 6. $k = 2$, $n = 5$, $\alpha = .01$

$c^2\sigma^2_j$ / $\sigma^2_{\beta_j}$	0.00	0.18	0.36	0.54	0.72	0.90	1.08	1.26	1.44	1.62	1.80
0.00	.008	.013	.012	.013	.020	.022					
0.04	.011	.006	.017	.013	.011	.021					
0.08	.006	.010	.011	.014	.013	.017					
0.12	.004	.009	.010	.013	.012	.015					
0.16	.007	.006	.009	.009	.011	.013					
0.20	.006	.005	.007	.007	.009	.014					
0.24	.006	.005	.009	.008	.011	.008					
0.28	.006	.004	.006	.005	.008	.009					
0.32											
0.36											

TABLE 3.4: Multiple Regression Weights for 6 Blocks of Sampling Distributions. Note: $\hat{\beta}_o$ = constant; $\hat{\beta}_1$ = estimated regression weight for $c^2_{\sigma^2_j}$; estimated regression weight for $\sigma^2_{\beta_j}$

BLOCK NO.	α	k	n	$\hat{\beta}_o$	$\hat{\beta}_1$	$\hat{\beta}_2$	R
1	.05	5	15	161.93	74.84	-383.87	0.976
2	.05	5	5	174.65	81.89	-421.07	0.965
3	.05	3	15	152.42	58.56	-318.50	0.964
4	.05	3	5	165.20	77.79	-355.58	0.963
5	.05	2	15	147.09	27.86	-266.37	0.914
6	.05	2	5	153.65	62.56	-302.33	0.959
1	.01	5	15	44.68	43.37	-167.48	0.960
2	.01	5	5	46.67	47.56	-178.44	0.946
3	.01	3	15	38.02	30.59	-123.40	0.934
4	.01	3	5	39.69	42.55	-145.22	0.937
5	.01	2	15	30.33	10.60	- 77.18	0.831
6	.01	2	5	31.30	25.79	- 88.00	0.877

TABLE 3.5: Multiple Regression Weights for 6 Blocks of Sampling Distributions

Note: $\hat{\beta}_o$ = constant; $\hat{\beta}_1$=estimated regression weight for $c^2_{\sigma^2_j}$; $\hat{\beta}_2$ estimated regression weight for $\sigma^2_{\beta_j}$; $\hat{\beta}_3$ = estimated regression weight for $\sigma^2_{\beta_j} \cdot c^2_{\sigma^2_j}$

BLOCK NO.	α	k	n	$\hat{\beta}_o$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	R
1	.05	5	15	156.14	81.28	-351.68	- 35.80	0.976
2	.05	5	5	155.74	102.91	-316.00	-116.80	0.972
3	.05	3	15	145.16	66.62	-278.17	- 44.80	0.965
4	.05	3	5	141.33	104.31	-223.00	-147.30	0.975
5	.05	2	15	156.14	38.06	-233.59	- 72.80	0.917
6	.05	2	5	155.74	77.20	-255.29	-104.50	0.963
1	.01	5	15	31.53	57.99	- 94.41	- 81.20	0.974
2	.01	5	5	25.79	70.54	- 63.51	-127.70	0.975
3	.01	3	15	27.16	42.66	- 63.08	- 67.00	0.951
4	.01	3	5	19.43	65.06	- 32.66	-125.10	0.973
5	.01	2	15	26.00	20.23	- 46.23	-68.80	0.856
6	.01	2	5	25.22	39.30	- 44.59	- 96.50	0.902

CHAPTER 4

DISCUSSION

INTERPRETATION OF RESULTS

The results of this study belong to that body of findings which attempts to apprise the scientist of the consequences of applying statistical tests to his data when the requirements of the underlying assumptions of the test are not fulfilled. For the most part, previous findings have related to situations where only one assumption is being violated with all other demands of the test satisfied. The present study arose out of the suggestion implicit in several separate findings that the simultaneous presence of violation of the homogeneity of regression and homogeneity of variance assumptions in ANCOVA would tend to cancel each other out in their effects.

The overall findings contained in Chapter 3 show that this tendency is in fact the case; for all 6 blocks of sampling distributions the effect on probability of Type 1 error of a given degree of violation of one assumption is strongly dependent on the degree of violation present in the other. Furthermore, the two violations are seen to have a neutralizing effect on one another. Inspection of Tables

3.2 and 3.3 reveals that many combinations of σ_β^2 and $c_{\sigma^2}^2$ have sampling distributions with Fcounts falling within the 95% confidence interval for P = 0.05 or P = 0.01 as the case may be (see Fcounts within bold lines). It is only those combinations involving a low level of violation of one assumption in combination with a high level of violation of the other that yield Fcounts extending beyond the confidence limits. Finally, for any given violation level of one assumption, the most serious departure from nominal probability levels occurs when the other assumption is fully satisfied (see Column 1 and Row 1 of each block).

From inspection of the multiple regression data of Tables 3.4 and 3.5, it is suggested that the inclusion of the product of the two violation measures as a third predictor is important not only because most of its regression estimates proved statistically significant but also because the resultant regression equations yield values of the constant term $\hat{\beta}_o$ which more readily approximate the expected numbers of Fcounts when $\sigma_{\beta_j}^2 = 0.0$ and $c_{\sigma_j^2}^2 = 0.0$ for both α levels (150 and 30 respectively) than those emerging from the two-predictor equations.

The differences between the 6 regression equations corresponding to α = .05 and those corresponding to α = 0.1 in Table 3.5 appear to be accounted for by changes in the relative weightings of $\sigma^2_{\beta_j}$ and $c^2_{\sigma^2_j}$. It is clear from inspection that the role of $c^2_{\sigma^2_j}$, the measure of violation of the homogeneity of variance assumption, is more important in the equations corresponding to $\alpha = .01$ than in those corresponding to $\alpha = .05$. This is also evident in the different locations of the areas circumscribed by bold lines in Tables 3.2 and 3.3 respectively. The confounding presence of the interaction term β_3 notwithstanding, it also appears that within each α level, significant differences between

1. regression equations corresponding to different values of κ (collapsed over n), and
2. pairs of regression equations corresponding to different values of n for a given value of κ,

are also attributable to changes in the relative weightings of $\sigma^2_{\beta_j}$ and $c^2_{\sigma^2_j}$. While it is difficult to interpret a trend in the former case, it is suggested that the latter differences are due to the increasing importance of $c^2_{\sigma^2_j}$ when n is relatively small (compare n = 5 with n = 15).

Referring again to Table 3.5, it can be seen from the column of multiple correlation coefficients (R), that the three predictors account for a very large proportion of the dependent variable (Fcounts); these figures testify to the high degree of linearity in the parameters.

QUALIFICATIONS TO RESULTS

Any Monte Carlo researcher who attempts to examine the effects of violating some specified assumption or set of assumptions underlying a statistical test is invariably faced with a methodological paradox: in order to evaluate the specific effects of violating the particular assumption or assumptions of his study, he must ensure that all other assumptions of the test are upheld. This precaution, however, reduces the generalizability of his results to situations where some of the other assumptions of the test are also not upheld. Consequently he can only hope that the general trend of his findings obtained in these latter situations. Often it is a hope that is not realized as the present study has attempted to show for one specific situation.

Apart from these general considerations, perhaps the greatest impediments to generalizability in this investigation are the restrictions on the behaviour of the covariate, X_{ij}. These were:

1. ΣX_{ij} and hence $\overline{X}_j$ were set to a constant for all j.
2. ΣX^2_{ij} was set equal to a constant for all j.

These two constraints on the covariate permitted a derivation of an invariant measure of violation of the homogeneity of re-regression assumption, while the second constraint permitted the extension of Box's approximate measure of violation of the homogeneity of variance assumption (derived for ANOVA) to the present ANCOVA context. Another qualification to the general aplicability of the present results is due to the arbitrary if systematic way in which,

1. sets of β_j and σ_j^2 were derived to correspond to given levels of $\sigma^2_{\beta_j}$ and $c^2_{\sigma^2_j}$ respectively, and

2. these two sets, so derived, were matched in the simulated ANCOVA experiments. More generally, even though it was possible to derive invariant (or nearly so) measures of violation of the two assumptions separately, there is no guarantee that the combination of these two measures constitutes an invariant measure of their joint violation.

APPLICATION OF RESULTS TO PRACTICAL SITUATIONS

Despite the qualifications outlined in the preceding section, the general results of the present study strongly points to the neutralizing effect on the F-test in ANCOVA of the simultaneous violation of the homogeneity of regression and homogeneity of variance assumptions. It is very unlikely that in any practical ANCOVA situation either of these two assumptions will be fully met, and consequently the ANCOVA F-test appears from the present findings to be even more robust in practice, than results from previous studies taking these violations one at a time have indicated. It is difficult to estimate the upper limits on the violations of these two assumptions as they occur in practice, but they are practically certain to fall within the range of violations encompassed by the present study, and unless the departure from one assumption is very big and is accompanied by only a slight departure from the other, it is safe to conclude that the ANCOVA F-test is only inconsequently affected by the presence of joint violations of the two assumptions. Finally, the findings of the present study suggest the limited usefulness of any test of homogeneity of regression slopes carried out on ANCOVA data without due regard being given to possible differences between treatment group variances.

REFERENCES

Atiqullah, M. "The Robustness of Covariance Analysis of a One-Way Classification", Biometrika 51, Dec. 1964, pp. 365-372.

Bartlett, M.S. "Properties of Sufficiency and Statistical Tests", Proc. Roy. Soc. (London), 1937, 160, 268-282.

Boneau, C. Alan "The Effects of Violations of Assumptions Underlying the t Test", Psychological Bulletin, 1960, 57, pp. 49-64.

Box, G.E.P. "Non-Normality and Tests on Variances", Biometrika, 1953, 40, pp. 318-335.

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", Ann. of Math. Statist., 1954, 25, pp. 290-302.

Cochran, W.G. "The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total", Annals of Eugenics, 11, 1941, pp. 47-52.

Cochran, William G. "Analysis of Covariance: Its Nature and Uses", Biometrics 13, Sept. 1957, pp. 261-281.

Cochran, William G. "Errors of Measurement in Statistics", Technometrics 10, Nov. 1968, pp. 637-666.

Daniels, H.E. "The Effect of Departures from Ideal Conditions Other than Non-Normality on the t and z Tests of Significance", Proc. Camb. Phil. Soc., 1938, 34, pp. 321-328.

Draper, N.R. and Smith H, Applied Regression Analysis, Wiley, 1966.

Elashoff, Janet D. "Analysis of Covariance: A Delicate Instrument", American Educational Research Journal 6, May 1969, pp. 383-401.

Evans, Selby H. and Anastasio, Ernest J. "Misuse of Analysis of Covariance When Treatment Effect and Covariate are Confounded", Psychological Bulletin 69, April 1968, pp. 225-234.

Graybill, Franklin A., Introduction to Matrices with Applications in Statistics, Wadsworth Publishing Company, Inc., Belmont, California, 1969.

Hartley, H.O. "The Maximum F-Ratio as a Short-Cut Test for Heterogeneity of Variance", Biometrika, 37, Dec. 1950, 308-312.

Horsnell, G. "The Effect of Unequal Group Variances on the F-test for the Homogeneity of Group Means", Biometrika, 1953, 40, pp. 128-136.

Keeping, E.S., Introduction to Statistical Inference, D. Van Nostrand Company, Inc., Princeton, New Jersey, 1962.

Kendall, M.G. and Stuart, A., The Advanced Theory of Statistics: Vol. II, Charles Griffin and Co. Ltd., London, 1961.

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

Lord, Frederic,M. "Large-Sample Covariance Analysis When the Control Variable is Fallible", Journal of the American Statistical Association 55, June 1960, pp. 307-321.

Lord, Frederic, M. "A Paradox in the Interpretation of Group Comparisons", Psychological Bulletin 68, Nov. 1967, pp. 304-305.

Marsaglia, G., Ananthanarayanan, K. and Paul, N., The McGill Random Number Package 'Super-Duper' (in preparation.)

Norton, D.W., An Empirical Investigation of Some Effects of Non-Normality and Heterogeneity on the F-Distribution, Unpublished Doctoral Dissertation, State University of Iowa, 1952.

Peckham, Percy D., An Investigation of the Effects of Non-Homogeneity of Regression Slopes Upon the F-Test of Analysis of Covariance, Doctoral Dissertation, 1968, University of Colorado.

Porter, Andrew C., The Effects of Using Fallible Variables in the Analysis of Covariance, Doctoral Dissertation, 1967, University of Wisconsin.

Potthoff, Richard F. "Some Scheffé-Type Tests for Some Behrens-Fisher Type Regression Problems", J.A.S.A. 60, Dec. 1965, pp. 1163-1190.

Rao, Radhakrishna C., Linear Statistical Inference and Its Applications, John Wiley and Sons, Inc., 1965.

Scheffé, Henry, The Analysis of Variance, New York, Wiley, 1969.

Welch, B.L. "The Significance of the Difference Between Two Means When the Populations Variances are Unequal", Biometrika 29, 1937, pp. 350-362.

Wilson, E.S. and Hilferty, M.M. "The Distribution of Chi-Square", Proc. Nat. Acad. Sci., 17, 1931, pp. 684-688.

Winer, B.J., Statistical Principles in Experimental Design, New York, McGraw-Hill, 1962.

APPENDIX A

1. Proof that $(R-Q)$ is symmetric and idempotent:

$R-Q$ is symmetric, since

$$(R-Q)' = R' - Q'$$
$$= R - Q \quad \text{using (2.4a) and (2.6)}$$

Let L be a matrix with 1 as its last element and zero's everywhere else, such that $SL = T$.

$$\text{Then } RQ = \left[I - SL(L'S'SL)^{-1}L'S'\right]\left[I - S(S'S)^{-1}S'\right]$$
$$= I - S(S'S)^{-1}S' - SL(L'S'SL)^{-1}L'S' + SL(L'S'SL)^{-1}L'S'S(S'S)^{-1}S'$$
$$= I - S(S'S)^{-1}S'$$
$$= Q \tag{A.1}$$

$(R-Q)$ is also idempotent, for

$$(R-Q)^2 = R^2 - RQ - QR + Q^2$$
$$= R - Q - Q + Q, \quad \text{since } R^2=R,\ Q^2=Q \text{ and } QR=Q'R'=(RQ)'=Q'=Q$$
$$= R - Q$$

2. Proof that $(R - Q)'Q = 0$:

$$\begin{aligned}(R - Q)'Q &= (R - Q)Q \\ &= (RQ - QQ \\ &= Q - Q \qquad \text{using (A1)} \\ &= 0\end{aligned}$$

3. Proof that under the constraints placed on the covariate X_{ij} as stated in Chapter 2,

$$\lambda_1 = 0 \quad \text{and} \quad \lambda_2 = n - 1 \left[\sum_j \beta_j^2 - k\,\overline{\beta}^2 \right]:$$

The constraints on X_{ij} are:

$$\Sigma X_{ij} = 0 \quad \text{for all } j \tag{A3}$$

$$\Sigma X_{ij}^2 = n - 1 \text{ for all } j \tag{A4a}$$

$$\sum_i \sum_j X_{ij}^2 = (n - 1)k \tag{A4b}$$

Let the design matrix X be partitioned into the following submatrices:

$$X = \left[A \,\vdots\, T \,\vdots\, B \right],$$

where A is the nk x k submatrix whose columns correspond to treatment assignment, T is the already defined nk x 1 vector of covariate values and B is the nk x k submatrix whose columns correspond to $(\beta_j - \overline{\beta})$, $j = 1,\ldots,k$.

then $A'T = T'A = 0$ (using A3) (A5)

$A'B = B'A = 0$ (using A3) (A6)

$$\begin{aligned}\lambda_1 &= \beta'X'(R - Q)\, X\beta \\ &= \beta'X'\left[I - T(T'T)^{-1}T' - I + S(S'S)^{-1}S'\right]X\beta \\ &= \beta'X'\left[S(S'S)^{-1}S' - T(T'T)^{-1}T'\right] X\beta\end{aligned}$$

Now

$$S = \left[A \,\vdots\, T\right] \quad \text{and} \quad S'S = \left[\begin{array}{c|c} A'A & A'T \\ \hline T'A & T'T \end{array}\right]$$

$$\left[\begin{array}{c|c} A'A & 0 \\ \hline 0 & T'T \end{array}\right], \quad \text{(using A5)}$$

$$(S'S)^{-1} = \left[\begin{array}{c|c} (A'A)^{-1} & 0 \\ \hline 0 & (T'T)^{-1} \end{array}\right],$$

using a theorem on the inverse of partitioned matrices (Graybill, 1969, p. 165).

$$S(S'S)^{-1}S' = \left[A \,\vdots\, T\right] \left[\begin{array}{c|c} (A'A)^{-1} & 0 \\ \hline 0 & (T'T)^{-1} \end{array}\right] \left[\begin{array}{c} A' \\ \hline T' \end{array}\right]$$

$$= A(A'A)^{-1}A' + T(T'T)^{-1}T' \qquad \text{(A7)}$$

Thus

$$\begin{aligned}\lambda_1 &= \beta'X'\left[A(A'A)^{-1}A' + T(T'T)^{-1}T' - T(T'T)^{-1}T'\right] X\beta \\ &= \beta'X'\left[A(A'A)^{-1}A'\right] X\beta\end{aligned}$$

APPENDIX A (cont'd)

$$A'X = \left[A'A \mid A'T \mid A'B \right]$$
$$= \left[A'A \mid 0 \mid 0 \right] \qquad \text{(using A5 and A6)}$$

$A'A$ is $k \times k$. Under the null hypothesis, $\tau_1 = \tau_2 \ldots \tau_k = 0$, the leading k elements of β are zero.

Hence $A'X\beta = 0$ (A8)

Hence $\lambda_1 = 0$ (A9)

$$\lambda_2 = \beta'X'\left[I - S(S'S)^{-1}S' \right] X\beta$$
$$= \beta'X'X\beta - \beta'X'\left[A(A'A)^{-1}A' \right] X\beta - \beta'X'\left[T(T'T)^{-1}T' \right] X\beta, \quad \text{(using A7)}$$
$$= \sum_j \beta_j^2 \sum_i X_{ij}^2 - 0 - \beta'X'T(T'T)^{-1}T'X\beta, \quad \text{(using A9)}$$

$$(T'T) = \sum_i \sum_j X_{ij}^2 \quad \text{and} \quad (T'T)^{-1} = 1/\sum_i \sum_j X_{ij}^2$$

Thus $\beta'X'T(T'T)^{-1}T'X\beta = 1/\sum_i \sum_j X_{ij}^2 \left[\beta'X'TT'X\beta \right]$

$$\beta'X'T = T'X\beta = \sum_j \beta_j \sum_i X_{ij}^2$$

Thus

$$\lambda_2 = \sum_j \beta_j^2 \sum X_{ij}^2 - 1/\sum \sum X_{ij}^2 \left(\sum_j \beta_j \sum X_{ij}^2 \right)^2$$
$$= n-1 \left(\sum_j \beta_j^2 \right) - \frac{(n-1)^2}{(n-1)k} \left(\sum_j \beta_j \right)^2$$
$$= n-1 \left[\sum_j \beta_j^2 - k\,\bar{\beta}^2 \right]$$

APPENDIX B

Values of β_j and σ_j^2 corresponding to levels of $\sigma_{\beta_j}^2$ and $c_{\sigma_j^2}^2$

k = 5

$\sigma_{\beta_j}^2$	β_1	β_2	β_3	β_4	β_5
0.00	1.000000	1.000000	1.000000	1.000000	1.000000
0.04	0.717157	0.858579	1.000000	1.141421	1.282843
0.08	0.600000	0.800000	1.000000	1.200000	1.400000
0.12	0.510102	0.755051	1.000000	1.244949	1.489898
0.16	0.434315	0.717157	1.000000	1.282843	1.565685
0.20	0.367544	0.683772	1.000000	1.316228	1.632456
0.24	0.307180	0.653590	1.000000	1.346410	1.692820
0.28	0.251669	0.625834	1.000000	1.374166	1.748331
0.32	0.200000	0.600000	1.000000	1.400000	1.800000
0.36	0.151472	0.575736	1.000000	1.424264	1.848528

$c_{\sigma_j^2}^2$	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2
0.00	1.000000	1.000000	1.000000	1.000000	1.000000
0.18	0.690374	0.838605	0.962725	1.098524	1.298665
0.36	0.558518	0.752084	0.923455	1.119061	1.420338
0.54	0.457049	0.677277	0.882294	1.124999	1.512725
0.72	0.372529	0.608515	0.839373	1.122283	1.589612
0.90	0.302851	0.546432	0.769713	1.113687	1.653676
1.08	0.244267	0.489461	0.754358	1.100861	1.708745
1.26	0.188417	0.429794	0.706622	1.082341	1.763235
1.44	0.143651	0.376756	0.661098	1.061314	1.809415
1.62	0.104365	0.324585	0.613175	1.036069	1.853195
1.80	0.071038	0.273825	0.563104	1.006574	1.894650

APPENDIX B - cont'd

$\kappa = 3$

$\sigma^2_{B_j}$	β_1	β_2	β_3
0.00	1.000000	1.000000	1.000000
0.04	0.755051	1.000000	1.244948
0.08	0.653590	1.000000	1.346410
0.12	0.575736	1.000000	1.424263
0.16	0.510102	1.000000	1.489898
0.20	0.452278	1.000000	1.547722
0.24	0.400000	1.000000	1.599999
0.28	0.351926	1.000000	1.648073
0.32	0.307180	1.000000	1.692820
0.36	0.265153	1.000000	1.734846

$c^2\sigma^2_j$	σ^2_1	σ^2_2	σ^2_3
0.00	1.000000	1.000000	1.000000
0.18	0.725208	0.961686	1.244632
0.36	0.594683	0.918516	1.342640
0.54	0.493581	0.873828	1.411666
0.72	0.400668	0.823106	1.470361
0.90	0.315367	0.766848	1.520687
1.08	0.236368	0.704075	1.564739
1.26	0.163549	0.633189	1.603846
1.44	0.98270	0.551836	1.638847
1.61	0.041036	0.449345	1.672246
1.80	0.005997	0.325273	1.701224

APPENDIX B (cont'd)

$k = 2$

$\sigma^2_{B_j}$	β_1	β_2
0.00	1.000000	1.000000
0.04	0.800000	1.200000
0.08	0.717157	1.282843
0.12	0.653590	1.346410
0.16	0.600000	1.400000
0.20	0.552786	1.447214
0.24	0.510102	1.489898
0.28	0.470850	1.529150
0.32	0.434315	1.565685
0.36	0.400000	1.600000

$c^2\sigma 2^{2}_{i\,j}$	σ^2_1	σ^2_2
0.00	1.000000	1.000000
0.18	0.755864	1.195270
0.36	0.635216	1.263527
0.54	0.512068	1.318251
0.72	0.390689	1.359177
0.90	0.223631	1.396420

APPENDIX C

LISTING OF SIMULATION PROGRAM AND SUPER-DUPER

```
C....................      SIMULATION  PROGRAM  ........................
C..
C..
      REAL *8 SUMD,SUMXD,SMIJX2,SMIJY2,SMIJXY,TJXI,TJYI,BIGTX,BIGTY,
     1SMTJX2,SMTJY2,SMTJXY,BIGTX2,BIGTY2,BIGTXY,ADJTOT,WITHIN,F
      DIMENSION TOTAL(10000),X(100),VARX(5),Y(100),
     1STANDE(5),BETAXY(5),FVALUE(3000)
      DATA  VARX(1),VARX(2),VARX(3),VARX(4),VARX(5)/5*1.0 /
    1 READ (5,100,END=200) NO,N,K,INTEG1,INTEG2,CRIT5,CRIT1,MUCH,NEED
  100 FORMAT (1X,I3,4X,I2,4X,I1,4X,2I3,5X,F6.2,4X,F6.2,4X,I4,4X,I2)
      READ (5,40) (STANDE(I),I=1,K)
      READ (5,40)  (BETAXY(L),L=1,K)
   40 FORMAT (5F9.6)
C..
C..
C..THE FOLLOWING BLOCK - PART 1 -  USES SUPERDUPER TO GENERATE
C..RANDOM NORMAL VARIATES, AND THEN ARRANGES THE X'S AND Y'S
C.. ACCORDING TO PARAMETRIC SPECIFICATIONS.
C..
C..
      SN=N
      SK=K
      NK=N*K
      SNK=NK
      CALL START (INTEG1,INTEG2)
      DO 18 I=1,NK
      X(I)=RNOR(0)
   18 CONTINUE
      LL=1
      INDEX=1
      L=1
      NN=N
    5 SUMD=0.0
      SUMXD=0.0
    6 SUMD=SUMD+X(L)
      SUMXD=SUMXD+X(L)*X(L)
      L=L+1
      IF(L.LE.NN)  GO TO 6
      AMEAN=SUMD/SN
      VARIAN=(SUMXD*SN-(SUMD*SUMD))/(SN*(SN-1.0))
      FACTOR=VARX(INDEX)/SQRT(VARIAN)
    7 X(LL)=(X(LL)-AMEAN)*FACTOR
      LL=LL+1
      IF(LL.LE.NN)  GO TO 7
      INDEX=INDEX+1
      NN=NN+N
      IF(NN.LE.NK)  GO TO 5
      LOT=MUCH*NK
      MORE=MUCH
      LEAP=1
      DO 20   I=1,NEED
      DO 19   J=1,LOT
      TOTAL(J)=RNOR(0)
   19 CONTINUE
      JUMP=0
   24 INDEX=1
      NN=N
      L=1
```

```
   10 J=L+JUMP
      Y(L)=TOTAL(J)*STANDE(INDEX)+BETAXY(INDEX)*X(L)
      L=L+1
      IF (L.LE.NN)  GO TO 10
      NN=NN+N
      INDEX=INDEX+1
      IF(NN.LE.NK)  GO TO 10
   11 CONTINUE
C..
C..
C.. THE FOLLOWING BLOCK - PART 2 - COMPUTES ANCOVA IN DOUBLE PRECISION
C..
C..
      SMIJX2=0.0
      SMIJY2=0.0
      SMIJXY=0.0
      L=1
   13 SMIJX2=SMIJX2+X(L)*X(L)
      SMIJXY=SMIJXY+X(L)*Y(L)
      SMIJY2=SMIJY2+Y(L)*Y(L)
      L=L+1
      IF(L.LE.NK)  GO TO 13
      SMTJX2=0.0
      SMTJY2=0.0
      SMTJXY=0.0
      BIGTX=0.0
      BIGTY=0.0
      L=1
      NN=N
   15 TJYI=0.0
      TJXI=0.0
   14 TJXI=TJXI+X(L)
      TJYI=TJYI+Y(L)
      L=L+1
      IF(L.LE.NN)  GO TO 14
      SMTJX2=SMTJX2+((TJXI*TJXI)/SN)
      SMTJXY=SMTJXY+((TJXI*TJYI)/SN)
      SMTJY2=SMTJY2+((TJYI*TJYI)/SN)
            BIGTX=BIGTX+TJXI
      BIGTY=BIGTY+TJYI
      NN=NN+N
      IF(NN.LE.NK)  GO TO 15
      BIGTX2=(BIGTX*BIGTX)/SNK
      BIGTXY=(BIGTX*BIGTY)/SNK
      BIGTY2=(BIGTY*BIGTY)/SNK
      ADJTOT=(SMIJY2-BIGTY2)-(((SMIJXY-BIGTXY)*(SMIJXY-BIGTXY))/
     1 (SMIJX2-BIGTX2))
      WITHIN=(SMIJY2-SMTJY2)-(((SMIJXY-SMTJXY)*(SMIJXY-SMTJXY))/
     1 (SMIJX2-SMTJX2))
      F=((ADJTOT-WITHIN)/(SK-1.0))/(WITHIN/(SNK-SK-1.0))
      FVALUE(LEAP) = F
      LEAP=LEAP+1
      JUMP=JUMP+NK
      IF(LEAP.LE.MORE) GO TO 24
      MORE=MORE+MUCH
   20 CONTINUE
C..
C..
```

```
C.. THE FOLLOWING BLOCK -  PART 3 - COMPARES THE 3000 F-VALUES
C.. WITH NOMINAL F-VALUES AT THE 0.05 AND 0.01 LEVELS RESPECTIVELY.
C..
C..
          INDEX5=0
      INDEX1=0
   17 DO 16  K=1,3000
      FVAL=FVALUE(K)
      IF(FVAL.LT.CRIT5)  GO TO 16
      INDEX5=INDEX5+1
      IF(FVAL.LT.CRIT1)  GO TO 16
      INDEX1=INDEX1+1
   16 CONTINUE
      WRITE (6,102)  NO,LEAP,INDEX5,INDEX1
  102  FORMAT (////,2X,I4,4X,I9,4X,I4,4X,I4)
      GO TO 1
  200 STOP
      END
```

```
*
*          MCGILL UNIVERSITY    SCHOOL OF COMPUTER SCIENCE
*
*          RANDOM NUMBER GENERATOR PACKAGE - 'SUPER-DUPER'
*          ------ ------ --------- -------    -----------
*
*          UNIFORM,NORMAL AND EXPONENTIAL RANDOM NUMBER GENERATOR
*
*          G. MARSAGLIA, K.ANANTHANARAYANAN, N.PAUL.
*
*        RANDOM NUMBER GENERATOR PACKAGE-REGISTER USAGE
*        ------ ------ --------- ------- -------- -----
*        GPR 0 - STORES RESULT OF IUNI,IVNI
*        GPR 1 - (REGB) CALCULATION OF RESULTS
*        GPR 2 - (REGC) CALCULATION OF RESULTS
*        GPR 3 - (REGD) CALCULATION OF RESULTS
*        GPR13 - ADDRESS OF SAVE AREA OF CALLING PROGRAM,OR OF THIS
*                PROGRAMS'S SAVE AREA ON CALL TO RNORTH OR REXPTH
*        GPR14 - CONTAINS RETURN ADDRESS.
*        GPR15 - USED AS BASE REGISTER.
*        FPR 0 - RESULT OF UNI,VNI,REXP,RNOR.
*
RANDOM START 0                   DEFINE ENTRY POINTS
       ENTRY START               CALL START(I1,I2)
       ENTRY UNI                 U=UNI(0)
       ENTRY VNI                 V=VNI(0)
       ENTRY RNOR                X=RNOR(0)
       ENTRY REXP                Y=REXP(0)
       ENTRY IUNI                K=IUNI(0)
       ENTRY IVNI                J=IVNI(0)
       EXTRN RNORTH              FORTRAN FUNCTIONS REQUIRED-RNORTH(I)
       EXTRN REXPTH                                         REXPTH(I)
REGB   EQU   1
REGC   EQU   2                   REGISTER EQUATES
REGD   EQU   3
*
*    CALL START(I1,I2)           I1,I2 ARE USED FOR STARTING THE TWO
*                                SEQUENCES 'MCGN' AND 'SRGN'.
       USING START,15
START  STM   REGB,REGD,24(13) SAVE REGISTERS 1,2,3
       LM    REGC,REGD,0(1)   LOAD ADDRESSES OF I1,I2 INTO REGC,REGD
       L     REGC,0(REGC)     LOAD VALUE OF I1 INTO REGC
       LTR   REGC,REGC
       BC    8,ST1            IF ZERO,STORE AT 'MCGN',ELSE
       O     REGC,X1          ENSURE ODD,TO KEEP PERIOD OF 'MCGN' LARGE
ST1    ST    REGC,MCGN        STORE AT 'MCGN'
       L     REGD,0(REGD)     LOAD I2 INTO REGD
       LTR   REGD,REGD
       BC    8,ST2            IF ZERO, STORE AT 'SRGN',ELSE
       N     REGD,X7FF        TAKE RESIDUE MODULO 2048
       O     REGD,X1          AND ENSURE NON-ZERO
ST2    ST    REGD,SRGN        AND STORE AT 'SRGN'.
RETRNO LM    REGB,REGD,24(13) RESTORE REGISTERS 1,2,3
       BCR   15,14            AND RETURN
*
*      U=UNI(0)                  RESULT IS NORMALIZED FLOATING POINT VALUE
*                                UNIFORMLY DISTRIBUTED ON (0.0,1.0).
       USING UNI,15
```

```
UNI     STM    REGB,REGD,24(13) SAVE REGISTERS 1,2,3
RDIGT1  L      REGB,SRGN        LOAD SRGN INTO REGB
        LR     REGC,REGB        AND INTO REGC
        SRL    REGC,15          SHIFT REGC RIGHT 15 BITS
        XR     REGB,REGC        AND XOR INTO REGB
        LR     REGC,REGB        COPY REGB INTO REGC
        SLL    REGC,17          SHIFT IT LEFT 17 BITS,
        XR     REGB,REGC        AND XOR INTO REGB
        ST     REGB,SRGN        SAVE THE NEW 'SRGN'
        L      REGD,MCGN        LOAD MCGN INTO REGD
        M      REGC,MULT        AND MULTIPLY BY 69069
        ST     REGD,MCGN        STORE RESULT,MODULO 2**32, AS NEW 'MCGN'
        XR     REGD,REGB        XOR NEW 'MCGN' AND 'SRGN' IN REGD
        SRL    REGD,8           SHIFT REGD RIGHT 8 BITS FOR F.P. FRACTION
        AL     REGD,CHAR        ADD CHARACTERISTIC X'40' INTO FIRST BYTE
        ST     REGD,FWD         STORE AT FWD, LOAD INTO FPR 0,
        LE     0,FWD            AND ADD NORMALIZED TO ZERO
        AE     0,Z              LEAVING RESULT 'UNI' IN FPR 0.
RETRN1  LM     REGB,REGD,24(13)
        BCR    15,14            RETURN
*
*       V=VNI(0)                RESULT IS NORMALIZED FLOATING POINT VALUE
*                               UNIFORM ON (-1.0,1.0)
        USING VNI,15
VNI     STM    REGB,REGD,24(13) SAVE REGISTERS 1,2,3
RDIGT2  L      REGB,SRGN        LOAD SRGN INTO REGB
        LR     REGC,REGB        AND INTO REGC
        SRL    REGC,15          SHIFT REGC RIGHT 15 BITS
        XR     REGB,REGC        AND XOR INTO REGB
        LR     REGC,REGB        COPY REGB INTO REGC
        SLL    REGC,17          SHIFT IT LEFT 17 BITS,
        XR     REGB,REGC        AND XOR INTO REGB
        ST     REGB,SRGN        SAVE THE NEW 'SRGN'
        L      REGD,MCGN        LOAD MCGN INTO REGD
        M      REGC,MULT        AND MULTIPLY BY 69069
        ST     REGD,MCGN        STORE RESULT,MODULO 2**32, AS NEW 'MCGN'
        XR     REGD,REGB        XOR NEW 'MCGN' AND 'SRGN' IN REGD
        SRA    REGD,7           SHIFT RIGHT 7 BITS PRESERVING SIGN BIT
        N      REGD,SIGN        ZERO OUT LAST 7 BITS  OF FIRST BYTE
        AL     REGD,CHAR        ADD CHARACTERISTIC X'40' TO FIRST BYTE
        ST     REGD,FWD         STORE AT FWD, LOAD INTO FPR 0
        LE     0,FWD            AND ADD NORMALIZED TO ZERO
        AE     0,Z              LEAVING RESULT 'VNI' IN FPR 0.
RETRN2  LM     REGB,REGD,24(13)
        BCR    15,14            RETURN
*
*       X=RNOR(0)               RESULT IS STANDARD NORMAL VARIATE.
*
*       METHOD
*       ------
*    1.  GENERATE H1H2H3H4H5H6H7H8,8 RANDOM HEXADECIMAL DIGITS.
*
*    2.  IF H1H2 .LT. 68, SET 'RNOR' TO
*                         (NTBL(H1H2)+.H3H4H5H6H7H8)/16, AND QUIT.
*    3.  IF H1H2 .LT. D0, SET 'RNOR' TO
*                         (-NTBL(H1H2-68)-.H3H4H5H6H7H8)/16, AND QUIT.
*    4.  IF H1H2H3 .LT. E2F, SET 'RNOR' TO
*                         (NTBL(H1H2H3-CE8)+.H4H5H6H7H8)/16, AND QUIT.
```

```
*     5.  IF H1H2H3 .LT. F5E, SET 'RNOR' TO
*                           (-NTBL(H1H2H3-E17)-.H4H5H6H7H8)/16, AND QUIT.
*     6.  ELSE,GENERATE 'RNOR' FROM THE NORMAL TOOTH-TAIL SUBPROGRAM.
*
*
           USING RNOR,15
RNOR       STM    REGB,REGD,24(13)  SAVE REGISTERS 1,2,3
RDIGT3     L      REGB,SRGN         LOAD SRGN INTO REGB
           LR     REGC,REGB         AND INTO REGC
           SRL    REGC,15           SHIFT REGC RIGHT 15 BITS
           XR     REGB,REGC         AND XOR INTO REGB
           LR     REGC,REGB         COPY REGB INTO REGC
           SLL    REGC,17           SHIFT IT LEFT 17 BITS,
           XR     REGB,REGC         AND XOR INTO REGB
           ST     REGB,SRGN         SAVE THE NEW 'SRGN'
           L      REGD,MCGN         LOAD MCGN INTO REGD
           M      REGC,MULT         AND MULTIPLY BY 69069
           ST     REGD,MCGN         STORE RESULT,MODULO 2**32, AS NEW 'MCGN'
           XR     REGD,REGB         XOR NEW 'MCGN' AND 'SRGN' IN REGD
NRCT       SLR    REGC,REGC         ZERO OUT REGC
           CL     REGD,X68          IF REGD GE 68000000,BRANCH TO 'ND2'
           BC     11,ND2
ND1        SLDL   REGC,8            SHIFT FIRST 2 HEX DIGITS INTO REGC
           IC     REGC,NTBL(REGC)   FETCH CORRESPONDING BYTE FROM NTBL
           STC    REGC,PSTWRD+1     STORE AS 2ND BYTE OF PSTWRD
           SRL    REGD,8            TAKE REMAINING 24 BITS OF REGD
           AL     REGD,PCHAR        FORM FLOATING POINT FRACTION,CHAR X'3F'
           ST     REGD,FRAC         AND STORE AT 'FRAC'
           LE     0,PSTWRD          ADD 'PSTWRD' AND 'FRAC'
           AE     0,FRAC            LEAVING RESULT IN FPR 0
           LM     REGB,REGD,24(13)
           BCR    15,14             RETURN
ND2        CL     REGD,XD0          IF REGD GE D0000000,BRANCH TO 'ND3'
           BC     11,ND3
           SLDL   REGC,8            SHIFT FIRST 2 HEX DIGITS INTO REGC
           SL     REGC,X68R         AND SUBTRACT 00000068
           IC     REGC,NTBL(REGC)   FETCH CORRESPONDING BYTE FROM NTBL
           STC    REGC,NSTWRD+1     STORE AS 2ND BYTE OF NSTWRD
           SRL    REGD,8            TAKE REMAINING 24 BITS OF REGD
           AL     REGD,PCHAR        FORM FLOATING POINT FRACTION,CHAR X'3F'
           ST     REGD,FRAC         AND STORE AT 'FRAC'
           LE     0,NSTWRD          SUBTRACT 'FRAC' FROM 'NSTWRD'
           SE     0,FRAC            LEAVING RESULT IN FPR 0
           LM     REGB,REGD,24(13)
           BCR    15,14             RETURN
ND3        CL     REGD,XE2F         IF REGD GE E2F00000,BRANCH TO 'ND4'
           BC     11,ND4
           SLDL   REGC,12           SHIFT FIRST 3 HEX DIGITS INTO REGC
           SL     REGC,XCE8         AND SUBTRACT 00000CE8
           IC     REGC,NTBL(REGC)   FETCH CORRESPONDING BYTE FROM NTBL
           STC    REGC,PSTWRD+1     STORE AS 2ND BYTE OF PSTWRD
           SRL    REGD,8            TAKE REMAINING 20 BITS OF REGD
           AL     REGD,PCHAR        FORM FLOATING POINT FRACTION,CHAR X'3F'
           ST     REGD,FRAC         AND STORE AT 'FRAC'
           LE     0,PSTWRD          ADD 'PSTWRD' AND 'FRAC'
           AE     0,FRAC            LEAVING RESULT IN FPR 0
           LM     REGB,REGD,24(13)
           BCR    15,14             RETURN
```

```
ND4      CL    REGD,XF5E            IF REGD GE XF5E00000,BRANCH TO 'NTTHTL'
         BC    11,NTTHTL
         SLDL  REGC,12              SHIFT FIRST 3 HEX DIGITS INTO REGC
         SL    REGC,XE17            AND SUBTRACT 00000E17
         IC    REGC,NTBL(REGC)      FETCH CORRESPONDING BYTE FROM NTBL
         STC   REGC,NSTWRD+1        STORE AS 2ND BYTE OF NSTWRD
         SRL   REGD,8               TAKE REMAINING 20 BITS OF REGD
         AL    REGD,PCHAR           FORM FLOATING POINT FRACTION,CHAR X'3F'
         ST    REGD,FRAC            AND STORE AT 'FRAC'
         LE    0,NSTWRD             SUBTRACT 'FRAC' FROM 'NSTWRD'
         SE    0,FRAC               LEAVING RESULT IN FPR 0
         LM    REGB,REGD,24(13)
         BCR   15,14                RETURN
NTTHTL   ST    REGD,ARG             STORE REGD AS ARGUMENT FOR RNORTH ROUTINE
         STM   14,0,12(13)          SAVE ALL REGISTERS FROM 14 TO 3.
         LR    3,13                 COPY PREVIOUS SAVE AREA ADDRESS TO GPR3
         LA    13,SVAREA            LOAD ADDRESS OF SVAREA INTO GPR13
         ST    13,8(0,3)            STORE ADDRESS OF SVAREA IN SAVE AREA
         ST    3,4(0,13)            STORE ADDRESS OF PREVIOUS SAVE AREA
         LA    1,ARGLST             PLACE ADDRESS OF ARGUMENT LIST IN GPR 1
         L     15,ADNTH
         BALR  14,15                BRANCH TO SUBPROGRAM
         LR    13,3                 RESTORE ADDRESS OF SAVE AREA IN GPR13
         MVI   12(13),X'FF'         SET RETURN INDICATOR
RETRN3   LM    14,REGD,12(13)       RESTORE ALL REGISTERS
         BCR   15,14                RETURN
*
*        Y=REXP(0)                  RESULT IS STANDARD EXPONENTIAL VARIATE.
*
*        METHOD
*        ------
*    1.  GENERATE H1H2H3H4H5H6H7H8, 8 RANDOM HEXADECIMAL DIGITS
*
*    2.  IF H1H2 .LT. D5, SET 'REXP' TO
*                            (ETBL(H1H2)+.H3H4H5H6H7H8)/16, AND QUIT.
*    3.  IF H1H2H3 .LT. F17, SET 'REXP' TO
*                            (ETBL(H1H2H3-CFF)+.H4H5H6H7H8)/16, AND QUIT
*    4.  ELSE,GENERATE 'REXP' FROM THE EXPONENTIAL TOOTH-TAIL SUBPROGRAM.
*
         USING REXP,15
REXP     STM   REGB,REGD,24(13)     SAVE REGISTERS 1,2,3
RDIGT4   L     REGB,SRGN            LOAD SRGN INTO REGB
         LR    REGC,REGB            AND INTO REGC
         SRL   REGC,15              SHIFT REGC RIGHT 15 BITS
         XR    REGB,REGC            AND XOR INTO REGB
         LR    REGC,REGB            COPY REGB INTO REGC
         SLL   REGC,17              SHIFT IT LEFT 17 BITS,
         XR    REGB,REGC            AND XOR INTO REGB
         ST    REGB,SRGN            SAVE THE NEW 'SRGN'
         L     REGD,MCGN            LOAD MCGN INTO REGD
         M     REGC,MULT            AND MULTIPLY BY 69069
         ST    REGD,MCGN            STORE RESULT,MODULO 2**32, AS NEW 'MCGN'
         XR    REGD,REGB            XOR NEW 'MCGN' AND 'SRGN' IN REGD
ERCT     SLR   REGC,REGC            ZERO OUT REGC
         CL    REGD,XD5             IF REGD GE D5000000,BRANCH TO 'ED2'
         BC    11,ED2
ED1      SLDL  REGC,8               SHIFT FIRST 2 HEX DIGITS INTO REGC
         IC    REGC,ETBL(REGC)      FETCH CORRESPONDING BYTE FROM ETBL
```

```
         STC   REGC,PSTWRD+1      STORE AS 2ND BYTE OF PSTWRD
         SRL   REGD,8             TAKE REMAINING 24 BITS OF REGD
         AL    REGD,PCHAR         FORM FLOATING POINT FRACTION,CHAR X'3F'
         ST    REGD,FRAC          AND STORE AT 'FRAC'
         LE    0,PSTWRD           ADD 'PSTWRD' AND 'FRAC'
         AE    0,FRAC             LEAVING RESULT IN FPR 0
         LM    REGB,REGD,24(13)
         BCR   15,14              RETURN
ED2      CL    REGD,XF17          IF REGD GE F1700000,BRANCH TO 'ETTHTL'
         BC    11,ETTHTL
         SLDL  REGC,12            SHIFT FIRST 3 HEX DIGITS INTO REGC
         SL    REGC,XCFF          AND SUBTRACT 00000CFF
         IC    REGC,ETBL(REGC)    FETCH CORRESPONDING BYTE FROM ETBL
         STC   REGC,PSTWRD+1      STORE AS 2ND BYTE OF PSTWRD
         SRL   REGD,8             TAKE REMAINING 20 BITS OF REGD
         AL    REGD,PCHAR         FORM FLOATING POINT FRACTION,CHAR X'3F'
         ST    REGD,FRAC          AND STORE AT 'FRAC'
         LE    0,PSTWRD           ADD 'PSTWRD' AND 'FRAC'
         AE    0,FRAC             LEAVING RESULT IN FPR 0
         LM    REGB,REGD,24(13)
         BCR   15,14              RETURN
ETTHTL   ST    REGD,ARG           STORE REGD AS ARGUMENT FOR REXPTH ROUTINE
         STM   14,0,12(13)        SAVE ALL REGISTERS FROM 14 TO 3.
         LR    3,13               COPY PREVIOUS SAVE AREA ADDRESS TO GPR 3
         LA    13,SVAREA          LOAD ADDRESS OF SVAREA INTO GPR13
         ST    13,8(0,3)          STORE ADDRESS OF SVAREA IN SAVE AREA
         ST    3,4(0,13)          STORE ADDRESS OF PREVIOUS SAVE AREA
         LA    1,ARGLST           PLACE ADDRESS OF ARGUMENT LIST IN GPR 1
         L     15,ADETH
         BALR  14,15              BRANCH TO SUBPROGRAM
         LR    13,3               RESTORE ADDRESS OF SAVE AREA IN GPR13
         MVI   12(13),X'FF'       SET RETURN INDICATOR
RETRN4   LM    14,REGD,12(13)     RESTORE ALL REGISTERS
         BCR   15,14              RETURN
*
*        K=IUNI(0)                UNIFORMLY DISTRIBUTED POSITIVE INTEGER.
*
         USING IUNI,15
IUNI     STM   REGB,REGD,24(13)   SAVE REGISTERS 1,2,3
RDIGT5   L     REGB,SRGN          LOAD SRGN INTO REGB
         LR    REGC,REGB          AND INTO REGC
         SRL   REGC,15            SHIFT REGC RIGHT 15 BITS
         XR    REGB,REGC          AND XOR INTO REGB
         LR    REGC,REGB          COPY REGB INTO REGC
         SLL   REGC,17            SHIFT IT LEFT 17 BITS,
         XR    REGB,REGC          AND XOR INTO REGB
         ST    REGB,SRGN          SAVE THE NEW 'SRGN'
         L     REGD,MCGN          LOAD MCGN INTO REGD
         M     REGC,MULT          AND MULTIPLY BY 69069
         ST    REGD,MCGN          STORE RESULT,MODULO 2**32, AS NEW 'MCGN'
         XR    REGD,REGB          XOR NEW 'MCGN' AND 'SRGN' IN REGD
         SRL   REGD,1             SHIFT LEFT 1 BIT,LEAVING SIGN BIT ZERO
         LR    0,REGD             AND MOVE RESULT 'IUNI' TO GPR 0.
RETRN5   LM    REGB,REGD,24(13)
         BCR   15,14              RETURN
*
*        J=IVNI(0)                UNIFORMLY DISTRIBUTED INTEGER.
*
```

```
*       METHOD                              THE BASIC RANDOM NUMBER IS  A COMBINATION
*       ------                              OF TWO SEPARATELY GENERATED NUMBERS,
*                                           'SRGN' & 'MCGN' AS FOLLOWS.
*       1.  TEMP=XOR(SRGN,SRGN SHIFTED RIGHT 15 BITS)
*       2.  SRGN=XOR(TEMP,TEMP SHIFTED LEFT 17 BITS)
*       3.  MCGN=MCGN*69069,MODULO 2**32
*       4.  RESULT=XOR(MCGN,SRGN)
*
         USING IVNI,15
IVNI     STM   REGB,REGD,24(13)  SAVE REGISTERS 1,2,3
RDIGT6   L     REGB,SRGN         LOAD SRGN INTO REGB
         LR    REGC,REGB         AND INTO REGC
         SRL   REGC,15           SHIFT REGC RIGHT 15 BITS
         XR    REGB,REGC         AND XOR INTO REGB
         LR    REGC,REGB         COPY REGB INTO REGC
         SLL   REGC,17           SHIFT IT LEFT 17 BITS,
         XR    REGB,REGC         AND XOR INTO REGB
         ST    REGB,SRGN         SAVE THE NEW 'SRGN'
         L     REGD,MCGN         LOAD MCGN INTO REGD
         M     REGC,MULT         AND MULTIPLY BY 69069
         ST    REGD,MCGN         STORE RESULT,MODULO 2**32, AS NEW 'MCGN'
         XR    REGD,REGB         XOR NEW 'MCGN' AND 'SRGN' IN REGD
         LR    0,REGD            LEAVE RESULT 'IVNI' IN GPRO
RETRN6   LM    REGB,REGD,24(13)
         BCR   15,14             RETURN
*       CONSTANTS SECTION
MULT     DC       F'69069'
SRGN     DC       F'01073'
X7FF     DC       X'000007FF'
MCGN     DC       F'12345'
X1       DC       X'00000001'
FWD      DC       F'0'
Z        DC       E'0.0'
CHAR     DC       X'40000000'
SIGN     DC       X'80FFFFFF'
XD5      DC       X'D5000000'
XF17     DC       X'F1700000'
XCFF     DC       X'00000CFF'
X68      DC       X'68000000'
XD0      DC       X'D0000000'
X68R     DC       X'00000068'
XE2F     DC       X'E2F00000'
XCE8     DC       X'00000CE8'
XF5E     DC       X'F5E00000'
XE17     DC       X'00000E17'
PSTWRD   DC       X'41AA0000'
NSTWRD   DC       X'C1AA0000'
PCHAR    DC       X'3F000000'
FRAC     DC       F'0'
ARG      DS       F
ADNTH    DC       A(RNORTH)
ADETH    DC       A(REXPTH)
ARGLST   DC       X'80'
         DC       AL3(ARG)
SVAREA   DS       18F
NTBL     DC     1X'00'            TABLE USED FOR NORMAL LOOK-UP
         DC     1X'01'            FIRST PART HAS 104 ELEMENTS
         DC     2X'02'
```

```
         DC      4X'03'
         DC      5X'04'
         DC      1X'09'
         DC      5X'0A'
         DC      3X'0E'
         DC      1X'12'
         DC      1X'17'
         DC      5X'00'              START OF SECOND PART OF NORMAL TABLE
         DC      5X'01'              223 ELEMENTS
         DC      4X'02'
         DC      2X'03'
         DC      1X'04'
         DC      5X'05'
         DC      5X'06'
         DC      5X'07'
         DC      5X'08'
         DC      4X'09'
         DC      4X'0B'
         DC      4X'0C'
         DC      4X'0D'
         DC      1X'0E'
         DC      3X'0F'
         DC      3X'10'
         DC      3X'11'
         DC      2X'12'
         DC      2X'13'
         DC      2X'14'
         DC      2X'15'
         DC      2X'16'
         DC      1X'17'
         DC      1X'18'
         DC      1X'19'
         DC      1X'1A'
         DC      1X'1B'
         DC      1X'1C'
         DC      1X'1D'
         DC     10X'05'
         DC      7X'06'
         DC      5X'07'
         DC      2X'08'
         DC      9X'0B'
         DC      5X'0C'
         DC      1X'0D'
         DC     10X'0F'
         DC      7X'10'
         DC      3X'11'
         DC     12X'13'
         DC      9X'14'
         DC      5X'15'
         DC      2X'16'
         DC     13X'18'
         DC     10X'19'
         DC      7X'1A'
         DC      5X'1B'
         DC      2X'1C'
         DC     15X'1E'
         DC     13X'1F'
         DC     12X'20'
```

```
            DC      10X'21'
            DC       9X'22'
            DC       8X'23'
            DC       7X'24'
            DC       6X'25'
            DC       5X'26'
            DC       4X'27'
            DC       3X'28'
            DC       3X'29'
            DC       2X'2A'
            DC       2X'2B'
    ETBL    DC      15X'00'          START OF TABLE FOR EXPONENTIALS
            DC      13X'01'          FIRST PART HAS 213 ELEMENTS
            DC       9X'02'
            DC       5X'03'
            DC       5X'06'
            DC       8X'08'
            DC       8X'0A'
            DC       6X'0C'
            DC       2X'0E'
            DC       2X'11'
            DC       4X'15'
            DC       1X'19'
            DC       2X'20'
            DC       1X'2B'
            DC       1X'01'          SECOND PART OF EXPONENTIAL TABLE
            DC       4X'02'          455 ELEMENTS
            DC       7X'03'
            DC      11X'04'
            DC      10X'05'
            DC       5X'06'
            DC       9X'07'
            DC       1X'08'
            DC       8X'09'
            DC       7X'0B'
            DC       1X'0C'
            DC       6X'0D'
            DC       4X'0E'
            DC       5X'0F'
            DC       5X'10'
            DC       3X'11'
            DC       4X'12'
            DC       4X'13'
            DC       4X'14'
            DC       3X'16'
            DC       3X'17'
            DC       3X'18'
            DC       2X'19'
            DC       2X'1A'
            DC       2X'1B'
            DC       2X'1C'
            DC       2X'1D'
            DC       2X'1E'
            DC       2X'1F'
            DC       1X'21'
            DC       1X'22'
            DC       1X'23'
            DC       1X'24'
```

```
         DC      1X'25'
         DC      1X'26'
         DC      1X'27'
         DC      1X'28'
         DC      1X'29'
         DC      1X'2A'
         DC      5X'05'
         DC      2X'07'
         DC      1X'09'
         DC      1X'0B'
         DC      4X'0D'
         DC      9X'0F'
         DC      3X'10'
         DC     10X'12'
         DC      5X'13'
         DC      9X'16'
         DC      6X'17'
         DC      2X'18'
         DC     13X'1A'
         DC     10X'1B'
         DC      7X'1C'
         DC      5X'1D'
         DC      2X'1E'
         DC     13X'21'
         DC     11X'22'
         DC      9X'23'
         DC      8X'24'
         DC      6X'25'
         DC      5X'26'
         DC      4X'27'
         DC      2X'28'
         DC      1X'29'
         DC     15X'2C'
         DC     14X'2D'
         DC     13X'2E'
         DC     12X'2F'
         DC     11X'30'
         DC     11X'31'
         DC     10X'32'
         DC      9X'33'
         DC      9X'34'
         DC      8X'35'
         DC      8X'36'
         DC      7X'37'
         DC      7X'38'
         DC      6X'39'
         DC      6X'3A'
         DC      6X'3B'
         DC      5X'3C'
         DC      5X'3D'
         DC      4X'3E'
         DC      4X'3F'
         END     RANDOM
C RNOR TOOTH FUNCTION
      FUNCTION RNORTH(K)
      DIMENSION C(45)
      DATA C/Z40FD2B5F,Z40FD2B5F,Z40FAA9AD,Z40F5A648,Z40F32496,
     $ Z40EE2131,Z40E69C1A,Z40E198B5,Z40DA139E,Z40D28E87,Z40C887BE,
```

```
     $ Z40C102A6,Z4086FBDD,Z40ACF513,Z40A2EE4A,Z4098E780,Z40916269,
     $ Z40875BA0,Z407D54D6,Z40734E0D,Z406BC8F6,Z4061C22C,Z405A3D15,
     $ Z405287FE,Z404B32E7,Z4043ADD0,Z403C28B9,Z40372554,Z402FA03D,
     $ Z402A9CD8,Z40259973,Z4020960E,Z401E145C,Z401910F7,Z40168F45,
     $ Z40140D93,Z40118BE0,Z3FF0A2E4,Z3FC887BE,Z3FA06C98,Z3F785172,
     $ Z3F785172,Z3F50364C,Z3F50364C,Z3F50364C/
      DATA I1/ZFBC35400/,I2/ZFE79702E/
      IF(K.GT.I1)GO TO 3
      S=UNI(0)
      T=UNI(0)
      B=AINT(7.*(S+T)+37.*ABS(S-T))
      X=UNI(0)-UNI(0)
      RNORTH=.0625*(X+SIGN(B,X))
      RETURN
 3    IF(K.GT.I2)GO TO 5
 4    RNORTH=2.75*VNI(0)
      J=16.*ABS(RNORTH)+1.
      IF(J-14) 6,6,7
   6  P=(J+J-1)*.1497466E-2
      GO TO 8
   7  P=(89-J-J)*.698817E-3
 8    IF(UNI(0).GT.79.78846*(EXP(-.5*RNORTH*RNORTH)
     $     -C(J)-P*(J-16.*ABS(RNORTH)))) GOTO4
      RETURN
 5    V=VNI(0)
      IF(V.EQ.0) GO TO 5
      X=SQRT(7.5625-2.*ALOG(ABS(V)))
      IF(UNI(0)*X.GT.2.75)GO TO 5
      RNORTH=SIGN(X,V)
      RETURN
      END
C     REXP TOOTH FUNCTION
      FUNCTION REXPTH(K)
      DIMENSION C(65)
      DATA C/Z40F00000,Z40E10000,Z40D40000,Z40C70000,Z40BB0000,
     $ Z40AF0000,Z40A50000,Z409B0000,Z40910000,Z40890000,Z40800000,
     $ Z40780000,Z40710000,Z406A0000,Z40640000,Z405E0000,Z40580000,
     $ Z40530000,Z404E0000,Z40490000,Z40440000,Z40400000,Z403C0000,
     $ Z40390000,Z40350000,Z40320000,Z402F0000,Z402C0000,Z40290000,
     $ Z40270000,Z40240000,Z40220000,Z40200000,Z401E0000,Z401C0000,
     $ Z401A0000,Z40190000,Z40170000,Z40160000,Z40150000,Z40130000,
     $ Z40120000,Z40110000,Z40100000,Z3FF00000,Z3FE00000,Z3FD00000,
     $ Z3FC00000,Z3FB00000,Z3FB00000,Z3FA00000,Z3F900000,Z3F900000,
     $ Z3F800000,Z3F800000,Z3F700000,Z3F700000,Z3F600000,Z3F600000,
     $ Z3F600000,Z3F500000,Z3F500000,Z3F400000,Z3F400000,Z3F400000/
      DATA I1/ZFB4FAA91/
      IF(K.GT.I1)GO TO 5
 1    U1=UNI(0)
      IF(U1.GT..7917049) GO TO 3
      T=1.-1.239962*U1
      REXPTH=-ALOG(T)
      J=16.*REXPTH+1.
      IF(UNI(0)*(.0604*T+.0039).GT.T-C(J))GOTO1
      RETURN
 3    REXPTH=19.20352*U1-15.20352
      J=16.*REXPTH+1.
      EX=EXP(-REXPTH)
      IF(UNI(0)*(.0604*EX+.0039).GT.EX-C(J))GOTO1
```

```
      RETURN
 5    REXPTH=4.-ALOG(UNI(0))
      RETURN
      END
```

APPENDIX D

Denotations ***: $p << .01$

**: $p < .01$

*: $p < .05$

TABLE D.1: Tests of Signifiance on the Parameter Estimates of the 12 Multiple Regression Equations of Table 3.5

BLOCK 1 $\alpha = 0.05$, $k = 5$, $n = 15$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_o$	106	33.99***
$\hat{\beta}_1$	106	18.84***
$\hat{\beta}_2$	106	-16.35***
$\hat{\beta}_3$	106	- 1.77

Multiple Correlation Coefficient (R)

$F = 1455.74$*** $DF_1 = 3$ $DF_2 = 106$

BLOCK 2 $\alpha = 0.05$, $k = 5$, $n = 5$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_o$	106	28.06***
$\hat{\beta}_1$	106	19.74***
$\hat{\beta}_2$	106	-12.16***
$\hat{\beta}_3$	106	- 4.78**

Multiple Correlation Coefficient (R)

$F = 1213.63$*** $DF_1 = 3$ $DF_2 = 106$

APPENDIX D (cont'd)

BLOCK 3 $\alpha = 0.05$, $k = 3$, $n = 15$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_o$	106	32.05***
$\hat{\beta}_1$	106	15.67***
$\hat{\beta}_2$	106	-18.12**
$\hat{\beta}_3$	106	- 2.25*

Multiple Correlation Coefficient (R)

$F = 983.54$*** $DF_1 = 3$ $DF_2 = 106$

BLOCK 4 $\alpha = 0.05$, $k = 3$, $n = 5$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_o$	106	30.20***
$\hat{\beta}_1$	106	23.74***
$\hat{\beta}_2$	106	-10.18***
$\hat{\beta}_3$	106	- 7.16***

Multiple Correlation Coefficient (R)

$F = 1387.57$*** $DF_1 = 3$ $DF_2 = 106$

BLOCK 5 $\alpha = 0.05$, $k = 2$, $n = 15$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_o$	44	25.80***
$\hat{\beta}_1$	44	3.76**
$\hat{\beta}_2$	44	- 7.08***
$\hat{\beta}_3$	44	- 1.20

Multiple Correlation Coefficient (R)

$F = 161.26$*** $DF_1 = 3$ $DF_2 = 44$

APPENDIX D (cont'd)

BLOCK 6 $\alpha = 0.05$, $k = 2$, $n = 5$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_o$	44	31.74***
$\hat{\beta}_1$	44	9.08***
$\hat{\beta}_2$	44	- 9.22***
$\hat{\beta}_3$	44	- 2.06*

Multiple Correlation Coefficient (R)

$F = 380.05^{***}$ $DF_1 = 3$ $DF_2 = 44$

BLOCK 1 $\alpha = 0.01$, $k = 5$, $n = 15$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_o$	106	12.68***
$\hat{\beta}_1$	106	24.8 ***
$\hat{\beta}_2$	106	- 8.1 ***
$\hat{\beta}_3$	106	- 7.4 ***

Multiple Correlation Coefficient (R)

$F = 1330.00^{***}$ $DF_1 = 3$ $DF_2 = 106$

BLOCK 2 $\alpha = 0.01$, $k = 5$, $n = 5$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_o$	106	9.50***
$\hat{\beta}_1$	106	27.69***
$\hat{\beta}_2$	106	- 5.00**
$\hat{\beta}_3$	106	-10.70***

Multiple Correlation Coefficient (R)

$F = 1356.42^{***}$ $DF_1 = 3$ $DF_2 = 106$

APPENDIX D (cont'd)

<u>BLOCK 3</u> $\alpha = 0.01,\ \kappa = 3,\ n = 15$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_0$	106	10.85***
$\hat{\beta}_1$	106	18.15***
$\hat{\beta}_2$	106	- 5.38***
$\hat{\beta}_3$	106	- 6.09***

Multiple Correlation Coefficient (R)

$F = 690.66^{***}$ $\quad DF_1 = 3$ $\quad DF_2 = 106$

<u>BLOCK 4</u> $\alpha = 0.01,\ \kappa = 3,\ n = 5$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_0$	106	7.91***
$\hat{\beta}_1$	106	28.18***
$\hat{\beta}_2$	106	- 2.84**
$\hat{\beta}_3$	106	-11.57***

Multiple Correlation Coefficient (R)

$F = 1260.22^{***}$ $\quad DF_1 = 3$ $\quad DF_2 = 106$

<u>BLOCK 5</u> $\alpha = 0.01,\ \kappa = 2,\ n = 15$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_0$	44	11.01***
$\hat{\beta}_1$	44	4.67**
$\hat{\beta}_2$	44	- 3.28**
$\hat{\beta}_3$	44	- 2.66**

Multiple Correlation Coefficient (R)

$F = 87.41^{***}$ $\quad DF_1 = 3$ $\quad DF_2 = 44$

APPENDIX D (cont'd)

BLOCK 6 $\alpha = 0.01$, $K = 2$, $n = 5$

Regression Coefficients	D.F.	T-value
$\hat{\beta}_o$	44	9.29***
$\hat{\beta}_1$	44	7.89***
$\hat{\beta}_2$	44	- 2.75**
$\hat{\beta}_3$	44	- 3.24**

Multiple Correlation Coefficient (R)

$F = 135.23$*** $DF_1 = 3$ $DF_2 = 44$

TABLE D.2: Analysis of Variance for Testing the Equality of Six Regression Equations at Each Level of α

$\alpha = .05$

Due to	DF	SS	F
Dev. from hypothesis	20	99017.96	480.32***
Separate regressions (residual)	512	101434.00	
Common regression (residual)	532	200452.00	

$\alpha = .01$

Due to	DF	SS	F
Dev. from hypothesis	20	43913.85	80.57***
Separate regressions (residual)	512	20452.55	
Common regression (residual)	532	64366.40	

APPENDIX D (cont'd)

TABLE D.3: Analysis of Variance for Testing the Equality of Three Regression Equations Corresponding to the Three Values of κ, at each Level of α

$\alpha = .05$

Due to	DF	SS	F
Dev. from hypothesis	8	36870.00	14.76***
Separate regressions (residual)	524	163582.00	
Common regression (residual)	536	200452.00	

$\alpha = .01$

Dev. from hypothesis	8	26030.5	44.47***
Separate regressions (residual)	524	38335.87	
Common regression (residual)	536	64366.4	

APPENDIX D (cont'd)

TABLE D.4: Analysis of Variance for Testing the Equality of Pairs of Regression Equations Corresponding to the Two Values of n at Each Level of κ, at Each Level of α

$\alpha = .05, \kappa = 2$

Due to	DF	SS	F
Dev. from hypothesis	4	10066.66	21.14***
Separate regressions (residual)	88	10477.24	
Common regression (residual)	92	20543.90	

$\alpha = .05, \kappa = 3$

Due to	DF	SS	F
Dev. from hypothesis	4	40227.00	52.20***
Separate regressions (residual)	212	40899.50	
Common regression (residual)	216	81126.50	

$\alpha = .05, \kappa = 5$

Due to	DF	SS	F
Dev. from hypothesis	4	11854.30	12.50***
Separate regressions (residual)	212	50057.30	
Common regression (residual)	216	61911.60	

APPENDIX D (cont'd)

$\alpha = .01,\ k = 2$

Due to	DF	SS	F
Dev. from hypothesis	4	1512.11	12.75***
Separate regressions (residual)	88	2609.66	
Common regression (residual)	92	4121.77	

$\alpha = .01,\ k = 3$

Due to	DF	SS	F
Dev. from hypothesis	4	7666.97	34.23***
Separate regressions (residual)	212	11869.63	
Common regression (residual)	216	19536.60	

$\alpha = .01,\ k = 5$

Due to	DF	SS	F
Dev. from hypothesis	4	1614.42	6.55***
Separate regressions (residual)	212	13063.08	
Common regressions (residual)	216	14677.50	