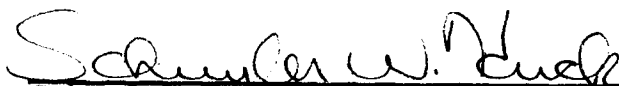
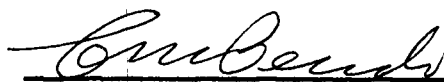


To the Graduate Council:

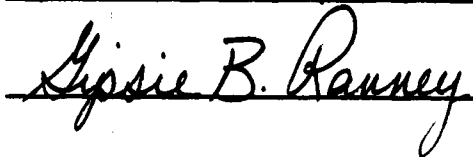
I am submitting herewith a dissertation written by Sheldon Brown Clark entitled "Increasing the Average Precision or Decreasing the Aggregate Variability Among Multiple A Priori Contrasts: Two Algorithms for Allocating Sample Material." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.


Schuyler W. Huck, Major Professor

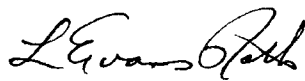
We have read this dissertation
and recommend its acceptance:







Accepted for the Council:



Vice Chancellor
Graduate Studies and Research

INCREASING THE AVERAGE PRECISION OR DECREASING THE
AGGREGATE VARIABILITY AMONG MULTIPLE A PRIORI
CONTRASTS: TWO ALGORITHMS
FOR ALLOCATING SAMPLE
MATERIAL

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Sheldon Brown Clark
March 1983

3065958

DEDICATION

This study is dedicated entirely and without hesitation or qualification, to the individual who has provided the author in this, as in all endeavors, with the strength and the will to continue. Though he chooses not to debase the extent of her contributions by pretending to enumerate them, let it be said, simply, that this effort is dedicated to the singlemost important force in the author's life, his wife, Fern Young Clark.

ACKNOWLEDGEMENTS

The author wishes to express his deepest appreciation to Dr. Schuyler Huck, not only for his assistance and many acts of kindness during the course of this study, but especially for being a friend and advisor, as well as a mentor--in the truest etymological sense of the word--throughout the course of his doctoral study. Also, the author is greatly indebted to Dr. Charles Peccolo for unselfishly providing opportunities to pursue professional growth activities, and, particularly for his willingness to share freely his knowledge and opinions; the author shall always strive to become as principled a person as he.

It has been a pleasure to work with Dr. Kathleen Lawler and Dr. Gipsie Ranney throughout this study. Their confidence in the professional integrity of the author is most gratifying, and their helpful comments and suggestions are appreciated.

Special thanks are extended to Dr. Jack Wilson, of the University of North Carolina at Asheville, who provided invaluable assistance in the verification of the mathematical algorithms. Acknowledgement is also made for the use of the facilities of the University of Tennessee Computing Center, in obtaining the results contained in this document.

ABSTRACT

In true experiments in which sample material can be randomly assigned to treatment conditions, most researchers presume that the condition of equal sample sizes is statistically desirable. When one or more a priori contrasts can be identified which represent a few overriding experimental concerns, however, allocating sample material unequally will lead to more powerful tests of certain hypotheses, when the assumptions underlying the analysis of variance are known to be met. The purpose of this study was to develop systematic procedures for assigning a fixed amount of sample material to treatment conditions using either of two optimality criteria: (a) the maximization of the average precision of a given set of contrasts, or (b) the equalization of precision across a set of such contrasts. On the basis of relevant literature, the squared standard error of the estimate was adopted as the appropriate measure to minimize or equalize.

Algorithms were developed mathematically to achieve each of these goals under specified conditions in single factor, fixed effects, analysis of variance applications. Comparisons of the average precision and variability of precision associated with equal allocation schemes and the two developed in the present study were made, using treatment levels of 3, 4, and 6, average cell sizes of 10

and 15, and representative sets of contrasts. It was determined that the algorithm which minimizes the average squared standard error does not, necessarily, result in the greatest average power, since its minimization cannot be equated to the maximization of power. The same algorithm leads to allocations which were found to have appreciable advantages over equal allocation in certain designs; that which equalizes precision across contrasts was found to result in less powerful analyses than either of the other schemes. Recommendations are made regarding situations in which the researcher might consider applying these algorithms in particular, and power analyses in general, when planning experiments.

TABLE OF CONTENTS

CHAPTER	PAGE
INTRODUCTION	1
I. STATEMENT OF PROBLEM	3
Planning Research	3
Equal Versus Unequal Allocation	16
Constraints of Study	17
Research Questions	19
Need for Study	20
II. ALLOCATING SAMPLE MATERIAL TO INCREASE PRECISION:	
MINIMIZING THE AVERAGE SE_{ψ}^2	22
III. ALLOCATING SAMPLE MATERIAL TO DECREASE VARIABILITY:	
EQUALIZING THE SE_{ψ}^2	27
IV. ALLOCATING SAMPLE MATERIAL AMONG TREATMENT	
CONDITIONS: A COMPARISON OF THREE ALTERNATIVES .	33
Rationale	34
Results	38
V. DISCUSSION	41
Maximizing Power	41
Applications	42
LIST OF REFERENCES	47
APPENDICES	56
Appendix A	57
Appendix B	60
Appendix C	62
VITA	79

LIST OF TABLES

TABLE	PAGE
1. A General Matrix of Coefficients for Planned Comparisons	11
2. Allocating Sample Material to Achieve Equal Precision in a Matrix of Helmert Contrasts . .	32
3. Precision Comparisons Among Three Allocation Schemes: Helmert, $k = 3$, $N = 30$	63
4. Precision Comparisons Among Three Allocation Schemes: Helmert, $k = 3$, $N = 45$	64
5. Precision Comparisons Between Two Allocation Schemes: Nonorthogonal, $k = 3$, $N = 30$	65
6. Precision Comparisons Between Two Allocation Schemes: Nonorthogonal, $k = 3$, $N = 45$	66
7. Precision Comparisons Among Three Allocation Schemes: Helmert, $k = 4$, $N = 40$	67
8. Precision Comparisons Among Three Allocation Schemes: Helmert, $k = 4$, $N = 60$	68
9. Precision Comparisons Between Two Allocation Schemes: Non-Helmert Orthogonal, $k = 4$, $N = 40$	69
10. Precision Comparisons Between Two Allocation Schemes: Non-Helmert Orthogonal, $k = 4$, $N = 60$	70

11.	Precision Comparisons Between Two Allocation	
	Schemes: Nonorthogonal, $k = 4$, $N = 40$	71
12.	Precision Comparisons Between Two Allocation	
	Schemes: Nonorthogonal, $k = 4$, $N = 60$	72
13.	Precision Comparisons Among Three Allocation	
	Schemes: Helmert, $k = 6$, $N = 60$	73
14.	Precision Comparisons Among Three Allocation	
	Schemes: Helmert, $k = 6$, $N = 90$	74
15.	Precision Comparisons Between Two Allocation	
	Schemes: Non-Helmert Orthogonal,	
	$k = 6$, $N = 60$	75
16.	Precision Comparisons Between Two Allocation	
	Schemes: Non-Helmert Orthogonal,	
	$k = 6$, $N = 90$	76
17.	Precision Comparisons Between Two Allocation	
	Schemes: Nonorthogonal, $k = 6$, $N = 60$	77
18.	Precision Comparisons Between Two Allocation	
	Schemes: Nonorthogonal, $k = 6$, $N = 90$	78

INTRODUCTION

In true experiments in which experimental material can be randomly assigned to treatment conditions, a common practice among researchers is to divide the pool of available experimental material equally among treatment conditions. Unfortunately, this condition of equal sample sizes is considered to be a desirable design characteristic regardless of the statistical questions being asked. The present study demonstrates how unequal allocation of sample material will lead to more POWERful¹ tests of hypotheses under certain conditions.

In addressing this issue, Chapter I proceeds as follows: first, a review of recommendations regarding those design decisions which can be made before any data are collected is presented, along with the benefits to be realized by doing so; next, those issues in statistics and research design which may impinge upon a researcher's ultimate decision to allocate his/her sample material unequally are considered; then, the constraints of the present study are delineated, followed by a statement of the research questions and a justification for the study. Chapters II and III contain the mathematical derivations of algorithms for allocating experimental material to

¹ "Power" is defined in the Glossary of Terms in Appendix A, along with other key words and phrases. The terms appearing in this glossary will be capitalized and underlined the first time they appear in the text.

treatment conditions using two different optimality criteria. Chapter IV examines the differences in the power with which certain hypotheses are tested for three different allocation schemes--the two resulting from the algorithms developed in Chapters II and III, and the more traditional equal allocation. Chapter V contains a discussion of the implications of the present study, as well as outlining areas in which further research is needed.

CHAPTER I

STATEMENT OF PROBLEM

Planning Research

The literature dealing with experimental design is replete with recommendations regarding what the focus of an investigator's effort should be when planning a study. Much of the concern among experts focuses on the four interrelated concepts of ALPHA level, sample size, EFFECT SIZE, and power. Since the four concepts constitute a "closed" system, in which varying one effects change in one or more of the others, separate consideration of each factor constitutes an imposed, rather than a natural, structure. Depending upon the nature of the investigations carried out by a targeted audience, or upon the relative importance ascribed to each by a particular author, experts have long exhorted researchers to consider one or more of these factors. Cohen (1969, pp. 7-8), for instance, focuses upon the role of power in experimental design:

Experimental design is an area of inquiry wholly devoted to the removal of irrelevant sources of variability for the increase of precision and therefore for the increase of the statistical power of tests of null hypotheses. . . . The greater the precision of the sample results, other things being equal, the greater the probability of detecting a nonnull state of affairs, i.e., the more clearly the phenomenon under test can manifest itself against the background of (experimentally irrelevant) variability.

Miller and Knapp underscore the importance of a priori consideration of such matters by posing the question,

Can experimental behavioral science really afford the dubious luxury of continuing to blunder upon significance, subject to the fortuitous, coincidental attainment of sufficient statistical power?
(p. 24)

Kerlinger (1964, p. 275) emphasizes that "research design has one grand purpose: to control variance." Let us consider what experts have said about each of the factors, alpha level, sample size, effect size, and power, as it relates to the question of allocation of sample material.

Alpha level. One of the most serious obstacles to the routine examination of the expected power associated with a particular research design has been the emphasis, historically, on the importance of avoiding a TYPE I ERROR. Since the probability of committing a Type I error is inversely related to that of committing a TYPE II ERROR, it is understandable why recommendations to assess the complement of Type II error--power--have often been overlooked. The relative seriousness of Type I and Type II errors has become an issue in itself in some areas of inquiry. As Cohen points out, however, by carrying out experimental investigations after setting alpha level and sample size, researchers have implicitly determined the limits for the values of the less commonly considered factors of effect size and power (1962, p. 151).

There seems to be increasing support for questioning the reasonableness of using conventional alpha rates regardless of the context of the investigation. Paul Games (1971b, pp. 561-563) criticizes sophisticated attempts to control the FAMILYWISE ERROR RATE in sets of planned CONTRASTS when no attention is given to the level of power that each contrast yields. He suggests that steps should be taken to ensure that the power of each contrast is at an acceptable level, even if it means that the Type I error risk is slightly higher than is conventional.

Sample size. One issue concerning the amount of experimental material included in a study revolves around the question of whether it is acceptable to purport to carry out confirmatory research when the number of experimental units is severely limited. Depending on the area of inquiry, the acquisition of additional sample material may or may not pose a problem to the investigator. A common constraint in this regard is the cost associated with the sample material itself, or the cost of applying treatments (Cleary & Linn, 1969; Cox, 1958; Finney, 1960; Scheffe, 1959). Some authors not only feel that this constraint is acceptable, but also that it tends to promote parsimonious investigations of well-thought-out hypotheses and fewer rejections of trivially false null hypotheses (Cohen, 1962; Overall & Dalal, 1968).

Another issue regarding sample size focuses on those situations in which the researcher has some flexibility in deciding how many subjects (or how much experimental material) to include in a study:

Decisions as to sample size are typically reached by recourse to local tradition, ready availability of data, unaided intuition, usually called 'experience,' . . . and rarely on the basis of Type II error analysis, which can always be performed prior to the collection of data. These non-rational bases for setting sample size must often result in investigations being undertaken which have little chance of success despite the actual falsity of the null hypothesis, and probably less often in the use of a far larger sample than is necessary. Either of these circumstances is wasteful of research effort. (Cohen, 1962, p. 145)

Similar views are expressed by Feldt (1973, p.221) and Overall and Dalal (1968, p. 363). Levin (1975, p. 107) goes further and states that, by not determining sample size through a priori consideration of power, "one of the few meaningful decisions associated with statistical hypothesis testing will have failed to have been exercised."

Effect size. Although it is commonly recognized as being one of the most difficult measures to establish a priori, most experts seem to agree that it behooves the researcher to posit an effect size (Cohen, 1962, p. 146, 1969, p. 278; Levin, 1971, p. 100; Overall & Dalal, 1968, p. 364). Cohen (1969, p. 12) also points out that one of the greatest benefits of establishing an effect size is

that the researcher is forced to think in more exact terms, i.e., in stating an alternate hypothesis. This process, according to Cohen, requires a "critical assessment of prior research in the area." Barcikowski (1973, p. 14) feels even more strongly:

It is the writer's firm belief that if a researcher is unable to select values [for effect size] then he should consider his experiment nothing more than a pilot study. Not knowing [the effect size] prevents one from being able to select an appropriate number of observations with which to test [the null hypothesis].

Power. At the present time most a priori design considerations tend to focus upon the power of the statistical questions, i.e., the PRECISION of the estimates. Although the primary focus of this document is the development of techniques for increasing power through unequal allocation of the experimental material among treatments, it can also be increased in other ways: increasing the total sample used, using more reliable measuring instruments, securing multiple measurements on each experimental unit, or stratifying samples into a larger or smaller number of levels according to values on one or more control variables, for example. Common to all of these techniques is the statistical goal--to reduce variability, thereby increasing the precision of the estimates.

Among proponents of a priori power analyses, Jacob Cohen has perhaps been the most zealous. He has decried

the low levels of power and/or large detectable effect sizes in abnormal-social psychological research (1962) and provided abundant amounts of material in the form of recommendations, charts, tables, and graphs to enable the practitioner to make use of his ideas (1965, 1968, 1969). Cohen directs his work primarily at the applied researcher, which is consistent with his belief that power analysis has heretofore been unavailable to this group, partially because the necessary materials for conducting such analyses were scattered over the periodical and hardcover literature, and partially because these analyses assume a degree of mathematical sophistication well beyond that of most behavioral scientists (1969, p. vii).

One of the strongest arguments for the routine assessment of power is couched in the relationship between sample size and power: if enough sample material is invested in an experiment, a researcher can virtually assure himself/herself of rejecting the null hypothesis, i.e., the power of the test will ultimately approach unity. This relationship is also important to the argument for the need to define a "meaningful difference" in treatments--effect size. As Cohen points out (1962, p. 153), the disparity in sample sizes alone is enough to explain seemingly incongruous results in apparently similar experiments. Cohen's efforts have not gone unnoticed. Witness the following excerpt from John Miller

and Thomas Knapp, in a monograph distributed by Phi Delta Kappa:

Given the importance of statistical power to the utility and integrity of difference testing as a decisive contributor to the advancement of knowledge, there is no room for equivocation. Estimating the common population variance, specifying a difference that can be regarded as meaningfully large, selecting a power level that enforces reasonable restraints upon Type II Error, and calculating the sample size required to satisfy the power specification should be as routine as the selection of a significance level. Moreover, like the selection of a significance level, these tasks should precede the collection of a single piece of data. (p. 23)

Given the four interrelated constructs of alpha level, sample size, effect size, and power, it is not surprising that Cohen has identified four different kinds of power analysis--one with each factor as the criterion. Although such analyses have been used to assess the attributes of completed research, their primary use lies in the comparison and selection of potential research designs. A technique which is also useful in increasing the power with which statistical questions are tested, but which places less burden upon the investigator (as commonly applied, at least), is that of planned COMPARISONS.

Planned comparisons. Essential to the use of planned comparisons is the ability of the researcher to state, a priori, a few hypotheses which dictate the interpretation of the whole experiment. By doing so, the investigator may diminish the probability of making a Type II error

(compared to post hoc procedures), while still maintaining a stringent control over Type I errors. The identification of situations in which planned comparisons are more appropriate than post hoc techniques has consumed much textbook and journal space (Table 1 depicts a general design for studies which employ planned comparisons). In general, experts agree that planned comparisons should be used in those studies for which the importance of not committing a Type II error is greater than the loss of information inherent in not being able to "data snoop" (Cox, 1958; Hays, 1981; Levin, 1975; Lindman, 1974). Some are more emphatic:

The first recommendation is to design an experiment so that the hypotheses of primary interest can be tested unambiguously by means of planned comparisons. The advantages of planned over post-hoc comparisons is so great that it is foolish not to decide upon some specific contrasts before the start of an experiment. (Keppel, 1973, p. 160)

Keppel goes on to suggest that hypotheses should not be evaluated statistically using post hoc procedures--that judgment should be withheld until further research is conducted based on those follow-up comparisons which prove to be of interest (pp. 162-163). A "compromise" position has been offered by some authorities who propose that the primary research questions be investigated using planned comparisons, followed by post hoc tests, if the F is significant after deducting the degrees of freedom and

Table 1

A General Matrix of Coefficients for Planned Comparisons^a

Contrast	Treatment Group					
	1	2	. . .	j	. . .	k
ψ_1	c_{11}	c_{12}	. . .	c_{1j}	. . .	c_{1k}
ψ_2	c_{21}	c_{22}	. . .	c_{2j}	. . .	c_{2k}
.	.	.		.		.
:	:	:		:		:
.	.	.		.		.
ψ_i	c_{i1}	c_{i2}	. . .	c_{ij}	. . .	c_{ik}
.	.	.		.		.
:	:	:		:		:
.	.	.		.		.
ψ_m	c_{m1}	c_{m2}	. . .	c_{mj}	. . .	c_{mk}

^aOf the general form,
 $c_{i1}\mu_1 + c_{i2}\mu_2 + \dots + c_{ij}\mu_j + \dots + c_{ik}\mu_k \quad (1.1)$

sums of squares that are consumed in the planned comparisons (Hays, 1981, p. 427; Kirk, 1968, p. 113).

A central issue in the wise use of planned comparisons is the identification of a few research questions. Due to the nature of the procedure, attention must be paid to the setting of the alpha level when more than one a priori hypothesis is tested. Awareness of the relationship among FAMILYWISE ERROR RATE, EXPERIMENTWISE ERROR RATE, and ERROR RATE PER COMPARISON is important to an understanding of the implications of testing multiple contrasts, on the probability of committing a Type I error. Generally, however, an accepted practice is to modify the alpha level for each comparison using a Bonferroni adjustment. This is true even when the contrasts are not mathematically orthogonal--as when sample sizes are unequal--if the degrees of freedom for the error term are large, i.e., greater than 40 (Games, 1971, p.533; Hsu, 1980, p. 32). Under the condition of nonorthogonality, the Bonferroni adjustment is conservative, so any error will not predjudice the conclusions toward rejection of the null hypothesis (Davis, 1969, p. 442; Perlmutter & Myers, 1973, p.182).

Although well-chosen planned comparisons (i.e., which are congruent with relationships found in the population) are more powerful than post hoc techniques (as long as the number of subjects is reasonably large), the advantage is

diminished by each additional contrast tested, according to the procedure cited above. Hence, much attention is devoted to the need for the researcher to limit his/her hypotheses:

Each esoteric issue posed to the data costs a [degree of freedom] which is lost from the error estimate, thus enfeebling the statistical power of the analysis. This, ultimately, is the reason that it is desirable in research that is to lead to conclusions to state hypotheses which are relatively few in number. (Cohen, 1968, p. 442)

After the researcher has framed a few well-chosen questions and translated them into a priori contrasts, the steps which should be taken next, in order to select the "best" design, depend on his/her (implicit or explicit) reactions to issues raised in the area of optimal design.

Optimal design. The area of optimal design is multifaceted and convoluted. An overview of the literature which appeared in an issue of Communications in Statistics devoted entirely to optimal design theory contains no fewer than 312 references, along with a caveat that several major areas in optimal design were excluded (Ash & Hedayat, 1978). A primary focus in this body of literature is the selection of the parameter which is to be optimized. The number of different criteria seems to be exceeded only by the number of authors who support each. Although the techniques are based on complex mathematical theory, the applications are iterative and sequential,

fitting models to empirical data ex post facto (Box & Wilson, 1951, p. 2; Hill & Hunter, 1966, pp. 576-585). Since most optimization research has developed outside the social sciences, many of the issues addressed by theorists in optimal design are of little concern to investigators of social and psychological phenomena.

The primary area of mutual concern seems to be the desirability of selecting experimental designs which maximize precision, i.e., minimize some measure of error variance (Ash & Hedayat, 1978, p. 1296; Bock, 1975, p. 307; Box & Wilson, 1951, p. 40; Feldt, 1958, p. 340; Finney, 1960, p. 11; Plackett & Burman, 1946, p. 307; Wald, 1943, p. 134):

By ruling out some of the error variance from the observations one is decreasing the standard error of the mean, and thus increasing the power of the test against whatever [alternate] hypothesis is true. Experiments in which the variability attributable to experimental or sampling error is small are said to be PRECISE; the result of such precision is that the experimenter is quite likely to be able to detect when something of interest is happening. (Hays, 1981, p. 254)

A less popular view is that the estimate which should be minimized is the standard error of the estimate (Cox, 1958, pp. 177, 283).

As pointed out previously, an appraisal of the power of a statistical test requires that the researcher establish a definite effect size of interest or that he/she adopt a reasonable relative effect size, as suggested

by Cohen (1962, p. 153). The former approach requires some knowledge of the magnitude of the within-group population variance. Though the feasibility of obtaining a reliable estimate of the error variance has been questioned, many authorities feel that such estimates are available. Possible sources for this information include previous research and pilot studies (Brooks, 1955, p. 414; Harris, Horvitz, & Mood, 1948, p. 399; Overall & Dalal, 1968, p. 364; Subkoviak & Levin, 1977, p. 50). Miller & Knapp suggest that the underlying problem may not be the unavailability of accurate estimates but a reluctance to use others' estimates. This, they feel, is a result of the customary emphasis on the uniqueness of an author's contribution, and it "tends to weaken, rather than strengthen, the continuity of science and the relevance of new discoveries" (pp. 22-23).

Assuming that a reasonably good estimate of the within-group population variance is secured and that power analyses can be conducted, the avenues to achieving increased precision are varied. Among the options available is the allocation of sample material across the treatment conditions. This is generally advisable only when planned comparisons constitute the primary focus of the experiment (Overall & Dalal, 1965, p. 345).

Equal Versus Unequal Allocation

Proponents of the practice of allocating experimental material equally across treatment conditions regardless of the research questions, typically, offer varied arguments for doing so: for ease of computation (Li, 1964, p. 197), to maintain the orthogonality of the statistical tests (Bock, 1975, p. 306), to maximize the power of the omnibus F (Appelbaum & Cramer, 1974, p. 337; Li, 1964, pp. 197-198), or to take advantage of the robustness of many statistical tests to violations of the assumptions underlying the analysis of variance (Bock, 1975, p. 306; Lachin, 1977, p. 321; Li, 1964, p. 197). On the other hand, some authorities feel that circumstances exist in which the investigator should consider allocating sample material unevenly. If there is differential interest in the research hypotheses, unequal sample sizes may be advisable (Appelbaum & Cramer, 1974, p. 336; Cox, 1958, p. 190; Kiefer, 1959, p. 279). If inconclusive results are due to excessive variability in one of the treatment conditions, the addition of sample material to that condition (i.e., "sequential sampling") may be in order (Harris et al., 1948, p. 392; Horsnell, 1953, p. 130). Because the greatest risk in unequal allocation lies in the increased importance of the assumptions underlying the analysis of variance (due to the robustness issue), it is imperative that the effects of such violations be examined.

The extensive literature on the effects of violations of the assumptions underlying the analysis of variance leaves little doubt that, by far, the most inviolable postulate (with the exception of the independence of observations) is that of homoscedasticity. Whereas departures from additivity and normality have negligible--or at least predictable--effects on outcomes, when within-group population variances are heterogeneous, the effect on Type I error rate and power depends on the distribution of n s relative to the magnitude of the within-group variances (Boneau, 1960, pp. 62-63; Games, 1971a, p. 101; Heilizer, 1964, p. 532; Lindman, 1974, p. 43; McNemar, 1969, p. 118). When comparisons involving more than two groups are involved, the effect becomes increasingly difficult to predict. The condition of heterogeneous variance coupled with unequal n s, then, poses a formidable problem for the investigator. Also, as Glass, Peckham, and Sanders point out, power analyses cannot be conducted under the condition of variance heterogeneity, since the computation of the noncentrality parameter is impossible (1972, p. 273).

Constraints of Study

Several general assumptions have been made in connection with this study which are consistent with recommendations made in the literature cited above:

1. The research focus is presumed to be confirmatory in nature, with a few planned comparisons

identified which represent the primary impetus of the investigation. Also, the study can be considered a true experiment, with sample material randomly assigned to treatment conditions.

2. The assumptions of homoscedasticity, normality, independence, and additivity are tenable.

Evidence as to the deleterious consequences of the failure to meet these assumptions in the condition of unequal sample sizes is overwhelming (Glass, Peckham, & Sanders, 1972).

3. There are constraints on the number of experimental units which can be utilized in the investigation.

Several additional conditions have been specified, either in an effort to limit the focus of the study or for the sake of mathematical convenience:

1. The cost of applying treatments is either constant across treatment conditions or it is not a factor.
2. The effect size of interest is constant across contrasts and is classified as "medium." This is consistent with recommendations made by Cohen (1962, p. 153, 1969, p. 279, 1970, p. 817) and others (Hsu, 1980, pp. 31-34) and represents a reasonable treatment effect for educational and psychological research.

3. The sum of the absolute values of the coefficients for each contrast is two. Although the choice of two is somewhat arbitrary, it is desirable that the sum of the absolute values of the coefficients be equal for all contrasts, so that the interpretation of effect size be the same regardless of the number of groups involved in the comparison (Hsu, 1980, p. 35).

Research Questions

In fixed effects, analysis of variance applications in which the researcher is able to formulate a priori contrasts that correspond to a few questions of primary interest:

1. If the researcher wishes to maximize the power of the aggregate investigation, how can a limited number of experimental units be apportioned across the cells of the design to maximize the precision of the set of contrasts?
2. If the researcher ascribes equal importance to every contrast, how may experimental units be distributed across treatment conditions to ensure that all contrasts are estimated with equal precision?
3. How does power differ between designs which use these algorithms to allocate sample material and

those which allocate sample material equally across treatments?

Need for Study

Basically, the first research question represents an extension of a principle which has long been in use, albeit in two very restricted applications: In experiments in which the researcher wishes to compare a control group with each of k experimental treatments, it has long been recognized that power is maximized by assigning $\sqrt{k}$ units to the control group for each unit assigned to an experimental treatment (Bock, 1975, p. 307; Cox, 1958, pp. 178-179; Finney, 1960, p. 12). If the researcher is interested in comparing the control group with the average of the k experimental treatments, maximum precision is achieved when half of the total experimental units are assigned to the control group and the other half are divided equally among the k treatments (Cox, 1958, pp. 178-179). The reasonableness of allocating sample material for other contrasts has been acknowledged, but the recommendation has previously been to find this optimal allocation through "trial and error" (Cox, 1958, p. 179; Harris et al., 1948, p. 393). The algorithm derived in Chapter II provides a systematic method for arriving at this optimal allocation, not only for a single contrast, but also for sets of contrasts.

Several authorities have expressed concern that power analyses restrict themselves typically to consideration of individual contrasts, and that methods are needed which control the simultaneous precision of an entire set of contrasts (Hsu, 1980, pp. 31-33; Kiefer, 1959, p. 305; Spjøtvoll, 1972, pp. 399-400). The algorithms developed herein provide a simple means for doing so, whether the researcher's goal is (a) to achieve the maximum average precision for a set of contrasts (Chapter II), or (b) to equalize the precision for all research questions (Chapter III). In either case, the techniques developed meet the criterion for utility expressed by Lindquist (1953, p. 3):

No efforts to increase the precision of an experiment will be of much avail unless one can also dependably describe the increased degree of precision attained.

In addition, the intent of this study is congruent with the fundamental goal of experimental design, which is, according to Cohen (1969, p. 8), "the removal of irrelevant sources of variability for the increase of precision and therefore for the increase of the statistical power of tests of null hypotheses."

CHAPTER II

ALLOCATING SAMPLE MATERIAL TO INCREASE PRECISION:

MINIMIZING THE AVERAGE SE_{ψ}^2

The matrix in Table 1 (p. 11) represents an investigation consisting of k treatment groups ($j = 1, 2, \dots, k$) and in which m a priori contrasts ($i = 1, 2, \dots, m$) have been defined. Let N represent a fixed amount of sample material to be allocated among the k treatment groups, and n_j be the amount allocated to the j th group, i.e.,

$$\sum_{j=1}^k n_j = N . \quad (2-1)$$

By definition,

$$\sum_{j=1}^k c_{ij} = 0 \quad \text{for all } i, \quad (2-2)$$

and, for the purposes of this study,

$$\sum_{j=1}^k |c_{ij}| = 2 \quad \text{for all } i. \quad (2-3)$$

The standard error of a contrast (SE_{ψ_i}) is as follows:

$$SE_{\psi_i} = \sqrt{\sigma_e^2 \sum_{j=1}^k (c_{ij}^2 / n_j)} , \quad (2-4)$$

where σ_e^2 is the population within-group variance. Under the constraints of the present study, it is evident that the standard error of a contrast becomes a function only of the coefficients of the contrast and the amount of sample material in each group.

In order to maximize the precision for a set of contrasts, it has been deemed appropriate to minimize the average error variance, i.e., $SE_{\psi_i}^2$ (Ash & Hedayat, 1978, p. 1296; Bock, 1975, p. 307; Box & Wilson, 1951, p. 40; Feldt, 1958, p. 340; Finney, 1960, p. 11; Plackett & Burman, 1946, p. 307; Wald, 1943, p. 134):

$$(1 / m) \sum_{i=1}^m \sum_{j=1}^k (c_{ij}^2 / n_j) . \quad (2-5)$$

By eliminating $(1 / m)$, since it is a constant, and summing the columns of the matrix first, 2-5 may be rewritten

$$\sum_{j=1}^k (\sum_{i=1}^m c_{ij}^2 / n_j) . \quad (2-6)$$

The goal, then, is to minimize the function F , where

$$F(n_1, n_2, \dots, n_k) = \sum_{j=1}^k (\sum_{i=1}^m c_{ij}^2 / n_j) . \quad (2-7)$$

Letting

$$d_j = \sum_{i=1}^m c_{ij}^2 , \quad (2-8)$$

we want to minimize

$$F(n_1, n_2, \dots, n_k) = \sum_{j=1}^k (d_j / n_j) , \quad (2-9)$$

subject to the constraint

$$n_1 + n_2 + \dots + n_k = N , \quad (2-10)$$

which is equivalent to

$$(n_1 / N) + (n_2 / N) + \dots + (n_k / N) = 1 . \quad (2-11)$$

Let

$$x_j = (n_j / N) , \text{ and} \quad (2-12)$$

$$n_j = N x_j . \quad (2-13)$$

Substituting 2-13 into 2-9 and eliminating the constant (N), we now want to minimize the function F,

$$F(x_1, x_2, \dots, x_k) = \sum_{j=1}^k (d_j / x_j) , \quad (2-14)$$

subject to the constraint

$$x_1 + x_2 + \dots + x_k = 1 . \quad (2-15)$$

Using a LaGrange multiplier (Chow, 1962; Umland & Smith, 1959), we form the function G:

$$G(x_1, x_2, \dots, x_k) = \sum_{j=1}^k (d_j / x_j) + \lambda (x_1 + x_2 + \dots + x_k) . \quad (2-16)$$

Taking the derivative of G with respect to each x_j and setting the results to zero,

$$\begin{aligned} \delta G / \delta x_1 &= (-d_1 / x_1^2) + \lambda = 0 \\ \delta G / \delta x_2 &= (-d_2 / x_2^2) + \lambda = 0 \\ &\vdots \\ &\vdots \end{aligned} \quad (2-17)$$

$$\delta G / \delta x_k = (-d_k / x_k^2) + \lambda = 0 .$$

Therefore,

$$\begin{aligned} \lambda &= (d_1 / x_1^2) = (d_2 / x_2^2) = \dots \\ &= (d_k / x_k^2) . \end{aligned} \quad (2-18)$$

Finding the reciprocals of each term and taking the square root,

$$\begin{aligned} y &= (x_1 / d_1^{\frac{1}{2}}) = (x_2 / d_2^{\frac{1}{2}}) = \dots \\ &= (x_k / d_k^{\frac{1}{2}}) , \end{aligned} \quad (2-19)$$

where

$$y = (1 / \lambda^{\frac{1}{2}}) . \quad (2-20)$$

This means that

$$\begin{aligned}
 \gamma &= (x_1 / d_1^{\frac{1}{2}}) \quad \text{and} \quad x_1 = \gamma d_1^{\frac{1}{2}} \\
 \gamma &= (x_2 / d_2^{\frac{1}{2}}) \quad \text{and} \quad x_2 = \gamma d_2^{\frac{1}{2}} \\
 &\vdots \\
 \gamma &= (x_k / d_k^{\frac{1}{2}}) \quad \text{and} \quad x_k = \gamma d_k^{\frac{1}{2}} .
 \end{aligned} \tag{2-21}$$

Since

$$\begin{aligned}
 x_1 + x_2 + \dots + x_k &= 1, \text{ by 2-15,} \\
 \gamma d_1^{\frac{1}{2}} + \gamma d_2^{\frac{1}{2}} + \dots + \gamma d_k^{\frac{1}{2}} &= 1,
 \end{aligned} \tag{2-22}$$

or

$$\gamma (d_1^{\frac{1}{2}} + d_2^{\frac{1}{2}} + \dots + d_k^{\frac{1}{2}}) = 1, \tag{2-23}$$

and

$$\gamma = 1 / (d_1^{\frac{1}{2}} + d_2^{\frac{1}{2}} + \dots + d_k^{\frac{1}{2}}). \tag{2-24}$$

Using 2-21 and substituting,

$$\begin{aligned}
 x_1 &= d_1^{\frac{1}{2}} / (d_1^{\frac{1}{2}} + d_2^{\frac{1}{2}} + \dots + d_k^{\frac{1}{2}}) \\
 x_2 &= d_2^{\frac{1}{2}} / (d_1^{\frac{1}{2}} + d_2^{\frac{1}{2}} + \dots + d_k^{\frac{1}{2}}) \\
 &\vdots
 \end{aligned} \tag{2-25}$$

$$x_k = d_k^{\frac{1}{2}} / (d_1^{\frac{1}{2}} + d_2^{\frac{1}{2}} + \dots + d_k^{\frac{1}{2}}),$$

or, in general,

$$x_j = d_j^{\frac{1}{2}} / (d_1^{\frac{1}{2}} + d_2^{\frac{1}{2}} + \dots + d_k^{\frac{1}{2}}). \tag{2-26}$$

By using 2-12, substituting in 2-26, and multiplying by N,

$$n_j = (N d_j^{\frac{1}{2}}) / (d_1^{\frac{1}{2}} + d_2^{\frac{1}{2}} + \dots + d_k^{\frac{1}{2}}). \tag{2-27}$$

Now, by substituting 2-8,

$$n_j = N \left(\sum^m c_{ij}^2 \right)^{\frac{1}{2}} \quad (2-28)$$

$$\div \left[\left(\sum^m c_{i1}^2 \right)^{\frac{1}{2}} + \left(\sum^m c_{i2}^2 \right)^{\frac{1}{2}} + \dots + \left(\sum^m c_{ik}^2 \right)^{\frac{1}{2}} \right] ,$$

or,

$$n_j = \frac{N \sqrt{\sum^m c_{ij}^2}}{\sum^k \sqrt{\sum^m c_{ij}^2}} . \quad (2-29)$$

Once an investigator has constructed a matrix of contrast coefficients, then, he/she may use equation 2-29 to determine how much sampling material must be allotted to each treatment group, n_j , in order to minimize the average squared standard errors for the set of contrasts represented in the matrix. By doing so, the precision of the aggregate investigation is maximized, according to the criterion employed.

CHAPTER III

ALLOCATING SAMPLE MATERIAL TO DECREASE VARIABILITY:

EQUALIZING THE SE_{ψ}^2

Using the generalized matrix of contrast coefficients contained in Table 1 (p. 11), along with the relationships described in Chapter II (represented by equations 2-1, 2-2, 2-3, and 2-4), a reasonable approach to equalizing the squared standard errors of all contrasts is an algebraic one. Under the constraints of this study, such equality can be achieved by setting $SE_{\psi_1}^2 = SE_{\psi_i}^2$, i.e.,

$$\sum^k (c_{1j}^2 / n_j) = \sum^k (c_{ij}^2 / n_j) , \quad (3-1)$$

for all i ($i = 2, 3, \dots, m$), subject to the constraint, as in 2-15,

$$x_1 + x_2 + \dots + x_k = 1 , \quad (3-2)$$

where

$$x_j = n_j / N . \quad (3-3)$$

It is apparent in such an approach that matrices of contrast coefficients which are not square (i.e., in which the number of contrasts is greater than or less than the number of treatment groups) may not yield usable solutions. If there are more rows than columns, the system is over-specified, and solutions can be derived only by happenstance. On the other hand, if there are more treatment groups than contrasts, solutions may not be unique, i.e., there could be infinite sets of solutions which satisfy

the equations. It would seem, therefore, that a modification of the approach should be used.

Accordingly, several varieties of matrices were tested in order to determine the characteristics of a usable matrix. It was discovered that nonorthogonal matrices yield usable unique or nonunique solutions only by chance, and that orthogonal matrices provide meaningful solutions--even nonunique--only in some cases. An example of an orthogonal matrix which yields no usable solutions is as follows:

$$\begin{bmatrix} 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 1/2 & 0 & -1/2 & -1/2 & 1/2 \\ 1/4 & -1 & 1/4 & 1/4 & 1/4 \end{bmatrix} \quad (3-4)$$

Though there is an infinite number of solutions with this matrix, all lead to zero and/or negative allocations to one or more groups.

Only those matrices of what shall be deemed the "Helmert" type seem to yield viable solutions consistently; for example, using five treatment groups again,

$$\begin{bmatrix} 1 & -1/4 & -1/4 & -1/4 & -1/4 \\ 0 & 1 & -1/3 & -1/3 & -1/3 \\ 0 & 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \quad (3-5)$$

This type of matrix seems to be common in the literature (Games, 1975b; Hays, 1981; Myers, 1979). Since the number

of columns--treatments--exceeds the number of rows--contrasts--the algebraic approach cited above provides non-unique solutions. It seems reasonable, therefore, to add an additional constraint, i.e., that the SE_{ψ}^2 not only be equal across contrasts, but that they also be at a minimum. The problem can be restated: we want to minimize

$$\sum_{j=1}^k (c_{1j}^2 / x_j) , \quad (3-6)$$

subject to 2-2, 2-3, 3-1, and 3-2.

By the method of LaGrange multipliers, we can define function F:

$$\begin{aligned} F(x_1, x_2, \dots, x_k) = & \sum_{j=1}^k (c_{1j}^2 / x_j) \\ & + \lambda_1 (x_1 + x_2 + \dots + x_k) \\ & + \lambda_2 \sum_{j=1}^k (c_{1j}^2 - c_{2j}^2) (1 / x_j) \\ & + \lambda_3 \sum_{j=1}^k (c_{1j}^2 - c_{3j}^2) (1 / x_j) + \dots \\ & + \lambda_m \sum_{j=1}^k (c_{1j}^2 - c_{mj}^2) (1 / x_j) . \end{aligned} \quad (3-7)$$

A complete solution would require that F be differentiated with respect to each x_j and each resulting equation be set equal to zero. For the present purposes, however, we need to consider only the last two variables, x_{k-1} and x_k , since the patterns of solutions among the nonunique solution sets seemed to indicate that those solutions which allocated material equally to these two groups tend to produce SE_{ψ}^2 which are smaller than other alternatives.

$$\begin{aligned}
\delta F / \delta x_{k-1} &= (-c_{1,k-1}^2) / x_{k-1}^2 + \lambda_1 \\
&+ \lambda_2 (c_{1,k-1}^2 - c_{2,k-1}^2) (-1 / x_{k-1}^2) \\
&+ \lambda_3 (c_{1,k-1}^2 - c_{3,k-1}^2) (-1 / x_{k-1}^2) + \dots \\
&+ \lambda_m (c_{1,k-1}^2 - c_{m,k-1}^2) (-1 / x_{k-1}^2) = 0 .
\end{aligned} \tag{3-8}$$

Likewise,

$$\begin{aligned}
\delta F / \delta x_k &= (-c_{1k}^2) / x_k^2 + \lambda_1 \\
&+ \lambda_2 (c_{1k}^2 - c_{2k}^2) (-1 / x_k^2) \\
&+ \lambda_3 (c_{1k}^2 - c_{3k}^2) (-1 / x_k^2) + \dots \\
&+ \lambda_m (c_{1k}^2 - c_{mk}^2) (-1 / x_k^2) = 0 .
\end{aligned} \tag{3-9}$$

Multiplying 3-9 by $(-x_k^2)$,

$$\begin{aligned}
c_{1k}^2 - \lambda_1 x_k^2 + \lambda_2 (c_{1k}^2 - c_{2k}^2) + \lambda_3 (c_{1k}^2 - c_{3k}^2) \\
+ \dots + \lambda_m (c_{1k}^2 - c_{mk}^2) = 0 .
\end{aligned} \tag{3-10}$$

In all matrices of the form 3-5, however,

$$|c_{j,k-1}| = |c_{jk}| , \tag{3-11}$$

so

$$c_{j,k-1}^2 = c_{jk}^2 . \tag{3-12}$$

Multiplying 3-8 by $(-x_{k-1}^2)$ and substituting 3-12,

$$\begin{aligned}
c_{1k}^2 - \lambda_1 x_{k-1}^2 + \lambda_2 (c_{1k}^2 - c_{2k}^2) + \lambda_3 (c_{1k}^2 - c_{3k}^2) \\
+ \dots + \lambda_m (c_{1k}^2 - c_{mk}^2) = 0 .
\end{aligned} \tag{3-13}$$

Subtracting 3-13 from 3-10,

$$-\lambda_1 x_k^2 + \lambda_1 x_{k-1}^2 = 0 . \tag{3-14}$$

Therefore,

$$x_{k-1}^2 = x_k^2 \text{ and } x_{k-1} = x_k . \tag{3-15}$$

Equation 3-15 also implies that, in order to meet the

criteria imposed on the problem, there must be equal sample material in groups k and $k-1$. This relationship provides the additional equation needed to determine unique solutions for Helmert matrices, in that it represents the k th equation for a problem which has k unknowns.

Because the pattern of Helmert matrices is regular regardless of the number of treatment groups, it is possible to devise a simple means for determining optimal allocation to achieve equal precision among contrasts. A FORTRAN computer program written by Rhyne (1970) was used to solve the systems of linear equations for designs consisting of up to 20 treatment groups. Table 2 provides values necessary to determine optimal allocation using the present criteria by applying the following formula:

$$n_j = \left[r_q / \sum^k r \right] N , \quad (3-16)$$

where

$$q = k - j + 1 , \quad (3-17)$$

and r represents the proportion of sample material to be included in group q for each unit included in group k . The use of 3-16 under the conditions cited ensures that each Helmert contrast is tested with equal precision, and that this precision is at a maximum.

Table 2

Allocating Sample Material^a to Achieve Equal Precision
in a Matrix of Helmert Contrasts^b

r^c	Group ^d q	k $\sum r$	r^c	Group ^d q	k $\sum r$
1.0000	1	1.0000	.5426	11	7.2283
1.0000	2	2.0000	.5391	12	7.7674
.6667	3	2.6667	.5361	13	8.3035
.6207	4	3.2874	.5336	14	8.8371
.5950	5	3.8824	.5314	15	9.3685
.5786	6	4.4610	.5295	16	9.8980
.5671	7	5.0281	.5186	17	10.4166
.5586	8	5.5867	.5172	18	10.9338
.5521	9	6.1388	.5159	19	11.4497
.5469	10	6.6857	.5147	20	11.9644

^a Using the formula $n_j = N (r_q / \sum^k r)$.

^b Arranged so that $\sum^m |c_{i1}| \leq \sum^m |c_{i2}| \leq \dots \leq \sum^m |c_{ik}|$.

^c Proportion of sample material in q for each unit in k .

^d $q = k - j + 1$.

CHAPTER IV

ALLOCATING SAMPLE MATERIAL AMONG TREATMENT CONDITIONS:

A COMPARISON OF THREE ALTERNATIVES

The alternative algorithms for assigning experimental material to treatment conditions which are developed in Chapters II and III are not presumed to be panaceas for the researcher. Certainly, the two imply that the objectives of the researcher are very different: (a) a desire to maximize the chances of rejecting at least one null hypothesis (which will be referred to as "Option II"), versus (b) a desire that each research question be tested with the same degree of precision, allowing (a) to be secondary (this will be called "Option III"). Since it is assumed that the researcher has formed several a priori hypotheses and that his/her objectives coincide with those of one of the alternative methods, the decision becomes a matter of weighing the relative advantages of equal and unequal allocation: Does the increase in precision offered by Option II--compared to allocating material equally (which will be referred to as "Option I")--compensate for the potential difficulties associated with unequal sample sizes? Is the appeal of testing each hypothesis with the same degree of precision (i.e., Option III), great enough to warrant the probable loss in power? Although these questions are clearly subjective and are

influenced by innumerable rational and philosophical factors far beyond the scope of the present study, a consideration of some of the measurable consequences of allocating sample material unevenly might prove to be instructive.

Rationale

Glass, Peckham, and Sanders, in a significant article dealing with the consequences of the failure to meet the assumptions underlying analyses of variance and covariance, argue strongly against the practice of using techniques of computer simulation to examine issues regarding the power of, and extent of agreement between, actual and theoretical distributions of test statistics (1972, pp. 239, 282). In the present context, since the relationships among alpha level, sample size, effect size, and power are well-defined, a comparison of the three allocation schemes can be carried out mathematically, making computer simulation "potentially misleading . . . [compared to] actual inferential applications" (Glass et al., 1972, p. 282).

Choice of k , N , and ψ . It is not the intention of this study to make specific recommendations regarding the combinations of k , N , and ψ which delineate sound and unsound applications of the two alternative algorithms. It would seem to be appropriate, nevertheless, to compare

the performance of these alternate allocation schemes to the more traditional equal allocation, using representative combinations of k , N , and ψ .

Accordingly, 15 studies which used computer-generated data to address issues in statistics and research design were reviewed to determine reasonable values for N and k . Ten different numbers of treatment groups were used in these investigations, ranging from 2 to 15, and 18 different values of n , with a range of 3 to 32. The most prevalent values of k were 3, 4, and 6, and were used in the present analysis. The most frequently occurring values for n were 10, 5, and 15, respectively. In the context of this study, however, the sample size of 5 was deemed inadequate, since the unequal n allocation schemes lead to treatment groups too small for reasonably reliable estimates. The size of the available pool of sample material (N) was calculated by multiplying k by n s of 10 and 15.

A similar approach was taken to identify potential sets of contrasts. Thirty-one different matrices of contrast coefficients were considered. Proportional allocations were computed for each, using Options II and III, after standardizing the coefficients to conform with the constraints of this study. The designs which differ negligibly from Option I were eliminated. Since Option III is only applicable to Helmert contrasts, Helmert

matrices were included for the k values of 3, 4, and 6. Also, included for each k value was one each of non-Helmert, orthogonal and nonorthogonal matrices. For $k = 3$, however, there is only one possible pair of orthogonal contrasts, i.e., the Helmert set.

Because of its theoretical focus and the limited ranges of N , k , and ψ included in the present study, the decision was made to allow fractional ns in the analyses, so that the advantages attributable to unequal allocation would not be masked. This decision is consistent with the distinction drawn between theoretical/mathematical distributions and those of actual samples, alluded to by McNemar (1969, pp. 39-40). The desirability of considering theoretical distributions based on fractional degrees of freedom and/or nondiscrete experimental units, rather than approximations reached via interpolation, is also supported (Hardison, Quade, & Langston, 1981, pp. 1-7; Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975, p. 270; Norusis, 1979, pp. 23-26).

Power analysis for planned comparisons. The nature of contrasts is such that they may be viewed as comparisons between two quantities--the weighted means represented by the positive coefficients and those represented by the negative coefficients. This is also apparent in the F -test which is applied to planned comparisons, i.e.,

there is one numerator degree of freedom. Hsu (1980) has used this fact to generate probability distributions for nonpairwise contrasts under the condition of equal ns. Using his notion of the effect size of a contrast as "the difference between two means expressed in σ units" (p. 36), an extension can be made to the case of unequal ns:

$$\phi = (\mu_p - \mu_n) / \left[\sum^k (c_{ij}^2 / n_j) \right]^{\frac{1}{2}}, \quad (4-1)$$

where ϕ is the noncentrality parameter in the t distribution, $(\mu_p - \mu_n)$ is the standardized difference between the positively weighted and negatively weighted means, and the denominator is the standard error of the contrast. The calculation of ϕ , along with Cohen's recommendation that, in the absence of other imperatives, an alpha rate of 0.05 and a "medium" effect size be adopted as reasonable standards (1965, p. 99), makes possible a mathematical determination of power. In this determination of power, however, it is necessary to ascertain probabilities associated with varying noncentral t distributions. This was accomplished in the present study using a Supplemental Procedure of the Statistical Analysis System (Hardison, Quade, & Langston, 1981), in conjunction with a computer program which performs the preliminary calculations and processes the results (Appendix B contains an example of one such program). The accuracy of the program was

checked by analyzing several examples from both Cohen and Hsu. A comparison of the output from the SAS program and that reported by Cohen and Hsu revealed that power computations differed only in the last significant digit. In order to maintain the experimentwise error rate at conventional levels (i.e., 0.05), the alpha level employed for each comparison was calculated using a Bonferroni adjustment.

Results

Tables 3 through 18 in Appendix C summarize the results of the comparisons of precision of the three different allocation schemes. Tables 3 through 6 incorporate three treatments groups; Tables 7 through 12, four; and Tables 13 through 18, six. All three levels of treatment groups have summaries for Helmert designs (Tables 3, 4, 7, 8, 13, and 14) and nonorthogonal designs (Tables 5, 6, 11, 12, 17, and 18), but only k-values of 4 and 6 have summaries for non-Helmert, orthogonal designs (Tables 9, 10, 15, and 16). Analyses are included for average ns of 10 (odd-numbered tables) and 15 (even-numbered tables). The standard error of the estimate and power for each contrast are reported, along with means and standard deviations for both, across each set of contrasts.

In all cases the mean standard error for Option II is less than that for Options I and III, but only in the

non-Helmert cases does the mean power for Option II exceed that for Option I. As the mean standard error is consistently lower for Option II, it is possible to compute the relative efficiency of the procedures; i.e., $\overline{SE}_{II}^2 / \overline{SE}_I^2$ or $\overline{SE}_{II}^2 / \overline{SE}_{III}^2$, where $\overline{SE}^2$ is the mean squared standard error and the subscript refers to the Option. For Option I, the range of relative efficiency is from .7441 (Tables 11 and 12) to .9997 (Table 4), with 87.5% exceeding .9000 and 75% exceeding .9500. For Option III, relative efficiency ranges from .9691 (Table 13) to .9792 (Tables 3 and 4).

By dividing the analyses into two groups--those for which Option I provides the greatest mean power (i.e., the Helmert designs) and those for which Option II provides the greatest mean power (i.e., the non-Helmert designs), a "power ratio" may be constructed by dividing the analysis with the greater power by that with the lesser power. Since the Helmert designs result in the greatest power for Option I, ratios which use Option I as the numerator are referred to as "PRI" and those which use Option II as the numerator are referred to as "PRII". For Option II, PRI range from 1.0046 (Table 4) to 1.0288 (Table 3) and for Option III, from 1.0292 (Table 6) to 1.0602 (Table 14). For Option I, PRII range from 1.0089 (Table 15) to 1.2513 (Table 12).

Variability in estimates across contrasts is, by definition, nonexistent for Option III. In the case of

Helmert matrices, it is consistently lower for Option II than for Option I, both for power and for the standard error of the estimate. For non-Helmert contrasts, however, no such pattern emerges.

CHAPTER V

DISCUSSION

Maximizing Power

One does not have to scrutinize the tables in Appendix C too closely to realize that the algorithm which presumes to maximize the average precision of a set of contrasts (Option II), results in a consistent shortfall in power--compared to equal allocations--when the contrasts are of the Helmert type (see Tables 3, 4, 7, 8, 13, and 14 in Appendix C). This seemingly incongruous outcome is due to a rather troublesome characteristic of measures of precision: whereas precision is commonly measured in standard deviation units (i.e., standard errors), most experts agree that efforts to maximize precision should focus on the squared standard error. Power, however, is a function of the standard error, rather than the squared standard error, and minimizing SE^2 is not equivalent to minimizing its square root. Although the mean standard error is lower for Option II than for Option I in all cases, the mean power associated with the set of contrasts is not. Power is intimately related to the separation of the central and noncentral t distributions, i.e., the distance between their means along the baseline. The more distance there is between them, the easier it is to detect that difference, i.e., the greater

the power. In the present context, a comparison of the baseline distances between the two distributions--quantified in the noncentrality parameter--reveals that the means of the noncentrality parameters are smaller for Option II than for Option I in the case of Helmert matrices. This indicates that the average distance between the two distributions is smaller for Option II; consequently, it is more difficult to detect differences and the power is lower. This immediately suggests a need to consider developing an algorithm comparable to Option II, which minimizes the standard error, rather than the squared standard error. The overwhelming consensus to do the opposite should be carefully considered, however, before such an effort is undertaken.

Another promising avenue of endeavor--which also runs counter to majority opinion--is an examination of the assertion that allocating sample material equally will maximize the omnibus F statistic (Appelbaum & Cramer, 1974, p. 337; Li, 1964, pp. 197-198). On the basis of this study it would seem that, working backwards from the coefficients of the maximum possible contrast, one could allocate N in such a manner that the value of F would be increased, thereby increasing the power of the omnibus F.

Applications

It is interesting to note that none of the research designs considered in Appendix C even approaches the

proposed minimum acceptable level for power, .80 (Cohen, 1965, pp. 98-99); in fact, the highest power value is little more than half this prescribed minimum. This apparent indictment of the combinations of k , N , and ψ which were inspected, might be tempered somewhat by the realization that the Bonferroni adjustment, with its concern for controlling the experimentwise error rate, is not widely practiced by many researchers, unfortunately. If each of the comparisons carried out in this study were tested at the nominal alpha level, the reported power levels would be measurably higher. On the other hand, the fact that the designs represent planned, rather than post hoc, comparisons--which would have even lower power values--strongly suggests that commonly-accepted values for n (i.e., 10 and 15) may provide grossly inadequate tests of even medium treatment effects. Thus, the routine assessment of the probable power of an investigation before data are collected, would seem to be in order--an opinion which only echoes that of most advocates of a priori precision analysis.

Should an investigator choose to use one of the alternate allocation schemes, close attention should be paid to the constraints, particularly the homogeneity of variance assumption. The analysis of variance model does not provide for nonadditive treatment effects (Glass et al., 1972, pp. 240-241; Kirk, 1968, p. 52;

Lindquist, 1953, pp. 75-78). Instead, they are hypothesized to be constant across individuals within groups, so the assumption of homoscedasticity is basically an assumption that the experimental material all comes from the same population. Since the homogeneity of variance assumption is central to both of these algorithms, anyone who implements them should decrease his/her chances of violating this fundamental assumption by randomly selecting sample material from a population and randomly assigning it to treatment conditions.

Though most of the advantages associated with Option II, in terms of relative efficiency, are similar, one notable exception occurs: a nonorthogonal pair of comparisons ($k = 4$, adapted from Lindman, 1974, p. 64) exhibits a relative efficiency of .7441 (see Tables 11 and 12, Appendix C). This seems to indicate that Option II offers substantial advantages in some experiments, particularly when the number of planned comparisons is limited. It also points out that the investigator who focuses his/her attention on fewer hypotheses may reap significant power rewards.

The criterion of forcing contrasts to be tested at the same level of power (Option III), though certainly not a popular concept, has a great deal of intuitive appeal. Whether the indifference to such a criterion stems from a conscious decision to reject it or from a lack of

awareness that multiple contrasts may be estimated with widely disparate precision, is unknown. Perhaps the failure on the part of most researchers in education and psychology to explore power issues before selecting a research design is indicative that the latter is probably the case. Regardless of etiology, it would seem that this indifference might be examined (or re-examined, as the case may be).

Thematic throughout the findings of this investigation is the need for more definitive "pure" research before the techniques developed are routinely applied in field research. An orderly process might be as follows: first, a broad investigation of a variety of coefficient matrices could establish the general characteristics of studies best suited for unequal allocation (i.e., which offer the greatest power advantage over equal allocation); next, the practical advantages to be realized through unequal allocation could be better defined by imposing the constraint of integer n s upon the allocation. This two-step process, coupled with a clear statement of the accompanying assumptions should permit reasonable guidelines to be established for the applied researcher who wishes to determine when unequal allocation will yield significant statistical benefits.

Beyond these primary concerns, there are several other issues which seem to warrant further study.

There are additional parameters which might be introduced into the models to broaden their usefulness, for example. The specification of different effect sizes for individual contrasts may be a reasonable goal (Hsu, 1980). Also, different costs associated with different treatment conditions is certainly not uncommon and might be taken into account (Overall & Dalal, 1965). Perhaps the reality of measurement error could be incorporated into the models (Levin & Subkoviak, 1977; Subkoviak & Levin, 1977).

Transcending all of these matters, however, is the considerable need for researchers in education and psychology to become aware of the fundamental procedures of power analysis, the a priori options implied by a consideration of these procedures, and, in particular, the implications of the failure to do so.

LIST OF REFERENCES

LIST OF REFERENCES

- Anderson, V. L., & McLean, R. A. Design of experiments. New York: Marcel Dekker, 1974.
- Appelbaum, M. I., & Cramer, E. M. Some problems in the nonorthogonal analysis of variance. Psychological Bulletin, 1974, 81(6), 335-343.
- Ash, A., & Hedayat, A. An introduction to design optimality with an overview of the literature. Communications in Statistics, 1978, A7(14), 1295-1325.
- Barcikowski, R. S. Optimum sample size and number of levels in a one-way random-effects analysis of variance. Journal of Experimental Education, 1973, 41(4), 10-16.
- Bernhardson, C. S. Type I error rates when multiple comparison procedures follow a significant F test of ANOVA. Biometrics, 1975, 31, 229-232.
- Bock, R. D. Multivariate statistical methods in behavioral research. New York: McGraw-Hill, 1975.
- Boneau, C. A. The effects of violations of assumptions underlying the t test. Psychological Bulletin, 1960, 57(1), 49-64.
- Box, G. E. P., & Draper, N. R. A basis for the selection of a response surface design. Journal of the American Statistical Association, 1959, 54, 335-352.
- Box, G. E. P., & Wilson, K. B. On the experimental attainment of optimum conditions. Journal of the Royal Statistical Society (Series B), 1951, 13(1), 1-45.
- Brooks, S. H. The estimation of an optimum subsampling number. Journal of the American Statistical Association, 1955, 50(270), 398-415.
- Budescu, David V. The power of the F test in normal populations with heterogeneous variances. Educational and Psychological Measurement, 1982, 42, 409-416.
- Carlson, J. E., & Timm, N. H. Analysis of nonorthogonal fixed-effects designs. Psychological Bulletin, 1974, 81(9), 563-570.

- Carmer, S. G., & Swanson, M. R. An evaluation of ten pairwise multiple comparison procedures by monte carlo methods. Journal of the American Statistical Association, 1973, 68(341), 66-74.
- Chew, V. Comparisons among treatment means in an analysis of variance (ARS Tech. Rep. H/6, Agricultural Research Service, U. S. Department of Agriculture). Washington, D. C.: U. S. Government Printing Office, 1978.
- Chow, W. M. A note on the calculation of certain constrained maxima. Technometrics, 1962, 4(1), 135-137.
- Cleary, T. A., & Linn, R. L. Error of measurement and the power of a statistical test. British Journal of Mathematical and Statistical Psychology, 1969, 22(1), 49-55.
- Cochrane, W. G., & Cox, G. M. Experimental designs. New York: John Wiley & Sons, 1950.
- Cohen, J. The statistical power of abnormal-social psychological research: a review. Journal of Abnormal and Social Psychology, 1962, 65(3), 145-153.
- Cohen, J. Some statistical issues in psychological research. In B. B. Wolman (Ed.), Handbook of clinical psychology. New York: McGraw-Hill, 1965.
- Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 70(6), 426-443.
- Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1969.
- Cohen, J. Approximate power and sample size determination for common one-sample and two-sample hypothesis tests. Educational and Psychological Measurement, 1970, 30, 811-831.
- Cox, D. R. Planning of experiments. New York: John Wiley & Sons, 1958.
- Cramer, E. M., & Appelbaum, M. I. Nonorthogonal analysis of variance--once again. Psychological Bulletin, 1980, 87(1), 51-57.
- Davies, O. L. (Ed.). Design and analysis of industrial experiments. New York: Hafner Publishing, 1956.

- Davis, D. J. Flexibility and power in comparisons among means. Psychological Bulletin, 1969, 71(6), 441-444.
- Dixon, W. J., & Massey, F. J., Jr. Introduction to statistical analysis (3rd ed.). New York: McGraw-Hill, 1969.
- Donaldson, T. S. Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. Journal of the American Statistical Association, 1968, 63, 660-676.
- Feldt, L. S. A comparison of the precision of three experimental designs employing a concomitant variable. Psychometrika, 1958, 23(4), 335-353.
- Feldt, L. S. What size samples for methods/materials experiments? Journal of Educational Measurement, 1973, 10(3), 221-226.
- Feldt, L. S., & Mahmoud, M. W. Power function charts for specification of sample size in analysis of variance. Psychometrika, 1958, 23(3), 201-210.
- Finney, D. J. The theory of experimental design. Chicago: University of Chicago Press, 1960.
- Fisher, R. A. The design of experiments. New York: Hafner Publishing, 1966.
- Fisher, R. A. Statistical methods and scientific inference (3rd ed.). New York: Hafner Press, 1973.
- Friedman, H. Simplified determinations of statistical power, magnitude of effect and research sample sizes. Educational and Psychological Measurement, 1982, 42, 521-526.
- Games, P. A. Inverse relation between the risks of type I and type II errors and suggestions for the unequal n case in multiple comparisons. Psychological Bulletin, 1971, 75(2), 97-102.
- Games, P. A. Multiple comparisons of means. American Educational Research Journal, 1971, 8(3), 531-565.
- Glass, G. V., Peckham, P. D., & Sanders, J. R. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research, 1972, 42(3), 237-288.

- Hammersley, J. M., & Handscomb, D. C. Monte carlo methods. New York: John Wiley & Sons, 1964.
- Hardison, C. D., Quade, D., & Langston, R. D. SAS functions to compute chi-squared, F, and t distributions and their inverses allowing for noncentrality and fractional degrees of freedom (Tech. Rep. S-122). Cary, N. C.: SAS Institute, January 1981.
- Harris, M., Horvitz, D. G., & Mood, A. M. On the determination of sample sizes in designing experiments. Journal of the American Statistical Association, 1948, 43(243), 391-402.
- Hays, W. L. Statistics (3rd ed.). New York: Holt, Rinehart & Winston, 1981.
- Heilizer, F. A note on variance heterogeneity in the analysis of variance. Psychological Reports, 1964, 14, 532-534.
- Herr, D. G., & Gaebelin, J. Nonorthogonal two-way analysis of variance. Psychological Bulletin, 1978, 85(1), 207-216.
- Hill, W. J., & Hunter, W. G. A review of response surface methodology: a literature survey. Technometrics, 1966, 8(4), 571-590.
- Horsnell, G. The effect of unequal group variances on the F-test for the homogeneity of group means. Biometrika, 1953, 40, 128-136.
- Howell, D. C., & McConaughy, S. H. Nonorthogonal analysis of variance: putting the question before the answer. Educational and Psychological Measurement, 1982, 42, 9-24.
- Hsu, L. M. On the power of multiple independent tests when the experimentwise error rate is controlled. Educational and Psychological Measurement, 1980, 40, 31-40.
- Huck, S. W., Sanders, W. L., & Clark, S. B. Planned comparisons, power, and sample size: equal n's may not be best. Paper presented at the meeting of the American Educational Research Association, Los Angeles, April 1981.
- Keppel, G. Design and analysis: a researcher's handbook. Englewood Cliffs, N. J.: Prentice-Hall, 1973.

- Kerlinger, F. N. Foundations of behavioral research. New York: Holt, Rinehart and Winston, 1964.
- Keselman, H. J. The statistic with the smaller critical value. Psychological Bulletin, 1974, 81(2), 130-131.
- Kiefer, J. Optimum experimental designs. Journal of the Royal Statistical Society (Series B), 1959, 21(2), 272-319.
- Kirk, R. E. Experimental design: procedures for the behavioral sciences. Belmont, California: Brooks/Cole, 1968.
- Kleijnen, J. P. C. Statistical techniques in simulation (2 vols.). New York: Marcel Dekker, 1974.
- Kurotschka, V. Optimal design of complex experiments with qualitative factors of influence. Communications in Statistics, 1978, A7(14), 1363-1378.
- Lachin, J. M. Sample size determinations for $r \times c$ comparative trials. Biometrics, 1977, 33, 315-324.
- Levin, J. R. Determining sample size for planned and post hoc analysis of variance comparisons. Journal of Educational Measurement, 1975, 12(2), 99-108.
- Levin, J. R., & Serlin, R. C. Approximate sample size and power determination for Scheffe contrasts. Journal of Educational Measurement, 1981, 18(1), 55-57.
- Levin, J. R., & Subkoviak, M. J. Planning an experiment in the company of measurement error. Applied Psychological Measurement, 1977, 1(3), 331-338.
- Li, J. C. R. Statistical inference (Vol. I). Ann Arbor, Michigan: Edwards Brothers, 1964.
- Lindman, H. R. Analysis of variance in complex experimental designs. San Francisco: W. H. Freeman & Company, 1974.
- Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.
- McNemar, Q. Psychological statistics (4th ed.). New York: John Wiley & Sons, 1969.

- Miller, J. K., & Knapp, T. R. The importance of statistical power in educational research. Occasional paper 13, Bloomington, Indiana: Phi Delta Kappa.
- Morgan, J. N., & Sonquist, J. A. Problems in the analysis of survey data, and a proposal. Journal of the American Statistical Association, 1963, 58, 415-435.
- Myers, J. L. Fundamentals of experimental design (3rd ed.). Boston: Allyn & Bacon, 1979.
- Neter, J., & Wasserman, W. Applied linear statistical models. Homewood, Illinois: Richard D. Irwin, Inc., 1974.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill, 1975.
- Norusis, M. J. SPSS statistical algorithms. Chicago: SPSS, Inc., 1979.
- Overall, J. E., & Dalal, S. N. Design of experiments to maximize power relative to cost. Psychological Bulletin, 1965, 64(5), 339-350.
- Overall, J. E., & Dalal, S. N. Empirical formulae for estimating appropriate sample sizes for analysis of variance designs. Perceptual and Motor Skills, 1968, 27, 363-367.
- Overall, J. E., Lee, D. M., & Hornick, C. W. Comparison of two strategies for analysis of variance in nonorthogonal designs. Psychological Bulletin, 1981, 90(2), 367-375.
- Overall, J. E., Spiegel, D. K., & Cohen, J. Equivalence of orthogonal and nonorthogonal analysis of variance. Psychological Bulletin. 1975, 82(2), 182-186.
- Perlmutter, J., & Myers, J. L. A comparison of two procedures for testing multiple contrasts. Psychological Bulletin, 1973, 79(3), 181-184.
- Petrinovitch, L. F., & Hardyck, C. D. Error rates for multiple comparison methods: some evidence concerning the frequency of erroneous conclusions. Psychological Bulletin, 1969, 71(1), 43-54.

- Plackett, R. L., & Burman, J. P. The design of optimum multifactor experiments. Biometrika, 1946, 33, 305-325.
- Ramsey, P. H. Power of univariate pairwise multiple comparison procedures. Psychological Bulletin, 1981, 90(2), 352-366.
- Rao, C. R. Linear statistical inference and its applications. New York: John Wiley & Sons, 1965.
- Rhyne, J. R. LYNSYS (Form U0101712). Knoxville, Tenn.: University of Tennessee Computing Center, April 1970.
- Scheffe, H. The analysis of variance. New York: John Wiley & Sons, 1959.
- Schmid, J. Meaningless complex posttest contrasts. Journal of Experimental Education, 1977, 46(1), 4-5.
- Silvey, S. D., Titterton, D. M., & Torsney, B. An algorithm for optimal designs on a finite design space. Communications in Statistics, 1978, A7(14), 1379-1389.
- Spjøtvoll, E. On the optimality of some multiple comparison procedures. Annals of Mathematical Statistics, 1972, 43(2), 398-411.
- Subkoviak, M. J., & Levin, J. R. Fallibility of measurement and the power of a statistical test. Journal of Educational Measurement, 1977, 14(1), 47-52.
- Toothaker, L. E., & Mallick, C. On "The statistic with the smaller critical value." Psychological Bulletin, 1975, 82(4), 541-542.
- Umland, A. W., & Smith, W. N. The use of LaGrange multipliers with response surfaces. Technometrics, 1959, 1(3), 289-292.
- Wald, A. On the efficient design of statistical investigations. Annals of Mathematical Statistics, 1943, 14, 134-140.
- Winer, B. J. Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill, 1971.

Wolf, G., & Cartwright, B. Rules for coding dummy variables in multiple regression. Psychological Bulletin, 1974, 81(3), 173-179.

Wu, C. F. Some iterative procedures for generating nonsingular optimal designs. Communications in Statistics, 1978, A7(14), 1399-1412.

APPENDICES

APPENDIX A

GLOSSARY OF TERMS

ALPHA: the probability that a statistical test will result in rejection of the null hypothesis when treatments do not actually differ, i.e., the probability that a mistake will be made in concluding that treatments differ when, in fact, observed differences are due to chance.

COMPARISON: See "contrast."

CONTRAST: a linear combination of treatment means in which the sum of the weights (coefficients) is zero. N.B., in this document the terms "comparison" and "contrast" are used interchangeably.

EFFECT SIZE: the degree to which a phenomenon is present in a population, i.e., a measure of the extent to which the null hypothesis is false.

ERROR RATE, EXPERIMENTWISE: the probability that at least one Type I error has been committed in the course of conducting the full range of statistical tests in an investigation.

ERROR RATE, FAMILYWISE: the probability that an entire set of statistical tests which are associated with a particular source of variance, contains at least one Type I error. In the case of a one-factor design, the familywise error rate is equivalent to the experimentwise error rate.

ERROR RATE PER COMPARISON: the probability that a single statistical test results in a Type I error, i.e., the nominal alpha level.

POWER (of a statistical test): the probability that the test will result in rejection of the null hypothesis when a true treatment difference of important magnitude does in fact exist.

PRECISION (of an estimate): the closeness with which a sample statistic can be expected to approximate the corresponding population parameter. The precision of an estimator is generally measured by the standard error of its sampling distribution, i.e., the standard error of the estimate.

TYPE I ERROR: the erroneous rejection of a true null hypothesis. The probability of committing a Type I error is commonly called the alpha level and is designated by α .

TYPE II ERROR: the failure to reject a false null hypothesis. The probability of committing a Type II error is commonly called beta and is designated by β .

APPENDIX B

EXAMPLE OF PROGRAM USED TO COMPUTE POWER

```

//KEQ6 JOB (48171,71328,RMTO,G-X),CLARK.SHELDON.B.
//S1 EXEC SAS
//SYSIN DD *
DATA A; INFILE DISK;
ARRAY P P1-P6;
ARRAY N N1-N6;
ARRAY C C1-C6;
ARRAY CN CN1-CN6;
INPUT COMMENT $ 1-30 K BIGN ALPHA NO_COMP / PSI P1-P6 C1-C6;
*
* ASSIGNS N'S TO GROUPS BASED ON PROPORTIONS;
DO OVER P;
    N = P*BIGN;
END;
*
* COMPUTES (CJ**2 / NJ);
DO OVER C;
    DO I =1 TO 6;
        CN = C**2 / N;
    END;
END;
*
* COMPUTES THE STANDARD ERROR OF THE CONTRAST;
SE = SQRT (SUM(OF CN1-CN6));
*
* COMPUTES THE NONCENTRALITY PARAMETER,
* USING MEDIUM EFFECT SIZE (0.5);
NC = (0.5/SE);
*
* COMPUTES DF FOR T-TEST, USING DF FOR MSE;
DF = BIGN-K-1;
*
*ADJUSTS ALPHA FOR THE NUMBER OF CONTRASTS TESTED;
ALPHA = ALPHA / NO_COMP;
*
* COMPUTES CRITICAL T-VALUE;
PROBEX = 1-ALPHA;
CRIT = TINV(PROBEX,DF,0);
*
*COMPUTES BETA AND POWER;
BETA = TPROB(CRIT,DF,NC);
POWER = 1-BETA;
DATA B; SET A;
KEEP COMMENT BIGN PSI SE POWER C1-C6 N1-N6;
PROC PRINT DATA=B;
TITLE ALL ANALYSES INVOLVE 6 GROUPS;
//DISK DD *

```

APPENDIX C

Table 3

Precision Comparisons Among Three Allocation Schemes:
Helmert, $k = 3$, $N = 30$

Contrast ^a	Allocation Criterion					
	Equal n_j^b		Min. Avg. SE^2^c		Equal SE^2^d	
	SE^e	Power ^f	SE^e	Power ^f	SE^e	Power ^f
1	.3873	.2369	.3951	.2294	.4216	.2068
2	.4472	.1886	.4393	.1940	.4216	.2068
Mean	.4173	.2178	.4272	.2117	.4216	.2068
Sigma	.0300	.0242	.0221	.0177	.0000	.0000

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & 1 & -1 \end{bmatrix}$$

^b $n_1=10$; $n_2=10$; $n_3=10$.

^c $n_1=9.27$; $n_2=10.365$; $n_3=10.365$.

^d $n_1=7.5$; $n_2=11.25$; $n_3=11.25$.

^e σ_e^2 is assumed to be 1 .

^f For medium effect size; global alpha is 0.05 .

Table 4

Precision Comparisons Among Three Allocation Schemes:
Helmert, $k = 3$, $N = 45$

Contrast ^a	Allocation Criterion					
	Equal n_j^b		Min.Avg. SE^2^c		Equal SE^2^d	
	SE^e	Power ^f	SE^e	Power ^f	SE^e	Power ^f
1	.3162	.3387	.3226	.3277	.3443	.2941
2	.3652	.2667	.3587	.2748	.3443	.2941
Mean	.3407	.3027	.3407	.3013	.3443	.2941
Sigma	.0245	.0360	.0181	.0265	.0000	.0000

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & 1 & -1 \end{bmatrix}$$

^b $n_1=15$; $n_2=15$; $n_3=15$.

^c $n_1=13.905$; $n_2=15.5475$; $n_3=15.5475$.

^d $n_1=11.25$; $n_2=16.875$; $n_3=16.875$.

^e σ_e^2 is assumed to be 1 .

^f For medium effect size; global alpha is 0.05 .

Table 5

Precision Comparisons Between Two Allocation Schemes:
Nonorthogonal, $k = 3$, $N = 30$

Contrast ^a	Allocation Criterion			
	Equal n_j^b		Min.Avg.SE ^{2 c}	
	SE ^d	Power ^e	SE ^d	Power ^e
1	.4472	.1450	.4581	.1393
2	.3873	.1860	.3841	.1888
3	.4472	.1450	.4297	.1552
Mean	.4272	.1587	.4240	.1611
Sigma	.0282	.0193	.0305	.0206

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1 & 0 \\ 1/2 & 1/2 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

^b $n_1=10$; $n_2=10$; $n_3=10$.

^c $n_1=11.16$; $n_2=8.319$; $n_3=10.521$.

^d σ_e^2 is assumed to be 1 .

^e For medium effect size; global alpha is 0.05 .

Table 6

Precision Comparisons Between Two Allocation Schemes:
Nonorthogonal, $k = 3$, $N = 45$

Contrast ^a	Allocation Criterion			
	Equal n_j ^b		Min.Avg. SE^2 ^c	
	SE^d	Power ^e	SE^d	Power ^e
1	.3652	.2129	.3740	.2039
2	.3162	.2774	.3136	.2818
3	.3652	.2129	.3509	.2290
Mean	.3489	.2344	.3462	.2382
Sigma	.0231	.0304	.0249	.0325

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1 & 0 \\ 1/2 & 1/2 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

^b $n_1=15$; $n_2=15$; $n_3=15$.

^c $n_1=16.74$; $n_2=12.4785$; $n_3=15.7815$.

^d σ_e^2 is assumed to be 1 .

^e For medium effect size; global alpha is 0.05 .

Table 7

Precision Comparisons Among Three Allocation Schemes:
Helmert, $k = 4$, $N = 40$

Contrast ^a	Allocation Criterion					
	Equal n_j ^b		Min.Avg. SE^2 ^c		Equal SE^2 ^d	
	SE^e	Power ^f	SE^e	Power ^f	SE^e	Power ^f
1	.3651	.2111	.3770	.1992	.4054	.1751
2	.3873	.1898	.3886	.1887	.4054	.1751
3	.4472	.1478	.4336	.1558	.4054	.1751
Mean	.3999	.1829	.3997	.1812	.4054	.1751
Sigma	.0347	.0263	.0244	.0185	.0000	.0000

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

^b $n_1=10$; $n_2=10$; $n_3=10$; $n_4=10$.

^c $n_1=9.116$; $n_2=9.612$; $n_3=10.636$; $n_4=10.636$.

^d $n_1=7.552$; $n_2=8.112$; $n_3=12.168$; $n_4=12.168$.

^e σ_e^2 is assumed to be 1 .

^f For medium effect size; global alpha is 0.05 .

Table 8

Precision Comparisons Among Three Allocation Schemes:
Helmert, $k = 4$, $N = 60$

Contrast ^a	Allocation Criterion					
	Equal n_j ^b		Min.Avg. SE^2 ^c		Equal SE^2 ^d	
	SE^e	Power ^f	SE^e	Power ^f	SE^e	Power ^f
1	.2981	.3137	.3078	.2956	.3310	.2583
2	.3162	.2812	.3173	.2794	.3310	.2583
3	.3651	.2157	.3541	.2282	.3310	.2583
Mean	.3265	.2702	.3264	.2677	.3310	.2583
Sigma	.0283	.0408	.0200	.0287	.0000	.0000

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

^b $n_1=15$; $n_2=15$; $n_3=15$; $n_4=15$.

^c $n_1=13.674$; $n_2=14.418$; $n_3=15.954$; $n_4=15.954$.

^d $n_1=11.328$; $n_2=12.168$; $n_3=18.252$; $n_4=18.252$.

^e σ_e^2 is assumed to be 1 .

^f For medium effect size; global alpha is 0.05 .

Table 9

Precision Comparisons Between Two Allocation Schemes:
Non-Helmert Orthogonal, $k = 4$, $N = 40$

Contrast ^a	Allocation Criterion			
	Equal n_j^b		Min.Avg. SE^2 ^c	
	SE^d	Power ^e	SE^d	Power ^e
1	.3651	.2111	.3330	.2498
2	.3162	.2750	.3222	.2656
3	.3162	.2750	.3222	.2656
Mean	.3325	.2537	.3258	.2603
Sigma	.0231	.0301	.0051	.0074

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \end{bmatrix}$$

^b $n_1=10$; $n_2=10$; $n_3=10$; $n_4=10$.

^c $n_1=13.724$; $n_2=8.760$; $n_3=8.760$;
 $n_4=8.760$.

^d σ_e^2 is assumed to be 1 .

^e For medium effect size; global alpha is 0.05 .

Table 10

Precision Comparisons Between Two Allocation Schemes:
Non-Helmert Orthogonal, $k = 4$, $N = 60$

Contrast ^a	Allocation Criterion			
	Equal n_j ^b		Min.Avg.SE ² ^c	
	SE ^d	Power ^e	SE ^d	Power ^e
1	.2981	.3137	.2719	.3718
2	.2582	.4085	.2631	.3948
3	.2582	.4085	.2631	.3948
Mean	.2715	.3769	.2660	.3871
Sigma	.0188	.0447	.0041	.0108

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \end{bmatrix}$$

^b $n_1=15$; $n_2=15$; $n_3=15$; $n_4=15$.

^c $n_1=20.586$; $n_2=13.140$; $n_3=13.140$;
 $n_4=13.140$.²

^d σ_e^2 is assumed to be 1 .

^e For medium effect size; global alpha is 0.05 .

Table 11

Precision Comparisons Between Two Allocation Schemes:
Nonorthogonal, $k = 4$, $N = 40$

Contrast ^a	Allocation Criterion			
	Equal n_j ^b		Min. Avg. SE ^{2 c}	
	SE ^d	Power ^e	SE ^d	Power ^e
1	.3651	.2650	.3405	.2975
2	.4472	.1915	.3603	.2709
Mean	.4062	.2283	.3504	.2842
Sigma	.0411	.0368	.0099	.0133

^aFor the coefficient matrix

$$\begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 1 & 0 & 0 & -1 \end{bmatrix}$$

^b $n_1=10$; $n_2=10$; $n_3=10$; $n_4=10$.

^c $n_1=18.044$; $n_2=4.252$; $n_3=4.252$;
 $n_4=13.448$.

^d σ_e^2 is assumed to be 1 .

^eFor medium effect size; global alpha is 0.05 .

Table 12

Precision Comparisons Between Two Allocation Schemes:
Nonorthogonal, $k = 4$, $N = 60$

Contrast ^a	Allocation Criterion			
	Equal n_j^b		Min.Avg. SE^2 ^c	
	SE^d	Power ^e	SE^d	Power ^e
1	.2981	.3775	.2780	.4235
2	.3651	.2694	.2941	.3860
Mean	.3316	.3235	.2861	.4048
Sigma	.0335	.0541	.0081	.0188

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 1 & 0 & 0 & -1 \end{bmatrix}$$

^b $n_1=15$; $n_2=15$; $n_3=15$; $n_4=15$.

^c $n_1=27.066$; $n_2=6.378$; $n_3=6.378$;
 $n_4=20.172$.

^d σ_e^2 is assumed to be 1 .

^e For medium effect size; global alpha
 is 0.05 .

Table 13

Precision Comparisons Among Three Allocation Schemes:
Helmert, $k = 6$, $N = 60$

Contrast ^a	Allocation Criterion					
	Equal n_j ^b		Min.Avg. SE^2 ^c		Equal SE^2 ^d	
	SE^e	Power ^f	SE^e	Power ^f	SE^e	Power ^f
1	.3464	.1789	.3600	.1653	.3856	.1440
2	.3535	.1716	.3631	.1625	.3856	.1440
3	.3651	.1607	.3678	.1583	.3856	.1440
4	.3873	.1428	.3810	.1475	.3856	.1440
5	.4472	.1081	.4261	.1185	.3856	.1440
Mean	.3799	.1524	.3796	.1504	.3856	.1440
Sigma	.0364	.0253	.0243	.0171	.0000	.0000

^a For the coefficient matrix

$$\begin{bmatrix} 1 & -1/5 & -1/5 & -1/5 & -1/5 & -1/5 \\ 0 & 1 & -1/4 & -1/4 & -1/4 & -1/4 \\ 0 & 0 & -1/3 & -1/3 & -1/3 & -1/3 \\ 0 & 0 & 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

^b $n_1=10$; $n_2=10$; $n_3=10$; $n_4=10$; $n_5=10$; $n_6=10$.

^c $n_1=9.102$; $n_2=9.282$; $n_3=9.558$; $n_4=10.026$; $n_5=11.016$; $n_6=11.016$.

^d $n_1=7.782$; $n_2=8.004$; $n_3=8.346$; $n_4=8.97$; $n_5=13.452$; $n_6=13.452$.

^e σ_e^2 is assumed to be 1 .

^f For medium effect size; global alpha is 0.05 .

Table 14

Precision Comparisons Among Three Allocation Schemes:
Helmert, $k = 6$, $N = 90$

Contrast ^a	Allocation Criterion					
	Equal n_j^b		Min.Avg. SE^2^c		Equal SE^2^d	
	SE^e	Power ^f	SE^e	Power ^f	SE^e	Power ^f
1	.2828	.2786	.2940	.2569	.3149	.2224
2	.2886	.2669	.2965	.2523	.3148	.2224
3	.2981	.2494	.3003	.2456	.3149	.2224
4	.3162	.2204	.3111	.2281	.3148	.2225
5	.3651	.1638	.3479	.1808	.3148	.2224
Mean	.3102	.2358	.3100	.2327	.3148	.2224
Sigma	.0297	.0410	.0198	.0278	.0000	.0000

^aFor the coefficient matrix

$$\begin{bmatrix} 1 & -1/5 & -1/5 & -1/5 & -1/5 & -1/5 \\ 0 & 1 & -1/4 & -1/4 & -1/4 & -1/4 \\ 0 & 0 & -1/3 & -1/3 & -1/3 & -1/3 \\ 0 & 0 & 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

^b $n_1=15$; $n_2=15$; $n_3=15$; $n_4=15$; $n_5=15$; $n_6=15$.

^c $n_1=13.653$; $n_2=13.923$; $n_3=14.337$; $n_4=15.039$;
 $n_5=16.524$; $n_6=16.524$.

^d $n_1=11.673$; $n_2=12.006$; $n_3=12.519$; $n_4=13.455$;
 $n_5=20.178$; $n_6=20.178$.

^e σ_e^2 is assumed to be 1 .

^fFor medium effect size; global alpha is 0.05 .

Table 15

Precision Comparisons Between Two Allocation Schemes:
Non-Helmert Orthogonal, $k = 6$, $N = 60$

Contrast ^a	Allocation Criterion			
	Equal n_j ^b		Min.Avg. SE^2 ^c	
	SE^d	Power ^e	SE^d	Power ^e
1	.2579	.3637	.2690	.3348
2	.3873	.1634	.3599	.1881
3	.4472	.1252	.4016	.1525
4	.3873	.1634	.4092	.1472
Mean	.3699	.2039	.3599	.2057
Sigma	.0691	.0936	.0557	.0762

^a For the coefficient matrix

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 & -1/3 & -1/3 & -1/3 \\ 1/2 & 1/2 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1/2 & -1/2 \end{bmatrix}$$

^b $n_1=10$; $n_2=10$; $n_3=10$; $n_4=10$;
 $n_4=10$; $n_5=10$; $n_6=10$.

^c $n_1=12.402$; $n_2=12.402$; $n_3=11.208$;
 $n_4=11.208$; $n_5=6.39$; $n_6=6.39$.

^d σ_e^2 is assumed to be 1 .

^e For medium effect size; global alpha is 0.05 .

Table 16

Precision Comparisons Between Two Allocation Schemes:
Non-Helmert Orthogonal, $k = 6$, $N = 90$

Contrast ^a	Allocation Criterion			
	Equal n_j ^b		Min.Avg. SE^2 ^c	
	SE^d	Power ^e	SE^d	Power ^e
1	.2106	.5386	.2197	.5001
2	.3162	.2470	.2939	.2858
3	.3651	.1860	.3279	.2298
4	.3162	.2470	.3341	.2214
Mean	.3020	.3047	.2939	.3093
Sigma	.0564	.1373	.0455	.1129

^a For the coefficient matrix

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 & -1/3 & -1/3 & -1/3 \\ 1/2 & 1/2 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1/2 & -1/2 \end{bmatrix}$$

^b $n_1=15$; $n_2=15$; $n_3=15$; $n_4=15$;
 $n_5=15$; $n_6=15$.

^c $n_1=18.603$; $n_2=18.603$; $n_3=16.812$;
 $n_4=16.812$; $n_5=9.585$; $n_6=9.585$.

^d σ_e^2 is assumed to be 1 .

^e For medium effect size; global alpha is 0.05 .

Table 17

Precision Comparisons Between Two Allocation Schemes:
Nonorthogonal, $k = 6$, $N = 60$

Contrast ^a	Allocation Criterion			
	Equal n_j^b		Min. Avg. SE^2^c	
	SE^d	Power ^e	SE^d	Power ^e
1	.3162	.2162	.3064	.2308
2	.4472	.1081	.4334	.1147
3	.4472	.1081	.4796	.0951
4	.4472	.1081	.4334	.1147
5	.2579	.3310	.2591	.3278
Mean	.3831	.1743	.3824	.1766
Sigma	.0806	.0888	.0844	.0896

^a For the coefficient matrix

$$\begin{bmatrix} 1/2 & 1/2 & 0 & 0 & -1/2 & -1/2 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 1/3 & 1/3 & 1/3 & -1/3 & -1/3 & -1/3 \end{bmatrix}$$

^b $n_1=10$; $n_2=10$; $n_3=10$; $n_4=10$;
 $n_5=10$; $n_6=10$.

^c $n_1=10.65$; $n_2=10.65$; $n_3=8.694$;
 $n_4=8.694$; $n_5=10.65$; $n_6=10.65$.

^d σ_e^2 is assumed to be 1 .

^e For medium effect size; global alpha is 0.05 .

Table 18

Precision Comparisons Between Two Allocation Schemes:
Nonorthogonal, $k = 6$, $N = 90$

Contrast ^a	Allocation Criterion			
	Equal n_j^b		Min.Avg. SE^2^c	
	SE^d	Power ^e	SE^d	Power ^e
1	.2582	.3368	.2502	.3593
2	.3651	.1638	.3538	.1746
3	.3651	.1638	.3916	.1422
4	.3651	.1638	.3538	.1746
5	.2106	.5037	.2116	.4993
Mean	.3128	.2664	.3122	.2700
Sigma	.0658	.1363	.0689	.1379

^a For the coefficient matrix

$$\begin{bmatrix} 1/2 & 1/2 & 0 & 0 & -1/2 & -1/2 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 1/3 & 1/3 & 1/3 & -1/3 & -1/3 & -1/3 \end{bmatrix}$$

^b $n_1=15; n_2=15; n_3=15;$
 $n_4=15; n_5=15; n_6=15$.

^c $n_1=15.975; n_2=15.975; n_3=13.041;$
 $n_4=13.041; n_5=15.975; n_6=15.975$.

^d σ^2 is assumed to be 1 .

^e For medium effect size; global alpha is 0.05 .

VITA

Sheldon Brown Clark was born in Asheville, North Carolina, on July 26, 1948. He received the Bachelor of Arts degree in Classics (Latin) from the University of North Carolina at Asheville, in June 1971. After teaching Latin for one year, he managed a commercial print shop for about four years, prior to returning to school. Subsequently he received the Master of Arts in Education degree in Guidance and Counseling from Western Carolina University in June 1977. As a graduate student at the University of Tennessee, Knoxville, he held a Graduate Teaching Assistantship in the department of Educational and Counseling Psychology for one year, a Graduate Research Assistantship in the Bureau of Educational Research and Service for two years, and was a Research Associate in the Office of Institutional Research for one year. He received the Doctor of Philosophy degree in March 1983, after accepting a position at the University of Mississippi as Assistant Director of the Bureau of Institutional Research and Planning, and Assistant Professor of Educational Research. He is married to the former Fern Hope Young, of Asheville, North Carolina.