1977

The Influence Of The Reliability Of The Dependent Variable On Statisticalpower.

Theresa M. Housden
University of the Pacific

THE INFLUENCE OF THE RELIABILITY OF THE DEPENDENT VARIABLE ON STATISTICAL POWER

A Dissertation

Presented to

the Graduate Faculty of the

University of the Pacific

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

---

by

Theresa M. Housden

May 1977

This dissertation, written and submitted by

THERESA M. HOUSDEN

is approved for recommendation to the Committee on Graduate Studies, University of the Pacific

Dean of the School or Department Chairman:

Oscar T. Jarvis

Dissertation Committee:

Roger L. Reimer, Chairman

Robert D. Morrow

B. R. Hopkins

[illegible]

Carolyn M. Fowle

Dated April 20, 1977

THE INFLUENCE OF THE RELIABILITY OF THE DEPENDENT VARIABLE ON STATISTICAL POWER

Abstract of Dissertation

PROBLEM: In the planning phase of developing quality educational research there are several critical decision points. Some of these decisions are made to ensure that the research will have adequate statistical power, that is, the capability to detect meaningful differences, if they do exist. In an unpublished article, K. D. Hopkins from the Laboratory of Educational Research, University of Colorado and B. R. Hopkins from the University of the Pacific commented that all theoretical and empirical options for increasing statistical power had been investigated except for the reliability of the dependent variable. In this article, they derived a method for estimating the statistical power of the analysis of variance (ANOVA) and the analysis of covariance (ANCOVA) which factored out the influence of the reliability of the dependent variable. This method involved modifying the non-centrality parameters used in estimating the power of the ANOVA and ANCOVA tests. This modification required some untested mathematical assumptions and therefore needed to be empirically verified.

PURPOSE: The purpose of this study was to empirically verify the method for estimating the statistical power of the ANOVA and the ANCOVA tests developed by Hopkins and Hopkins. An additional purpose was to investigate the influence of the reliability of the dependent variable on the statistical power of the ANOVA and ANCOVA tests.

PROCEDURES: A Monte Carlo computer simulation method was used to estimate the actual power values of the ANOVA and ANCOVA tests. A comparison was then made to the power values derived by using the Hopkins and Hopkins method. By using the power values required for this comparison, it was also possible to determine the influence of the reliability of the dependent variable on statistical power for the ANOVA and ANCOVA tests.

FINDINGS: It was found that the method developed by Hopkins and Hopkins to estimate the statistical power of the ANOVA and ANCOVA tests provides a good estimate of power. It was also found that as test reliability increases by .1, the statistical power of ANOVA increases by about .06 or .07 and the statistical power of the ANCOVA increases by approximately .10 or .12 when using the pretest as a covariate.

RECOMMENDATIONS:

For Researchers

1. Greater emphasis should be placed on developing and using highly reliable dependent variables in educational research.

2. If a covariate with an anticipated correlation with the

variate of .5 or more is available, the ANCOVA test rather than the ANOVA test should be considered whenever appropriate.

3. Statistical power should be estimated routinely in the planning phase of behavioral research.

For Future Studies

1. Computer programs to estimate the statistical power of the ANOVA and ANCOVA tests should be included in the commonly used computer software packages.

2. Power charts for .10 level of significance need to be developed and included in research textbooks along with .05 and .01 levels of significance.

ACKNOWLEDGMENTS

The author is deeply indebted to Dr. B. R. Hopkins for the opportunity to conduct this study of statistical power. It has been exceptionally rewarding to work with Dr. Hopkins on such a challenging topic. His enthusiastic guidance and technical support during the study facilitated its successful completion.

I offer sincere appreciation to Dr. Roger L. Reimer, my dissertation chairman, for his continuous support, assistance and console throughout this research process.

To other members of my dissertation committee--Dr. Robert Morrow, Dr. Carolyn Fowle, and Dr. Lee C. Fennell, I gratefully acknowledge their helpful suggestions during the writing of the dissertation.

I would like to offer a special thanks to Dr. Thomas C. Coleman, Chairman of the Department of Educational Administration, for his never ending personal support during my program of doctoral studies.

I would also like to thank my parents and my sister for their patience and loving support during the period of this study.

To my husband, Dr. Jack Housden, who encouraged me to pursue doctoral studies and provided continuous encouragement throughout the doctoral program, I dedicate this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION TO THE STUDY

In 1923, W. A. McCall published How to Experiment in Education.[1] Two years later, R. A. Fisher published his classic, Statistical Methods for Research Workers.[2] These landmark works established the foundation for current behavioral research and its methodology. About forty years later, Donald T. Campbell and Julian C. Stanley wrote a chapter for the Handbook of Research on Teaching. That chapter, "Experimental and Quasi-experimental Designs for Research on Teaching," quickly became the guidebook for those involved in behavioral research and specifically educational research. They stated that the experiment is

> . . . the only means for settling disputes regarding educational practice, as the only way of verifying educational improvements, and as the only way of establishing a cumulative tradition in which improvements can be introduced without the danger of a faddish discard of old wisdom in favor of inferior novelties.[3]

Furthermore, Isaac and Michael stated that "in educational assessment and decision-making, it [research] is the only way to make rational choices

---

[1] W. A. McCall, How to Experiment in Education (New York: Macmillan, 1923).

[2] R. A. Fisher, Statistical Methods for Research Workers (1st ed.; London: Oliver & Boyd, 1925).

[3] Donald T. Campbell and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," Handbook of Research on Teaching, ed. N. L. Gage (Chicago: Rand McNally, 1963), p. 172.

between alternative practices, to validate educational improvements . . ."[4]

Educational experiments, therefore, potentially play a significant role in educational decision making. Because of this role, the quality of these experiments becomes crucial. High quality experiments may lead to better educational decisions and a more effective educational process.

RATIONALE FOR THE STUDY

In the planning phase of developing quality educational research there are several critical decision points. Some of these decisions are made to ensure that the research will have adequate statistical power, that is, the ability to detect meaningful differences, if they do exist. J. K. Brewer stated, "It is almost universally accepted by educational researchers that the power of a statistical test is important and should be substantial."[5] However, according to Brewer, the power of the statistical tests used in research studies published in the American Educational Research Journal as well as other journals in the behavioral sciences tends to be dangerously low.[6] This criticism identified a need for more attention to the power of statistical tests used in educational research.

In 1973, Kenneth D. Hopkins outlined several methods of increasing statistical power. One method was to increase the reliability of the

---

[4]Steven Isaac and William B. Michael, Handbook in Research and Evaluation (San Diego: Robert R. Knapp, 1972), p. iii.

[5]J. K. Brewer, "On the Power of Statistical Tests in the American Educational Research Journal," American Education Research Journal, IX, No. 3 (1972), p. 391.

[6]Ibid., pp. 394-95.

criteria or dependent variable. He commented that:

> Although methods for quantifying the precise effects on power of changes in the reliability of the dependent variable are not readily available, it is certain that power is increased by a gain in reliability of the dependent variable.[7]

In an unpublished article, K. D. Hopkins and B. R. Hopkins commented that all theoretical and empirical options for increasing statistical power had been investigated except for the reliability of the dependent variable. In this article they derived a method for estimating the power of statistical tests (specifically for the analysis of variance [ANOVA] and the analysis of covariance [ANCOVA]) which included the reliability of the dependent variable.[8]

THE PROBLEM AND ITS SIGNIFICANCE

The interaction between the reliability of the dependent variable and power was expressed in theoretical terms by Hopkins and Hopkins. This theoretical relationship was based on several mathematical assumptions and because of these assumptions, an empirical verification of the derived relationship was required.[9] The purpose of this study was to empirically verify the hypothesized relationship between the reliability of the dependent variable and statistical power. Other factors considered

---

[7]Kenneth D. Hopkins, "Research Design and Analysis Clinic: Preventing the Number-One Misinterpretation of Behavioral Research, or How to Increase Statistical Power," The Journal of Special Education, VII, No. 1 (1973), p. 106.

[8]K. D. Hopkins and B. R. Hopkins, "The Effect of the Reliability of the Dependent Variable on the Power of the Analysis of Variance" (unpublished paper, University of Colorado, Boulder; and University of the Pacific, Stockton, 1974), pp. 1-12.

[9]Ibid., pp. 5-12.

were the sample size, number of groups, the level of significance, and the magnitude of the treatment effect.

Frequently, the educational researcher has either ignored the issue of statistical power when conducting research or has been poorly armed to deal with the problem.[10] Several authors have specified techniques for increasing statistical power; however, as previously noted, the influence of the reliability of the dependent variable had generally not been considered.

It was hoped that this study would provide adequate information so that a researcher could easily identify and maximize the statistical power of his ANOVA and ANCOVA tests. If statistical power were estimated routinely during the planning phase of research, studies without adequate statistical power could be identified. Also, unnecessarily large and elaborate studies could be reduced with substantial financial savings and still maintain the ability to reject the null hypothesis when it is false, that is, adequate statistical power.[11] Specifically, researchers want their statistical tests to have high statistical power. They also want to know the minimum effort required to achieve this level of statistical power.[12] This study focused on the effects on statistical power of increasing or decreasing the reliability of the dependent variable and of varying other relevant parameters concurrently (see Table 1, page 5).

METHOD

The method used to accomplish this study involved "Monte Carlo"

---

[10]Brewer, loc. cit. [11]Hopkins and Hopkins, op. cit., p. 1.

[12]Ibid., p. 12.

techniques. Samples of random numbers were generated by the Burroughs B-6700 computer, manipulated, and analyzed by both ANOVA and ANCOVA procedures. This process was repeated one thousand times for each unique combination of parameters and the results were accumulated. Systematically, the relevant parameters were manipulated and the resulting statistical power estimated. In Table 1, the parameters that were varied and their respective minimums, maximums and increment levels are identified.

Table 1

Parameters of the Study and Their Minimums, Maximums and Increment Levels

	Minimum	Maximum	Increment level
1. Reliability of the dependent variable	.3	.9	.2
2. Sample size	6	100	varies
3. Noncentrality parameter	1.0	varies	varies
4. Level of significance	.01	.10	varies
5. Number of groups	2	5	varies

DELIMITATIONS OF THE STUDY

This study was delimited by the number of cases examined. Since it was impractical to vary every parameter infinitely, the following cases were selected because they represented the majority of practical applications: the reliability of the dependent variable varied from 0.3 to 0.9; the sample size varied from 6 to 100; the non-centrality parameter increased upward from 1.0; the level of significance varied from 0.01 to

0.05; and the number of groups varied from 2 to 5.

DEFINITIONS

The following operational definitions are provided for this study. Other definitions will be given in the text as required.

The "degrees of freedom" is the number of independent observations. That is, the total number of observations minus the number of restrictions on the observations.[13]

The "level of significance" is the probability of obtaining a test statistic that falls within the critical region when the null hypothesis is true.[14]

The "Monte Carlo" technique consists of simulating an experiment to determine some probabilistic property of a population of objects or events by the use of random sampling applied to the components of the objects of events.[15]

The "non-centrality parameter" (ϕ)[16] is: $\phi = \sqrt{\dfrac{n\Sigma_i\ (\mu_i-\mu)^2/J}{\sigma_\omega^2}}$

where n = group sample size,
μ_i = mean of group i and i varies from 1 to J,
μ = grand mean,
J = number of groups, and
σ_ω^2 = within group variance.

The "null hypothesis" (H_0) is the hypothesis of "no relationship or difference."[17] It is the one actually tested statistically.

The "power" of a statistical test is the probability of rejecting the null hypothesis when it is false. It is mathematically, a function of the degrees of freedom, the level of significance and the

---

[13] Jerome L. Myers, Fundamentals of Experimental Design (2d ed.; Boston: Allyn and Bacon, Inc., 1972), p. 82.

[14] Ibid., p. 48.

[15] C. West Churchman, Russel L. Ackoff, and E. Leonard Arnoff, Introduction to Operations Research (New York: John Wiley & Sons, Inc., 1957), p. 175.

[16] Myers, op. cit., p. 89.

[17] Isaac and Michael, op. cit., p. 142.

non-centrality parameter.[18]

"Reliability" is an index of the consistency between measurements in a series. The reliability coefficient tells what proportion of the test variance is nonerror variance.[19]

OVERVIEW

The purpose of this study was to empirically investigate the interaction between the reliability of the dependent variable and the statistical power of the analysis of variance and the analysis of covariance. The influence of other relevant parameters, namely, sample size, non-centrality parameter, level of significance and number of groups were considered. This purpose was accomplished by using "Monte Carlo" methods and the Burroughs B-6700 computer. Systematically, the relevant parameters were manipulated, and the resulting statistical power of the ANOVA and ANCOVA procedures estimated.

---

[18] Myers, op. cit., p. 49.

[19] J. C. Stanley and K. D. Hopkins, Educational and Psychological Measurement and Evaluation (Englewood Cliffs, New Jersey: Prentice Hall, 1972), p. 456.

Chapter 2

REVIEW OF THE LITERATURE

A purpose of this study, as noted in Chapter 1, was to empirically verify the hypothesized relationship between the reliability of the dependent variable and statistical power. Statistical power is not a well understood phenomena, and consequently, there is a scarcity of published research on the topic. The intention of the review of literature in this chapter, therefore, was to acquaint the reader with existing studies of statistical power, as well as to establish the major operational concepts required by the study. The sources consulted in this review are summarized in Appendix A.

The first portion of this chapter deals with statistical power and its relation to the purpose of the study. An overview of statistical power, beginning with the Pearson-Hartley power charts of 1951,[1] is presented which emphasizes the concept of statistical power, followed by attention to research on how reliability (errors of measurement) relates to power. The second portion of the chapter addressed two major research tools, reliability and Monte Carlo methods. These two concepts are fundamental to the methodology of the study.

---

[1]E. S. Pearson and H. O. Hartley, "Charts of the Power Function for Analysis of Variance Tests, Derived from the Non-Central F-Distribution," Biometrika, XXXVIII, No. 1 (1951), 112-30.

STATISTICAL POWER

The first major contribution to the field of statistical power was made by E. S. Pearson and H. O. Hartley in 1951 when they published the analysis of variance power charts.[2] From this point on, statistical power was expressed as a function of sample size, number of groups, level of significance, and magnitude of the treatment effect (the non-centrality parameter). Various applications and implications of this approach were later expressed in publications about power by J. K. Brewer,[3] J. Cohen,[4] K. D. Hopkins,[5] as well as others.

Since a major purpose in conducting educational research is to detect differences between sample groups when a difference does exist, high statistical power is important to educators. J. Cohen, a leading statistical power theorist, proposed that .80 should be the minimum level of power of statistical tests for research in the behavioral sciences.[6] Unfortunately, most research in the behavioral science, has been conducted with power estimates far below the .80 level. For example, Brewer surveyed all articles published in the American Educational Research

---

[2] Ibid.

[3] J. K. Brewer, "On the Power of Statistical Tests in the American Educational Research Journal," American Educational Research Journal, IX, No. 3 (1972), 391-401.

[4] J. Cohen, "Approximate Power and Sample Size Determination for Common One-Sample and Two-Sample Hypothesis Tests," Educational and Psychological Measurement, XXX, No. 4 (1970), 811-31.

[5] K. D. Hopkins, "Research Design and Analysis Clinic: Preventing the Number-One Misinterpretation of Behavioral Research, or How to Increase Statistical Power," The Journal of Special Education, VII, No. 1 (1973), 105-6.

[6] Cohen, op. cit., p. 825.

Journal from November 1969 to May 1971. He studied the power of the 373 statistical tests which were reported as statistically significant and found that the average power values for 205 F and t tests were .13, .47, and .73 for small, medium, and large effect sizes.[7] Although these studies did report significant differences, the probability of discovering these differences was low and below Cohen's .80 level of adequate power.

An Overview of Statistical Power

Statistical power is a complex phenomenon to understand. Conceptually, statistical power is "the probability of detecting a [specified] difference in sample values when indeed the difference exists at some level in the population,"[8] that is, the probability of rejecting the null hypothesis, when there is a difference. Power can be represented graphically as the shaded area in Figure 1, page 11.

Computationally, power can be estimated by a two step process. First the non-centrality parameter (ϕ) is computed and then the power value is read from the Pearson-Hartley power charts using the appropriate degrees of freedom and level of significance.

Methods to increase statistical power have been outlined by various authors. K. D. Hopkins presented several ways to increase power such as, increasing the total n, relaxing the level of significance,

---

[7] Brewer, op. cit., pp. 394-95; see also J. Cohen, "The Statistical Power of Abnormal Social Psychological Research," Journal of Abnormal and Social Psychology, LXV, No. 1 (1962), 145-53; see also R. F. Haase, "Power Analysis in Research in Counselor Education," Counselor Education and Supervision, XIV, No. 2 (1974), 124-32; see also B. J. Jones and J. K. Brewer, "An Analysis of the Power of Statistical Tests Reported in the 'Research Quarterly'," The Research Quarterly, XLIII, No. 1 (1972), 7-15.

[8] Haase, op. cit., p. 126.

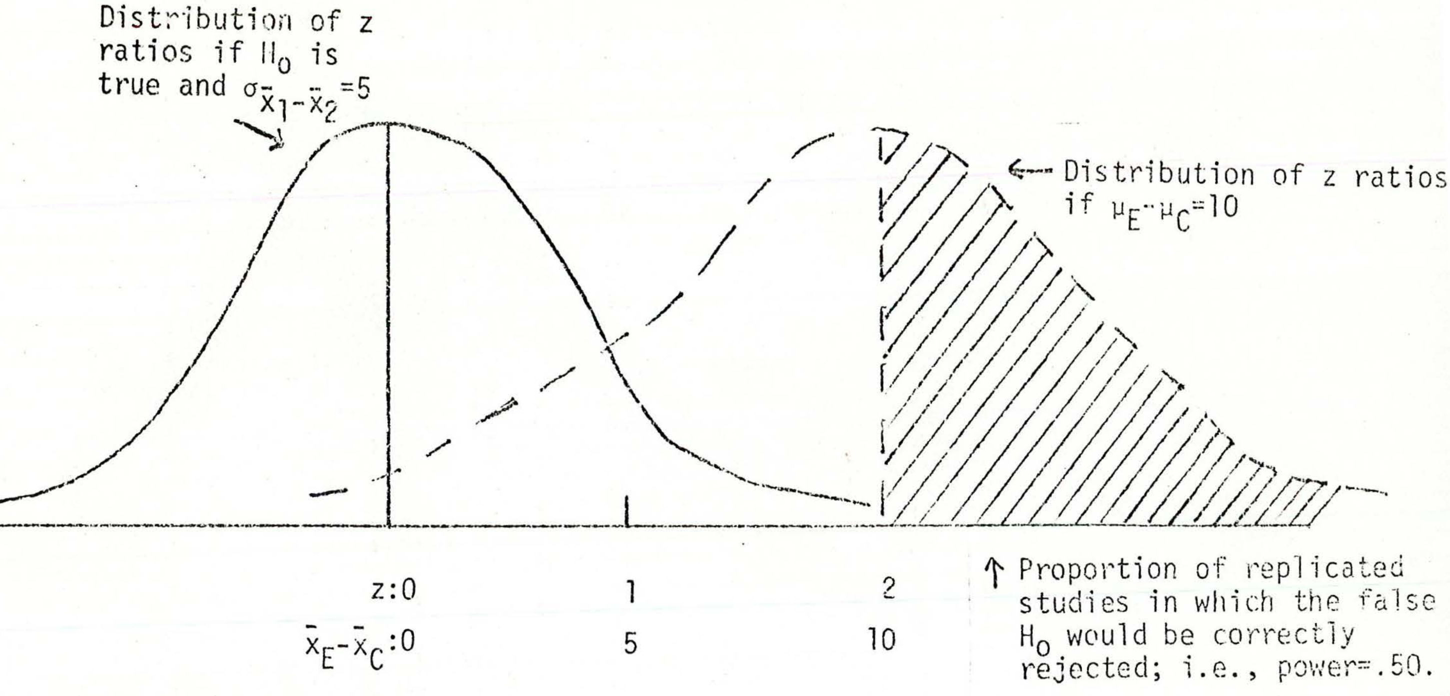


Figure 1

An Illustration of the Probability of Failing to Reject a False Null Hypothesis When the True Difference in Population Means is $2\sigma_{\bar{X}_1-\bar{X}_2}$[9]

In this illustration $\sigma_{\bar{X}_1-\bar{X}_2}= 5$ and there is 10 IQ-point treatment effect, yet the probability that H_0 will be rejected is only .5.

---

[9]Hopkins, op. cit., p. 104.

making a directional hypothesis (use a one-tailed test), using a more efficient experimental design, increasing the potency of the treatment, and increasing the reliability of the dependent variable. He also indicated that all theoretical and empirical options for increasing statistical power had been investigated except for the reliability of the dependent variable.[10]

The Influence of Test Reliability on Power

Traditionally, statistical procedures were developed for models where the variables were assumed to be free from the errors of measurement.[11] Specifically, discussions of statistical power implicitly assumed that the observations were errorless or "true" measures.[12] Unfortunately, most educational research involves the measurement of variables with considerable error. In 1958, J. P. Sutcliffe indicated that measurement error does in fact decrease statistical power.[13] However, he gave no way to assess this loss of power. The reliability of the dependent variable is inversely related to the errors of measurement.[14] Consequently, the

---

[10]Hopkins, op. cit., pp. 105-6.

[11]A. C. Porter, How Error of Measurement Affect ANOVA, Regression Analysis, ANCOVA and Factor Analysis. U.S. Educational Resources Information Center, ERIC Document ED 050 172, January, 1971, p. 23.

[12]T. A. Cleary and R. L. Linn, "Error of Measurement and the Power of a Statistical Test," British Journal of Mathematical and Statistical Psychology, XXII, No. 1 (1969), 49.

[13]J. P. Sutcliffe, "Error of Measurement and the Sensitivity of a Test of Significance," Psychometrika, XXIII, No. 1 (1958), 9.

[14]J. C. Stanley and K. D. Hopkins, Educational and Psychological Measurement and Evaluation (Englewood Cliffs, New Jersey: Prentice Hall, 1972), p. 118.

reliability of the dependent variable must also affect power.

Various approaches have been used to augment power by increasing the reliability of the dependent variable. In 1969, T. A. Cleary and R. L. Linn developed a cost model for the F-test and provided a procedure to optimize power by varying the number of subjects and the test length.[15] Since the length of a test directly influenced the test's reliability, this was one of the first studies to hint at the influence of the reliability of the dependent variable on power. In 1973, K. D. Hopkins indicated that methods for quantifying the precise effects of changes in the reliability of the dependent variable on statistical power were not readily available.[16] In an unpublished article written in 1974, Hopkins and Hopkins addressed this problem. In this article they derived formulas for estimating the non-centrality parameters and, consequently, the power of the ANOVA and ANCOVA statistical tests which were a function of test reliability.[17]

Hopkins and Hopkins Methods

The methods developed by Hopkins and Hopkins to estimate the power of the ANOVA and ANCOVA tests as a function of reliability required a reformulation of the computation of the non-centrality parameter. For the ANOVA, a non-centrality parameter (ϕ_A) is computed as a function of the reliability of the dependent variable (r_{XX}) and the non-centrality

---

[15] Cleary and Linn, op. cit., pp. 49-50.

[16] Hopkins, op. cit., p. 106.

[17] K. D. Hopkins and B. R. Hopkins, "The Effect of the Reliability of the Dependent Variable on the Power of the Analysis of Variance" (unpublished paper, University of Colorado, Boulder; and University of the Pacific, Stockton, 1974), pp. 1-12.

parameter for perfect test reliability ($\phi_{1.0}$). Specifically, the formula for ϕ_A is:

$$\phi_A = \phi_{1.0} \sqrt{r_{xx}}$$

where r_{xx} = the reliability of the dependent variable,

$$\phi_{1.0} = \sqrt{\frac{n\sum_{i=1}^{J}(\mu_i-\mu)^2}{J\,\sigma_t^2}}$$

(That is, the non-centrality parameter assuming perfect reliability of the dependent variable.),

n = group sample size,

$\sigma_t^2 = \sigma_\omega^2 - \sigma_e^2$, where $\sigma_e^2 = 0$,

J = number of groups,

μ_i = mean of group i, and

μ = grand mean.[18]

After the non-centrality parameter (ϕ_A) is computed, the regular procedures for estimating power are applied. That is, the derived non-centrality parameter is used with the Pearson-Hartley power charts to estimate power.

The non-centrality parameter for the ANCOVA test was based on the following computational formula:

$$\phi_C = \sqrt{\frac{\phi_A^2}{1-r_{xy}^2}}$$

where r_{xy} = the correlation between the variate and covariate (frequently the covariate is the pretest and the variate is the posttest), and

ϕ_A = the non-centrality parameter for the ANOVA test at a specific reliability level.[19]

---

[18]Ibid., p. 6. [19]Ibid., p. 11.

During the derivation of the formulas for the non-centrality parameters for the ANOVA and ANCOVA, several assumptions were necessary. Most of these assumptions were the common assumptions required for the ANOVA and ANCOVA tests such as, normally and independently distributed observations, and homogeneity of variance. However, in order to make the non-centrality parameter a function of the reliability of the dependent variable, an additional assumption was required.

One of the complex formulas required in order to derive ϕ_A and ϕ_C was:

$$E\ (F) = \left[\frac{J\ (n-1)}{J\ (n-1)-2} \right] \left[1 + \frac{n \sum_{i=1}^{J} (\mu_i - \mu)^2}{\sigma_\omega^2\ (J-1)} \right]$$

This formula was then viewed as two factors,

$$\text{factor A} = \frac{J\ (n-1)}{J\ (n-1)-2}, \text{ and factor B} = 1 + \frac{n\sum_{i=1}^{J} (\mu_i - \mu)^2}{\sigma_\omega^2\ (J-1)}$$

As nJ (where n is the number of observations per group and J is the number of groups) became larger, factor A approached one (factor B was not similarly affected). The additional assumption Hopkins and Hopkins invoked was that as nJ approached infinity, factor A approached one. Consequently, the formula for E (F) approximated factor B only as reflected in the formula:

$$E\ (F) \cong 1 + \frac{n \sum_{i=1}^{J} (\mu_i - \mu)^2}{\sigma_\omega^2\ (J-1)}$$

Hopkins and Hopkins indicated that this approximation was at most a 2 percent underestimate for nJ greater than 100. But when n and J were not large, the error in the formula was greater. For example, if n=11 and J=4, factor A would be 1.05 which is a 5 percent error from the assumed value of 1.00.

Consequently, because of this assumption which was required to make the non-centrality parameters for ANOVA and ANCOVA functions of the reliability of the dependent variable, the validity of the formulas was in question for small n and J. Therefore, an empirical investigation of the power of the ANOVA and ANCOVA test for small n and J became desirable. Since the reason for deriving the new non-centrality parameters was to account for the influence of test reliability on power, varying reliability values needed to be investigated concurrently.

RESEARCH TOOLS

There are two major research tools which the reader needs to understand in order to adequately visualize the methodology of the study. They are reliability and the Monte Carlo method. This portion of the review of the literature was developed to introduce these basic concepts and, consequently, does not purport to represent a review of the gamut of literature available on these research topics.

A concise definition of the reliability of a measure offered by Stanley and Hopkins is, "reliability . . . is that proportion of the total variance which is true variance." Symbolically, reliability (r_{xx}) equals the ratio of true variance (σ_t^2) and total variance (σ_ω^2).[20]

$$r_{xx} = \frac{\sigma_t^2}{\sigma_\omega^2}$$

This definition of reliability is standard among textbooks and periodical literature and will be used in this study.

---

[20]Stanley and Hopkins, op. cit., p. 119.

The Monte Carlo method is a scientific procedure for simulating an experiment and evaluating the results under prespecified conditions.[21] The increased use of computer simulations in problem solving has resulted in the development of the Monte Carlo technique in business, social science and education.[22]

Churchman said:

> In essence, the Monte Carlo technique consists of simulating an experiment to determine some probabilistic property of a population of objects or events by the use of random sampling applied to the components of the objects or events.[23]

He further noted that one of the most practical uses of the Monte Carlo method was to obtain approximate evaluations of mathematical expressions which are built on probability distributions. His statement in this regard is particularly relevant to establishing a rationale for using the technique in the current study.

A recent Monte Carlo study by A. C. Porter investigated the effects of sample size, the reliability of the covariable and other parameters upon the ANCOVA test.[24] His research task was to empirically compare Lord's U statistic with the conventional analysis of covariance for varying levels of reliability and sample size. Porter's use of Monte Carlo procedures further validates the appropriateness of these procedures for this study.

---

[21] C. West Churchman, Russel L. Ackoff, and E. Leonard Arnoff, Introduction to Operations Research (New York: John Wiley & Sons, Inc., 1957), p. 12.

[22] Ibid., p. 184. [23] Ibid., p. 175.

[24] A. C. Porter, "The Effects of Using Fallible Variables in the Analysis of Covariance" (unpublished Ph.D. dissertation, University of Wisconsin, 1967), p. 39.

SUMMARY

A review of literature revealed significant advancement in the study of statistical power beginning with the appearance of the 1951 Pearson-Hartley power charts. Pearson and Hartley succeeded in describing power for the analysis of variance as a function of sample size, number of treatment groups, level of significance, and treatment effect (non-centrality parameter).

The lack of use of current knowledge about statistical power to improve educational research has been of concern to some researchers. Jacob Cohen, an authority on statistical power, has suggested that behavioral research should possess power of at least .80. However, James Brewer and others discovered that studies in the behavioral sciences have had power far less than .80.

The literature indicated that several scholars were interested in pursuing the study of statistical power beyond its traditional model where the variables were assumed to be free of errors of measurement. Sutcliffe, in 1958, demonstrated that measurement error does in fact decrease statistical power. In 1969, Cleary and Linn offered a cost model for maximizing the benefits of the F-test by considering the effects of varying test length, which, of course, is known to be directly related to test reliability. As recently as 1974, Hopkins and Hopkins addressed the problem of test reliability and statistical power. They developed a method for new non-centrality parameters for ANOVA and ANCOVA which were a function of test reliability. However, the new method required an additional assumption about the non-centrality parameter which resulted in a threat to the validity of the formulas for small n and J. Consequently, an empirical investigation of the approach using Monte Carlo methods became desirable. The literature review also included the basic concepts of test reliability and Monte Carlo methods.

Chapter 3

PROCEDURES

The procedures for this study involved three major steps. The first step was to determine the power of ANOVA and ANCOVA from Pearson-Hartley power charts based on the Hopkins and Hopkins formula for the non-centrality parameters for fallible measures (The Formula Method). The second step involved the determination of the power of ANOVA and ANCOVA using a Monte Carlo simulation method (The Monte Carlo Method). The final step required the comparison of the power of ANOVA and ANCOVA derived from step 1 and step 2 to determine the validity of the Hopkins and Hopkins formula approach (Comparison of Methods).

THE FORMULA METHOD

A principal task of this study was to compare the power values resulting from the Hopkins and Hopkins formula method for determining the non-centrality parameter for fallible measures to the power values resulting from a Monte Carlo computer simulation method. Therefore, estimates of statistical power for the ANOVA and ANCOVA F-tests based on the Hopkins and Hopkins method needed to be determined.

Estimating the power of ANOVA and ANCOVA test from the 1951 Pearson-Hartley power charts required the determination of several statistical parameters. These parameters were: level of significance (α), numerator degrees of freedom (df_1), denominator degrees of freedom (df_2), non-centrality parameter for infallible measures ($\phi_{1.0}$), reliability of the dependent variable (r_{xx}), non-centrality parameter for ANOVA for fallible measures (ϕ_A), non-centrality parameter for ANCOVA for fallible measures (ϕ_C), and the correlation between the covariate and the variate (r_{xy}). These parameters are explained further in Table 2, page 22 and Table 3, page 23.

The following steps were used in estimating the power values for the ANOVA test. For each unique set of parameters:

1. $\phi_{1.0}$ was computed (in this study, this value was assigned).
2. ϕ_A was computed from $\phi_{1.0}$ and r_{xx}.
3. Power was estimated from the Pearson-Hartley power charts using ϕ_A, α, df_1, and df_2 for the ANOVA test.

For the ANCOVA test, the following steps were employed in estimating the power values:

1. ϕ_A was computed from $\phi_{1.0}$ and r_{xx} for ANOVA.
2. ϕ_C was computed from ϕ_A and r_{xy} (for this study, $r_{xy} = r_{xx}$).
3. Power was estimated from the Pearson-Hartley power charts using ϕ_C, α, df_1, and df_2 for the ANCOVA test.

The power values obtained through this process were subsequently used to illustrate the relationship between reliability and statistical power for the ANOVA and ANCOVA tests.

THE MONTE CARLO METHOD

According to Kerlinger:

> A Monte Carlo procedure is an empirical study of statistics using random numbers. In the behavioral sciences, the term Monte Carlo is usually applied to an empirical study of some method or

Table 2

Parameters Required in Order to Use the Pearson-Hartley Power Charts to Estimate Power for the Analysis of Variance (ANOVA)

Parameter	Definition	How determined
α	Level of significance	Arbitrarily fixed: $\alpha = .10$, $\alpha = .05$, $\alpha = .01$
df_1	Degrees of freedom for the numerator of the ANOVA F-test	$df_1 = J - 1$ where J = number of groups. For example, J = 2, $df_1 = 1$.
df_2	Degrees of freedom for the denominator of the ANOVA F-test	$df_2 = N - J$ where N = the total number of observations in the ANOVA. For example, N = 12, J = 2, $df_2 = 10$.
$\phi_{1.0}$	Non-centrality parameter assuming perfect reliability of the dependent variable (i.e., $\sigma_e^2 = 0$)	$\phi_{1.0} = \sqrt{\frac{n\ \Sigma_i(\mu_i-\mu)^2}{J\ \sigma_t^2}}$ (If J, σ_t^2, and n are held constant, and $\phi_{1.0}$ is increased, this represents a greater treatment effect of difference between means) where μ_i = treatment group mean, μ = grand mean, σ_t^2 = within group variance for the ANOVA when $\sigma_e^2=0$, and n = number per group.
r_{xx}	Reliability of the dependent variable in the ANOVA	$r_{xx} = \frac{\sigma_t^2}{\sigma_\omega^2}$
ϕ_A	Non-centrality Parameter for the ANOVA, as a function of reliability of the dependent variable (i.e., $\sigma_e^2 = 0$)	$\phi_A = \phi_{1.0}\sqrt{r_{xx}}$

Table 3

Parameters Required in Order to Use the Pearson-Hartley Power Charts to Estimate Power for the Analysis of Covariance (ANCOVA)

Parameter	Definition	How determined
α	Level of significance	Arbitrarily fixed: $\alpha = .10$, $\alpha = .05$, $\alpha = .01$.
df_1	Degrees of freedom for the numerator of the ANCOVA F-test	$df_1 = J - 1$ where J = number of groups. For example, $J = 2$, $df_1 = 1$.
df_2	Degrees of freedom for the denominator of the ANCOVA F-test	$df_2 = N - J - 1$ where N = total number of observations in the ANCOVA. For example, $N = 12$, $J = 2$, $df_2 = 9$.
r_{xy}	Correlation between the covariate and variate in the ANCOVA	For this study, $r_{xy} = r_{xx}$ where r_{xx} is the reliability of the dependent variable.
ϕ_C	Non-centrality parameter for the ANCOVA as a function of the reliability of the dependent variable.	$\phi_C = \dfrac{\phi_A}{\sqrt{1 - r_{xy}^2}}$ where ϕ_A is the non-centrality parameter for the ANOVA for $r_{xx} = r_{xy}$.

model that a scientist wishes to explore.[1]

Since this study involved an empirical verification of a statistical method, the Monte Carlo procedure was considered appropriate. As noted in Chapter 2, this procedure was used in other studies which involved a simulation of various aspects of the ANOVA and ANCOVA statistical tests.

In this study, a computer simulation model was developed using Monte Carlo procedures. This computer program was written and partially tested by B. R. Hopkins at the University of the Pacific. As a part of this study, this computer program was further tested and the various computer runs accomplished. The computer used for this investigation was the Burroughs B-6700 model and the language used was FORTRAN IV. A copy of the computer program is in Appendix B.

In general terms, the computer program randomly selected ten numbers between zero and one for small groups and four numbers between zero and one for large groups. The arithmetic mean was taken and the resulting value was used as an observation for a group. This process was used so that the within group observations would approximate a normal distribution by the Central Limit Theorem.[2] At this point, error factors were systematically added to each observation according to usual Monte Carlo procedures so that after J groups of n observations each were generated, the groups approximated the desired characteristics under study. That is, they had, on the average, the required level of reliability, and the required non-centrality parameter value. For cases with

---

[1]F. N. Kerlinger, Foundations of Behavioral Research (2d ed.; New York: Holt, Rinehart and Winston, 1973), p. 204.

[2]Ibid., p. 207.

more than two groups, the distance between adjacent means was approximately the same. Next, the ANOVA and ANCOVA statistical tests were conducted and the decisions (reject the null hypothesis or do not reject the null) were recorded for the three levels of significance, which were .10, .05, and .01. This process of randomly selecting numbers, adjusting them according to Monte Carlo procedures, conducting ANOVA and ANCOVA statistical tests, and recording the results of these tests was repeated 1,000 times for each unique set of parameters (J, n, $\phi_{1.0}$, α, r_{xy}, r_{xx}). One thousand replications was determined to be adequate and an accepted procedure based on two previously conducted Monte Carlo studies investigating the power of statistical tests.[3]

The number of times the null hypothesis was rejected at the various levels of significance for each unique set of parameters was recorded. The proportion of rejections of the null hypothesis out of 1,000 chances for the ANOVA and ANCOVA tests were used as estimates of statistical power.

COMPARISON OF METHODS

The strategy used in this study to compare the power values derived from the Hopkins and Hopkins Formula Method with the power values derived from the Monte Carlo Method involved a graphic comparison procedure. This procedure was selected primarily because it would

---

[3]A. C. Porter, "The Effects of Using Fallible Variables in the Analysis of Covariance" (unpublished Ph.D. dissertation, University of Wisconsin, 1967), p. 46; see also F. B. Baker and R. O. Collier, Jr., "Some Empirical Results on Variance Ratios Under Permutations in the Completely Randomized Design," Journal of the American Statistical Association, LXI, No. 5 (1966), 818.

demonstrate to what extent the data from the Formula Method approximated the data derived from the Monte Carlo Method.

Graphs were created using power estimates from both Methods. In these graphs, power was presented as a function of test reliability. Separate graphs were created for ANOVA and for ANCOVA comparisons.

SUMMARY

A purpose of this study was to empirically verify the Methods to estimate statistical power for the ANOVA and ANCOVA tests developed by Hopkins and Hopkins. The procedures used to accomplish this task involved three phases.

The first phase was to outline the steps required to estimate the statistical power of the ANOVA and ANCOVA F-tests using the Hopkins and Hopkins formula method. This method involved the computation of non-centrality parameters for the ANOVA (ϕ_A) and ANCOVA (ϕ_C) and the estimations of power based on the Pearson-Hartley power charts.

In the second phase, the general procedures of the Monte Carlo method were explained. In brief, a computer program was used which simulated the conditions and parameters of interest in this study. As a result of this simulation, power values were generated for the ANOVA and ANCOVA tests.

A comparison of the power estimates resulting from these two methods was required in order to determine the accuracy of the formula method developed by Hopkins and Hopkins. The method chosen was the graphic comparison method.

Chapter 4

FINDINGS

This chapter is divided into four sections. The first section contains the power estimates and non-centrality parameter values for the ANOVA and ANCOVA tests based on the Hopkins and Hopkins Formula Method. The second section contains the power estimates for ANOVA and ANCOVA based on the Monte Carlo Method. Graphs comparing the results of both methods are presented in the third section. Conclusions and observations based on the data presented in sections one through three are summarized in the final section.

RESULTS OF THE FORMULA METHOD

Arithmetic procedures for the Hopkins and Hopkins Formula Method were outlined in previous chapters. In capsule form, for the ANOVA test, a $\phi_{1.0}$ is assigned and ϕ_A is computed for a specific r_{xx}. This ϕ_A value is used with the Pearson-Hartley power charts along with the appropriate J, n, and α level to estimate power.

Power estimates for the ANCOVA test require the computation of ϕ_C which is a function of ϕ_A and r_{xy} (correlation between variate and covariate). This ϕ_C value is used with the Pearson-Hartley power charts to estimate the power of the ANCOVA for each specific J, n. and α.

ANOVA Tables

The computed ϕ_A values are presented in Table 4, page 28. The

Table 4

Selected Non-Centrality Parameter Values (ϕ_A) Based on the Formula* Developed by Hopkins and Hopkins for the ANOVA Test

Reliability of dependent variable (r_{xx})	ϕ_A Values		
	$\phi_{1.0}=1.5$	$\phi_{1.0}=2.0$	$\phi_{1.0}=2.5$
.3	.82	1.10	1.37
.4	.95	1.26	1.58
.5	1.06	1.41	1.77
.6	1.16	1.55	1.94
.7	1.25	1.67	2.09
.8	1.34	1.79	2.24
.9	1.42	1.90	2.37

*$\phi_A = \phi_{1.0} \sqrt{r_{xx}}$

$\phi_{1.0}$ values were selected to provide a basis of comparison with the $\phi_{1.0}$ values used by the Monte Carlo Method. Reliability values from .3 to .9 with intervals of .1 were used so that when the resulting ϕ_A values were used to estimate power, the curves created would be fairly smooth.

Estimates of the statistical power of the ANOVA test for two groups are presented in Table 5, page 30. The $\phi_{1.0}$ values of 2.0 and 2.5 were chosen because, for α=.05, the Pearson-Hartley power charts start at 1.0. The $\phi_{1.0}$ value of 1.5 was not selected because, as can be observed from Table 4, not all ϕ_A values were above the minimum 1.0 required for the power charts. Since a purpose of computing the ϕ_A values was to later graph the power estimates, only $\phi_{1.0}$ values with ϕ_A values all above 1.0 were selected, that is, $\phi_{1.0}$ values of 2.0 and 2.5 only. For two groups and α=.01, the Pearson-Hartley power charts start at 2.0. Therefore, no power values were estimated for two groups with α=.01.

Power estimates for three and five groups are offered in Tables 6 and 7, pages 31 and 32. For groups of three or more, the Pearson-Hartley power charts begin at 1.0 for α=.05 and α=.01, therefore, power estimates were made for $\phi_{1.0}$=2.0 and $\phi_{1.0}$=2.5 with α=.05 and α=.01. The degrees of freedom chosen for all cases were selected to provide power estimates to correspond to the J (number of groups) and n (number per group) used in the Monte Carlo Method.

It can be observed from Table 5, that as test reliability increases, so does statistical power. It can also be noted that power increases as the number per group and degrees of freedom increases. However, the reader must keep in mind that these findings are for the specific cases where n=10, 30, or 100 and where $\phi_{1.0}$ equals either 2.0 or 2.5.

Table 5

Estimations of the Statistical Power of the ANOVA Test for Two Groups Based on the Hopkins and Hopkins Formula for ϕ_A and the Pearson-Hartley Power Curves

Reliability of dependent variable (r_{xx})	Power for $\phi_{1.0}$=2.0			Power for $\phi_{1.0}$=2.5		
	df=(1,10) n=6	df=(1,30) n=16	df=(1,198) n=100	df=(1,10) n=6	df=(1,30) n=16	df=(1,198) n=100
	α=.05	α=.05	α=.05	α=.05	α=.05	α=.05
.3	.30	.31	.33	.42	.47	.49
.4	.38	.41	.43	.53	.58	.61
.5	.43	.49	.51	.62	.68	.70
.6	.50	.57	.59	.70	.76	.79
.7	.56	.63	.65	.76	.82	.84
.8	.63	.69	.72	.82	.87	.89
.9	.68	.74	.77	.85	.90	.92

Table 6

Estimations of the Statistical Power of the ANOVA Test for Three Groups Based on the Hopkins and Hopkins Formula for ϕ_A and the Pearson-Hartley Power Curves

Reliability of dependent variable (r_{xx})	Power for $\phi_{1.0}$=2.0				Power for $\phi_{1.0}$=2.5			
	df=(2,9) n=4		df=(2,30) n=11		df=(2,9) n=4		df=(2,30) n=11	
	α=.05	α=.01	α=.05	α=.01	α=.05	α=.01	α=.05	α=.01
.3	.29	.10	.34	.13	.41	.17	.50	.26
.4	.38	.13	.45	.21	.52	.25	.63	.38
.5	.43	.19	.54	.29	.62	.31	.73	.50
.6	.52	.23	.63	.37	.72	.40	.82	.60
.7	.57	.29	.68	.43	.78	.47	.88	.68
.8	.64	.32	.78	.51	.84	.55	.93	.76
.9	.70	.38	.81	.57	.88	.60	.95	.82

Table 7

Estimations of the Statistical Power of the ANOVA Test for Five Groups Based on the Hopkins and Hopkins Formula for ϕ_A and the Pearson-Hartley Power Curves

Reliability of dependent variable (r_{xx})	Power for $\phi_{1.0}$=2.0 df=(4,10) n=3		Power for $\phi_{1.0}$=2.5 df=(4,10) n=3	
	α=.05	α=.01	α=.05	α=.01
.3	.30	.10	.47	.19
.4	.40	.15	.60	.29
.5	.50	.21	.70	.38
.6	.58	.25	.80	.48
.7	.65	.32	.86	.56
.8	.72	.40	.91	.65
.9	.78	.45	.93	.72

As test reliability increases from .3 to .9, the average increase in power for the data provided in Table 5, is .42. It can also be observed from Table 5, that with high test reliability, small samples can provide the basis for experiments with adequate power. An example of this is $\phi_{1.0}$=2.5, n=6, α=.05, $r_{xx} \geq .8$ where the resulting power is greater than .80. Another observation is the increase in power as $\phi_{1.0}$ increases from 2.0 to 2.5 (when J, n, α, r_{xx} are held constant and $\phi_{1.0}$ is increased, $\phi_{1.0}$ reflects an increase in treatment effect, that is, a greater difference between treatment group means). For example, for J=2, n=16, α=.05 and r_{xx}=.6, power changes from .57 to .76 as $\phi_{1.0}$ increases from 2.0 to 2.5.

In Table 6, page 31, the power estimates for three groups with two sample sizes, 4 and 11 are presented. It can be observed that the level of significance (α) influences power. That is, as α increases from .01 to .05, power increases. For the samples provided, power increases an average of .25 as α goes from .01 to .05. It can also be observed that test reliability influences power. As test reliability increases from .3 to .9, power increases an average of .45 for α=.05, and an average of .43 for α=.01 for the samples provided in Table 6. It should be noted, however, that these observations are restricted to α=.05 and α=.01, and for n=4 and n=11.

Power estimates based on the Formula Method for five groups are presented in Table 7, page 32. To correspond with the Monte Carlo data for five groups, the power for n=3 with $\phi_{1.0}$=2.0 and $\phi_{1.0}$=2.5 was estimated. As was observed in Tables 5 and 6, as α increases from .01 to .05, power is augmented. Also, as reliability varies from .3 to .9 power increases. For example, as reliability goes from .3 to .9, for $\phi_{1.0}$=2.5, n=3, and α=.01, power increases from .19 to .72.

ANCOVA Tables

The non-centrality parameters for ANCOVA, ϕ_C, are presented in Table 8, page 35, for varying $\phi_{1.0}$ values. For this study, $r_{xy}=r_{xx}$, that is, the correlation between the variate and the covariate was set equal to test reliability. For Tables 9, 10, and 11, pages 36, 37, and 38 respectively, $\phi_{1.0}$ values of 2.0 and 2.5 are used because the corresponding ϕ_C values from Table 8 were not all above 1.0 for $\phi_{1.0}$ values less than 2.0.

In Table 9, page 36, the power estimates for ANCOVA for two groups with sample sizes varying from 6 to 100 are presented. Power estimates for α=.05 are provided. It can be noted that power increases as sample size (n) increases. This finding is consistent with known relationships between sample size and power. It can also be noted that power increases as the correlation between the variate and the covariate increases. For example, as r_{xy} and r_{xx} vary from .3 to .9, power increases from .30 to .90 for α=.05, n=6, and $\phi_{1.0}$=2.0. As was previously indicated, power of .80 or greater is desirable in the behavioral sciences. From Table 9 it can be observed that even for n=6, power of .83 is obtained with r_{xy}=.7 for α=.05 and $\phi_{1.0}$=2.0. When $\phi_{1.0}$ is increased to 2.5, r_{xy} of only .6 is required to obtain a power value of .87.

Estimations of the statistical power of the ANCOVA test for three groups are presented in Table 10, page 37. As has been noted previously, power increases as n, α, r_{xy}, and $\phi_{1.0}$ increase. Increasing one or more parameters can change the power estimate from below .8 to above the desired minimum .8 value. For example, for n=4, $\phi_{1.0}$=2.0, r_{xy}=.7, the

Table 8

Selected Non-Centrality Parameter Values (ϕ_C) Based on the Formula* Developed by Hopkins and Hopkins for the ANCOVA Test

Correlation between variate and covariate where $r_{xy}=r_{xx}$ (r_{xy})	ϕ_C Values		
	$\phi_{1.0}=1.5$	$\phi_{1.0}=2.0$	$\phi_{1.0}=2.5$
.3	.86	1.15	1.44
.4	1.04	1.37	1.72
.5	1.22	1.63	2.04
.6	1.45	1.94	2.43
.7	1.75	2.34	2.93
.8	2.33	2.98	3.73
.9	3.26	4.36	5.44

$*\phi_C = \sqrt{\dfrac{\phi_A^2}{1-r_{xy}^2}}$, where $r_{xx} = r_{xy}$

Table 9

Estimations of the Statistical Power of the ANCOVA Test for Two Groups Based on the Hopkins and Hopkins Formula for ϕ_C and the Pearson-Hartley Power Charts

Correlation between variate and covariate where $r_{xy}=r_{xx}$ (r_{xy})	Power for $\phi_{1.0}=2.0$			Power for $\phi_{1.0}=2.5$		
	df=(1,9) n=6	df=(1,29) n=16	df=(1,197) n=100	df=(1,9) n=6	df=(1,29) n=16	df=(1,197) n=100
	$\alpha=.05$	$\alpha=.05$	$\alpha=.05$	$\alpha=.05$	$\alpha=.05$	$\alpha=.05$
.3	.30	.33	.38	.44	.51	.53
.4	.40	.48	.50	.57	.66	.68
.5	.53	.62	.64	.71	.78	.81
.6	.67	.75	.78	.87	.91	.93
.7	.83	.89	.91	.95	.98	.98
.8	.96	.98	.99	.99	.99	.99
.9	.99	.99	.99	.99	.99	.99

Table 10

Estimations of the Statistical Power of the ANCOVA Test for Three Groups Based on the Hopkins and Hopkins Formula for ϕ_C and the Pearson-Hartley Power Charts

Correlation between variate and covariate where $r_{xy}=r_{xx}$ (r_{xy})	Power for $\phi_{1.0}=2.0$				Power for $\phi_{1.0}=2.5$			
	df=(2,8) n=4		df=(2,29) n=11		df=(2,8) n=4		df=(2,29) n=11	
	α=.05	α=.01	α=.05	α=.01	α=.05	α=.01	α=.05	α=.01
.3	.30	.10	.33	.12	.43	.18	.55	.28
.4	.40	.15	.49	.23	.60	.28	.71	.49
.5	.55	.25	.65	.40	.72	.40	.86	.63
.6	.70	.38	.82	.59	.88	.60	.96	.85
.7	.86	.55	.95	.80	.97	.79	.99	.97
.8	.97	.82	.99	.97	.99	.96	.99	.99
.9	.99	.99	.99	.99	.99	.99	.99	.99

Table 11

Estimations of the Statistical Power of the ANCOVA Test for Five Groups Based on the Hopkins and Hopkins Formula for ϕ_C and the Pearson-Hartley Power Charts

Correlation between variate and covariate where $r_{xy}=r_{xx}$ (r_{xy})	Power for $\phi_{1.0}=2.0$ df=(4,9) n=3		Power for $\phi_{1.0}=2.5$ df=(4,9) n=3	
	$\alpha=.05$	$\alpha=.01$	$\alpha=.05$	$\alpha=.01$
.3	.30	.08	.48	.19
.4	.42	.12	.63	.30
.5	.58	.26	.78	.44
.6	.74	.40	.92	.65
.7	.89	.60	.98	.84
.8	.99	.86	.99	.98
.9	.99	.99	.99	.99

power for α=.01 is estimated as .55 but the power for α=.05 is an acceptable .86. As r_{xy} varies from .3 to .9 for $\phi_{1.0}$=2.0, n=4, and α=.01, power increases from .10 to .99. As n goes from 6 to 100, the increase in power as shown in Table 10 is a maximum of .25. For three groups with only 11 per group, α=.05, and $\phi_{1.0}$=2.5, r_{xy} of only .5 can provide the basis for an experiment with power estimated at .86.

In Table 11, page 38, the power estimates based on the Hopkins and Hopkins formula for ANCOVA for five groups with a sample size of 3 are presented. As was observed for previous cases, power becomes greater as $\phi_{1.0}$, r_{xy}, and α increase. It can be noted that even for n=3, an experiment with adequate power can be conducted with r_{xy} of .6 or greater.

RESULTS OF THE MONTE CARLO METHOD

In order to examine and verify the formulas developed by Hopkins and Hopkins to estimate the statistical power of the ANOVA and ANCOVA test, a computer program using Monte Carlo procedures was used. Data resulting from this computer program were generated for the parameters relevant to this study and were summarized in Tables 12 through 17, pages 40 through 45.

ANOVA and ANCOVA Tables

Power values for ANOVA and ANCOVA for two groups with $\phi_{1.0}$ varying from 1.0 to 2.5, α varying from .01 to .10, r_{xx} varying from .3 to .9, and n varying from 6 to 100 are in Tables 12 and 13. Power values with three significant digits are provided since the number represents the number of times the null hypothesis was rejected out of 1000 trials.

As can be observed from Tables 12 and 13, for both the ANOVA and

Table 12

Estimations of Statistical Power for the ANOVA and the ANCOVA Tests for Two Groups Based upon the Monte Carlo Simulation for $\phi_{1.0}$ Equal to 1.0 and 1.5

		$\phi_{1.0}$=1.0 ANOVA			$\phi_{1.0}$=1.0 ANCOVA			$\phi_{1.0}$=1.5 ANOVA			$\phi_{1.0}$=1.5 ANCOVA		
n	r_{xx}= r_{xy}	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01
6	.3	.170	.104	.022	.183	.103	.025	.287	.187	.051	.286	.170	.054
6	.5	.259	.147	.050	.281	.166	.045	.379	.242	.086	.432	.288	.098
6	.7	.294	.197	.070	.437	.294	.113	.505	.354	.142	.693	.552	.280
6	.9	.320	.196	.060	.835	.723	.415	.566	.418	.187	.992	.970	.813
16	.3	.214	.128	.032	.209	.134	.027	.328	.207	.069	.333	.224	.071
16	.5	.267	.160	.058	.328	.224	.069	.424	.307	.120	.524	.384	.173
16	.7	.300	.204	.061	.467	.326	.147	.530	.395	.174	.761	.657	.370
16	.9	.364	.252	.097	.890	.815	.602	.623	.485	.238	.997	.991	.954
100	.3	.194	.125	.040	.195	.134	.039	.336	.216	.091	.352	.232	.091
100	.5	.243	.156	.056	.290	.190	.075	.455	.331	.144	.531	.395	.197
100	.7	.339	.230	.080	.491	.388	.187	.534	.393	.209	.818	.723	.465
100	.9	.397	.272	.114	.913	.850	.670	.644	.486	.269	1.000	.999	.983

Table 13

Estimations of Statistical Power for the ANOVA and the ANCOVA Tests for Two Groups Based upon the Monte Carlo Simulation for $\phi_{1.0}$ Equal to 2.0 and 2.5

		$\phi_{1.0}=2.0$ ANOVA			$\phi_{1.0}=2.0$ ANCOVA			$\phi_{1.0}=2.5$ ANOVA			$\phi_{1.0}=2.5$ ANCOVA		
n	$r_{xx}=r_{xy}$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$
6	.3	.434	.303	.108	.438	.287	.098	.536	.392	.152	.526	.388	.151
6	.5	.545	.417	.176	.610	.464	.215	.751	.607	.290	.795	.669	.353
6	.7	.735	.602	.274	.907	.814	.512	.888	.778	.471	.982	.955	.741
6	.9	.780	.660	.363	1.000	.999	.977	.937	.885	.630	1.000	1.000	.998
16	.3	.470	.339	.144	.479	.343	.153	.601	.456	.233	.631	.501	.257
16	.5	.643	.497	.254	.721	.608	.360	.778	.658	.389	.876	.783	.534
16	.7	.740	.603	.362	.935	.878	.684	.908	.846	.571	.988	.976	.903
16	.9	.848	.748	.482	1.000	1.000	.999	.958	.907	.729	1.000	1.000	1.000
100	.3	.491	.373	.168	.502	.389	.182	*	*	*	*	*	*
100	.5	.642	.525	.289	.748	.635	.389	*	*	*	*	*	*
100	.7	.746	.634	.402	.942	.907	.757	*	*	*	*	*	*
100	.9	.845	.747	.539	1.000	1.000	1.000	*	*	*	*	*	*

*Values for $\phi_{1.0}=2.5$ for n=100 were not computed.

Table 14

Estimations of Statistical Power for the ANOVA and the ANCOVA Tests for Three Groups Based upon the Monte Carlo Simulation for $\phi_{1.0}$ Equal to 1.0 and 1.5

		$\phi_{1.0}$=1.0 ANOVA			$\phi_{1.0}$=1.0 ANCOVA			$\phi_{1.0}$=1.5 ANOVA			$\phi_{1.0}$=1.5 ANCOVA		
n	r_{xx}= r_{xy}	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01
4	.3	.186	.116	.034	.206	.113	.033	.270	.158	.048	.261	.163	.046
4	.5	.258	.158	.051	.262	.158	.037	.404	.264	.091	.442	.305	.103
4	.7	.281	.178	.042	.414	.255	.085	.517	.332	.131	.702	.555	.211
4	.9	.332	.202	.060	.858	.734	.413	.593	.435	.172	.994	.978	.845
11	.3	.205	.123	.033	.207	.130	.041	.339	.227	.093	.355	.238	.089
11	.5	.273	.171	.047	.321	.214	.064	.445	.304	.132	.513	.383	.182
11	.7	.320	.203	.070	.518	.378	.163	.578	.448	.201	.832	.717	.458
11	.9	.384	.262	.087	.960	.911	.719	.677	.548	.277	1.000	.999	.987

Table 15

Estimations of Statistical Power for the ANOVA and the ANCOVA Tests for Three Groups Based upon the Monte Carlo Simulation for $\phi_{1.0}$ Equal to 2.0 and 2.5

		$\phi_{1.0}$=2.0 ANOVA			$\phi_{1.0}$=2.0 ANCOVA			$\phi_{1.0}$=2.5 ANOVA			$\phi_{1.0}$=2.5 ANCOVA		
n	$r_{xx}=r_{xy}$	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01
4	.3	.413	.264	.087	.398	.256	.082	.592	.437	.181	.583	.419	.164
4	.5	.612	.427	.181	.646	.475	.197	.766	.630	.324	.820	.680	.365
4	.7	.717	.568	.265	.901	.801	.498	.885	.773	.443	.978	.943	.730
4	.9	.837	.690	.362	.997	.996	.981	.956	.881	.619	1.000	.999	.998
11	.3	.483	.360	.164	.501	.364	.166	.651	.521	.251	.673	.541	.279
11	.5	.675	.549	.285	.755	.637	.376	.848	.741	.475	.934	.852	.627
11	.7	.804	.694	.418	.970	.931	.789	.941	.892	.700	.997	.994	.966
11	.9	.880	.801	.554	1.000	1.000	1.000	.972	.936	.803	1.000	1.000	1.000

Table 16

Estimations of Statistical Power for the ANOVA and the ANCOVA Tests for Five Groups Based upon the Monte Carlo Simulation for $\phi_{1.0}$ Equal to 1.0 and 1.5

		$\phi_{1.0}=1.0$ ANOVA			$\phi_{1.0}=1.0$ ANCOVA			$\phi_{1.0}=1.5$ ANOVA			$\phi_{1.0}=1.5$ ANCOVA		
n	$r_{xx}=r_{xy}$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$
3	.3	.172	.109	.025	.176	.099	.023	.309	.202	.054	.302	.195	.051
3	.5	.259	.151	.041	.287	.171	.052	.404	.263	.092	.462	.305	.108
3	.7	.304	.192	.058	.452	.303	.105	.588	.413	.163	.799	.647	.305
3	.9	.370	.238	.078	.930	.819	.498	.655	.486	.202	1.000	.994	.937

Table 17

Estimations of Statistical Power for the ANOVA and the ANCOVA Tests for Five Groups Based upon the Monte Carlo Simulation for $\phi_{1.0}$ Equal to 2.0 and 2.5

		$\phi_{1.0}=2.0$ ANOVA			$\phi_{1.0}=2.0$ ANCOVA			$\phi_{1.0}=2.5$ ANOVA			$\phi_{1.0}=2.5$ ANCOVA		
n	$r_{xx}=r_{xy}$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$
3	.3	.459	.308	.105	.463	.312	.108	.624	.476	.205	.617	.466	.196
3	.5	.656	.485	.199	.719	.565	.256	.820	.688	.378	.891	.769	.440
3	.7	.813	.667	.359	.950	.878	.614	.941	.847	.541	.993	.984	.830
3	.9	.871	.765	.432	1.000	1.000	.996	.980	.938	.713	1.000	1.000	1.000

ANCOVA tests, power is augmented as n, α, r_{xx} and $\phi_{1.0}$ increase. For the ANOVA test, as $\phi_{1.0}$ varies from 1.0 to 2.0, for n=100, r_{xx}=.9 and α=.10, power increases from .397 to .845. The level of significance (α) also influences power. An increase in α from .01 to .10 results in dramatic increases in power values, especially for r_{xx} below .7. For example, as α varies from .01 to .10, for $\phi_{1.0}$=1.5, n=16, r_{xx}=.5, power is more than tripled as it increases from .120 to .424. Except for a few cases where $r_{xx}=r_{xy}$=.3, the power of the ANCOVA is greater than the power of the ANOVA test. For $\phi_{1.0}$=1.5, α=.01, and n=100, the power increases from .269 for ANOVA to .983 for ANCOVA.

The Monte Carlo Method power estimates for three groups for ANOVA and ANCOVA are in Tables 14 and 15, pages 42 and 43. Power values were computed for sample sizes of 4 and 11. Similar observations about power can be made concerning the increase in power as α, n, $\phi_{1.0}$, r_{xx}, and r_{xy} become larger. It is possible from Tables 12 to 15 to compare what happens to power when 3 groups of 4 each (n=12) are used instead of 2 groups of 6 each (n=12). While controlling for $\phi_{1.0}$, α, r_{xx}, and r_{xy}, the power is generally greater when three groups are used instead of two.

The statistical power estimates for five groups with three per group based upon the Monte Carlo simulation are offered in Tables 16 and 17, pages 44 and 45. It can be observed that for the ANOVA test with n=3, $r_{xy}=r_{xx}$=.9, α=.05 and $\phi_{1.0}$ of 2.5 power is greater than .8. However, for the ANCOVA test with the same parameters, a $\phi_{1.0}$ of only 1.0 is necessary to achieve a comparable level of power.

Difference Between the Power Estimates of ANCOVA and ANOVA

According to Kerlinger,

> Analysis of Covariance [ANCOVA] is a form of analysis of variance [ANOVA] that tests the significance of the differences between means of final experimental data by taking into account the correlation between the dependent variable and one or more covariates . . . .[1]

The objective of this section is to investigate the increase in statistical power resulting from the use of the ANCOVA test rather than the ANOVA test. As indicated by Kerlinger, the major difference between ANOVA and ANCOVA is the influence of the correlation of the variate and covariate which ANCOVA includes and ANOVA does not include. The data in Tables 18, 19, and 20, pages 48, 49, and 50, represent the influence of this correlation. The data in Table 18 are the result of subtracting the ANCOVA power values from the ANOVA power values taken from Tables 12 and 13 while controlling for J, n, $\phi_{1.0}$, r_{xx}, and α. For example, from Table 13, for $\phi_{1.0}=2.0$, $\alpha=.01$, n=16, and $r_{xx}=.9$, the ANCOVA value is .999 and the ANOVA value is .482. The difference between these two values is .517 which is presented in Table 18 for the same parameters. This number, .517, represents the increase in power when the ANCOVA test is used instead of the ANOVA test when the correlation between the variate and the covariate is .9.

Examining the data in Table 18, it can be observed that for correlations .5 or below, the increase in power is at most .145 and generally less than .100. For correlations of .7 and .9, increases in power are more substantial and vary from .080 to .716. In general, smaller levels of significance have the greatest increase in power, that is, $\alpha=.01$ has a larger increase in power than $\alpha=.05$ when ANCOVA is used instead of

---

[1]F. N. Kerlinger, Foundations of Behavioral Research (2d ed.; New York: Holt, Rinehart, and Winston, 1973), p. 370.

Table 18

Difference in Power Values Between the ANCOVA and the ANOVA Test
Based on the Monte Carlo Simulation Values for Two Groups

n	$r_{xy}=r_{xx}$	$\phi_{1.0}=1.0$			$\phi_{1.0}=1.5$			$\phi_{1.0}=2.0$			$\phi_{1.0}=2.5$		
		$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$
6	.3	.013	-.001	.003	-.001	-.017	.003	.004	-.016	-.010	-.010	-.004	-.001
6	.5	.022	.022	-.005	.053	.046	.012	.065	.047	.039	.044	.062	.063
6	.7	.143	.097	.043	.188	.198	.138	.172	.212	.238	.094	.177	.270
6	.9	.515	.527	.355	.426	.552	.626	.220	.339	.614	.063	.115	.368
16	.3	-.005	.006	-.005	.005	.017	.002	.009	.004	.009	.030	.045	.024
16	.5	.061	.064	.011	.100	.077	.053	.078	.111	.106	.098	.125	.145
16	.7	.167	.122	.086	.231	.262	.196	.195	.275	.322	.080	.130	.332
16	.9	.526	.563	.505	.374	.506	.716	.152	.252	.517	.042	.093	.271
100	.3	.001	.009	-.001	.016	.016	.000	.011	.016	.014	*	*	*
100	.5	.047	.034	.019	.076	.064	.053	.106	.110	.100	*	*	*
100	.7	.152	.158	.107	.284	.330	.256	.196	.273	.355	*	*	*
100	.9	.516	.578	.556	.356	.513	.714	.155	.253	.461	*	*	*

*Power values for ANOVA and ANCOVA for n=100 were not computed.

Table 19

Difference in Power Values Between the ANCOVA and the ANOVA Test Based on the Monte Carlo Simulation Values for Three Groups

		$\phi_{1.0}$=1.0			$\phi_{1.0}$=1.5			$\phi_{1.0}$=2.0			$\phi_{1.0}$=2.5		
n	r_{xy}= r_{xx}	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01	α=.10	α=.05	α=.01
4	.3	.020	-.003	-.001	-.009	.005	-.002	-.015	-.008	-.005	-.009	-.018	.017
4	.5	.004	.000	-.014	.038	.041	.012	.044	.048	.016	.054	.050	.041
4	.7	.133	.077	.043	.185	.223	.080	.184	.233	.233	.093	.170	.287
4	.9	.526	.532	.353	.401	.543	.673	.160	.306	.619	.044	.118	.379
11	.3	.002	.007	.008	.016	.011	-.004	.018	.004	.002	.022	.020	.028
11	.5	.048	.043	.017	.068	.079	.050	.080	.088	.091	.086	.111	.152
11	.7	.198	.175	.093	.254	.269	.257	.166	.237	.371	.056	.102	.266
11	.9	.576	.649	.632	.323	.451	.710	.120	.199	.446	.028	.064	.197

Table 20

Difference in Power Values Between the ANCOVA and the ANOVA Test Based on the Monte Carlo Simulation Values for Five Groups

		$\phi_{1.0}=1.0$			$\phi_{1.0}=1.5$			$\phi_{1.0}=2.0$			$\phi_{1.0}=2.5$		
n	$r_{xy}= r_{xx}$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$	$\alpha=.10$	$\alpha=.05$	$\alpha=.01$
3	.3	.004	-.010	-.002	-.007	-.007	-.003	.004	.004	.003	-.007	-.010	-.009
3	.5	.028	.020	.011	.058	.042	.016	.063	.080	.057	.071	.081	.062
3	.7	.148	.111	.047	.211	.234	.142	.137	.211	.255	.052	.137	.289
3	.9	.560	.581	.420	.345	.508	.735	.129	.235	.564	.020	.062	.287

ANOVA. A similar result occurs when the increase in power for α=.05 is compared with the increase in power for α=.10. As $\phi_{1.0}$ increases, the difference in power between ANCOVA and ANOVA becomes greater up to the point where both are close to 1.000. For example, from Table 18, where n=16 and r_{xy}=.7, the difference for $\phi_{1.0}$ of 1.0 is .086, for $\phi_{1.0}$ of 1.5 it is .196, for $\phi_{1.0}$ of 2.0 it is .322, and for $\phi_{1.0}$ of 2.5 it is .332. It can also be observed from Table 18 that as the correlation increases, the difference in ANCOVA and ANOVA becomes greater. This increase continues until both values approach 1.000.

The data for the differences in power between the ANCOVA and ANOVA for three groups are presented in Table 19. Observations similar to those for two groups can be made. In general, the difference in power between ANCOVA and ANOVA becomes greater as the level of significance becomes smaller, as the correlation becomes greater, as sample size increases, and as the treatment effect ($\phi_{1.0}$) becomes larger. Variations from these generalizations result from the stochasticity of power when n, r_{xx}, r_{xy} and $\phi_{1.0}$ are low. Also, when both ANOVA and ANCOVA power values are close to one, the difference becomes small.

Differences in power between the ANCOVA and ANOVA tests for five groups which are presented in Table 20 form similar patterns to the differences noted for two and three groups. It is interesting to note, that with only n=3, $\phi_{1.0}$=1.5, α=.01, and r_{xy}=.9, power increases .735 when ANCOVA is used instead of ANOVA.

The change in power when ANCOVA is used instead of ANOVA for varying correlation (r_{xy}) and reliability (r_{xx}) levels is presented in Figure 2, page 52. As can be observed from the graph, for r_{xy} and r_{xx} greater than .5, ANCOVA is a more powerful statistic than ANOVA except

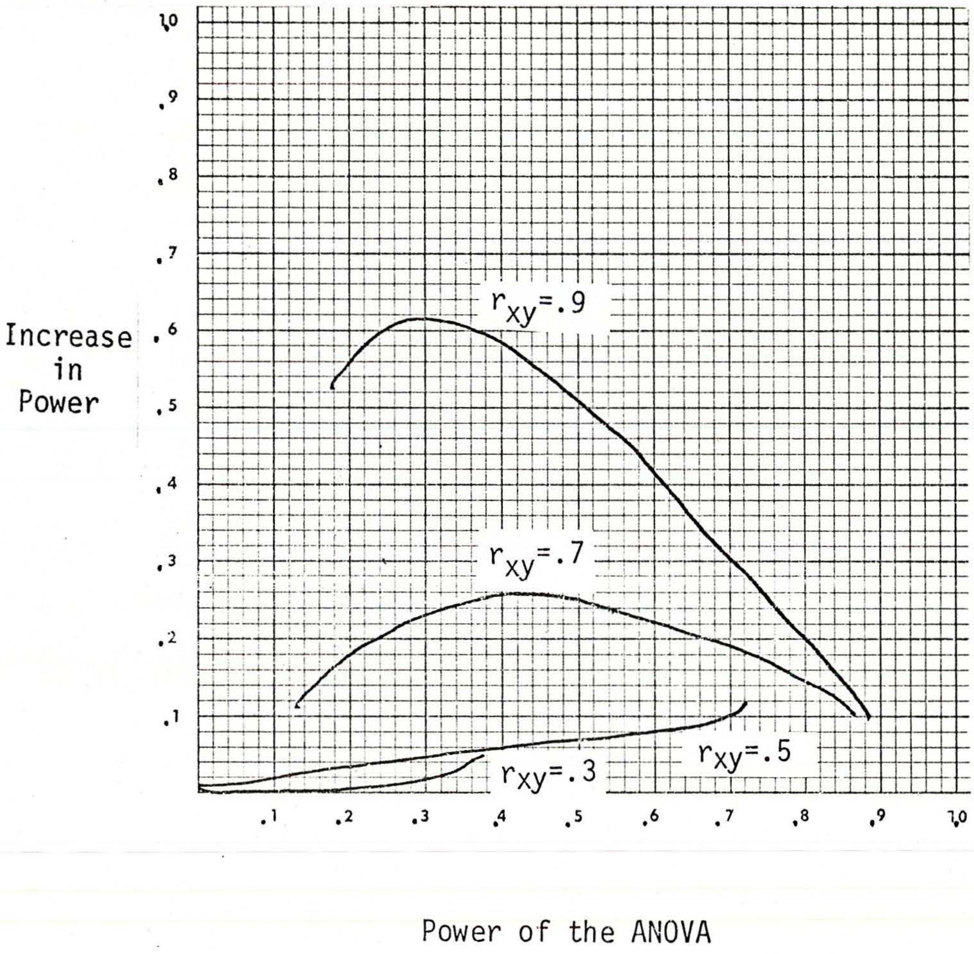


Figure 2

An Approximation of the Increase in Power When ANCOVA Is Used Instead of ANOVA for Varying Levels of r_{xy}* with α=.05

*In this study, $r_{xy}=r_{xx}$.

when the power of the ANOVA is .8 or more. It can also be noted that when r_{xy} and r_{xx} are less than or equal to .3, ANCOVA is comparable in power to the ANOVA test.

COMPARISON OF THE HOPKINS AND HOPKINS FORMULA METHOD TO THE MONTE CARLO METHOD

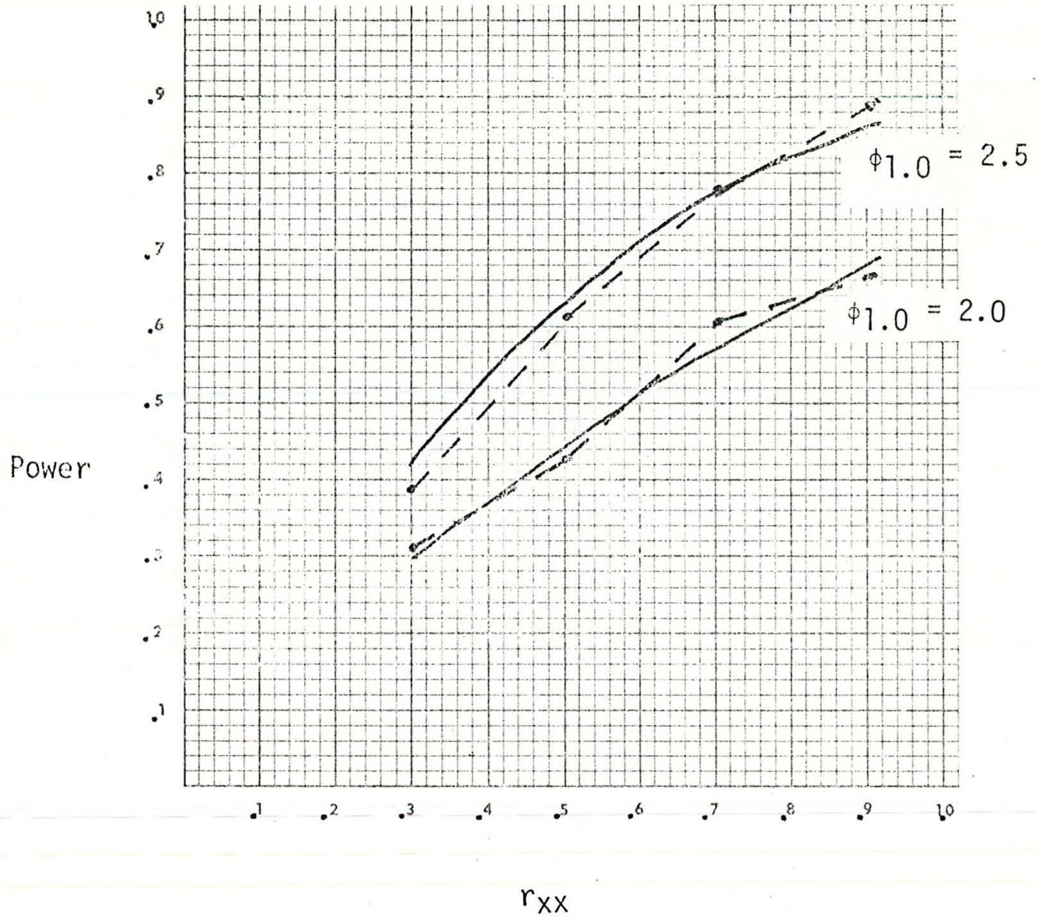
The graphic representations presented in this section summarize selected data derived from the first two sections. The graphs are presented to facilitate the comparison of the Formula and the Monte Carlo Methods for estimating the statistical power of the ANOVA and the ANCOVA tests.

The graphs presented in Figures 3 through 20, pages 54 through 71, provide for the comparison of the Formula Method and the Monte Carlo Method for J varying from 2 to 5, for n varying from 3 to 100, for α varying from .01 to .05 for $\phi_{1.0}$ varying from 2.0 to 2.5, and for r_{xx} and r_{xy} varying from .3 to .9.

ANOVA Graphs

The graphs presented in Figures 3 through 11 provide the basis for comparing the Formula Method to the Monte Carlo Method. Since the Monte Carlo Method simulates the actual power values, the accuracy of the Formula Method can be assessed by this process.

By examining the graphs, it can be observed that the Monte Carlo derived power values vary less than .05 from the formula derived power values. In general, the variation between the two methods is less than .02. This is an exceptionally close fit since the formula derived values were based on estimates from the Pearson-Hartley power charts which may contain some interpolation error.

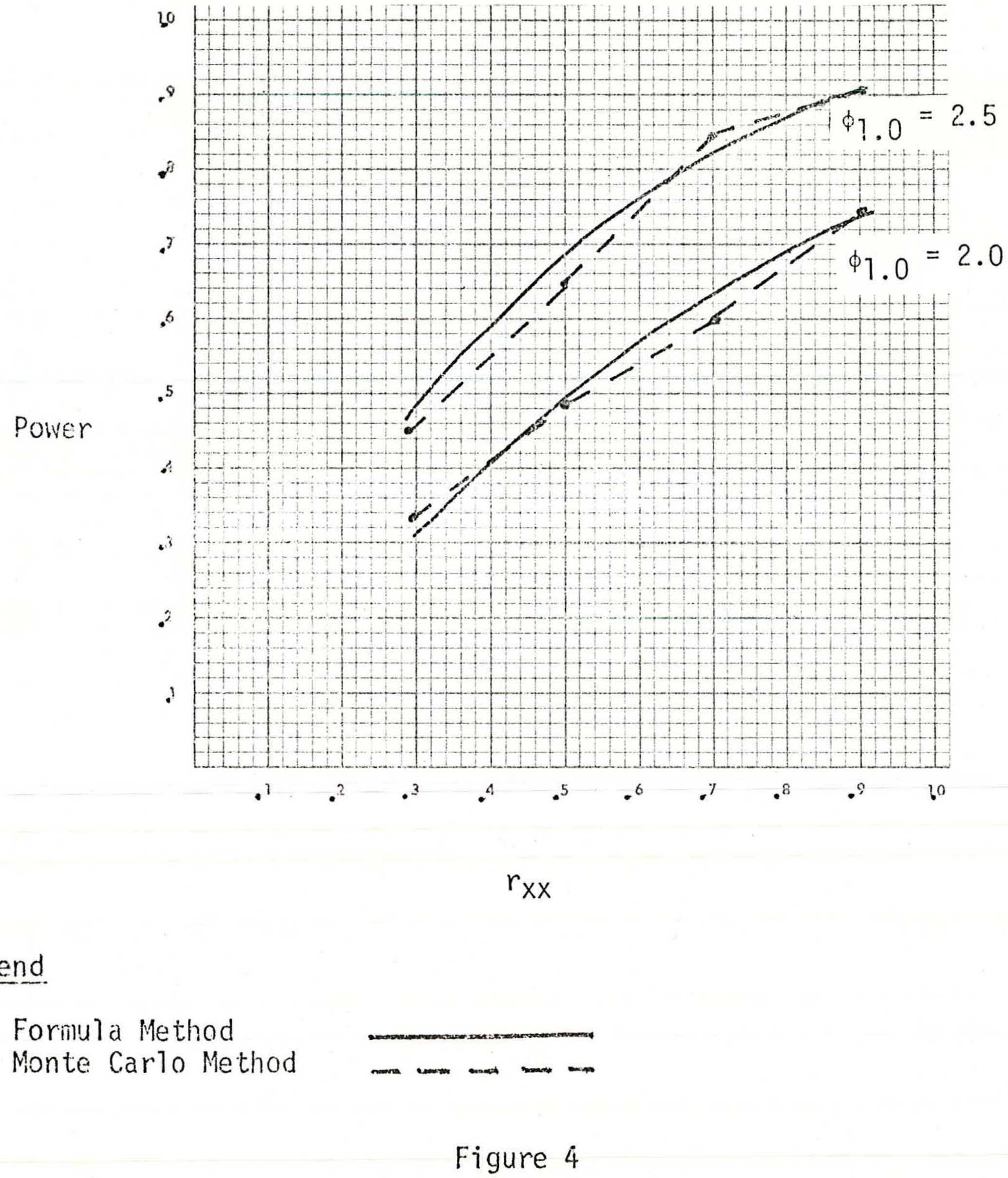


Legend

Formula Method
Monte Carlo Method

Figure 3

A Comparison of the Power of the ANOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=2, n=6, and α=.05



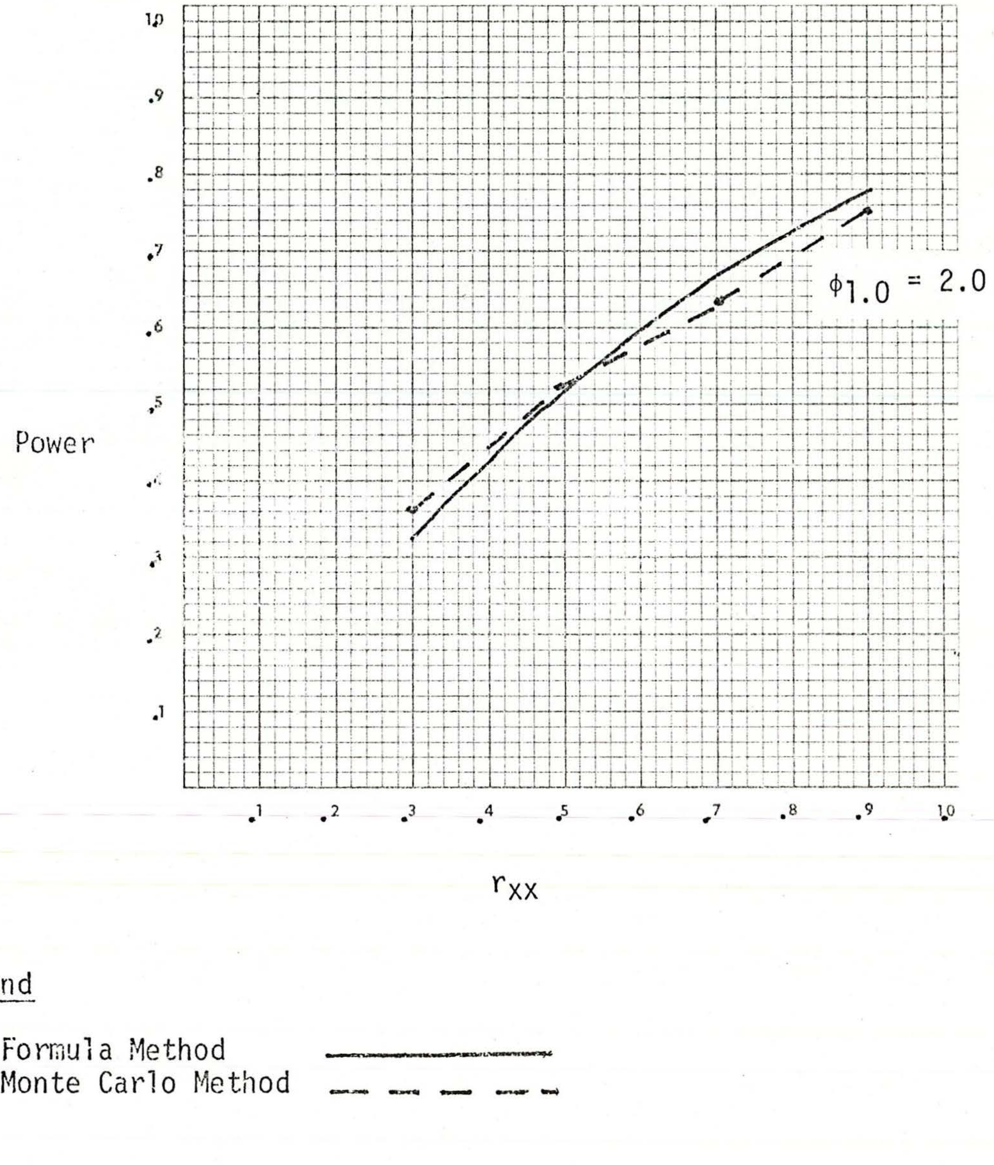
Legend

Formula Method ————

Monte Carlo Method — — — —

Figure 4

A Comparison of the Power of the ANOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=2, n=16, and α=.05



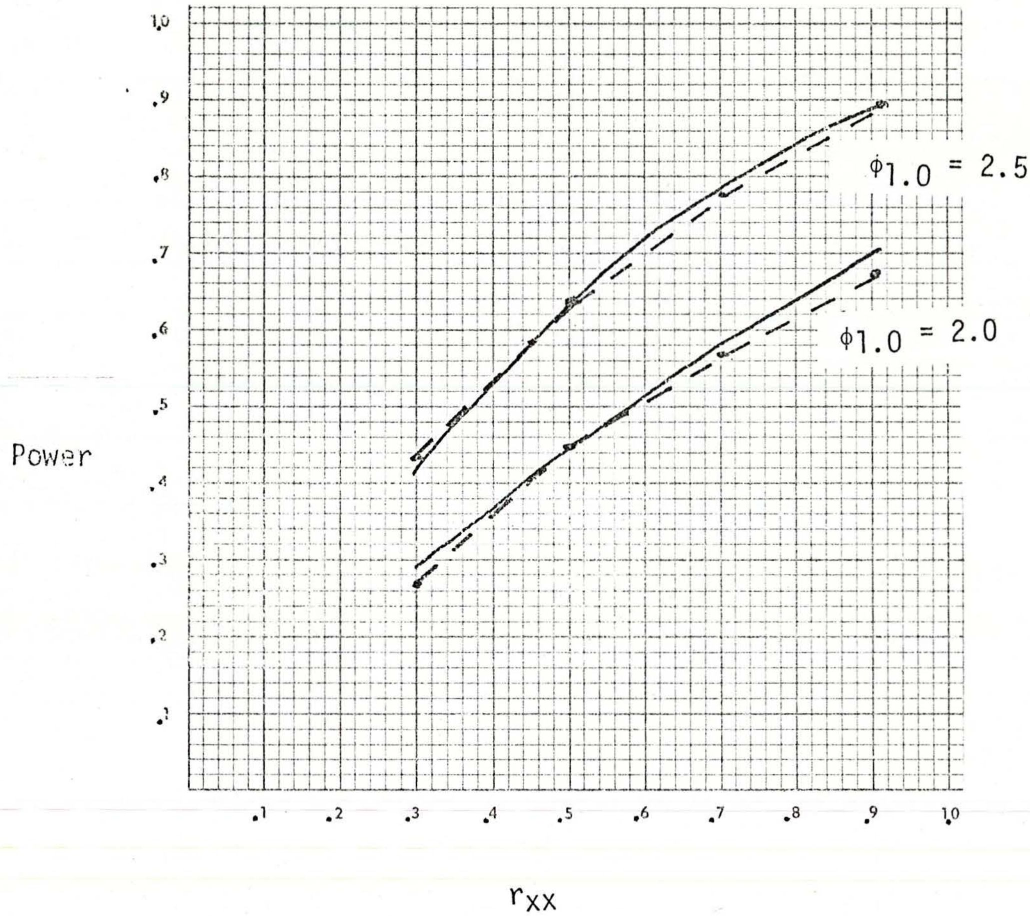
Legend

Formula Method ————

Monte Carlo Method — — — —

Figure 5

A Comparison of the Power of the ANOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=2, n=100, and α=.05



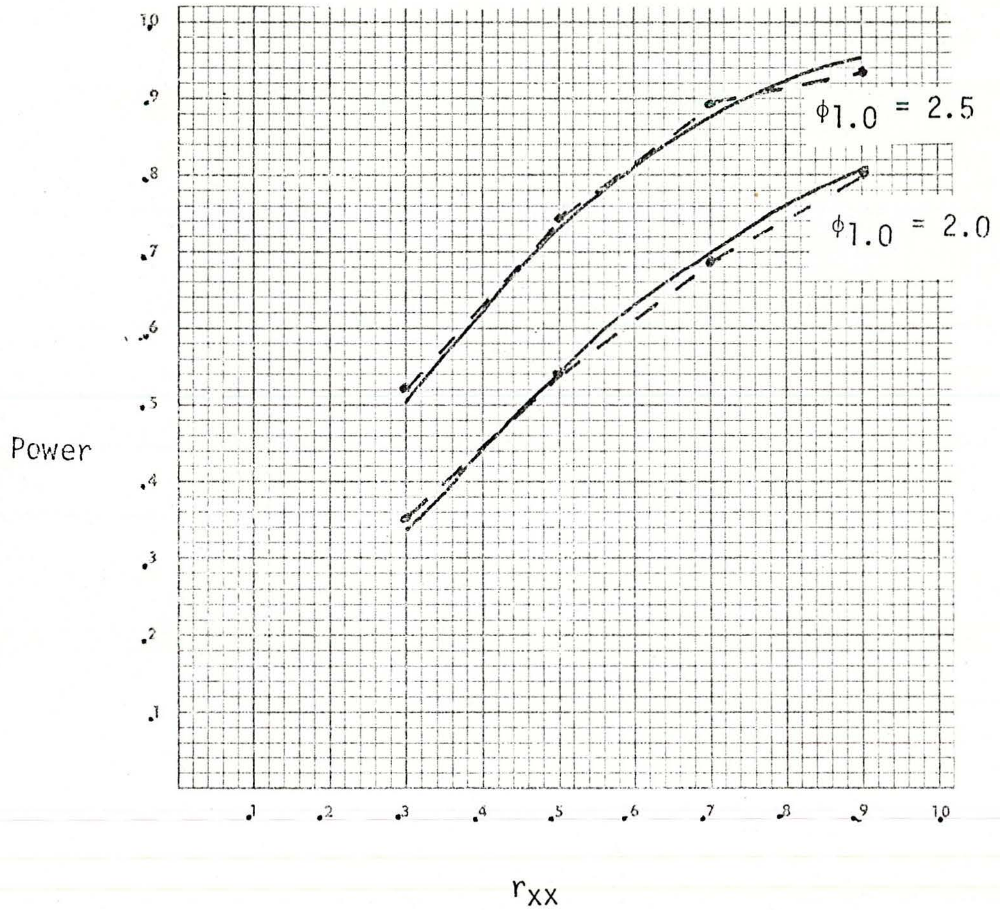
Legend

Formula Method ————————

Monte Carlo Method — — — — —

Figure 6

A Comparison of the Power of the ANOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=3, n=4, and α=.05



Legend

Formula Method ————————

Monte Carlo Method — — — — —

Figure 7

A Comparison of the Power of the ANOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=3, n=11, and α=.05

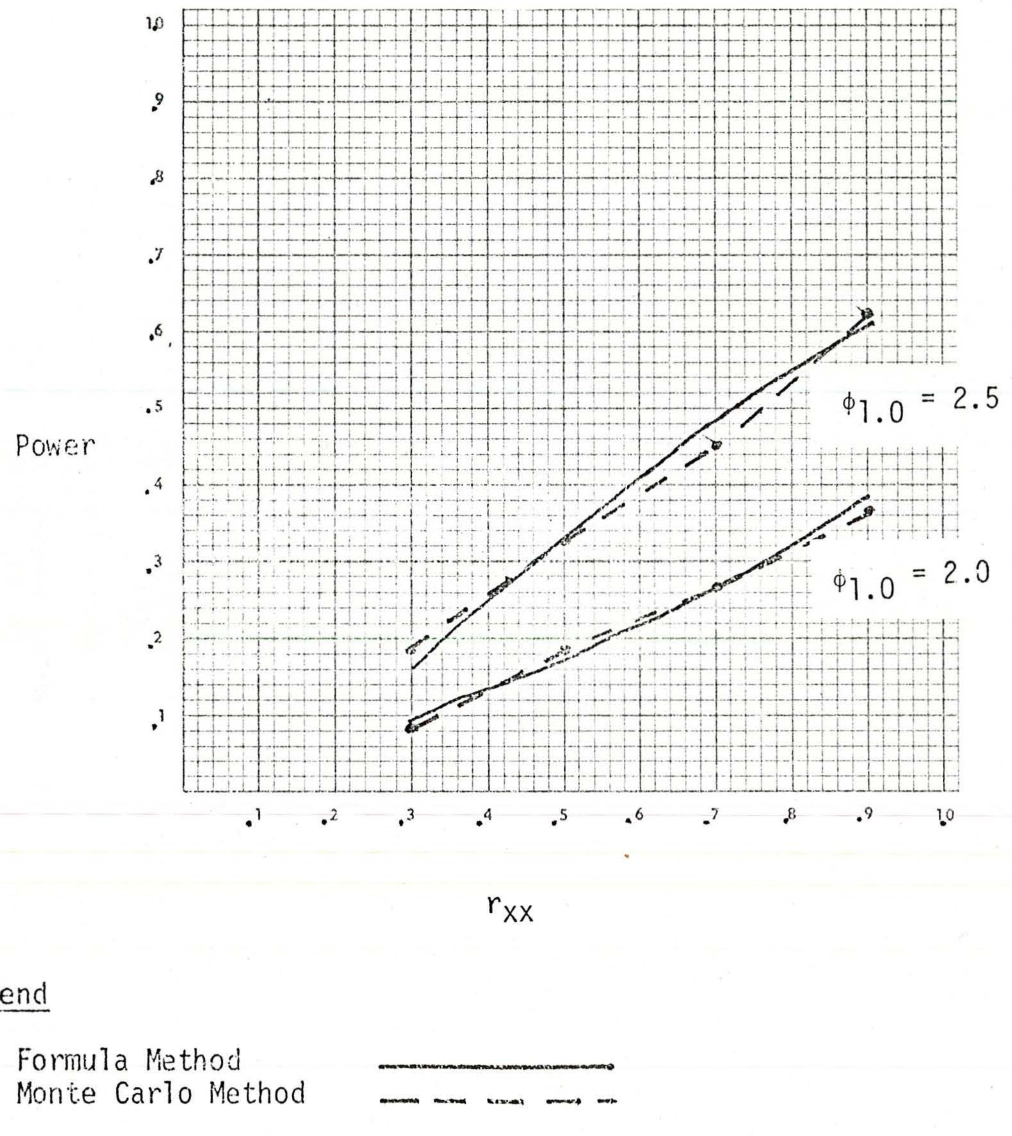
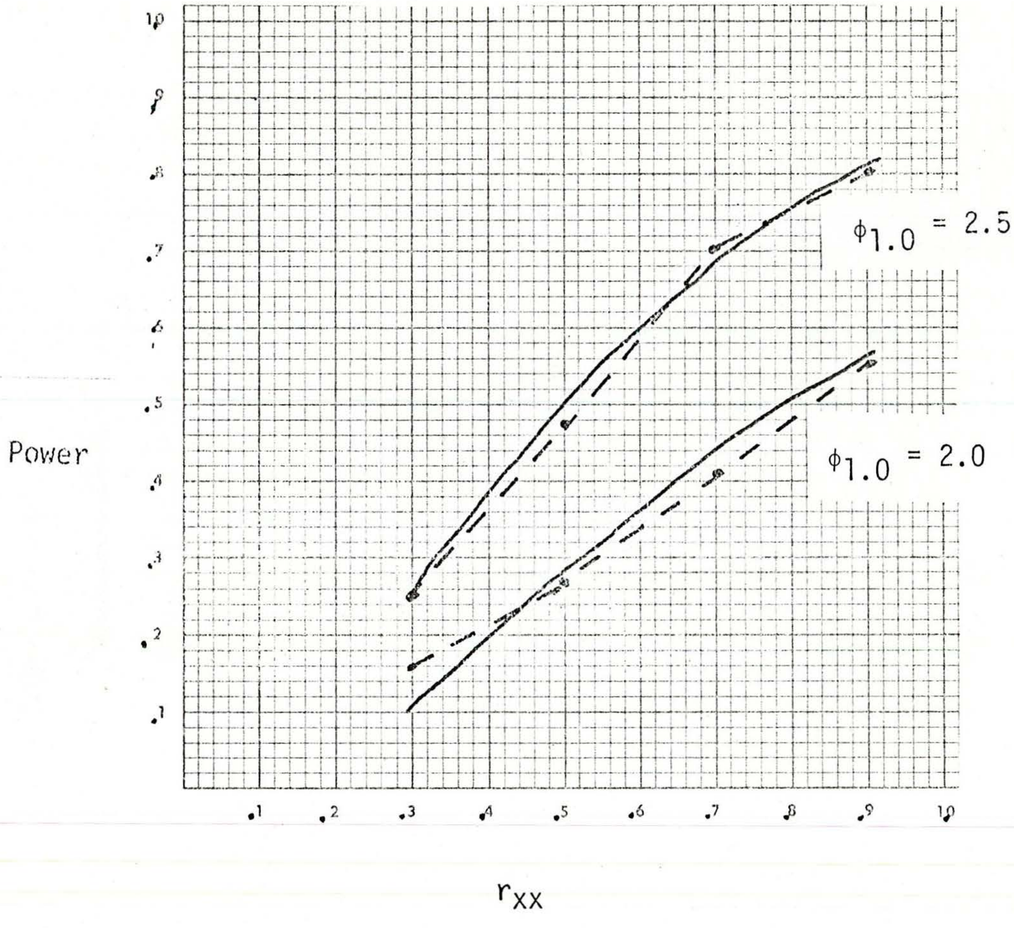


Figure 8

A Comparison of the Power of the ANOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=3, n=4, and α=.01



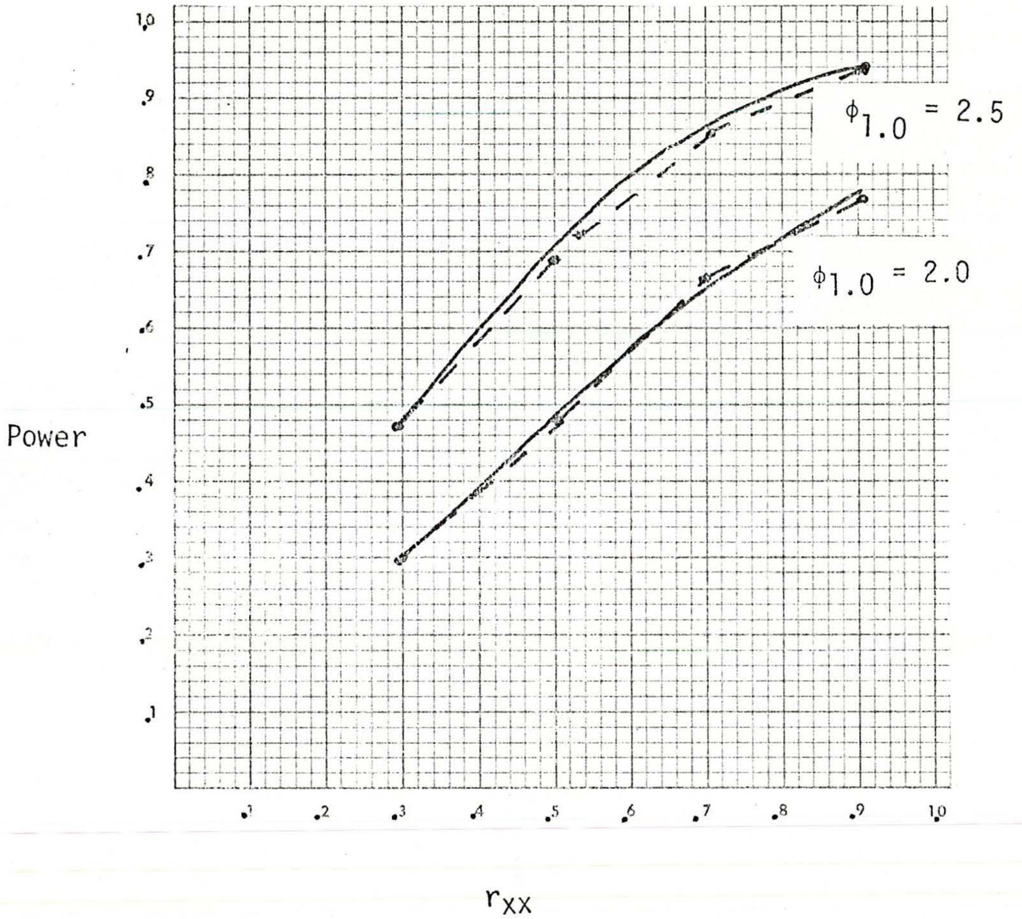
Legend

Formula Method ————————

Monte Carlo Method — — — — —

Figure 9

A Comparison of the Power of the ANOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=3, n=11, and α=.01

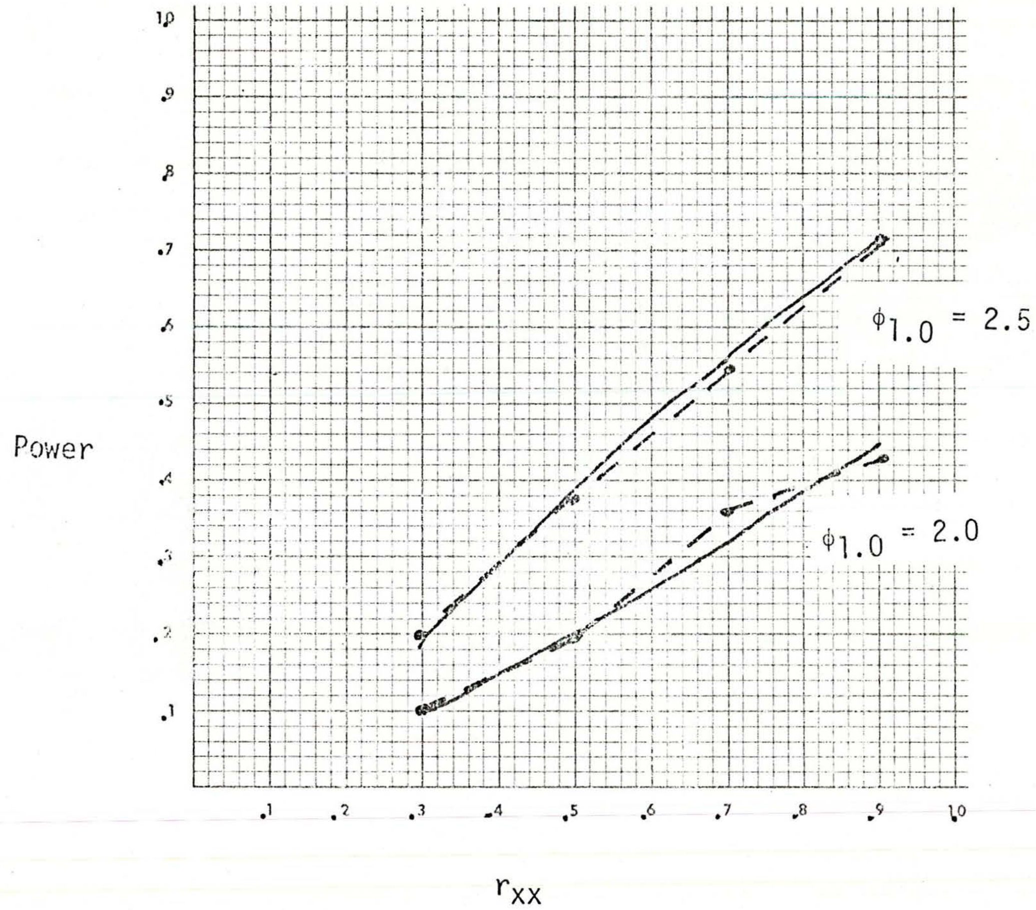


Legend

Formula Method
Monte Carlo Method

Figure 10

A Comparison of the Power of the ANOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=5, n=3, and α=.05

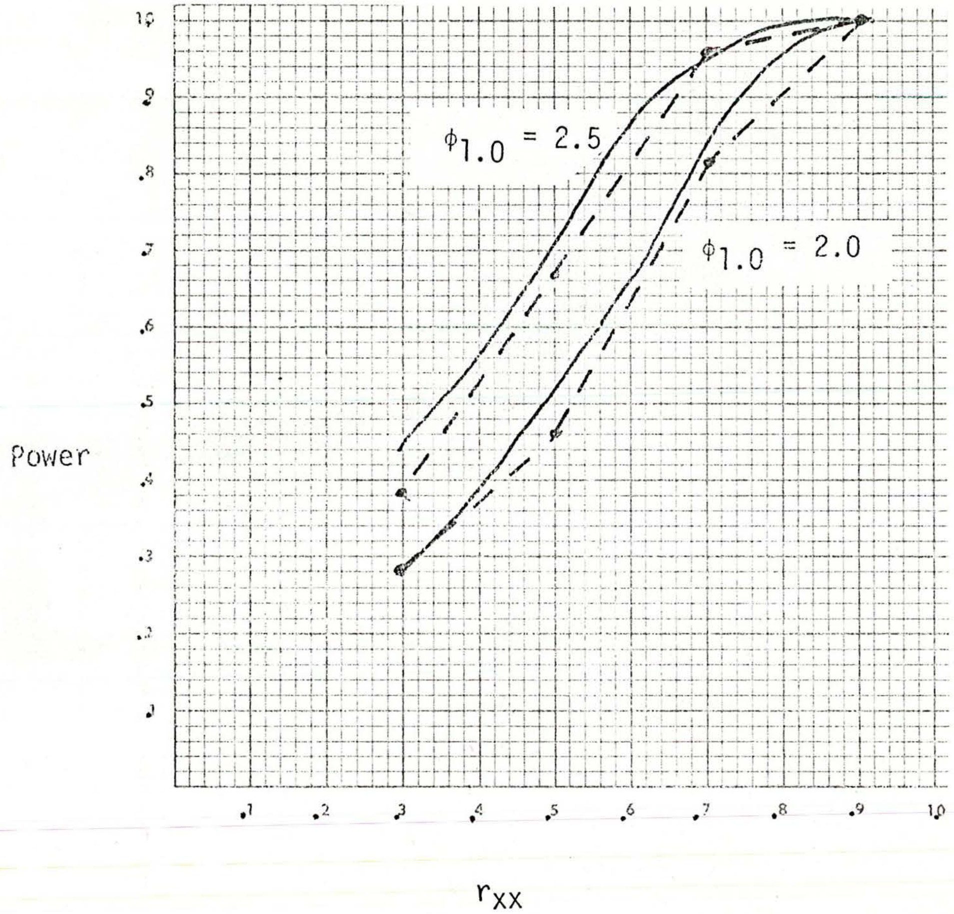


Legend

Formula Method ————————
Monte Carlo Method — — — — —

Figure 11

A Comparison of the Power of the ANOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=5, n=3, and α=.01



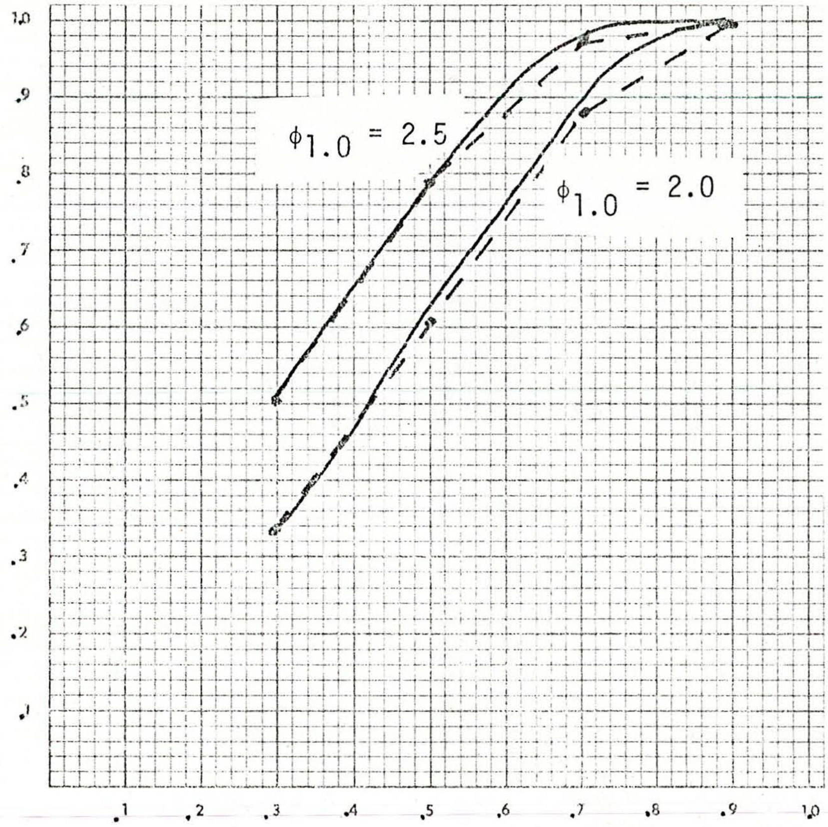
Legend

Formula Method ———————
Monte Carlo Method — — — — —

Figure 12

A Comparison of the Power of the ANCOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=2, n=6, and α=.05

Power



r_{xx}

Legend

Formula Method ————————

Monte Carlo Method — — — — —

Figure 13

A Comparison of the Power of the ANCOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=2, n=16, and α=.05

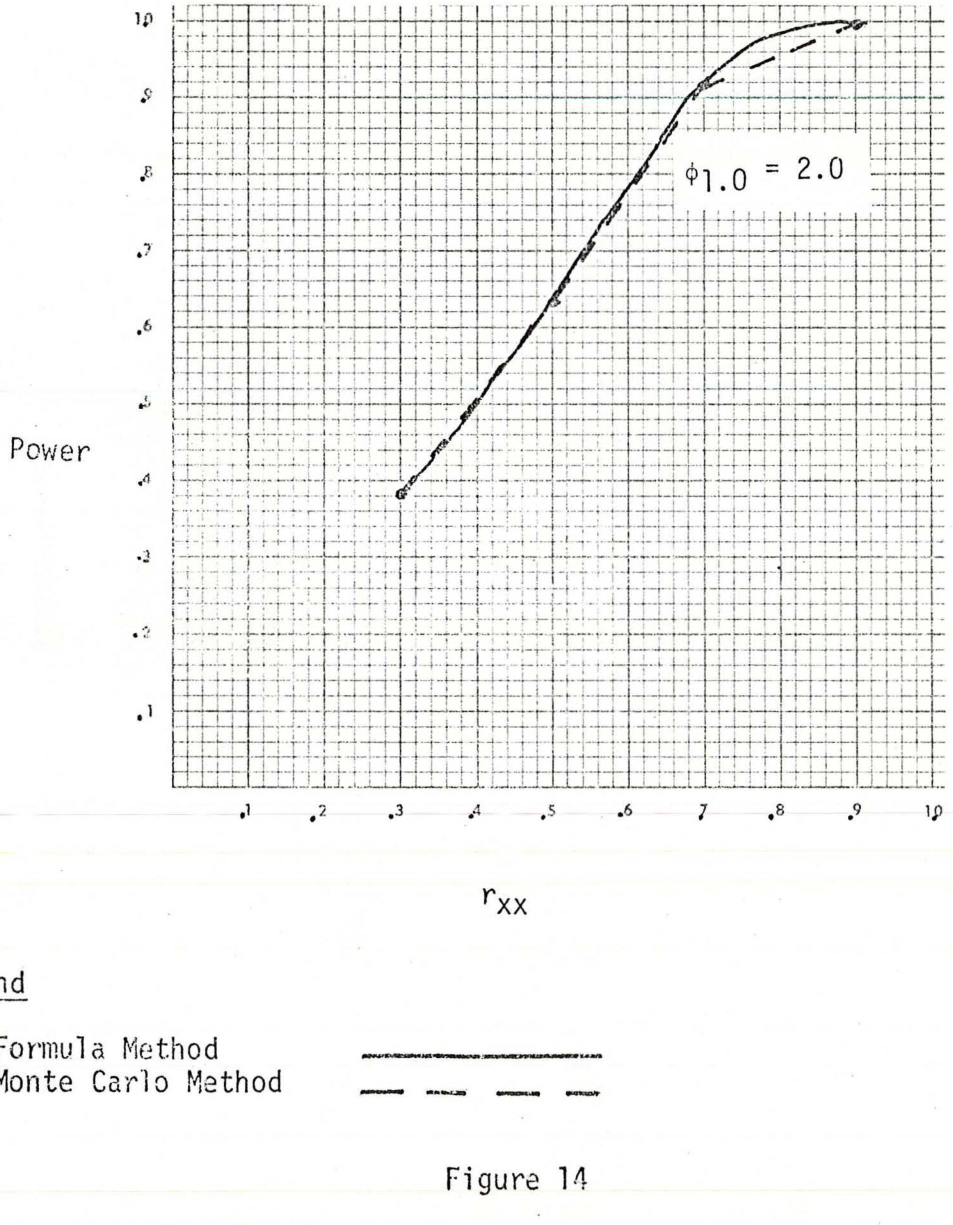
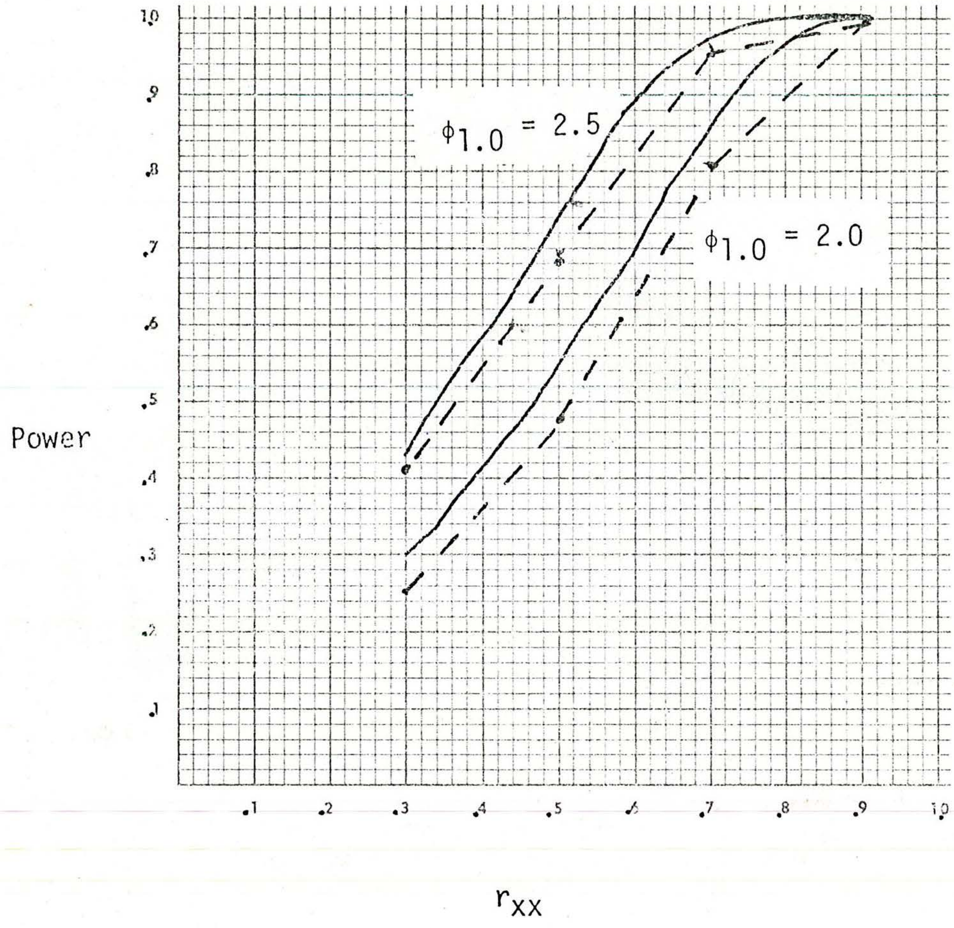


Figure 14

A Comparison of the Power of the ANCOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=2, n=100, and α=.05

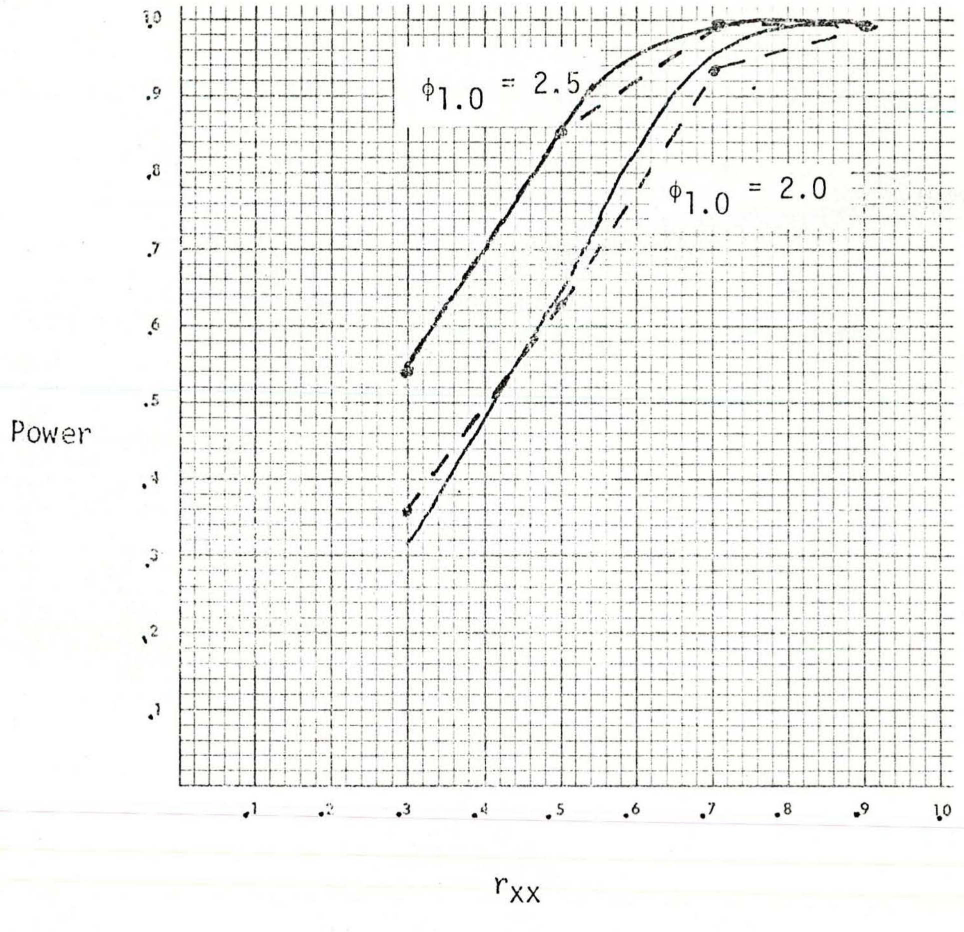


<u>Legend</u>

Formula Method ————————
Monte Carlo Method — — — — —

Figure 15

A Comparison of the Power of the ANCOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=3, n=4, and α=.05



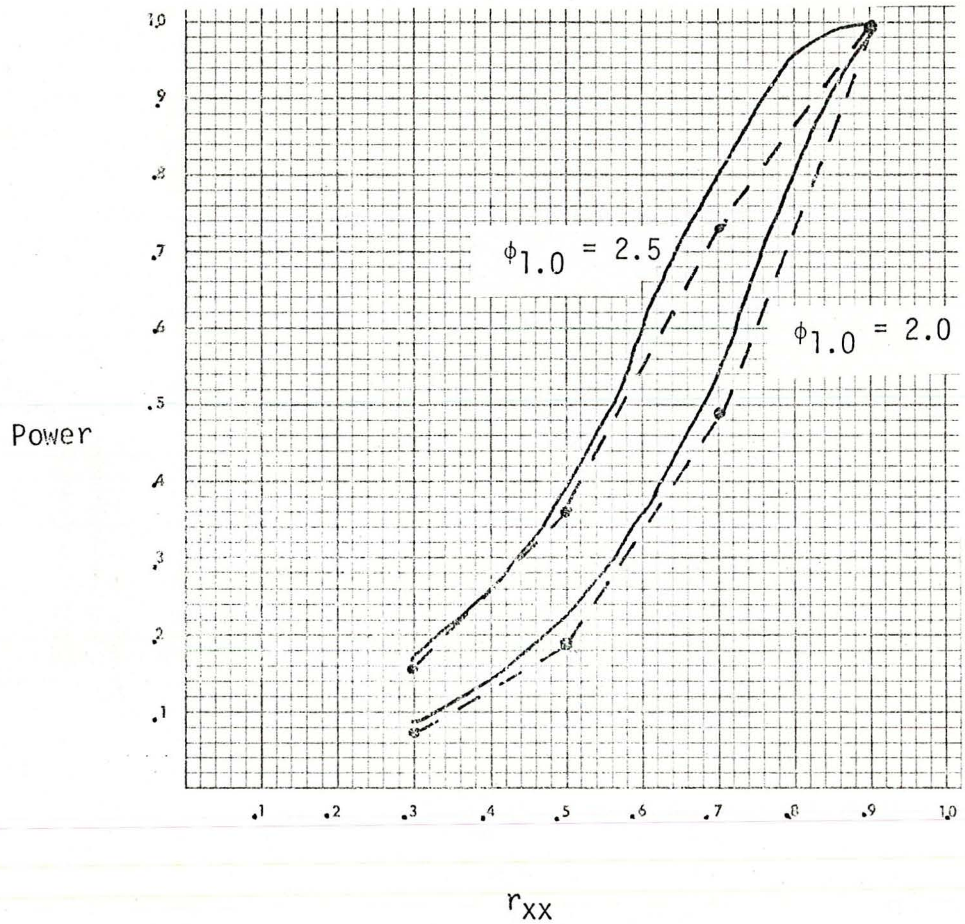
Legend

Formula Method ————————

Monte Carlo Method — — — —

Figure 16

A Comparison of the Power of the ANCOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=3, n=11, and α=.05

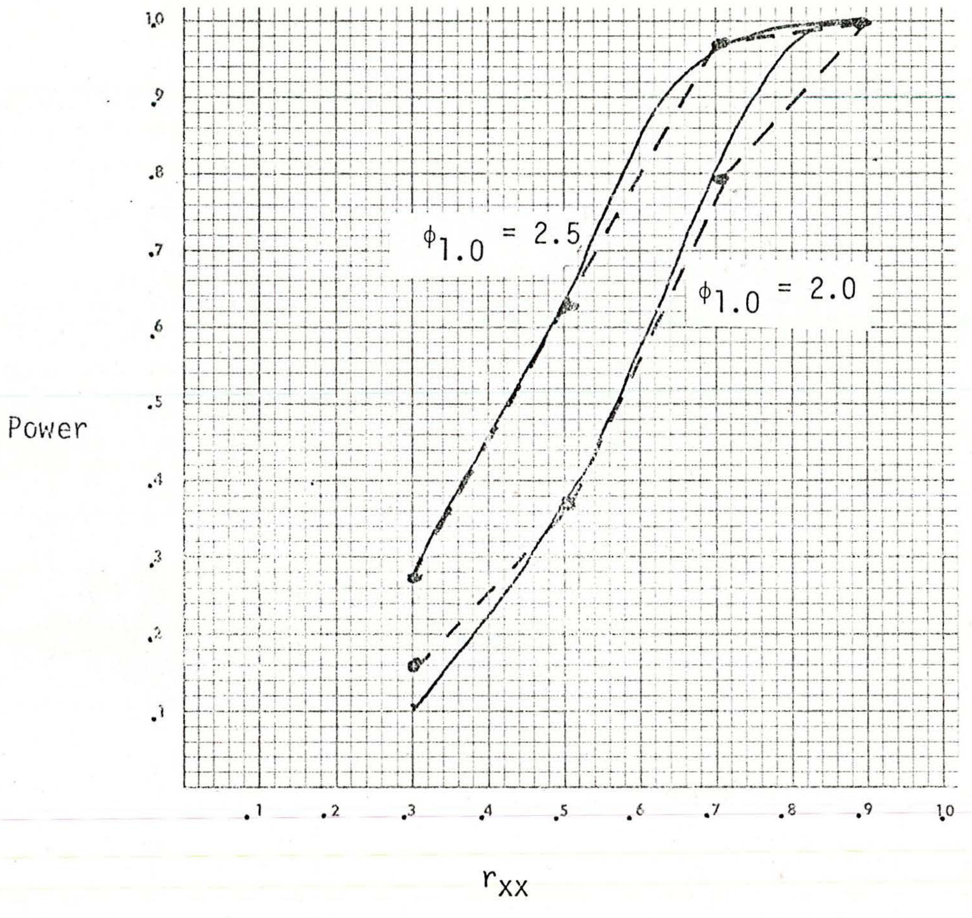


Legend

Formula Method ————————
Monte Carlo Method — — — — —

Figure 17

A Comparison of the Power of the ANCOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=3, n=4, and α=.01

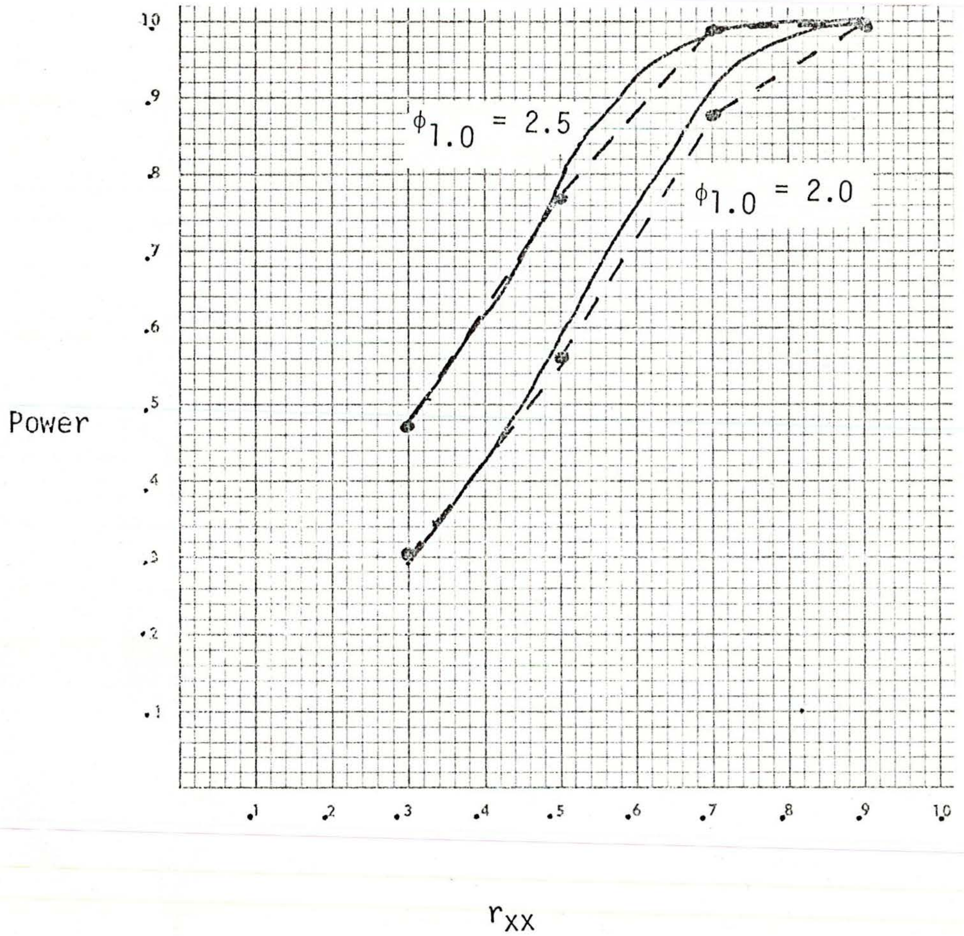


Legend

Formula Method
Monte Carlo Method

Figure 18

A Comparison of the Power of the ANCOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=3, n=11, and α=.01



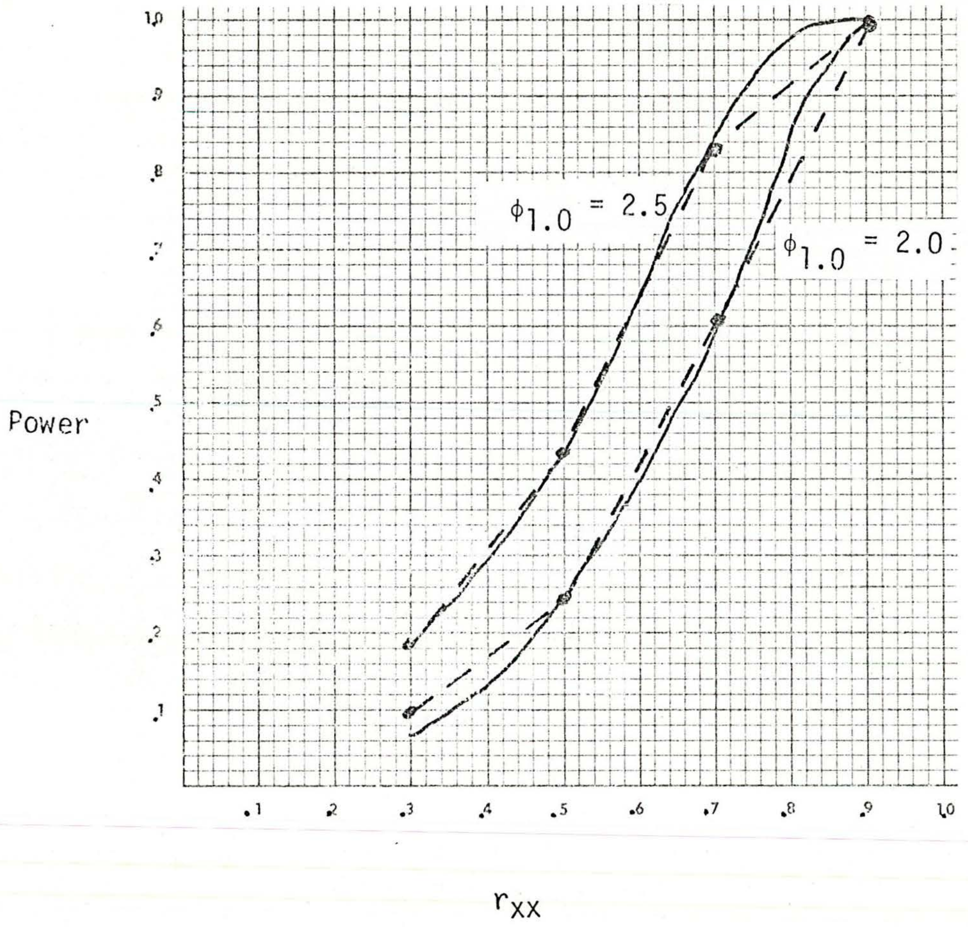
Legend

Formula Method ————

Monte Carlo Method — — — —

Figure 19

A Comparison of the Power of the ANCOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=5, n=3, and α=.05



Legend

Formula Method ————————
Monte Carlo Method — — — — —

Figure 20

A Comparison of the Power of the ANCOVA as Estimated by the Formula Method and the Monte Carlo Method with Power as a Function of Test Reliability (r_{xx}) for J=5, n=3, and α=.01

It can be observed from Figures 3, 4, and 5 that for two groups with α=.05, $\phi_{1.0}$=2.0, power never reaches the desired .8 level even for n=100. However, for two groups, when $\phi_{1.0}$=2.5, with only n=6, adequate power is achieved when $r_{xx} \geq .75$. Figures 6 and 7 demonstrate that for three groups with $\phi_{1.0}$=2.5 and α=.05, adequate power can be obtained for $r_{xx} \geq .75$. For three groups, n=11, $\phi_{1.0}$=2.0, and r_{xx}=.9 the desired .8 minimum level of power is also achieved. Figures 8 and 9 present the results for three groups with α=.01 for n=4 and n=11. When α=.01, adequate power is reached for n=11, and $\phi_{1.0}$=2.5. Figures 10 and 11 provide the power estimates for five groups, n=3, for α=.05 and α=.01. When α=.05, sufficient power is obtained when $\phi_{1.0}$=2.5 and reliability is greater than .6. However, for α=.01, power is less than .8 for both $\phi_{1.0}$=2.0 and $\phi_{1.0}$=2.5.

By examining Figures 3 through 11, it can be observed that the influence of test reliability is fairly consistent. Since the curves in these graphs are somewhat linear, a generalization can be made concerning the increase in power as test reliability increases. For each .10 increase in test reliability, power increases by approximately .06 or .07. This finding is for the ANOVA test and is independent of $\phi_{1.0}$, n, J, α, and r_{xx}.

ANCOVA Graphs

The data for ANCOVA based on the Formula Method and the Monte Carlo Method are presented graphically in Figures 12 through 20. A comparison between the two methods for the ANCOVA test can be made from these graphs. The purpose of this comparison is to determine the accuracy of the Formula Method for the ANCOVA test.

As can be noted from Figures 12 through 20, the Formula Method values sometimes overestimate power but vary no more than .07 from the Monte Carlo derived values. In most cases, the differences are less than .03.

For all cases examined in this study where $\phi_{1.0}$=2.0 and $\phi_{1.0}$=2.5, adequate power was achieved for ANCOVA with moderate (.5 to .7) to high (above .7) correlation and reliability levels. This finding further supports the superiority of the ANCOVA test over the ANOVA test. For r_{xy} and r_{xx} less than .9, the curves presented in Figures 12 through 20 provide the basis for a generalization about the influence of r_{xy} and r_{xx} on statistical power. For each .10 increase in the correlation between the variate and the covariate and test reliability, statistical power increases by approximately .10 to .12. This finding is for the ANCOVA test and is independent of $\phi_{1.0}$, n, J, α, r_{xx} and r_{xy}.

SUMMARY

The purpose of Chapter 4 was to present the findings of this study. These findings were presented in the form of tables, figures and narrative interpretation. The Formula Method was compared to the Monte Carlo Method; also the ANCOVA test was compared to the ANOVA test. The influence of other parameters were also examined. They were number of groups (J), number per group (n), level of significance (α), test reliability (r_{xx}), correlation between the variate and the covariate (r_{xy}), and treatment effect ($\phi_{1.0}$).

As a result of this study, it was found that as test reliability increases by .1, the statistical power of the ANOVA increases approximately .06 or .07. For ANCOVA, it was found that as the correlation between the variate and the covariate and test reliability increase by .1, the statistical power is increased by about .10 or .12. Another finding was that for correlations greater than or equal to .5, ANCOVA is considerably more powerful than ANOVA.

The graphs illustrated that the Formula Method provides a good estimate of statistical power even for small J and n. Various levels of significance and reliability were examined and it was observed that the Formula Method maintained its stability. It was concluded that the assumption referred to in Chapter 2 made by Hopkins and Hopkins during the derivation of the Formula for ϕ_A and ϕ_C did not seriously invalidate the resulting power estimates.

Chapter 5

SUMMARY, IMPLICATIONS AND RECOMMENDATIONS

In this chapter, an overview of the study is presented along with a summary of the findings. Implications and recommendations of these findings are also discussed. Specifically, recommendations are made concerning areas for future study and general recommendations to those interested in the administration of educational research and evaluation.

OVERVIEW

In the planning phase of developing quality educational research there are several critical decision points. Some of these decisions are made to ensure that the research will have adequate statistical power, that is, the ability to detect meaningful differences, if they do exist. In an unpublished article, K. D. Hopkins and B. R. Hopkins commented that all theoretical and empirical options for increasing statistical power had been investigated except for the reliability of the dependent variable. In this article they derived a method for estimating the power of statistical tests (specifically ANOVA and ANCOVA) which included the reliability of the dependent variable. However, this method was based on some untested mathematical assumptions and therefore needed to be empirically verified.

The purpose of this study was to empirically verify the method developed and to investigate the interaction between the reliability of the dependent variable and the statistical power of the analysis of variance and the analysis of covariance. The influence of other relevant parameters, namely, sample size, non-centrality parameter, level of significance, and number of groups were considered.

This purpose was accomplished by using Monte Carlo methods and the Burroughs B-6700 computer. Systematically, the relevant parameters were manipulated, and the resulting statistical power of the ANOVA and ANCOVA procedures estimated.

FINDINGS

It was found that the formula developed by Hopkins and Hopkins to estimate statistical power, which included as a factor the reliability of the dependent variable, provides a stable estimate of power even for small values J and n (J=number of groups and n=number in each group). As test reliability increases by .1, the statistical power of the ANOVA increases by about .06 or .07. It was also found that, for ANCOVA, as the correlation level and reliability level increase by .1, the resulting statistical power is increased by approximately .10 or .12. It was also observed that for correlations greater than or equal to .5, the ANCOVA test is more powerful than the ANOVA test. When the correlation was .9, the increase in power when ANCOVA was used rather than ANOVA was greatest when the original power of ANOVA was between .2 and .8. It was also observed that when the correlation was low, for example, .3, little or no gain in power was achieved by using ANCOVA instead of ANOVA.

In general it was observed that when the reliability of the dependent variable was high, the ANCOVA test had high power even for groups with small sample size (n). It was also observed that when the reliability was low, power was consistently low. Even with large sample sizes, experiments with dependent variables having low reliability had low power.

IMPLICATIONS AND RECOMMENDATIONS

The power of a statistical test should be estimated during the planning stage of a research study. There are three major implications of doing this. First, if the derived power is low, the researcher may decide not to conduct the study or may decide to change some of the parameters (that is, increase n, α, or r_{xx}). Second, whether the null hypothesis is rejected or is not rejected, the researcher needs to know how much confidence can be given to this result. If the study had power of .90 and the null hypothesis was not rejected, the researcher could feel fairly confident with the results. However, if the power was .30, and the null hypothesis was not rejected, the researcher would be less sure of the result. There may have been no difference or there may have been a difference, but the statistical procedures used may not have had enough power to detect the difference. Another reason for computing power during the planning phase of an experiment is to identify unnecessarily large and elaborate studies. For example, if an n of 20 will provide high power, it would be costly and unreasonably cumbersome to use an n of 100.

Recommendations for Future Studies

The following recommendations for future study are made. Since test reliability strongly influences the power of the ANOVA and ANCOVA test, the influence of reliability on other statistical tests should be investigated. It would be helpful to the researcher who wants to consider .10 level of significance in studies if power charts were created for .10 level of significance. Currently, charts for only .01 and .05 are available. Another aid to researchers would be to include computer programs to compute statistical power in the commonly used computer software packages such as SPSS (Statistical Package for the Social Sciences).

Recommendations for Researchers

As a result of this study, a few recommendations are made for researchers and evaluators. Because of the influence of test reliability on power, more emphasis should be placed on developing and using highly reliable dependent variables. Also, whenever defensible, a .10 level of significance should be considered since it is a convenient way to increase the power of the ANOVA and ANCOVA tests. If a covariate with an anticipated correlation with the variate of .5 or greater is available, the ANCOVA test rather than the ANOVA test should be considered whenever appropriate.

As previously indicated, statistical power should be estimated routinely in the planning phase of research. If this was done, high power could be achieved by varying the relevant parameters to optimize power. Parameters of studies without adequate power could be altered and unnecessarily large and elaborate studies could be reduced.

BIBLIOGRAPHY

A. BOOKS

Churchman, C. West, Russell L. Ackoff, and E. Leonard Arnoff. Introduction to Operations Research. New York: John Wiley & Sons, Inc., 1957.

Cohen, Jacob. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1969.

Fisher, R. A. Statistical Methods for Research Workers. London: Oliver & Boyd, 1925.

Gage, N. L., ed. Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Hays, William L. Statistics. New York: Holt, Rinehart and Winston, 1963.

Issac, Steven, and William B. Michael. Handbook in Research and Evaluation. San Diego: Robert R. Knapp, 1972.

Kerlinger, F. N. Foundations of Behavioral Research. 2d ed. New York: Holt, Rinehart and Winston, 1973.

Kirk, Roger E. Experimental Design: Procedures for the Behavioral Sciences. Belmont: Brooks/Cole, 1968.

McCall, W. A. How to Experiment in Education. New York: Macmillan, 1923.

Meyer, J. L. Fundamentals of Experimental Design. Boston: Allyn and Bacon, 1972.

Naylor, Thomas H. The Design of Computer Simulation Experiments. Durham, North Carolina: Duke University Press, 1969.

Roscoe, John T. Fundamental Research Statistics for the Behavioral Sciences. 2d ed. New York: Holt, Rinehart and Winston, 1975.

Stanley, Julian C., and Kenneth D. Hopkins. Educational and Psychological Measurement and Evaluation. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

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

B. PERIODICALS

Baker, F. B., and R. O. Collier, Jr. "Some Empirical Results on Variance Ratios under Permutations in the Completely Randomized Design," Journal of the American Statistical Association, LXI, No. 5 (1966), 814-20.

Borich, Gary D., and Robert C. Godbout. "Extreme Groups Designs and the Calculations of Statistical Power," Educational and Psychological Measurement, XXXIV (Autumn, 1972), 663-75.

Brewer, J. K. "On the Power of Statistical Tests in the American Educational Research Journal," American Educational Research Journal, IX, No. 3 (1972), 391-401.

_______. "Issues of Power: Clarification," American Educational Research Journal, XI, No. 2 (1974), 189-92.

Cleary, T. A., and R. L. Linn. "Error of Measurement and the Power of a Statistical Test," British Journal of Mathematical and Statistical Psychology, XXII, No. 1 (1969), 49-55.

Cochran, W. G. "Errors of Measurement in Statistics," Technometrics, X, No. 4 (1968), 637-66.

_______. "Some Effects of Errors of Measurement on Multiple Correlations," Journal of the American Statistical Association, LXV, No. 329 (1970), 22-34.

Cohen, J. "Approximate Power and Sample Size Determination for Common One-Sample and Two-Sample Hypothesis Tests," Educational and Psychological Measurement, XXX, No. 4 (1970), 811-31.

_______. "Statistical Power Analysis and Research Results," American Educational Research Journal, X, No. 3 (1973), 225-29.

_______. "The Statistical Power of Abnormal Social Psychological Research," Journal of Abnormal and Social Psychology, LXV, No. 1 (1962), 145-53.

Haase, Richard F. "Power Analysis in Research in Counselor Education," Counselor Education and Supervision, XIV, No. 2 (1974), 124-32.

Hopkins, K. D. "Research Design and Analysis Clinic: Preventing the Number-One Misinterpretation of Behavioral Research, or How To Increase Statistical Power," The Journal of Special Education, VII, No. 1 (1973), 103-106.

Jones, Billie J., and James K. Brewer. "An Analysis of the Power of Statistical Tests Reported in The Research Quarterly," The Research Quarterly, XLIII, No. 1 (1972), 7-15.

Meyer, D. L. "Issues of Power: Rejoinder," American Educational Research Journal, XI, No. 2 (1974), 193-94.

_______. Statistical Tests and Surveys of Power: A Critique," American Educational Research Journal, XI, No. 2 (1974), 179-88.

Overall, John E., and Sudhir N. Salal. "Design of Experiments to Maximize Power Relative to Cost," Psychological Bulletin, LXIV, No. 5 (1965), 339-50.

Pearson, E. S., and H. O. Hartley. "Charts of the Power Function for Analysis of Variance Tests, Derived from the Non-Central F-Distribution," Biometrika, XXXVIII, No. 1 (1951), 112-30.

Penick, J., and J. K. Brewer. "The Power of Statistical Tests in Science Teaching Research," Journal of Research in Science Teaching, IX, No. 4 (1972), 377-79.

Rodger, R. S. "Multiple Contrasts, Factors, Error Rate and Power," The British Journal of Mathematical and Statistical Psychology, XXVII, No. 2 (1974), 199-204.

Sutcliffe, J. P. "Error of Measurement and the Sensitivity of a Test of Significance," Psychometrika, XXIII, No. 1 (1958), 9-17.

C. OTHER SOURCES

Hopkins, K. D., and B. R. Hopkins. "The Effect of the Reliability of the Dependent Variable on the Power of the Analysis of Variance," (unpublished paper, University of Colorado, Boulder; and University of the Pacific, Stockton, 1974).

Porter, Andrew C. How Errors of Measurement Affect ANOVA, Regression Analysis, ANCOVA and Factor Analyses. U.S., Educational Resources Information Center, ERIC Document Ed 050 172, January, 1971.

_______. "The Effects of Using Fallible Variables in the Analysis of Covariance." Unpublished Ph.D. dissertation, University of Wisconsin, 1967.

APPENDIX A

A Summary of the Literature Search
Conducted for this Study

Source	V.I.--to date included
Current Index to Journals in Education	X
Current Index to Statistics: Applications, Methodology and Theory	X
Business Periodicals Index	X
Psychological Abstracts	X
Research in Education (ERIC)	X
DATRIX II Computer Search	
Current Statistical Textbooks and their Bibliographies	
Bibliographies of related periodical literature	

APPENDIX B

Computer Program Using Monte Carlo Procedures to Estimate the Statistical Power of the ANOVA and ANCOVA F-Tests for Varying J, n, $\phi_{1.0}$, α, and r_{xx}

```
     B6700/B7700    F O R T R A N   C O M P I L A T I O N   M A R K

FILE  5=JIN, UNIT = READER

FILE  6= JOUT,  UNIT = PRINTER
C    THE 1ST CARD-NO. OF GROUPS, N, NREPS, IWANTR, IWANTF, NRANOR,NCOPY
C       IN THIS FORMAT I2, I3, I4, 2I1, 10I2)
C    2ND- XSTART,NO. OF PHI VALUES , INITIAL RXX, RXXINC, RXXMAX, F90,
C       F95, F99, FCOV90, FCOV95 FCOV99 IN THE FORMAT - F5.0,I5, 14F5.2
C    3RD- THE SD OF THE UNIFORM RANDOM NUMBER POPULATION, FOLLOWED BY T
C       DESIRED PHI VALUES IN THIS FORMAT- 16F5.0


        DIMENSION XZCOEF(1000)
        DIMENSION PHI(10)
        DIMENSION X(10,100), FY(1000), FZ(1000), FYZ(1000), Y(10,100)
        DIMENSION Z(10,100), ALPHAS(10,10), YZCOFF(1000),  XYCOEF(1000)
        JIN = 5
        JOUT = 6
        READ(JIN,500) NGRUPS, N, NREPS , IWANTR, IWANTF,NRANOR ,NCOPY
500     FORMAT(I2, I3, I4, 2I1,9I2)
         RANORM = NRANOR
        READ(JIN,501)XSTART,NALFST,RXX,RXXINC,RXXMAX,F90,F95,F99,
       2 FCOV90, FCOV95, FCOV99, COVRYZ
501     FORMAT(F5.0,I5, 14F5.2)
        READ(JIN,504) SIGMA, (PHI(I), I=1,NALFST)
504     FORMAT(16F5.0)
        SIGMA = SIGMA/SQRT(RANORM)
        XGRUPS = NGRUPS
        XN=N
        DO 37 I = 1,NALFST
        XK = XGRUPS*(PHI(I)*SIGMA)**2/XN
        GO TO (37,23,24,25,26,27),NGRUPS
 23     D= SQRT(XK/2.0)
        ALPHAS(I,1) = -1.0*D
        ALPHAS(I,2) =D
        GO TO 37
24      D= SQRT(XK/2.0)
        ALPHAS(I,1) = -1.0*D
        ALPHAS(I,2)= 0
        ALPHAS(I,3) = D
        GO TO 37
25      D=SQRT(XK/20.0)
        ALPHAS(I,1) = -3.0*D
        ALPHAS(I,2)= -1.0*D
        ALPHAS(I,3)= D
        ALPHAS(I,4)= 3.0*D
        GO TO 37
```

```
26      D=SQRT(XK/10.0)
        ALPHAS(I,1) = -2.0*D
        ALPHAS(I,2) = -1.0*D
        ALPHAS(I,3) = 0
        ALPHAS(I,4) = D
        ALPHAS(I,5) = 2.0*D
        GO TO 37
27      D=SQRT(XK/70.0)
        ALPHAS(I,1) = -5.0*D
        ALPHAS(I,2) = -3.0*D
        ALPHAS(I,3) = -1.0*D
        ALPHAS(I,4) = D
        ALPHAS(I,5) = 3.0*D
        ALPHAS(I,6) =         5.0*D
37      CONTINUE
        DO 20 L = 1,NALFST
        RXXA = RXX - RXXINC
10      CONTINUE
        RXXA = RXXA + RXXINC
        IF(RXXA.GT.RXXMAX) GO TO 20
        SEFAC = SQRT(RXXA/(1.0-RXXA))
        RYZ = COVRYZ
        RXY = COVRYZ**2/RXXA
        IF(RXY.GT.0.999999) RXY = .999999
        SEFACY = SQRT(RXY/(1.0-RXY))
        IF(COVRYZ.EQ.0.0) SEFACY = SEFAC
        IF(COVRYZ.EQ.0.0) RYZ=RXXA
        NYF90=0
        NYF95=0
        NYF99=0
        NYLT90 = 0
        NZF90=0
        NZF95=0
        NZF99=0
        NZLT90=0
        NYZF90=0
        NYZF95=0
        NYZF99=0
        NYZLT9=0
        DO 101 K = 1,NREPS
        DO 103 J = 1,NGRUPS
        DO 103 I = 1,N
        TEMP = 0
        DO 32 NRAN=1,NRANOR
        TEMP = TEMP + RANDOM(XSTART)
32        CONTINUE
        X(J,I) = TEMP/RANORM + ALPHAS(L,J)
        YTEMP=0
        DO 33 NRAN = 1,NRANOR
        YTEMP = YTEMP + RANDOM(XSTART)
33      CONTINUE
         YTEMP = YTEMP/RANORM
        Y(J,I) = X(J,I) +(YTEMP - .5) / SEFACY      - ALPHAS(L,J)
        ZTEMP =0
        DO 34 NRAN = 1,NRANOR
        ZTEMP = ZTEMP + RANDOM(XSTART)
34      CONTINUE
        ZTEMP = ZTEMP/RANORM
        Z(J,I) = X(J,I) +  (ZTEMP - .5) / SEFAC
```

```
103   CONTINUE
      IF(IWANTR.EQ.0) GO TO 11
      CALL CORREL(X,Y,NGRUPS,N,COEFF)
      XYCOEF(K) = COEFF
      CALL CORREL(Y,Z,NGRUPS,N,COEFF)
      YZCOEF(K) = COEFF
      CALL CORREL(X,Z,NGRUPS,N,COEFF)
      XZCOEF(K) = COEFF
11     CONTINUE
      CALL ANOVA(Y,Z,NGRUPS,N,FYVAL,FZVAL,FYZVAL)
      FY(K) = FYVAL
      FZ(K) = FZVAL
      FYZ(K) = FYZVAL
101   CONTINUE
      DO 30 M = 1,NREPS
      IF(FY(M).GT.F90) NYF90 = NYF90 + 1
      IF(FY(M).GT.F95) NYF95 = NYF95 + 1
      IF(FY(M).GT.F99) NYF99 = NYF99 + 1
      IF(FY(M).LT.F90) NYLT90=NYLT90 + 1
      IF(FZ(M).GT.F99) NZF99 = NZF99 + 1
      IF(FZ(M).GT.F95) NZF95 = NZF95 + 1
      IF(FZ(M).GT.F90) NZF90 = NZF90 +1
      IF(FZ(M).LT.F90) NZLT90 = NZLT90+1
      IF(FYZ(M).GT.FCOV99) NYZF99 = NYZF99 + 1
      IF(FYZ(M).GT.FCOV95) NYZF95 = NYZF95 + 1
      IF(FYZ(M).GT.FCOV90) NYZF90 = NYZF90 +1
30    IF(FYZ(M).LT.FCOV90) NYZLT9 = NYZLT9 +1
      DO 35 NCOP = 1,NCOPY
      WRITE(JOUT,601) PHI(L), RXXA, RYZ, (ALPHAS(L,J), J = 1,NGRUPS)
601   FORMAT('1PHI =', F7.2/ '0RZZ (RELIABILITY) =', F6.3/ 10X, 'RYZ
     2RRELATION BETWEEN COVARIATE AND VARIATE) = ', F6.3/ '0TRT EFFEC
     3ARRAY =', 10F10.5)
      WRITE(JOUT,600) NGRUPS, N, NREPS
600   FORMAT('0NO. OF GROUPS(J)=', I10, 10X, 'NO. OF SCORES/GROUP(N)
     2 I10, 10X, 'REPLIC. OF J GROUPS OF N=', I10)
       IF(IWANTR.EQ.0) GO TO 12
      SUMRXY=0
      SUMRXZ=0
      SUMRYZ = 0
      DO 40 I = 1,NREPS
      SUMRXY = SUMRXY + XYCOFF(I)
      SUMRXZ = SUMRXZ + XZCOFF(I)
      SUMRYZ = SUMRYZ + YZCOFF(I)
40    CONTINUE
      XNREPS = NREPS
      RXYBAR = SUMRXY/XNREPS
      RXZBAR = SUMRXZ / XNREPS
      RYZBAR = SUMRYZ/XNREPS
      WRITE(JOUT,602) RXYBAR, RXZBAR, RYZBAR
602   FORMAT('0 MEAN RXY =', F10.3, 10X, 'MEAN RXZ =', F10.3, 10X, 'M
     2 RYZ =', F10.3)
      WRITE(JOUT,603) (XYCOFF(K), K = 1,NREPS)
```

```
603    FORMAT('0COMPUTED RXY COEFFICIENTS'/ ( ' ', 20F6.2))
       WRITE(JOUT,700)  (XZCOEFF(K), K = 1,NREPS)
700    FORMAT('0COMPUTED RXZ COEFF', / ('0', 20F6.2))
       WRITE(JOUT,604)  (YZCOEFF(K), K = 1,NREPS)
604    FORMAT('0COMPUTED RYZ COEFFICIENTS'/ (' ', 20F6.2))
12     CONTINUE
       IF(IWANTF.EQ.0) GO TO 13
       WRITE(JOUT,606)  (FY(K), K = 1,NREPS)
606    FORMAT('0 F-RATIOS FOR ANOVA OF Y SCORES'/('0', 10F12.3))
616    FORMAT('0 F-RATIOS FOR ANOVA OF Z SCORES'/('0', 10F12.3))
       WRITE(JOUT,616)  (FZ(K), K = 1,NREPS)
        WRITE(JOUT,626)  (FYZ(K), K = 1,NREPS)
626    FORMAT('0 F-RATIOS FOR ANCOVA OF Z SCORES WITH Y AS COVARIATE'/
     2 ('0', 10F12.3))
13     CONTINUE
       WRITE(JOUT,608) F90, F90, F95, F99
608    FORMAT(5(/), 5X, ' F''S L.T. F90 (', F5.2,  ')           F''S G.T
     2F90 (', F5.2, ')           F''S G.T. F95 (', F5.2, ')           F'
     3 G.T. F99 (', F5.2, ')' / 1X, 110('-'))
       WRITE(JOUT,605) NYLT90, NYF90, NYF95, NYF99
605    FORMAT('0Y', I16, 3I30)
       WRITE(JOUT,609) NZLT90, NZF90, NZF95, NZF99
609    FORMAT('0Z', I16, 3I30)
       WRITE(JOUT,608) FCOV90, FCOV90, FCOV95, FCOV99
       WRITE(JOUT,610) NYZLT9, NYZF90, NYZF95, NYZF99
610    FORMAT('0YZ', I15, 3I30)
35     CONTINUE
       GO TO 10
20     CONTINUE
       STOP
       END
```

```
       SUBROUTINE CORREL(X,Y,NGRUPS,N,COEFF)
       DIMENSION X(10,100), Y(10,100)
       SUMX=0
       SUMY =0
       SUMXX=0
       SUMYY=0
       SUMXY=0
        DO 1 J = 1,NGRUPS
       DO1 I=1,N
       SUMX = SUMX + X(J,I)
       SUMY = SUMY + Y(J,I)
       SUMXX = SUMXX + X(J,I)**2
       SUMYY = SUMYY + Y(J,I)**2
       SUMXY = SUMXY + X(J,I) * Y(J,I)
1      CONTINUE
       XN = N * NGRUPS
       XNUM = SUMXY - SUMX*SUMY/XN
       SSX = SUMXX - SUMX*SUMX/XN
       SSY = SUMYY - SUMY*SUMY/XN
       COEFF = XNUM/SQRT(SSX*SSY)
       RETURN
       END
```

```
      SUBROUTINE ANOVA(Y,Z,NGRUPS,N,FYVAL,FZVAL,FYZVAL)
      REAL MSBY, MSWY, MSBZ,MSWZ,MSBYZ,MSWYZ
      DIMENSION Y(10,100), Z(10,100)
      XN=N
      TSUMZ=0
      TSUMZZ=0
      TN = N*NGRUPS
      TSUMY=0
      TSUMYY=0
      TSUMYZ=0
      SSWY=0
      SSWZ=0
      SSWYZ =0
      DO1 J=1,NGRUPS
      SUMZ=0
      SUMZZ=0
      SUMY=0
      SUMYY=0
      SUMYZ=0
      DO 2I=1,N
      SUMZ = SUMZ + Z(J,I)
      SUMY = SUMY + Y(J,I)
      SUMZZ = SUMZZ + Z(J,I)**2
      SUMYY = SUMYY + Y(J,I)**2
      SUMYZ = SUMYZ + Y(J,I)*Z(J,I)
      TSUMZ = TSUMZ + Z(J,I)
      TSUMY = TSUMY + Y(J,I)
      TSUMYY = TSUMYY + Y(J,I)**2
      TSUMZZ = TSUMZZ + Z(J,I)**2
      TSUMYZ = TSUMYZ + Y(J,I)*Z(J,I)
2     CONTINUE
      SSWY = SUMYY - SUMY*SUMY/XN + SSWY
      SSWZ = SSWZ + SUMZZ -SUMZ*SUMZ/XN
      SSWYZ = SSWYZ + SUMYZ - SUMY*SUMZ/XN
1     CONTINUE
      TSSY = TSUMYY - TSUMY*TSUMY/TN
      TSSZ = TSUMZZ - TSUMZ*TSUMZ/TN
      TSSYZ = TSUMYZ - TSUMY*TSUMZ/TN
      TSSZAD = TSSZ - TSSYZ*TSSYZ/TSSY
      SSWZAD = SSWZ - SSWYZ**2/SSWY
      SSBZAD = TSSZAD - SSWZAD
      SSBY = TSSY - SSWY
      SSBZ = TSSZ - SSWZ
      DFB = NGRUPS -1
      DFW = NGRUPS * (N-1)
      MSBY = SSBY/DFB
      MSWY=SSWY/DFW
      MSBZ = SSBZ/DFB
      MSWZ = SSWZ/DFW
      MSBYZ = SSB7AD/DFB
      MSWYZ = SSWZAD/ (DFW-1.0)
      FYVAL = MSBY/MSWY
      FZVAL = MSBZ/MSWZ
      FYZVAL = MSBYZ/MSWYZ
      RETURN
      END
```