ABSTRACT

AN INVESTIGATION OF THE EFFECTS OF SELECTED METHODS OF POST HOC BLOCK FORMATION ON THE ANALYSIS OF VARIANCE

By

David John Wright

Post hoc blocking refers to the imposition of a generalized randomized blocks design on observations from a simple, one-factor design with fixed treatments when additional observations are available on some concomitant variable. The three post hoc blocking methods were:

(a) post hoc blocking within treatment groups (PHB/T): within each treatment group, the experimental units are ranked on the concomitant variable and divided into b blocks of equal size;

(b) fixed-range post hoc blocking (FRPB): b blocks of equal size are formed in the concomitant variable's population distribution and experimental units are assigned to blocks within treatment groups according to the b-1 population quantiles of the concomitant variable;

(c) sampled-range post hoc blocking (SRPB): the experimental units in all treatment groups are pooled, ordered on the concomitant variable,

and then assigned to blocks within treatment groups according to the b-1 pooled-sample quantiles of the concomitant variable.

For purposes of comparability to earlier research, attention was restricted to the case of dependent and concomitant variables with homogeneous bivariate normal distributions across treatment populations. For each post hoc blocking method, the major topics were: testing the hypothesis of null treatment effects under conditions of block-treatment additivity; testing the hypothesis of block-treatment additivity; and the precision of the method in the additive condition, relative to the one-factor analyses of variance and covariance.

The investigation relied heavily on Monte Carlo methods to generate empirical distributions of mean squares and F ratios where analytic solutions were either unknown or unreasonably difficult. For each combination of treatments, blocks, average treatment-block cell frequency, and population correlation (ρ) between the dependent and concomitant variables, 1000 samples of observations were generated. Within each sample, equal numbers of observations were randomly assigned to treatments and then assigned to blocks within treatments according to the three post hoc blocking methods. A one-factor analysis of variance was computed on the unblocked observations,

followed by unweighted means analyses for the three post hoc blocking methods. If any of the methods had zero cell frequencies, the sample was dropped for that method and additional samples were later generated to bring the total number of samples to 1000.

The PHB/T method was of no value in testing the hypothesis of null treatment effects in the presence of either block or interaction effects and provided no gain in precision when both were zero. Treatment, block, and interaction effects were comfounded in the test of the null treatments hypothesis. The PHB/T method is excluded from further remarks about the post hoc blocking methods.

Blocks, in the FRPB method, were a fixed factor and the ratio of treatments over residual mean squares followed the central F distribution under the null treatments hypothesis, regardless of the degree of block-treatment non-additivity. The FRPB method had greater precision than the one-factor design only for $\rho \geq 0.6$. Unless block-treatment cell sizes happen to be very nearly equal, the FRPB method does not offer appreciable gains in precision. A priori blocking or the analysis of covariance, when applicable, are typically more powerful. The FRPB ratio of interaction over residual mean squares did, however, provide a means for investigating block-treatment interaction or, equivalently, regression heterogeneity

in the analysis of covariance. For the specific cases generated, the test of the block-treatment additivity hypothesis showed reasonable power, particularly for samples with nearly equal cell sizes.

Blocks, in the SRPB method, behaved like a finite factor. In the presence of increasing block-treatment non-additivity, the ratio of treatments over residual mean squares was increasingly liberal, while the ratio of treatments over interaction mean squares was even more conservative than the former ratio was liberal. The same conclusions appeared to be valid for the a priori sampled range method where equal-sized blocks are formed prior to assignment of experimental units to treatment groups within blocks. Neither sampled-range method provides a clear test of the treatments null hypothesis unless sample sizes are very large, little or no interaction is present, or the test is conditioned on a prior test for non-additivity. For the SRPB method, the test for additivity had reasonable power when the interaction was great enough to make the test for treatment effects markedly liberal.

The SRPB method, like the FRPB method had greater precision than the one-factor analysis of variance for $\rho \geq 0.6$. For purposes of increasing precision, both a

priori blocking and the analysis of covariance are preferable. Unless cell sizes are very nearly equal, even the one-factor analysis of variance is generally more precise. Investigation of block-treatment non-additivity on an after the fact basis is the only clear asset of the two post hoc blocking methods.

AN INVESTIGATION OF THE EFFECTS OF SELECTED METHODS OF POST HOC BLOCK FORMATION ON THE ANALYSIS OF VARIANCE

By

David John Wright

A THESIS

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department of Educational Psychology

1971

To my grandfather, Harley Z Wooden

ACKNOWLEDGEMENTS

I would like to thank the members of my doctoral committee, Professors Andrew C. Porter, Robert C. Craig, Maryellen McSweeney, Judith Henderson, and John Vinsonhaler for their encouragement and assistance throughout my graduate program. Particular thanks are due to Andrew Porter, who has been a stimulating teacher, an uncommonly attentive advisor, and a good friend. I would also like to acknowledge the contributions to my graduate studies of Professor Charles F. Wrigley and my colleagues Howard S. Teitlebaum and John F. Draper. Charlotte Hayes provided patient and accurate assistance with the typing and proofing of the final draft. Above all, I am grateful for the understanding and support of my wife, Carole, who still can't believe it is finally finished.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Researchers are often interested in testing an hypothesis that t treatments or levels of an independent variable have had an identical, additive effect on some population as reflected in a dependent variable. A common and straightforward way to test this hypothesis is to draw t random samples from the population of interest (or randomly assign elements of a random sample to each of t subsamples) and subject the samples to the treatments. If, in this completely randomized experiment, treatment effects are additive in the metric of the dependent variable, experimental units operate independently, and the dependent variable is normally distributed with equal variances across treatments, then the one-factor analysis of variance F statistic provides a test of the experimenter's hypothesis.

Given a basic research paradigm of randomly equivalent treatment groups, independence, additivity, and a dependent variable with homogeneous normal distributions, much of experimental design is concerned with increasing the precision of the experiment by minimizing the standard errors of the treatment means. Cox (1958, pp. 8-9)

considers precision as a function of

(a) the intrinsic variability of experimental units and the accuracy of the experiment[1];

(b) the number of experimental units;

(c) the experimental design and method of analysis.

The first of these factors, intrinsic variability and experimental accuracy, is generally least amenable to manipulation without affecting the experiment's external validity. For example, a restricted range of experimental units might be used to decrease intrinsic variability or materials might be presented by videotape to increase the accuracy of presentation but then the experiment may no longer generalize to a larger population with live presentation of materials. Experimental precision is approximately proportional to the number of experimental units per treatment group; increasing the number of units is a direct but often prohibitively expensive way to reach a particular level of precision. Precision can sometimes be increased appreciably if additional information on concomitant variables, available prior to the experiment, is incorporated into the experimental design and analysis (e.g., formation of homogeneous blocks of experimental units) or into the analysis alone (e.g., analysis of covariance). This increase

---

[1] Intrinsic variability and experimental accuracy refer to unit and technical errors, respectively, in the terminology of Wilk and Kempthorne (1955, 1956a, 1956b) and Scheffe' (1959).

in precision can be by a factor as great as one minus the squared correlation between the dependent and concomitant variables.

While a number of ways exist for utilizing information on concomitant variables in experimental design and analysis, three methods are of interest here:

(a) analysis of covariance;

(b) generalized randomized block designs;

(c) post hoc blocking, the major topic of this paper.

The analysis of covariance statistically removes variation associated with one or more concomitant variables and is quite thoroughly covered by Cochran (1957) in the lead article for an issue of Biometrics devoted entirely to that topic.[2] Some of the analysis of covariance's principal disadvantages involve stringent assumptions about the relationship between dependent and concomitant variables, i.e. that the relationship is adequately specified (linear, quadratic, exponential, etc.), that the regression is constant across treatment populations (Peckham, 1969), and that components of the model are additive. A further assumption about error free measurement of the covariate is of little consequence when groups are randomly equivalent (Lord, 1963; Porter, 1967, 1971; DeGracie, 1969).

---

[2] See Elashoff (1969) for a more recent review and summary of research.

Blocks or levels designs refer to the class of experimental designs where blocks are formed by homogeneously grouping experimental units according to some antecedent characteristic, then randomly assigning units to treatment groups within blocks. Typically, blocks are formed or sampling is carried out in such a manner that an equal number of experimental units are assigned to each block and treatment group combination. No distinction is made here between blocks, levels, and strata according to the concomitant variable's type of measurement scale as did Glass and Stanley (1970, pp. 493-494), for example. Also, while the blocks terminology has sometimes been reserved for the case of a single experimental unit per block-treatment combination, Wilk's (1955) "generalized randomized blocks design" terminology will be followed with n_{ij} experimental units per block-treatment combination, depicted in Figure 1. Y_{ijk} denotes the value of the dependent variable associated with experimental unit k in the i^{th} block and j^{th} treatment combination where i = 1, 2,...,b; j = 1, 2,...,t; and k = 1, 2,..., n_{ij} = n, a constant, in most cases.

Cox (1957) and Feldt (1958) investigated the precision of designs utilizing concomitant variables, including analysis of covariance and two forms of blocking. They agreed that blocking tended to be more precise for low correlations between dependent and concomitant variables while the analysis of covariance was more precise for

Block	Treatments				
	1	. . .	j	. . .	t
1	Y_{111} ⋮ Y_{11k} ⋮ $Y_{11n_{11}}$		Y_{1j1} ⋮ Y_{1jk} ⋮ $Y_{1jn_{1j}}$		Y_{1t1} ⋮ Y_{1tk} ⋮ $Y_{1tn_{1t}}$
⋮					
i	Y_{i11} ⋮ Y_{i1k} ⋮ $Y_{i1n_{i1}}$		Y_{ij1} ⋮ Y_{ijk} ⋮ $Y_{ijn_{ij}}$		Y_{it1} ⋮ Y_{itk} ⋮ $Y_{itn_{it}}$
⋮					
b	Y_{b11} ⋮ Y_{b1k} ⋮ $Y_{b1n_{b1}}$		Y_{bj1} ⋮ Y_{bjk} ⋮ $Y_{bjn_{bj}}$		Y_{bt1} ⋮ Y_{btk} ⋮ $Y_{btn_{bt}}$

Figure 1

The Generalized Randomized Blocks Design

very high correlations, but did not agree on the transition point for the relative precisions of the two methods. Pingel (1968), in a Monte Carlo study of generalized randomized blocks designs, distinguished four methods of block formation:

(a) fixed-value blocks: b different values of the blocking variable are selected and then, for each of the b values, tn experimental units are randomly sampled;

(b) sampled-value blocks: b values of the blocking value are randomly sampled and then, for each of the b values, tn experimental units are randomly sampled;

(c) fixed-range blocks: specify b mutually exclusive ranges of equal probability in the blocking variable's population distribution and then, within each of the b ranges, randomly sample tn experimental units;

(d) sampled-range blocks: randomly sample btn experimental units, rank them according to their values on the blocking variable, and then randomly assign the first tn units to treatments within block one, the next tn units to treatments within block two, etc.

Pingel concluded that design precision was a function of block formation with fixed- and sampled-value methods most precise and the widely used sampled-range method least precise.

Post Hoc Blocking Methods

Post hoc blocking, a term coined by Porter and McSweeney (1970), refers to a class of blocking methods which impose a generalized randomized blocks design on data collected from a completely randomized experiment. Block formation is accomplished after experimental units have been assigned to treatment groups. The post hoc blocking methods utilize information on a concomitant variable which is antecedent to, or at least independent of, treatment effects. This independence restriction is the same as that imposed on potential covariates in the analysis of covariance.

When data are available on some concomitant variable in a completely randomized experiment, a post hoc blocking method might be considered under one of the following conditions:

(a) Heterogeneous regression: An analysis of covariance was planned but preliminary analyses revealed that the regression slopes for the dependent variable and covariable varied across treatments, implying nonadditivity of treatment effects, or a blocks by treatments interaction if the covariable were used to form blocks. In this case the blocking method is post hoc both in the

sense of being applied after units were assigned to treatments and also as a method of investigating the violation of the additivity assumption.

(b) Nonlinearity: The form of the dependent-concomitant variable relationship is not linear, but not well enough known to use nonlinear analysis of covariance techniques. Nonlinearity could be discovered in preliminary analyses or from sources external to the experiment.

(c) Metric: The concomitant variable is measured in a metric (e.g., nominal) not amenable to covariance adjustment. Examples are sex, some socio-economic status indicators, curriculum groups, or percentile ranks.
If the decision to utilize the concomitant variable is made after units have been assigned to treatments, the blocking is post hoc.

(d) Precision: If, under some conditions, post hoc block methods yield greater precision than other methods such as covariance adjustment, then the post hoc blocking methods would be preferred on those grounds. For a range of low correlations between dependent variable and covariable, some methods of a priori block formation are more precise than the analysis of covariance (Cox, 1957; Feldt, 1958) and the same results may hold for post hoc blocking methods.
Clearly, then, post hoc block formation methods are not only of academic interest, but may also prove to be of

practical value for increasing experimental precision or investigation of treatment effect nonadditivity.

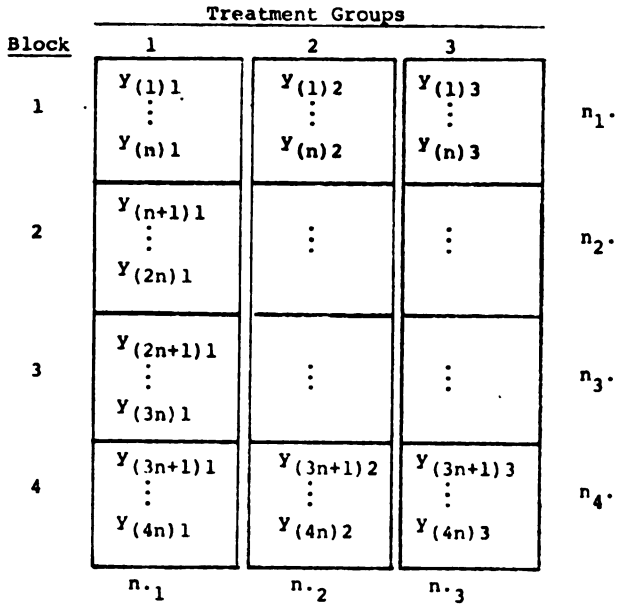
The three types of post hoc block formation considered in this paper are

(a) post hoc blocking within treatment groups (PHB/T);

(b) fixed-range post hoc blocking (FRPB);

(c) sampled-range post hoc blocking (SRPB).

The post hoc blocking within treatment groups method ranks experimental units within each treatment group with respect to a concomitant variable. The n highest units in each treatment group are assigned to block one, the n next highest in each group are assigned to block two, and so on until all experimental units have been assigned to blocks, as depicted in Figure 2A. The dependent variable is then analyzed by a two-factor analysis of variance. Porter and McSweeney (1970) investigated post hoc blocking in the context of a Monte Carlo investigation of the relative efficiency of the Kruskal-Wallis and Friedman test statistics, nonparametric analogs of the one-factor and randomized block analyses of variance. The found, for several combinations of treatment groups and sample sizes, that as the correlation between dependent and concomitant variables increased, the Friedman analysis of variance of ranks, when performed on data which was post hoc blocked within treatment groups, yielded high Type-I error rates and low power for noncentral



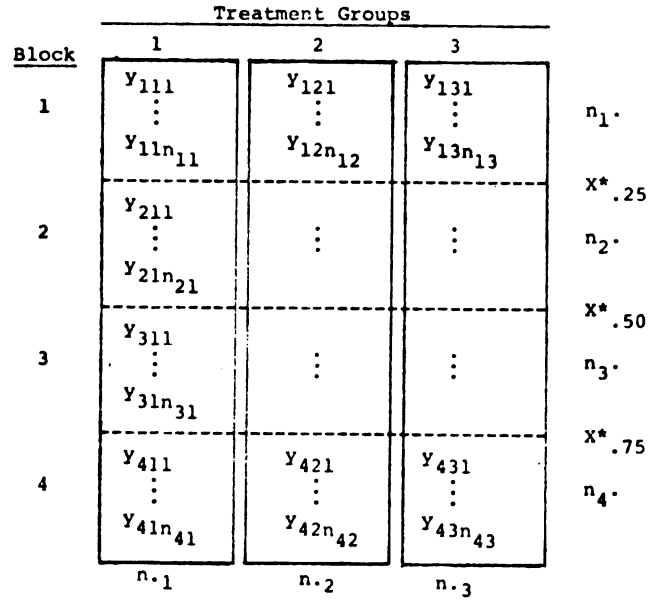
A. Post Hoc Blocking Within Treatment Groups Method

Units are ordered on x and assigned to blocks of size n, separately for each treatment group.

$y_{(i)j}$ = i^{th} y-observation (ordered on x) in group j

$n_{i\cdot} = 3n = N/4$

$n_{\cdot j} = 4n = N/3$



B. Fixed- and Sampled-Range Post Hoc Blocking Methods

Fixed-Range Method: Units are assigned to blocks within treatment groups according to the population quartiles of X.

$E(n_{ij}) = n$

$E(n_{i\cdot}) = N/4$

$X^*_{.50} = X_{.50}$ = Population Median of X

Sampled-Range Method: Units are assigned to blocks within treatments according to the pooled-samples quartiles of X.

$E(n_{ij}) = n$

$n_{i\cdot} = N/4$

$X^*_{.50} = \hat{X}_{.50}$ = Pooled-Samples Median of X

Figure 2

Three Post Hoc Blocking Methods with Four Blocks Each Imposed on a One-Factor Design with Three Treatment Groups and 4n Experimental Units per Treatment Group

cases, relative to both the Friedman on a priori blocks and the Kruskal-Wallis on simple randomly assigned data. That is, as the correlation between blocking and dependent variables increased, post hoc blocking within treatment groups led increasingly to rejection of the null hypothesis whether it was true (central case) or false (non-central case) but the power was low compared to the other methods studied. A preliminary Monte Carlo investigation of post hoc blocking within treatment groups for the one-factor and randomized blocks analyses of variance suggested results similar to those of Porter and McSweeney for the nonparametric methods, leading to this broader investigation of post hoc blocking.

Fixed-range post hoc blocking, the second form considered here, refers to the case of dividing a concomitant variable's population distribution into b ranges such that an equal proportion of the distribution is contained in each range (e.g., quartiles, deciles, etc.). Experimental units are assigned to blocks within treatment groups on the basis of their score on the concomitant or blocking variable and an analysis of variance is computed on the dependent variable. For example, if four blocks are desired, then the concomitant variable's population quartiles determine the range of scores which belong in each block. Experimental units in the concomitant variable's first quartile are assigned to block

one, those in the second quartile are assigned to block two, and so on, as shown in Figure 2B. The dependent variable is then analyzed by an unweighted means analysis of variance with four blocks and t treatments.

Sampled-range post hoc blocking is similar to the fixed-range method except that the pooled treatment group samples are divided into b blocks of equal size with respect to the concomitant variable rather than splitting the population into b segments or blocks of equal size. That is, the block endpoints are statistics (sample median, quartiles, deciles, etc.) rather than population parameters. This form of post hoc blocking is similar to the most common type of a priori block formation where all experimental units in a sample are ordered with respect to a concomitant variable and divided into b blocks of size tn each. The tn units in each block are then randomly assigned in equal numbers to each of the t treatments, whereas in the post hoc method the cell frequencies, are in general, unequal. Pingel (1968) found that while this method of a priori block formation is widely used and often presented in introductory texts as the method of block formation, it is also the least precise of the four methods he studied.

Other possible methods of block formation, such as equal sized intervals, either fixed or sampled, were deemed unfeasible as post hoc blocking methods for two

reasons. First, the blocks would have to be weighted in some fashion in order to obtain unbiased estimates of population parameters--a procedure which is possible in principle but very difficult in practice. The second, more serious, drawback involves the problem of empty cells for blocks that represent seqments with low frequencies, such as at the extremes of a normal distribution. These shortcomings can be overcome when blocks are part of the experimental design, but are virtually insurmountable after the fact.

As methods for increasing experimental precision, the analysis of covariance and generalized randomized blocks models have been studied thoroughly by other investigators but post hoc blocking methods have received virtually no systematic consideration. In view of their potential utility for incorporating concomitant variables into a design in order to increase precision or investigate nonadditivity of treatment effects (as in the case of regression heterogeneity), the remainder of this investigation is principally concerned with methods of post hoc block formation.

Scope of the Investigation

The principal topic in the investigation of post hoc block formation methods is testing the hypothesis of null treatment effects when the data fit a normal theory model and are analyzed by an unweighted means analysis of variance. The choice of a test statistic is determined by considering, for the case of null treatment effects:

(a) the expected values of mean squares when all assumptions are true;

(b) the expected values of mean squares when the assumption of block-treatment additivity is violated;

(c) empirical Type-I error rates when all assumptions are met; and

(d) empirical Type-I error rates when the assumption of block-treatment additivity is violated.

Two secondary topics in the investigation are:

(a) tests for block-treatment non-additivity (interaction) and

(b) the relative precision of the one-factor analysis of variance, the post hoc blocking methods, and the one-factor analysis of covariance, as well as the empirical power of the one-factor analysis of variance and the

unweighted means analyses of the post hoc blocking methods for one case of noncentral treatments.

The investigation is limited to a normally distributed dependent variable with constant variance across treatment groups and a joint bivariate normal density with a concomitant variable. Topics are investigated analytically where feasible or empirically with Monte Carlo techniques when analytic techniques are either unknown or unreasonably difficult. Where Monte Carlo techniques are required, cases are sampled from combinations of two and four treatment levels; two and four blocks; correlation of 0, 0.2, 0.4, 0.6, and 0.8 between dependent and concomitant variables; average block-treatment cell sizes between three and twenty; null and non-null treatment effects; and homogeneous and heterogeneous regression.

One further limitation should be noted. The analyses for the fixed- and sampled-range post hoc blocking methods are inapplicable if any block-treatment combinations have zero frequencies. For that reason, all investigations of the two methods are limited to cases where post hoc blocking resulted in non-zero cell frequencies.

CHAPTER II

DESCRIPTION OF THE INVESTIGATION

The Normal Theory Model

The basic model in this investigation is the classical normal theory fixed effects model. Under the null hypothesis, the t treatment populations are the only populations of interest and, for each treatment population, the values of the dependent variable, Y, have a normal distribution with common variance σ_y^2 and mean μ_y. The sample observations constitute independent random samples and treatment effects, if any, are strictly additive. The model is then

$$Y_{jk} = \mu + \beta_j + e_{jk}$$

where $j = 1,\ldots,t$

$k = 1,\ldots,n_j$, the number of units in sample j.

$\mu = \frac{1}{t} \sum_{j=1}^{t} \mu_j$, the average of t population means.

$\beta_j = \mu_j - \mu$

$$\sum_{j=1}^{t} \beta_j = 0$$

and the e_{jk} are independently $N(0, \sigma_y^2)$.

The hypothesis of major interest is

$$H_o: \mu_1 = \ldots = \mu_t = \mu$$

or equivalently,

$$H_o: \frac{1}{t-1} \sum_{j=1}^{t} \beta_j^2 = 0.$$

The appropriate test statistic is the well known F ratio of mean squares for treatments and residual, as listed in Table 1.

Table 1. Expected Values of Mean Squares in the One-Factor, Fixed Effects Analysis of Variance.

Source of Variation	Degrees of Freedom	Mean Square	Expected Value of Mean Square
Treatments	$(t-1)$	MS_T	$\sigma_y^2 + \frac{\sum_{j=1}^{t} n_j \beta_j^2}{t-1}$
Residual	$\sum_{j=1}^{t} (n_j-1)$	MS_R	σ_y^2

While a concomitant variable could have any kind of functional relationship with the dependent variable and be measured in any kind of metric, it is assumed for both theoretical and practical reasons that the concomitant variable X and the dependent variable Y have a joint bivariate normal density with constant means, variances and correlation ρ_{xy} in each treatment

population[3]. This assumption appears to be reasonable for much educational research and facilitates comparison with earlier research of Cox (1957), Feldt (1958), and Pingel (1968), among others.

Thus, it is assumed that for all t treatment populations, the dependent and concomitant variables, Y and X, respectively, have a common bivariate normal density function with parameters μ_{y_j}, μ_x, σ_y^2, σ_x^2, and ρ_{yx}. This implies that the effects of treatment are additive and equal under the null hypothesis. The only alternative hypothesis entertained is a shift alternative, i.e., $\mu_{y_j} \neq \mu_{y_j}$, for two or more populations but all other parameters are unchanged.

The Generalized Randomized Blocks Model

When blocks are added to the basic normal theory model, the design is as depicted in Figure 1. If blocks are considered fixed, as in the fixed-range post hoc blocking method, then the n_{ij} observations in the i^{th} block and j^{th} treatment group combination are assumed to be drawn from a normal population with mean μ_{ij} and variance σ_e^2. The fixed effects analysis of variance model is

---

[3]This assumption holds unless an exception is specifically noted. One noncentral treatments case and one case of regression heterogeneity is included in the empirical portion of this investigation.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

where $i = 1,\ldots,b;$

$j = 1,\ldots,t;$

$k = 1,\ldots,n_{ij}$

$$\mu = \mu.. = \sum_{i=1}^{b} \sum_{j=1}^{t} \frac{\mu ij}{bt};$$

$$\alpha_i = \mu_i. - \mu = \sum_{j=1}^{t} \frac{\mu_{ij}}{t} - \mu;$$

$$\beta_j = \mu._j - \mu = \sum_{i=1}^{b} \frac{\mu_{ij}}{b} - \mu;$$

$$(\alpha\beta)_{ij} = \mu_{ij} - \mu_i. - \mu._j + \mu;$$

$$e_{ijk} = Y_{ijk} - \mu_{ij};$$

e_{ijk} are independently $N(0, \sigma_e^2)$ for all i and j and $\sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0.$

When blocks are considered finite or random but treatment effects remain fixed, as may be the case in the sampled range post hoc blocking method, the generalized randomized blocks analysis of variance model becomes

$$Y_{ijk} = \mu + a_i + \beta_j + (a\beta)_{ij} + e_{ijk}$$

where i, j, k are defined above,

$$\sum_j \beta_j = \sum_j (a\beta)_{ij} = 0;$$

e_{ijk} are independently $N(0, \sigma_e^2)$ for all i and j;

$a_i \sim N(0, \sigma_a^2)$;

$(a\beta)_{ij} \sim N(0, \sigma^2_{(a\beta)})$;

and all random components are jointly normal.

The hypothesis of principal interest in this set of experimental designs remains that of equal treatment means

H_o: $\mu_{\cdot 1} = \ldots = \mu_{\cdot t}$ or equivalently $\frac{1}{t-1} \Sigma\beta_j^2 = 0$.

Note that block effects are assumed to be non-zero since blocking is included to increase precision, while the assumption of homogeneous bivariate normal distributions precludes non-zero interaction effects.

In the generalized randomized blocks design, the appropriate ratio of mean squares for testing a treatment main effect depends on whether blocks are fixed or random (Scheffe', 1959, Collier, 1960), as well as whether subclass frequencies are equal, proportional, or disproportional. While many references, such as Anderson and Bancroft (1952) or Snedecor and Cochran (1967) treat unequal subclass frequencies as a nuisance due to accident or poor planning, the reason for the unequal frequencies determines the appropriate analysis, if any. Bancroft (1968), in a monograph on the topic of unequal subclass frequencies, points out that if the subclass frequencies are proportional to the marginal frequencies then the sums of squares remain orthogonal but a direct test of the null hypothesis may not exist since treatment and error mean squares have disparate expectations. The same appears to be true when unequal

population sizes require proportional sampling.[4] If subclass frequencies are disproportionate due to treatment effects, population frequencies, etc., weighted least squares may be appropriate, while if the subclass frequencies would be proportional (or equal in the cases of interest here) except for sampling fluctuation, an expected subclass frequencies analysis may be employed. In the expected subclass frequencies analysis, subclass means are multiplied by the corresponding expected frequencies and the resultant totals substituted into the usual proportional subclass analysis of variance formulas. Table 2 contains the expected values of mean squares for the case of equal expected frequencies.[5] Finally, if subclass frequencies are unequal only because of sampling fluctuation, (as is the case in the post hoc blocking methods considered here) then the unweighted means analysis may be used, where all between classes sums of squares are computed on subclass means, then adjusted by the harmonic mean of the observed subclass frequencies, and within subclass data form the pooled estimate of residual variation as usual. Expected values for mean

---

[4]These conclusions were drawn from Wilk and Kempthorne (1956a, 1956b) but erroneously attributed to their 1955 paper.

[5]Other treatments of expected mean squares can be found in Wilk (1955) and Collier (1960) for randomized blocks designs, Cornfield and Tukey (1956) and the Wilk and Kempthorne series for factorial designs, and a generalized extension in Millman and Glass (1967).

Table 2. Expected Values of Mean Squares in the Generalized Randomized Blocks Design with Fixed Treatment Effects and Fixed, Finite, or Random Blocks. Experimental Units are Random.

Source of Variation	Degrees of Freedom	Mean Square	Expected Value	
			Fixed Blocks	Finite or Random Blocks
Blocks	$(b-1)$	MS_B	$\sigma_e^2 + \tilde{n}t\frac{\Sigma\alpha_i^2}{b-1}$	$\sigma_e^2 + \tilde{n}t\sigma_a^2$
Treatments	$(t-1)$	MS_T	$\sigma_e^2 + \tilde{n}b\frac{\Sigma\beta_j^2}{t-1}$	$\sigma_e^2 + (1-\frac{b}{B})\tilde{n}\sigma_{a\beta}^2 + \tilde{n}b\frac{\Sigma\beta_j^2}{t-1}$
Block-Treatment Interaction	$(t-1)(b-1)$	MS_{TB}	$\sigma_e^2 + \tilde{n}\frac{\Sigma\Sigma(\alpha\beta)_{ij}^2}{(t-1)(b-1)}$	$\sigma_e^2 + \tilde{n}\sigma_{a\beta}^2$
Residual (within cell)	$(n_{ij}-1)$	MS_R	σ_e^2	σ_e^2

$\tilde{n}$ = $\overline{E[n_{ij}]}$ in the expected subclass frequencies analysis; the harmnoic mean of n_{ij}s in the unweighted means analysis for fixed combinations of cell sizes or the expected value of the harmonic mean when cell sizes vary, as in the empirical sampling distributions of this investigation; $\underline{n}$ for equal cell frequencies.

B = the number of blocks in the population of blocks.

squares for the unweighted means analysis of the generalized randomized block design with fixed treatment effects, fixed, finite or random block effects, and random experimental units are presented in Table 2. The relative efficiencies of the expected subclass frequencies and unweighted means analyses are believed to be quite similar (Bancroft, 1968), so only the unweighted means analysis is considered in the sequel.

Inspection of Table 2 suggests that in the fixed effects model, under the null hypothesis, the expected mean square for treatments is σ_e^2, the residual within cell population variance, while for the mixed effects model the expected mean square for treatments is $\sigma_e^2 + \tilde{n}\sigma_{a\beta}^2$, the residual population variance plus a component due to treatments by blocks interaction, if any. The appropriate F ratios for testing the hypotheses of equal treatment effects are, then, the mean square for treatments over the mean square for residual in the fixed effects model and the mean square for treatments over the mean square for treatments by blocks interaction in the mixed effects model.

If, however, blocks are a random sample from a finite population of blocks, the expected mean square for treatments under the null hypothesis is $\sigma_e^2 + (1-b/B)\sigma_{a\beta}^2$, where β denotes the number of blocks in the population of blocks. Neither the ratio of treatment over residual mean squares

nor the ratio of treatment over interaction mean squares has a central F distribution under the null hypothesis when interaction is present. That is, if blocks are finite, interaction is present, and central F distribution tables are used to define critical values for significance testing, then the actual probability of a Type-I error will be greater than the nominal α level in the table for the ratio of treatments over residual mean squares and less than the nominal α level for the ratio of treatments over interaction mean squares. In the sampled-range method blocks are defined by sample statistics and the variability of block endpoints may cause blocks to resemble a finite or random factor, rather than a fixed factor as in the fixed-range method.

Precision

The precision of an experiment, as noted earlier, is inversely related to residual population variance and directly related to sample size. Thus, decreasing the population variance or increasing sample sizes yields a more precise design. Unfortunately, the best known and most useful attempts to quantify precision are referred to in terms of imprecision, the reciprocal of precision. For that reason, the remainder of this section refers to measures and indices of imprecision.

Imprecision, the average variability of the difference between two treatment means, is essentially a function of the population variance and sample size. Following Cox (1957) the <u>true</u> average imprecision of a design is

$$V_t = \underset{j \neq j'}{\text{Ave}} V(Y_{\cdot j \cdot} - Y_{\cdot j' \cdot}) = 2\sigma_e^2/n,$$

the variance of the estimated difference between treatment means, averaged over all pairs of treatments where

$Y_{\cdot j \cdot}$, $Y_{\cdot j' \cdot}$ = means for treatments j and j',

σ_e^2 = residual population variance for the design,

and n = the number of experimental units per treatment.

If it is assumed that a concomitant variable X has a homogeneous bivariate normal density with Y, then the conditional variance of Y for fixed X is denoted

$$\sigma_o^2 = \sigma_y^2(1 - \rho_{xy}^2)$$

where ρ denotes the population correlation between X and Y, and the minimum true average imprecision is then

$$\text{Min}(V_t) = 2\sigma_o^2/n.$$

An index of <u>true</u> imprecision, I_t, can be formed from the ratio of true average imprecision for a design to the minimum true average imprecision, i.e.,

$$I_t = \frac{V_t}{\text{Min}(V_t)} = \frac{2\sigma_e^2/n}{2\sigma_o^2/n}$$

or,

$$I_t = \frac{\sigma_e^2}{\sigma_o^2},$$

the ratio of the residual variance for a particular design to residual variance when all variance due to a concomitant variable is removed.

In an actual experiment, the residual variance is estimated from sample data rather than being a known parameter value. Since the accuracy of the variance estimate depends on the number of degrees of freedom available for estimation, and that number varies with the type of experimental design, the index of imprecision can be modified to reflect that variability by use of Fisher's (1935) factor for information lost in estimation. The resulting index of apparent imprecision is

$$I_a = I_t \frac{(f + 3)}{(f + 1)}$$

$$= \frac{\sigma_e^2 (f + 3)}{\sigma_o^2 (f + 1)}$$

where f = degrees of freedom for estimating σ_e^2. The index of apparent imprecision was used by Cox (1957) and Feldt (1958) to study the effects of incorporating concomitant variables into experimental designs. So long as the total number of experimental units per treatment group remains fixed and the number of units per block-treatment cell is constant, the index of apparent imprecision reflects the relative power of the F test statistics for each design. The power of the F statistic (for

t treatment groups) is a function of the degrees of freedom for estimating residual variance and a non-centrality parameter

$$\phi = \left[\frac{K\Sigma\beta_j^2}{(t-1)}/\sigma_e^2 \right]^{1/2}$$

Where $k = \begin{cases} \bar{n}. & \text{in the one-factor design,} \\ \tilde{n}b & \text{in the generalized randomized blocks design,} \\ nb & \text{for equal cell frequencies.} \end{cases}$

ϕ_o^2 is simply the treatment mean square component for treatment effects divided by the residual variance. When an experimental design affects only the residual variance and its degrees of freedom, the index of apparent imprecision is monotonically related to power.

A Generalized Index of Apparent Imprecision: I_g

Two of the post hoc blocking methods, the fixed-range and sampled-range methods, do not maintain equal cell frequencies. The non-centrality parameter for these methods is reduced by a factor of $(\tilde{n}/n)^{1/2}$ from the case of equal cell frequencies, a reduction which is reflected in neither the index of true imprecision nor the index of apparent imprecision. Thus, neither index is directly applicable to the present investigation. It is clearly desirable to use an index which is identical to the index of apparent imprecision when cell frequencies are equal, yet reflects the loss of power inherent in the disparate

cell frequencies of the unweighted means analysis.

Consider the generalized index of imprecision

$$I_g = \frac{\phi_o^2}{\phi_a^2} \cdot \frac{f+3}{f+1}$$

where $\phi_o^2 = \frac{nb\Sigma\beta_j^2}{(t-1)}/\sigma_o^2$, the squared non-centrality parameter for equal cell frequencies and minimum residual variance σ_o^2,

$\phi_a^2 = \frac{\tilde{n}b\Sigma\beta_j^2}{(t-1)}/\sigma_e^2$, the squared non-centrality parameter for a design with non-zero cell frequencies and actual residual variance σ_e^2, and $\frac{f+3}{f+1}$ = Fisher's adjustment for estimation of σ_e^2. Thus,

$$I_g = \frac{\frac{nb\Sigma\beta_j^2}{(t-1)}/\sigma_o^2}{\frac{\tilde{n}b\Sigma\beta_j^2}{(t-1)}/\sigma_e^2} \cdot \frac{f+3}{f+1}$$

$$= \frac{n\sigma_e^2}{\tilde{n}\sigma_o^2} \cdot \frac{f+3}{f+1}$$

and if cell freqeuncies are equal--as in the one-factor design, analysis of covariance for the one-factor design, and the post hoc blocking within treatments method--then $\tilde{n}$ is identical to n and the generalized index I_g reduces

to the index of apparent imprecision used by Cox and Feldt.

Precision of the One- and Two-Factor Designs

Inspection of the expected mean squares in Tables 1 and 2 for the one- and two-factor designs suggests that the relative precision of the two types of designs is a function of the relative magnitudes of σ_y^2, the residual variance of the one-factor design without blocking or covariance adjustment, the ratio $n/\tilde{n}$, and σ_e^2 or $\sigma_e^2 + (1-b/B)n\sigma_{a\beta}^2$ in the two-factor design with fixed, finite, or random blocks, respectively. Since $\sigma_{a\beta}^2 = 0$ under the assumption of homogeneous regression across treatment populations, that component is not considered further.

The imprecision of the one-factor, completely randomized design is a straightforward function of the sample sizes and the correlation between the dependent variable and the (unused) covariable. The unadjusted average variance of treatment mean differences is

$$V_t = 2\sigma_y^2/n,$$

while the minimum average variance is

$$\text{Min}(V_t) = 2\sigma_o^2/n,$$

but $\sigma_o^2 = (1-\rho^2)\sigma_y^2$

and thus the generalized index of apparent imprecision

is

$$I_g = \frac{1}{(1 - \rho^2)} \cdot \frac{(f + 3)}{(f + 1)}$$

Thus, when $\rho^2 = 0$, the one-factor, completely randomized analysis of variance is the most precise design available, with maximum degrees of freedom for estimating the residual variance. Clearly, too, as ρ^2 becomes large, this design is increasingly less precise and blocking or covariance adjustment becomes more advantageous.

Cox (1957) showed that when the analysis of covariance is applied to the one-factor design, the index of apparent imprecision

$$I_a = \frac{\sigma_a^2}{\sigma_o^2} \cdot \frac{f + 3}{f + 1}$$

$$= \frac{\sigma_y^2(1 - \rho^2)(1 + \frac{1}{f - 2})}{\sigma_y^2(1 - e^2)} \cdot \frac{f + 3}{f + 1}$$

$$= \frac{f - 1}{f - 2} \cdot \frac{f + 3}{f + 1}$$

Since the number of experimental units per treatment is not altered, the generalized index I_g is identical to I_a. In the analysis of covariance, I_g has the interesting property of being independent of all parameters but the residual degrees of freedom. As is demonstrated in the following sections, it is the only index considered in

this paper which does not vary with the correlation between dependent and concomitant variables.

The relative imprecision of the two-factor or generalized randomized blocks design is rather less straightforward than for the one-factor design, depending on the within block variability of the concomitant variable, as well as ρ and the sample sizes. Feldt (1958) has shown that given the assumptions of bivariate normal densities, homogeneous regression and homoscedasticity, the variance of the dependent variable in block i is

$$\sigma_{y_i}^2 = \sigma_y^2[1 - \rho^2(1 - \frac{\sigma_{xi}^2}{\sigma_x^2})]$$

where $\sigma_{x_i}^2$ = variance of X in block i.

Since σ_e^2 is simply the average of $\sigma_{y_i}^2$ over blocks,

$$\sigma_e^2 = \sigma_y^2[1 - \rho^2(1 - \bar{\sigma}_x^2/\sigma_x^2)]$$

where $\bar{\sigma}_x^2$ denotes the average within block variance of X. The generalized index of apparent imprecision is then

$$I_g = \frac{n\sigma_y^2[1 - \rho^2(1 - \bar{\sigma}_x^2/\sigma_x^2)]}{\tilde{n}\sigma_y^2(1 - \rho^2)} \cdot \frac{(f + 3)}{(f + 1)}$$

$$= \frac{n[1 - \rho^2(1 - \bar{\sigma}_x^2/\sigma_x^2)]}{\tilde{n}(1 - \rho^2)} \cdot \frac{(f + 3)}{(f + 1)} \cdot$$

Clearly, as $\bar{\sigma}_x^2$, the within block variance of the concomitant variable, approaches 0, the residual variance for this design approaches σ_o^2, its minimum value and, as $\bar{\sigma}_x^2$ approaches σ_x^2, the residual variance approaches σ_y^2, its maximum value.

Alternatively, if σ_x^2 is considered to be the sum of between and within block variances,

$$\sigma_x^2 = \sigma_{\bar{x}}^2 + \bar{\sigma}_x^2,$$

then maximizing $\sigma_{\bar{x}}^2$, the between blocks variance of X, maximizes the precision of the design if all other factors are held constant. Conversely, small between block variance of the concomitant variable leads to a less precise design. Thus, it is desirable to have both a strong correlation between Y and X and minimum average within blocks variance on X.

Within Block Variation of the Concomitant Variable

The residual variance, σ_e^2, as noted in the previous section, is a function of the correlation with the concomitant blocking variable and the concomitant variable's average within block variance, while the generalized index of apparent imprecision is also affected by the residual degrees of freedom and the variability of cell sizes. Since σ_y^2, σ_x^2, and ρ are assumed to be fixed for any population or set of populations, the generalized

index of imprecision and thus the precision of the three methods of block formation depends on $\bar{\sigma}_x^2$, the within blocks variance of X, as well as the number of degrees of freedom available for estimating σ_e^2 and the harmonic mean of cell frequencies. Thus, the value of $\bar{\sigma}_x^2$ plays an important role in this investigation.

The first method of post hoc block formation, post hoc blocking within treatments, forms equal sized blocks within samples on the concomitant variable. For this method, the within blocks variance, $\bar{\sigma}_x^2$, is a function of order statistics.[6] Within any one treatment group j,

$$\sigma^2_{x_{ij}} = \begin{cases} \frac{1}{n}\sum_{k=r}^{s} E[X_{(k)}^2] - \frac{2}{n(n-1)}\sum_{k=r}^{s-1}\sum_{k'=r+1}^{s} E[X_{(k)}\,X_{(k')}] & \text{for } n > 1, \\ E[X_{(k)}^2] - E[X_{(k)}]^2 & \text{for } n = 1 \end{cases}$$

where n = number of units per block-treatment combination,

i = block number,

r = (i-1)n, s = i(n),

$X_{(k)}$ = k^{th} order statistic, i.e., the k^{th} ordered value in the sample of size nb = N/t.

All treatment groups are random samples of the same size

---

[6]The following is adapted from Pingel (1968).

from populations with identical distributions on X and thus the within block variance, $\sigma^2_{x_{ij}}$, is constant across treatment populations for each block i and equal to the marginal within block variance $\sigma^2_{x_i}$. The average within block variance is then

$$\bar{\sigma}^2_x = \frac{1}{b}\Sigma\sigma^2_{xi},$$

the average over blocks of the marginal within block variances.

The expected values of $X_{(k)}$, $X_{(k)}^2$, and $X_{(k)}X_{(k')}$ have no simple expression, but tables of numerical approximations for normal deviates from samples of size 20 or less have been computed by Teichroew (1956) and reproduced in Sarhan and Greenberg (1962). Except for small sample sizes, Monte Carlo estimation is the most feasible approach.

For the fixed-range post hoc blocking method, where all experimental units are assigned to blocks according to the b-1 population quantiles of X (median, quartiles, etc.), the marginal within block variance of block i is

$$\sigma^2_{x_i} = 1 + \frac{X_{(i-1)}f[X_{(i-1)}] - X_{(i)}f[X_{(i)}]}{b} - \left(\frac{f[X_{(i-1)}] - f[X_{(i)}]}{b}\right)^2$$

where $X_{(i)}$ and $X_{(i-1)}$ are the endpoints of block i,

$f[X_{(i)}]$ = ordinate at $X_{(i)}$,

b = the number of blocks of equal area,

and $\bar{\sigma}_x^2 = \frac{1}{b}\Sigma\sigma_{x_i}^2$.

The actual pooled estimate of $\bar{\sigma}_x^2$ is more variable because it is pooled across treatment groups within each block, where n_{ij}, the number of units in block i and treatment group j, is a random variable, subject to the restriction that the sum across blocks is constant for all treatment groups and the sum of expected cell sizes across treatment groups is constant for all blocks. In order to make the estimates comparable over post hoc blocking methods, $\bar{\sigma}_x^2$ is also estimated by Monte Carlo methods.

Finally, in the sampled-range method of post hoc block formation, where all samples are pooled and divided into equal sized blocks according to the pooled sample quantiles, the within block variance of X is a function of order statistics as in the post hoc blocking within groups method. The principal difference between the two methods is that for the sampled-range method, the treatment groups have common block endpoints, while for the post hoc blocking within treatments method, block endpoints are defined separately for each treatment group. Thus the within block variances are functions of samples of size tn rather than t samples of size n, as in the post hoc blocking within treatment method. Since available estimates are limited to pooled sample sizes of twenty or less, estimates of $\bar{\sigma}_x^2$ are included in the

Monte Carlo investigation for the sampled-range method as well.

Data Generation

All data for this investigation were generated by several versions of a single program. The program generated empirical distributions of mean squares and F ratios for central, noncentral, or non-additive, central cases of selected combinations of design parameters. Each distribution was based on 1000 samples of pseudorandom normal deviates (PRNDs) which were used to form observations on the dependent and concomitant variables. Each sample was analyzed by the one-factor analysis of variance, as well as by the unweighted means analysis of the post hoc blocking methods.

Random Number Generator

Each pseudorandom normal deviate was the mean of sixteen pseudorandom numbers with a uniform distribution, generated by the multiplicative congruential method, and rescaled to have a population mean and variance of 0.0 and 1.0, respectively. The PRNDs were generated with a Compass-coded routine written by William Silverman and adapted by Richard Wolfe (1963). The routine was further modified to generate vectors of PRNDs and to

optimize calculation speed by making it specific to the Michigan State University CDC 3600 installation. Appendix A contains summary statistics for ten sets of 1000 samples each, which were generated by the modified routine. Porter (1967) has a more detailed summary of the same routine's properties.

Generation of Dependent Variables

For each sample, the program generated two vectors of PRNDs--the covariable (X) and an independent variable (Z)--each with expected value 0.0 and population variance 1.0. Central treatments, additive case dependent variables were constructed by the formula

$$Y_i = \rho X_i + (1 - \rho^2)^{1/2} . Z_i$$

where Y is the dependent variable, ρ is the population correlation between Y and X, and Y has an expected value of 0.0 and population variance of 1.0.

Noncentral, additive cases were generated by adding a constant (β) to each observation on the dependent variable in one-half of the treatment groups and subtracting the same constant from the remaining dependent variable observations. In all cases, a value of β was chosen such that

$$\phi^2 = \frac{nb\sum_{j}^{t} \beta_j^2}{t-1} / \sigma_y^2 = 1.0$$

in the one-factor design and

$$\phi'^2 = \frac{\tilde{n}b\sum_{j}^{t}\beta_j^2}{t-1} / \sigma_e^2 = \frac{\tilde{n}}{n} \cdot \frac{\sigma_y^2}{\sigma_e^2} \cdot \phi^2$$

in the fixed- and sampled-range methods. Thus, ϕ, the non-centrality parameter was 1.0 for all noncentral treatments one-factor analyses, but varied for the fixed- and sampled-range analyses, depending on the harmonic mean of cell frequencies and the magnitude of residual within cell variance.

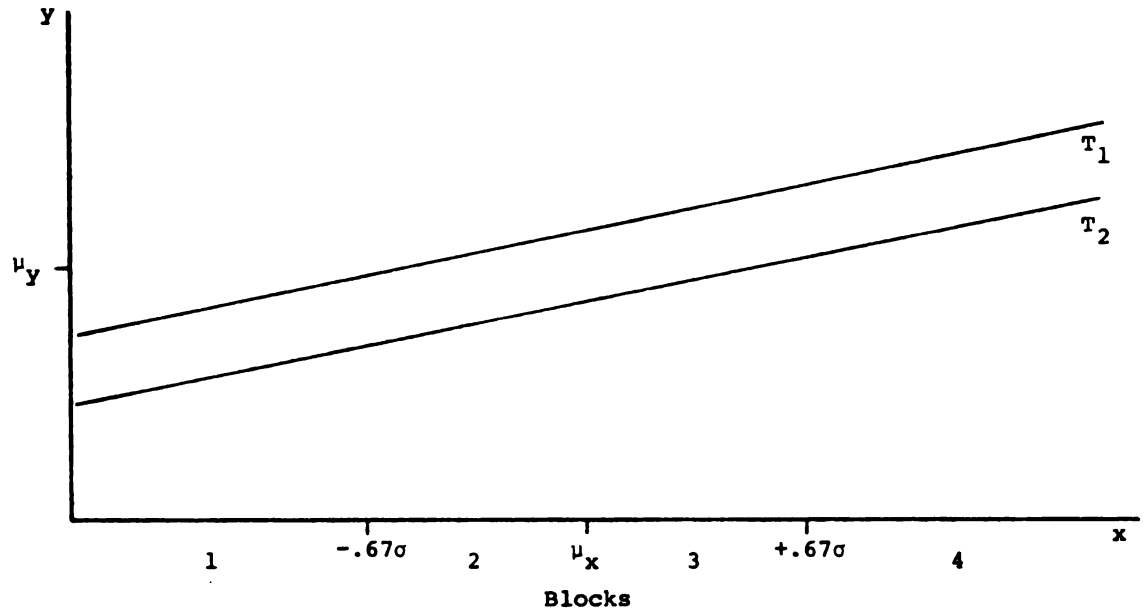
Non additivity was introduced by generating observations on the dependent variable for one-half of the treatment groups by

$$Y_i = \rho X_i + (1 - \rho^2)^{1/2} \cdot Z_i$$

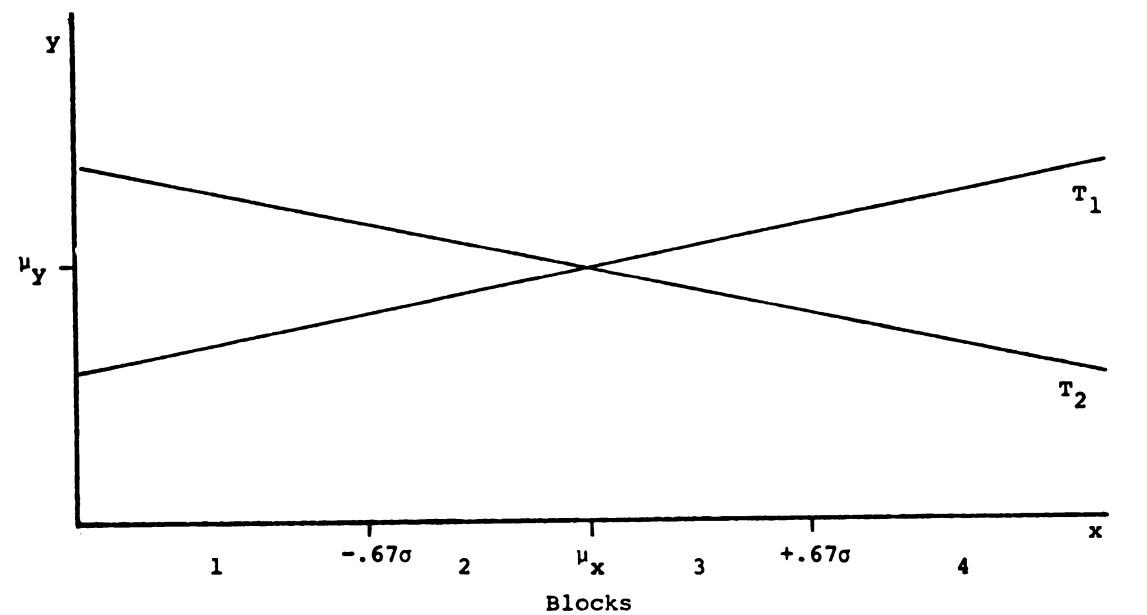
and the remaining observations by

$$Y_i = - \rho X_i + (1 - \rho^2)^{1/2} \cdot Z_i .$$

Feldt (1958) demonstrated that regression heterogeneity is both a necessary and sufficient condition for non-additivity (block-treatment interaction) and this method made the non-additivity a function of the correlation between the dependent and concomitant variables. As shown in Figure 3, the treatment populations had regression slopes of equal magnitude but opposite sign. Thus, when $\rho = 0.0$, there was no interaction; as ρ approached 1.0, the interaction became increasingly extreme.



A. Additive Case: Regression lines are parallel and coincide in the absence of treatment effects.



B. Non-Additive Case: Regression Slopes = $+\rho$, $-\rho$. Regression lines intersect at (μ_x, μ_y) in the absence of treatment effects.

Figure 3

Regression lines in the Additive and Non-Additive Conditions: t=2, b=4

Block Formation and Analysis

Once the observations on the dependent and concomitant variables had been generated, bn observations were randomly assigned to each of the t treatment groups and a one-factor analysis of variance was computed on Y. The observations in each treatment group were then ordered on X and assigned to blocks by the post hoc blocking methods. For the post hoc blocking within treatments method, the n smallest X observations were assigned to block one, the next n smallest were assigned to block two, and so on, until b blocks of size n had been formed within each treatment group. For the fixed-range post hoc blocking method, all X observations less than the first population quantile (e.g., quartiles for four blocks, the median for two blocks) of the unit normal distribution were assigned to block one, those less than the second quantile were assigned to block two, and so on until each block-treatment combination contained n_{ij} observations with $\sum_{i}^{b} n_{ij} = nb$ for each treatment group.

In the sampled-range post hoc blocking method, the tn smallest X observations in the pooled treatment groups were assigned to block one, the tn next smallest X observations were assigned to block two, and so on, until all observations had been assigned to block-treatment combinations with $\sum_{i}^{b} n_{ij} = nb$ for each treatment group and

$\sum_{j}^{t} n_{ij} = nt$ for each block.

After the observations were assigned to block-treatment combinations for each of the post hoc blocking methods, unweighted means analyses of variance were computed on the Y observations for each method. If zero cell frequencies were encountered for either the fixed- or sampled-range method, the analysis was not computed for that method and an additional sample was generated for it later.

Type-I error rates for central (null) treatment effects in the additive and non-additive conditions, as well as power for non-central treatment effects in the additive condition were computed by counting the number of times each F ratio exceeded selected percentiles (0.25, 0.50, 0.75, 0.90, 0.95, and 0.99) of the theoretical central F distribution with appropriate degrees of freedom.

Design Parameters Used in the Investigation

The empirical results of the investigation are a function of the post hoc blocking method, five parameters, and two conditions:

(a) the number of treatments, t;

(b) the number of blocks, b;

(c) the average number of experimental units per block-treatment combination, n;

(d) the correlation between dependent and concomi-

tant variables, ρ;

(e) the non-centrality parameter, ϕ; and

(f) additive/non-additive conditions.

The Number of Treatments

The number of treatment groups in this investigation was either two or four. Two treatments is the logical lower limit in a comparative experiment, while four treatment groups is a reasonable upper limit for most small-scale research studies. Further, there was no theoretical reason to expect different conclusions about precision or appropriate ratios of mean squares from any other number of treatment groups. An odd number of treatments was excluded for a purely practical reason, the computational algorithm was more efficient if it could be assumed that only even numbers of treatment groups would be used.

The Number of Blocks

The number of blocks was limited to either two or four. Two blocks is the minimum number possible, while four blocks again appeared to be a reasonable upper limit for small studies for several reasons. First, two and four blocks are common to other studies, such as Feldt

(1958) and Pingel (1968). Also, unless the number of observations per treatment is large, forming more blocks is likely to result in zero cell frequencies which precludes use of the fixed- and sampled-range methods. Finally, forming more blocks would have the effect of producing smaller within block variance on the blocking variable and thus decrease the dependent variable's residual variance and increase precision, but only if the correlation is great enough to offset the loss in degrees of freedom and probable increased variability in cell frequencies.

The Average Number of Units Per Cell

The average number of experimental units per block-treatment combination was three, five or ten. Three units per cell was included to see if that small a number is feasible with methods that require at least one observation per cell. It was originally planned to include up to twenty units per cell, but no clear advantage was apparent after the smaller cases had been generated. Adding more observations would simply increase the stability of the variance component estimators slightly, while causing a disproportionate increase in computer time. Doubling the number of observations more than quadruples the execution time, due primarily to the extensive amount of sorting required by the post hoc blocking methods.

The Correlation Between Dependent and Covariables

The correlation between dependent and blocking variables was 0.0, 0.2, 0.4, 0.6, and 0.8. The one-factor analysis of variance is known to have greatest precision when the correlation is zero, while the analysis of covariance has greatest relative precision for high correlations. The correlations found in behavioral-science research are rarely greater than 0.8 and are more often in the range of 0.4 to 0.6. Small values (0.0 and 0.2) were included primarily to gain insight into the relative imprecision of the post hoc blocking methods in the absence of strong block effects.

The Non-Centrality Parameter

All noncentral cases had the same non-centrality parameter, $\phi = 1.0$, for the one-factor design, while ϕ varied with the ratio $(\tilde{n}\ \sigma_y^2/n\ \sigma_e^2)$ for the post hoc blocking methods. The residual population variance, σ_y^2, was also held constant, resulting in identical expected treatments mean squares for all one-factor analyses while power varied with degrees of freedom. Other possible data generation rules included holding constant: the magnitude of treatment effects (β_j); the sum of squared treatment effects $(\Sigma\ \beta_j^2)$; or $\phi' = \frac{t-1}{t}\phi$, the non-centrality para-

meter definition commonly used with power charts.[7] Interpretation of empirical results was judged sufficiently difficult to outweigh the other advantages peculiar to each approach.

Additive and Non-Additive Conditions

Monte Carlo investigations like the present study typically place heavy reliance on ideal conditions, where all assumptions of the model are met, to investigate average values of mean squares, indices of imprecision, and degree of fit to theoretical reference distributions. However, in the present study, of all assumptions are met, either the mean square for block-treatment interaction or the residual mean square would be appropriate for testing the hypothesis of null treatment effects. It is only under conditions of non-additivity that one mean square or the other would be clearly preferable.

A secondary topic in the investigation was the test for block-treatment interaction. Empirical and theoretical error rates should be in close agreement for the ratio of interaction and residual mean squares in the additive condition and the ratio should have reasonable power in the non-additive condition.

---

[7] See, for example, Pearson and Hartley (1951), Fox (1956), and Feldt and Mahmoud (1958), all of which are widely reproduced in statistics texts and handbooks.

Empirical Cases Generated

The specific combinations of parameters used in the generation of empirical data are summarized in Figure 4. In the table, C denotes a central treatments case with all assumptions met; N denotes the introduction of non-additivity in the central case where half of the treatment populations have regression slopes equal to ρ, while the remainder have regression slopes equal to $-\rho$; and ϕ denotes non-central treatments with all other assumptions met.

t	b	n	ρ 0.0	0.2	0.4	0.6	0.8
2	2	5	Cϕ			Cϕ	
2	4	5	CNϕ	CN	CN	CNϕ	CN
4	2	5	CNϕ			CNϕ	
4	4	3	C			C	
4	4	5	CNϕ	CNϕ	CNϕ	CNϕ	CNϕ
4	4	10	Cϕ			Cϕ	

C = central case, all assumptions met
N = non-additivity introduced into central case
ϕ = treatments non-centrality parameter = 1.0, all assumptions met

Figure 4

Summary of Cases Investigated Empirically

CHAPTER III

RESULTS OF THE INVESTIGATION

The empirical phase of the investigation, as indicated in Chapter II, involved the generation of empirical distributions of mean squares and empirical F distributions for the one-factor design and the three methods of post hoc block formation: post hoc blocking within treatments; fixed-range post hoc block formation; and sampled-range post hoc block formation. The three cases considered were the central case with all assumptions met, non-central treatments with the non-centrality parameter $\phi = 1.0$, and regression heterogeneity with regression slopes equal to ρ and $-\rho$.

Following a brief digression on, and dismissal of, the post hoc blocking within treatments method, results are presented for average mean squares and empirical F distributions of the one-factor design analysis of variance and the two remaining post hoc blocking methods. Finally, results for precision and power are presented.

The Case Against Post Hoc Blocking Within Treatments

It became apparent, well before the start of the empirical investigation that the post hoc blocking within treatments (PHB/T) method is not a viable way to gain experimental precision. With this method, equal-sized blocks are formed on some concomitant variable within each treatment group and the data are analyzed according to the usual two-factor analysis of variance, which is identical to the unweighted means analysis when cell frequencies are equal. However, so long as equal-sized blocks are formed within treatment groups, blocking can have no possible effect on the variability of treatment means, which is the key to increased precision. The method does have an effect on the residual variance though. While the treatments mean square is identical to the one-factor treatments mean square, the one-factor design's residual mean square is partitioned into components which estimate variance due to blocks and within-block residual (see Table 3). The within-block residual makes a reasonable estimate of what the treatments mean square would be if the null hypothesis were true and blocks were formed across treatments rather than within them. Except in the trivial case of $\rho = 0.0$ and additivity ($\sigma_b^2 = \sigma_{tb}^2 = 0.0$), neither the ratio MS_T/MS_R, nor the ratio MS_T/MS_{TB} have central F distributions under the null hypothesis, since the expected mean square for treatments is greater

Table 3. Expected Values of Mean Squares for the Post Hoc Blocking Within Treatments Method

Source of Variation	Degrees of Freedom	Mean Square	Expected Value
Treatments	(t-1)	MS_T	$\sigma_y^2 + \frac{nb\sum_{j}^{t}\beta_j^2}{(t-1)}$
Blocks	(b-1)	MS_B	$\sigma_{r'}^2 + nt\sigma_a^2$
TxB Interaction	(t-1)(b-1)	MS_{TB}	$\sigma_{r'}^2 + n\sigma_{a\beta}^2$
Residual (within cell)	tb(n-1)	$MS_{R'}$	$\sigma_{r'}^2$

$$\sigma_{r'}^2 = \sigma_y^2[1 - \rho^2(\bar{\sigma}_x^2/\sigma_x^2)]$$

where Y = dependent variable
X = blocking variable
$\bar{\sigma}_x^2$ = average within block variance of X
ρ = population correlation of X and Y

than the expected mean squares for either block-treatment interaction or within-cell residual.

Tables 4 and 5 present empirical evidence of the inadequacy of the post hoc blocking within treatments method for the case of t = 4, b = 4, n = 5, and ρ = 0.0, 0.2, 0.4, 0.6, 0.8. In Table 4, the one-factor average mean squares for treatments and residual were approximately equal as ρ varied from 0.0 to 0.8, as did the PHB/T average treatments mean square. The average mean squares for interaction and residual were approximately equal and decreased as ρ increased. Clearly, neither mean square provided an unbiased estimate of the treatments mean square. In Table 5, the observed Type-I error rates for the one-factor F ratio were close to expected error rates under the null hypothesis for all values of ρ. On the other hand, observed error rates for the PHB/T rations MS_T/MS_R, and MS_T/MS_{TB} became increasingly liberal with increases in ρ. If a strong relation exists between the dependent and blocking variables and the post hoc blocking within treatments method is used to test the hypothesis of null treatment effects, the actual probability of falsely rejecting the null hypothesis is much greater than the nominal level in the F tables.

Table 4. Empirical Average Mean Squares for the One-Factor Design and the Post Hoc Blocking Within Treatments Method: Additive Case, $\phi = 0.0$, $t = 4$, $b = 4$, $n = 5$, $\rho = 0.0 - 0.8$

Source of Variation	ρ				
	0.0	0.2	0.4	0.6	0.8
One-Factor Design					
Treatments	1.001	.998	.996	.991	1.027
Residual	.999	1.003	1.000	1.003	1.000
Post Hoc Blocking Within Treatments					
Treatments	1.001	.998	.996	.991	1.027
Blocks	1.017	1.891	4.343	8.213	14.055
TxB Interaction	1.008	.968	.872	.725	.500
Residual	.994	.966	.861	.705	.458

Table 5. Observed Type-I Error Rates for the One-Factor Design and the Post Hoc Blocking Within Treatments Method: Additive Case, $\phi = 0.0$, $t = 4$, $b = 4$, $n = 5$, $\rho = 0.0 - 0.8$

Design and Ratio	ρ	Expected Rate (1000α)					
		750	500	250	100	50	10
One-Factor Design							
	0.0	742	493	245	100	58	16
	0.2	743	500	238	92	45	9
MS_T/MS_R	0.4	740	491	258	107	56	11
	0.6	748	496	242	89	49	15
	0.8	750	500	269	108	62	14
	Median	743	496	245	100	56	14
Post Hoc Blocking Within Treatments							
	0.0	746	489	240	100	55	15
	0.2	749	517	260	103	52	9
$MS_T/MS_{R'}$	0.4	784	550	310	146	89	20
	0.6	830	649	387	214	130	39
	0.8	897	775	586	405	311	156
	Median	784	550	310	146	89	20
	0.0	751	483	246	104	51	6
	0.2	753	516	256	112	49	10
MS_T/MS_{TB}	0.4	771	553	303	126	78	19
	0.6	826	619	365	168	100	26
	0.8	886	755	522	323	211	67
	Median	771	553	303	126	78	19

The Hypothesis of Null Treatment Effects

For the remaining two types of post hoc block formation, two kinds of evidence were used to determine the appropriate variance ratios for testing the hypothesis of null treatment effects: average mean squares and Type-I error rates of empirical F distributions in both additive and non-additive conditions. Both the fixed- and sampled-range post hoc blocking (FRPB, SRPB) methods are special cases of the generalized randomized blocks design with fixed treatments. The appropriate test statistic for the design is, at least in part, a function of whether blocks are a fixed, random or finite factor. When all assumptions, including the assumption of block-treatment additivity, are met, the treatment, residual, and interaction mean squares have identical expected values and both MS_T/MS_R and MS_T/MS_{TB} are distributed as central F variates under the null treatments hypothesis.

If, however, the assumption of block-treatment additivity is violated, the expected values of the mean squares vary with the nature of the blocking factor. For fixed blocks, the interaction mean square contains an interaction component and only the ratio MS_T/MS_R follows the central F distribution under the null hypothesis. When blocks are random, the treatment and interaction mean squares contain identical interaction components and only the ratio MS_T/MS_{TB} follows a central F distribution under

the null hypothesis. If the blocks factor is finite, none of the expected mean squares have identical values. The residual mean square contains no interaction component, the treatments mean square has an interaction component, reduced by the factor (1 - b/B), while the interaction mean square has an unreduced interaction component. In this case, neither MS_T/MS_{TB} nor MS_T/MS_R follow the central F distribution under the null hypothesis and no direct test of the type considered here exists.

Empirical Average Mean Squares

Tables 6 - 9 present empirical average mean squares for the one-factor design, the fixed-range post hoc blocking method, and the sampled-range post hoc blocking method in the additive condition. The tables also contain average mean squares for noncentral treatments and the average within block variance of the covariable (X), as well as the average harmonic means and the number of additional iterations required to generate 1000 samples with non-zero cell frequencies. These variables are considered in later sections. Table 6 contains the cases of t = 2 and 4, b = 2, n = 5, and ρ = 0.0 and 0.6. Tables 7 and 8 contain the cases of two and four treatments, respectively, with b = 4, n = 5, and ρ = 0.0 - 0.8 in steps of 0.2. Table 9 contains the cases of t = 4, b = 4, n = 3 and 10, and ρ = 0.0 and 0.6.

Table 6. Empirical, Mean Squares for the One-Factor Design and the Fixed- and Sampled-Range Post Hoc Blocking Methods, Additive Condition: t = 2, 4; b = 2; n = 5; ρ = 0.0, 0.6

Source of Variation	Average Mean Square			
	t = 2 b = 2 n = 5		t = 4 b = 2 n = 5	
	ρ = 0.0	ρ = 0.6	ρ = 0.0	ρ = 0.6
One-Factor Design				
Treatments (T)	.962	1.007	.997	1.006
T: Non-central	1.962	2.007	1.997	2.006
Residual (R)	.997	1.010	1.009	.998
Fixed-Range PHB Method				
Treatments (T)	.962	.786	.997	.772
T: Non-central	1.854	1.681	1.880	1.653
Blocks (B)	1.030	4.986	.981	8.813
TxB Interaction	1.050	.762	.979	.759
Residual (R)	.991	.774	1.012	.770
X-Residual	.360	.356	.358	.362
Harmonic Mean	4.458	4.478	4.416	4.406
Extra Iterations*	5	3	6	10
Sampled-Range PHB Method				
Treatments (T)	.963	.769	1.000	.781
T: Non-central	1.913	1.719	1.913	1.693
Blocks (B)	1.067	4.856	.954	8.830
TxB Interaction	1.025	.757	.983	.771
Residual (R)	.991	.786	1.012	.771
X-Residual	.379	.374	.368	.369
Harmonic Mean	4.749	4.748	4.567	4.561
Extra Iterations*	0	0	1	2

*Additional iterations required to generate 1000 cases with non-zero cell frequencies.

Table 7. Empirical Mean Squares for the One-Factor Design and the Fixed- and Sampled-Range Post Hoc Blocking Methods, Additive Condition: $t = 2$; $b = 4$; $n = 5$; $\rho = 0.0 - 0.8$

Source of Variation	Average Mean Square				
	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
One-Factor Design					
Treatments (T)	.907	.981	1.001	1.010	1.022
T: Non-central	1.907	1.981	2.001	2.010	2.022
Residual (R)	1.003	1.002	1.002	.999	1.001
Fixed-Range PHB Method					
Treatments (T)	.958	.938	.850	.684	.459
T: Non-central	1.850	-	-	1.473	-
Blocks (B)	1.038	1.384	2.448	4.099	6.522
TxB Interaction	1.027	.986	.872	.676	.437
Residual (R)	.999	.963	.856	.693	.450
X-Residual	.141	.135	.139	.137	.136
Harmonic Mean	4.167	4.157	4.130	4.183	4.139
Extra Iterations*	27	19	26	21	21
Sampled-Range PHB Method					
Treatments (T)	.915	.916	.860	.677	.471
T: Non-central	1.869	-	-	1.539	-
Blocks (B)	1.000	1.380	2.560	4.330	6.911
TxB Interaction	1.026	.987	.871	.691	.446
Residual (R)	1.003	.965	.858	.693	.454
X-Residual	.148	.145	.147	.146	.146
Harmonic Mean	4.561	4.573	4.570	4.551	4.554
Extra Iterations*	1	4	0	4	0

*Additional iterations required to generate 1000 cases with non-zero cell frequencies.

Table 8. Empirical Mean Squares for the One-Factor Design and the Fixed- and Sampled-Range Post Hoc Blocking Methods, Additive Condition: $t = 4$; $b = 4$; $n = 5$; $\rho = 0.0 - 0.8$

	Average Mean Square				
SOURCE	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
One-Factor Design					
Treatments (T)	1.001	1.003	.998	.991	1.027
T: Non-central	2.001	2.003	1.998	1.991	2.027
Residual (R)	.997	.988	1.000	1.003	1.000
Fixed-Range PHB Method					
Treatments (T)	1.003	.955	.871	.700	.447
T: Non-central	1.834	1.809	1.697	1.569	1.229
Blocks (B)	1.030	1.743	3.930	7.420	12.548
TxB Interaction	1.012	.975	.852	.706	.448
Residual (R)	.999	.962	.857	.694	.444
X-Residual	.137	.136	.157	.139	.137
Harmonic Mean	4.125	4.108	4.125	4.124	4.129
Extra Iterations*	35	49	35	66	51
Sampled-Range PHB Method					
Treatments (T)	1.012	.950	.864	.692	.450
T: Non-central	1.853	1.876	1.726	1.587	1.281
Blocks (B)	1.033	1.783	4.003	7.611	13.059
TxB Interaction	1.024	.972	.870	.710	.449
Residual (R)	.992	.965	.859	.699	.446
X-Residual	.141	.139	.137	.142	.141
Harmonic Mean	4.326	4.328	4.326	4.326	4.338
Extra Iterations*	16	17	16	21	18

*Additional Iterations required to generate 1000 cases with non-zero cell frequencies.

Table 9. Empirical Mean Squares for the One-Factor Design and the Fixed- and Sampled-Range Post Hoc Blocking Methods, Additive Condition: t = 4; b = 4; n = 3, 10; ρ = 0.0, 0.6

Source of Variation	Average Mean Square			
	t = 4 b = 4 n = 3		t = 4 b = 4 n = 10	
	ρ = 0.0	ρ = 0.6	ρ = 0.0	ρ = 0.6
One-Factor Design				
Treatments (T)	.976	.964	.987	.978
T: Non-central	-	-	1.987	1.978
Residual (R)	1.007	1.002	1.002	1.001
Fixed-Range PHB Method				
Treatments (T)	.996	.715	.976	.706
T: Non-central	-	-	1.883	1.618
Blocks (B)	.996	4.556	1.015	15.699
TxB Interaction	1.012	.698	1.006	.688
Residual (R)	1.005	.701	1.001	.692
X-Residual	.137	.139	.137	.139
Harmonic Mean	2.336	2.321	9.140	9.145
Extra Iterations*	730	744	0	0
Sampled-Range PHB Method				
Treatments (T)	1.002	.705	.980	.704
T: Non-central	-	-	1.914	1.631
Blocks (B)	.981	4.583	1.017	16.067
TxB Interaction	1.019	.708	1.003	.688
Residual (R)	1.001	.689	1.001	.692
X-Residual	.147	.148	.139	.140
Harmonic Mean	2.441	2.434	9.369	9.376
Extra Iterations*	302	346	0	0

*Additional iterations required to generate 1000 cases with non-zero cell frequencies.

Tables 10 - 12 present empirical average mean squares for the one-factor design and the FRPB and SRPB methods for a smaller set of non-additive cases, with the degree of non-additivity a function of ρ, as described in Chapter II. Table 10 contains the cases of t = 4, b = 2, n = 5, and ρ = 0.6. Tables 11 and 12 contain the cases of t = 2 and 4, respectively, with b = 4, n = 5, and ρ = 0.2 - 0.8 in steps of 0.2.

Inspection of Tables 6 - 12 suggests that for the one-factor design, the average treatments and residual mean squares were approximately equal for all combinations of t, b, n, and ρ in both the additive and non-additive conditions. As summarized in Table 13, the ratio of average MS_T/MS_R ranged from 0.90 to 1.03 with a median of 0.99 in the additive cases and from 0.90 to 1.05 with a median of 1.00 in the non-additive cases. None of the design parameters should affect the expected mean squares of the one-factor design and none did in the empirical cases generated.

The average blocks mean square for both the FRPB and SRPB methods increased with t, b, n, and ρ for $\rho \geq 0.2$ in the additive condition and was approximately equal to the average residual mean square in all non-additive cases, as well as for those additive cases with ρ = 0.0. The results were as expected, since the expected mean square for blocks is $\sigma_e^2 + \tilde{n}t\sigma_a^2$, as shown in Table 2, with σ_a^2 a function of both b and ρ in the additive case

Table 10. Empirical Mean Squares Under Conditions of Non-Additivity for the One-Factor Design and the Fixed- and Sampled-Range Post Hoc Blocking Methods: $t = 4$; $b = 2$; $n = 5$; $\rho = 0.6$; $\phi = 0.0$

Source of Variation	Average Mean Square
One-Factor Design	
Treatments (T)	1.021
Residual (R)	1.000
Fixed-Range PHB Method	
Treatments (T)	.770
Blocks (B)	.778
TxB Interaction	3.452
Residual (R)	.771
X-Residual	.362
Extra Iterations*	10
Sampled-Range PHB Method	
Treatments (T)	.854
Blocks (B)	.775
TxB Interaction	3.421
Residual (R)	.776
X-Residual	.369
Extra Iterations*	2

*Additional iterations required to generate 1000 cases with non-zero cell frequencies.

Table 11. Empirical Mean Squares Under Conditions of Non-Additivity for the One-Factor Design and the Fixed- and Sampled-Range Post Hoc Blocking Methods: $t = 2$; $b = 4$; $n = 5$; $\rho = 0.2 - 0.8$; $\phi = 0.0$

Source of Variation	Average Mean Square			
	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
One-Factor Design				
Treatments (T)	.970	1.039	1.010	.911
Residual (R)	.998	.991	1.010	1.010
Fixed-Range PHB Method				
Treatments (T)	.941	.876	.697	.419
Blocks (B)	1.003	.887	.723	.445
TxB Interaction	1.330	2.335	4.191	6.612
Residual (R)	.963	.853	.694	.448
X-Residual	.135	.139	.137	.135
Extra Iterations*	24	26	21	21
Sampled-Range PHB Method				
Treatments (T)	.957	.998	.984	.879
Blocks (B)	.980	.865	.727	.459
TxB Interaction	1.363	2.456	4.413	7.022
Residual (R)	.963	.857	.694	.454
X-Residual	.145	.147	.146	.146
Extra Iterations*	2	0	4	0

*Additional iterations required to generate 1000 cases with non-zero cell frequencies.

Table 12. Empirical Mean Squares Under Conditions of Non-Additivity for the One-Factor Design and the Fixed- and Sampled-Range Post Hoc Blocking Methods: t = 4; b = 4; n = 5; ρ = 0.2 - 0.8 ϕ = 0.0

Source of Variation	Average Mean Square			
	ρ = 0.2	ρ = 0.4	ρ = 0.6	ρ = 0.8
One-Factor Design				
Treatments (T)	1.001	1.006	1.027	.996
Residual (R)	1.002	.999	1.014	1.007
Fixed-Range PHB Method				
Treatments (T)	.957	.867	.703	.434
Blocks (B)	.957	.860	.730	.436
TxB Interaction	1.237	1.841	2.985	4.540
Residual (R)	.962	.863	.698	.447
X-Residual	.136	.136	.139	.137
Extra Iterations*	49	49	66	51
Sampled-Range PHB Method				
Treatments (T)	.959	.913	.786	.601
Blocks (B)	.968	.879	.727	.447
TxB Interaction	1.247	1.891	3.094	4.679
Residual (R)	.965	.862	.700	.450
X-Residual	.139	.139	.142	.141
Extra Iterations*	17	17	21	18

*Additional iterations required to generate 1000 cases with non-zero cell frequencies.

and σ_a^2 arbitrarily 0.0 in the empirical non-additive cases.

The average treatments, interaction, and residual mean squares were approximately equal in the additive condition for both the FRPB and SRPB methods, decreasing with increases in t, b, n, and ρ for $\rho \geq 0.2$ (as is necessary if precision is to be gained by the post hoc blocking methods) and unaffected by t, b, or n for $\rho = 0.0$. As shown in Table 13, the ratios of average mean squares, MS_T/MS_R and MS_T/MS_{TB}, had median values of approximately 1.0 for both methods, while the ratio MS_T/MS_{TB} had slightly wider ranges for both methods. The greater range of the latter ratio was probably due to the fact that MS_{TB} was always estimated with fewer degrees of freedom than MS_R.

Results are less readily summarized for the fixed- and sampled-range methods in the non-additive condition. Comparison of Tables 6 and 10, 7 and 11, and 8 and 12 shows that the average residual mean squares were unaffected by the introduction of non-additivity. The same tables show that the average interaction mean square increased with ρ across all combinations of b and t.

For the FRPB method, the average treatments mean square was not sensitive to non-additivity. The ratio of average mean squares MS_T/MS_R had a median of 1.0 and a range of 0.09 (see Table 13), essentially the same as in the additive condition. The ratio of average mean squares MS_T/MS_{TB} had a median of 0.24 and a range of

Table 13. Medians and Ranges of the Ratios of Average Mean Squares in the Additive and Non-Additive Conditions: One-Factor Design and the Fixed- and Sampled-Range Post Hoc Blocking Methods, $\phi = 0.0$

Ratio	Additive Case		Non-Additive Case ($\rho \geq 0.2$)	
	Median	Range	Median	Range
One-Factor Design				
MS_T/MS_R	.99	.90 - 1.03	1.00	.90 - 1.05
Fixed-Range PHB Method				
MS_T/MS_R	1.00	.96 - 1.02	1.00	.94 - 1.03
MS_T/MS_{TB}	1.00	.92 - 1.05	.24	.06 - .77
Sampled-Range PHB Method				
MS_T/MS_R	.99	.91 - 1.02	1.10	.99 - 1.94
MS_T/MS_{TB}	.99	.89 - 1.02	.25	.12 - .77

Additive cases computed from Tables 6 - 9.
Non-Additive cases computed from Tables 10 - 12.

0.71, with most of the variability due to ρ rather than variations of b and t at each level of ρ. These results, while not definitive, suggest that the residual mean square is the appropriate error term for testing the null-treatment effects hypothesis.

For the SRPB method, the average treatments mean square was highly variable across combinations of ρ, t, and b. The ratio of average mean squares MS_T/MS_R had a median of 1.1 and a range of 0.95, nearly nine times as variable as the additive case. The average mean squares ratio MS_T/MS_{TB} had a median of 0.25 and a range of 0.65. These results, when compared to the additive condition results, suggest that MS_T contained some fractional component due to block-treatment interaction, as would be the case if blocks represented a finite factor.

Empirical F Distributions

While the average values of mean squares are of some theoretical interest, for practical purposes the degree of fit between empirical and reference distributions is far more important. In particular, if a ratio of mean squares has empirical frequencies close to theoretical frequencies at the usual α levels (0.10, 0.05, 0.01) in both the additive and non-additive conditions, then that ratio and the design for which it was computed is a candidate for actual use. If it shows high power,

relative to logical alternative designs, then it should be put to use.

Table 14 presents empirical Type-I error rates of the central F ratio MS_T/MS_R for the one-factor design, with the medians and ranges of empirical frequencies at the bottom of the table. These distributions were included because the one-factor design is the basic design which the post hoc blocking methods modify and because it provided both a check on the accuracy of data generation and some indication of the amount of variability to be expected in the empirical distributions. The median error rates tended to be slightly conservative, i.e., with only one exception the median empirical error rates were less than the theoretical error rates of the reference distributions. The range of error rates varied from sixty-eight at the fiftieth percentile to thirteen at the ninety-ninth percentile, with no apparent trends across combinations of t, b, n, or ρ. The variability and lack of exact fit between empirical and theoretical distributions was most likely because the empirical distributions were based on a finite number of samples, while the theoretical distributions assume an infinite population of samples.

Empirical Type-I error rates under the null hypothesis for the central F ratio MS_T/MS_R, additive condition, are presented in Tables 15 and 16 for the FRPB and SRPB methods, respectively. For both methods, median observed

Table 14. Observed Type-I Error Rates for the Ratio MS_T/MS_R: One-Factor Design, Additive Condition, $\phi = 0.0$

					Expected Rate (1000α)					
ρ	t	N*	df_1	df_2	750	500	250	100	50	10
0.0	2	10	1	18	740	491	238	90	38	10
	2	20	1	38	735	451	227	89	43	7
	4	10	3	36	752	519	248	97	41	8
	4	12	3	44	733	500	241	83	33	3
	4	20	3	76	742	493	245	100	58	16
	4	40	3	156	765	494	248	101	41	5
0.2	2	20	1	38	734	491	253	82	46	9
	4	20	3	76	743	500	238	92	45	9
0.4	2	20	1	38	754	473	234	101	56	12
	4	20	3	76	740	491	258	107	56	11
0.6	2	10	1	18	749	509	265	97	51	10
	2	20	1	38	765	491	250	97	53	10
	4	10	3	36	750	486	235	100	54	13
	4	12	3	44	744	481	244	90	44	6
	4	20	3	76	748	496	242	89	49	15
	4	40	3	156	735	496	243	91	43	11
0.8	2	20	1	38	736	484	249	103	51	11
	4	20	3	76	750	500	269	108	62	14
	Median				743	492	245	97	48	11
	Range				32	68	42	26	29	13

*N = bxn in the Post Hoc Blocking Methods.

Table 15. Observed Type-I Error Rates for the Ratio MS_T/MS_R: Fixed-Range Post Hoc Blocking Method, Additive Condition, $\phi = 0.0$

						Expected Rate (1000α)					
ρ	t	b	n	df_1	df_2	750	500	250	100	50	10
0.0	2	2	5	1	16	725	478	236	97	50	6
	2	4	5	1	32	749	484	247	90	46	10
	4	2	5	3	32	739	509	242	99	38	9
	4	4	3	3	32	748	481	255	97	45	9
	4	4	5	3	64	736	512	258	108	53	12
	4	4	10	3	144	764	488	253	88	37	9
0.2	2	4	5	1	32	752	502	239	90	42	7
	4	4	5	3	64	758	487	233	107	58	8
0.4	2	4	5	1	32	755	501	237	98	56	6
	4	4	5	3	64	739	483	239	111	58	15
0.6	2	2	5	1	16	760	500	256	109	54	8
	2	4	5	1	32	750	505	248	93	55	15
	4	2	5	3	32	747	497	246	99	47	12
	4	4	3	3	32	771	493	262	109	60	16
	4	4	5	3	64	740	495	259	95	61	18
	4	4	10	3	144	758	519	263	100	49	10
0.8	2	4	5	1	32	751	508	262	100	55	14
	4	4	5	3	64	744	494	245	105	59	13
	Median					750	496	248	100	54	10
	Range					32	58	42	26	29	13

Table 16. Observed Type-I Error Rates for the Ratio MS_T/MS_R: Sampled-Range Post Hoc Blocking Method, Additive Condition, $\phi = 0.0$

						Expected Rate (1000α)					
ρ	t	b	n	df_1	df_2	750	500	250	100	50	10
0.0	2	2	5	1	16	741	487	235	92	37	8
	2	4	5	1	32	744	484	236	87	40	5
	4	2	5	3	32	746	495	256	100	42	8
	4	4	3	3	32	769	496	273	86	36	9
	4	4	5	3	64	743	510	253	111	58	15
	4	4	10	3	144	773	495	251	83	43	5
0.2	2	4	5	1	32	751	500	242	86	41	7
	4	4	5	3	64	744	492	244	104	50	6
0.4	2	4	5	1	32	731	481	252	99	56	12
	4	4	5	3	64	750	489	239	109	57	18
0.6	2	2	5	1	16	738	498	250	103	54	10
	2	4	5	1	32	745	519	241	89	51	14
	4	2	5	3	32	752	495	255	108	53	11
	4	4	3	3	32	762	513	259	111	57	18
	4	4	5	3	64	739	507	257	97	50	12
	4	4	10	3	144	761	505	258	112	43	7
0.8	2	4	5	1	32	761	519	241	109	57	12
	4	4	5	3	64	762	507	268	113	48	6
	Median					750	497	252	102	50	10
	Range					38	38	38	30	22	13

error rates fluctuated around the expected error rates with variability comparable to that observed for the one-factor design reported in Table 14. The variability in error rates revealed no apparent trends across combinations of t, b, n, or ρ.

Empirical Type-I error rates for the central F ratio MS_T/MS_{TB}, additive condition, are presented in Tables 17 and 18 for the FRPB and SRPB methods, respectively. The median observed error rates for both methods were close to expected error rates for both methods, with the fixed-range method's error rates slightly more conservative for $\alpha \leq 0.25$. Variability in error rates was comparable to the variability for the one-factor design in Table 14. As was the case for Tables 14 - 16, there was no apparent trend associated with changes in t, b, n, or ρ.

The observed Type-I error rates in the null treatments, additive condition were consistent with the empirical average mean squares in Tables 6 - 9. Both of the ratios MS_T/MS_R and MS_T/MS_{TB} had empirical α levels close to nominal α levels for both the fixed- and sampled-range post hoc blocking methods. A clear choice of post hoc blocking method and test statistic depends on the effects of violating the assumption of non-additivity and the precision of the method.

The importance of including non-additive cases can be clearly illustrated by the example of Pingel's (1968)

Table 17. Observed Type-I Error Rates for the Ratio MS_T/MS_{TB}: Fixed-Range Post Hoc Blocking Method, Additive Condition, $\phi = 0.0$

						Expected Rate (1000α)					
ρ	t	b	n	df_1	df_2	750	500	250	100	50	10
0.0	2	2	5	1	1	718	482	237	85	41	4
	2	4	5	1	3	743	488	234	88	51	11
	4	2	5	3	3	758	494	243	89	44	13
	4	4	3	3	9	752	502	248	90	44	9
	4	4	5	3	9	743	508	241	96	50	8
	4	4	10	3	9	766	493	241	93	47	7
0.2	2	4	5	1	3	777	510	219	74	31	6
	4	4	5	3	9	757	503	244	89	38	7
0.4	2	4	5	1	3	755	491	243	97	43	7
	4	4	5	3	9	754	510	257	99	54	9
0.6	2	2	5	1	1	758	500	247	96	42	6
	2	4	5	1	3	771	506	232	84	42	11
	4	2	5	3	3	775	523	247	100	57	9
	4	4	3	3	9	780	515	257	90	42	7
	4	4	5	3	9	743	500	243	83	38	6
	4	4	10	3	9	776	513	270	112	48	7
0.8	2	4	5	1	3	764	523	271	119	61	10
	4	4	5	3	9	751	498	254	105	50	8
	Median					758	502	244	92	44	8
	Range					38	41	51	45	30	9

Table 18. Observed Type-I Error Rates for the Ratio MS_T/MS_{TB}: Sampled-Range Post Hoc Blocking Method, Additive Condition, $\phi = 0.0$

						Expected Rate (1000α)					
ρ	t	b	n	df_1	df_2	750	500	250	100	50	10
0.0	2	2	5	1	1	734	476	226	78	34	11
	2	4	5	1	3	738	464	230	91	54	14
	4	2	5	3	3	754	491	252	102	55	10
	4	4	3	3	9	772	517	246	90	42	8
	4	4	5	3	9	730	496	258	100	48	10
	4	4	10	3	9	776	486	258	94	45	8
0.2	2	4	5	1	3	747	497	229	78	24	5
	4	4	5	3	9	758	499	225	91	41	15
0.4	2	4	5	1	3	724	489	236	99	45	12
	4	4	5	3	9	740	494	239	95	52	13
0.6	2	2	5	1	1	746	502	246	95	46	9
	2	4	5	1	3	751	505	256	97	51	11
	4	2	5	3	3	771	517	249	93	46	12
	4	4	3	3	9	756	516	246	86	49	6
	4	4	5	3	9	737	477	265	96	49	13
	4	4	10	3	9	758	516	263	114	55	10
0.8	2	4	5	1	3	772	517	269	123	70	13
	4	4	5	3	9	757	505	247	99	55	8
	Median					752	498	246	96	49	10
	Range					52	53	44	36	46	10

investigation of a priori block formation methods. He considered only the case of block-treatment additivity in his Monte Carlo study and concluded that the sampled-range method had a random blocks factor and thus the appropriate test of the null treatment effects hypothesis in the ratio of mean squares for treatments and block-treatment interaction, MS_T/MS_{TB}. However, the tables in Appendix B show that under conditions of severe non-additivity the average mean square for treatments was much less than the average mean square for interaction and the ratio MS_T/MS_{TB} was extremely conservative. The ratio MS_T/MS_R provided a better, but slightly liberal fit to the theoretical F distribution.

Empirical Type-I error rates for the ratio MS_T/MS_R under conditions of non-additivity are presented in Table 19 for the fixed- and sampled-range post hoc blocking methods. For the FRPB method, observed and empirical error rates were in close agreement over all combinations of ρ, t, and b. The median observed error rates were close to theoretical error rates and the ranges were comparable to those of the additive cases.

Results for the SRPB method were quite different. For that method, the observed error rates became increasingly liberal as ρ varied from 0.2 to 0.8. Error rates tended to be less liberal for t = 4 than for t = 2, which was due more to the larger pooled sample sizes and the method of generating non-additive cases than the change

Table 19. Observed Type-I Error Rates for the Ratio MS_T/MS_R: Fixed- and Sampled-Range Post Hoc Blocking Methods, Non-Additive Condition, $\phi = 0.0$

						Fixed-Range Post Hoc Blocking Method Expected Rate (1000α)				Sampled-Range Post Hoc Blocking Method Expected Rate (1000α)			
ρ	t	b	n	df_1	df_2	250	100	50	10	250	100	50	10
0.2	2	4	5	1	32	243	91	40	6	242	88	39	6
	4	4	5	3	64	260	91	49	11	261	97	50	12
	Median					252	91	44	8	252	92	44	9
0.4	2	4	5	1	32	248	89	51	11	258	123	75	15
	4	4	5	3	64	230	86	50	11	260	110	55	15
	Median					239	88	50	11	259	116	65	15
0.6	4	2	5	3	32	248	100	51	13	275	125	68	19
	2	4	5	1	32	231	104	50	16	304	158	97	32
	4	4	5	3	64	261	120	60	18	319	142	72	16
	Median					248	104	51	16	304	142	72	16
0.8	2	4	5	1	32	252	101	50	9	448	271	187	71
	4	4	5	3	64	243	104	45	8	358	186	113	36
	Median					248	102	48	8	403	228	150	54
Median						248	100	50	11	275	125	75	16
Range						30	39	20	12	206	193	148	65

in the number of treatment groups. Median observed error rates were clearly greater than the theoretical error rates and were far more variable than any of the cases in Tables 14 - 18. Under conditions of extreme non-additivity ($\rho \geq 0.4$), actual α levels were higher than the nominal α levels of the theoretical F distribution.

Table 20 contains non-additive condition, empirical Type-I error rates for the ratio MS_T/MS_{TB} for both the fixed- and sampled-range post hoc blocking methods. For both methods, the observed error rates were conservative with $\rho = 0.2$ and rapidly approached zero as ρ increased. The SRPB error rates tended to be slightly less conservative than those of the FRPB method but not to any appreciable degree. For all cases generated with $\rho > 0$, the SRPB ratio MS_T/MS_{TB} was more conservative than the corresponding ratio MS_T/MS_R was liberal.

The empirical average mean squares and the empirical F distributions support the conclusion that blocks represent a fixed effect in the fixed-range post hoc blocking method and that the ratio MS_T/MS_R has a distribution reasonably close, if not identical to, the central F distribution, even under conditions of extreme non-additivity.

Blocks in the sampled-range method, however, appeared to behave like a finite factor. The factor is clearly not finite in the usual sense of a random sample from a finite population, but rather each sample of blocks

Table 20. Observed Type-I Error Rates for the Ratio MS_T/MS_{TB}: Fixed- and Sampled-Range Post Hoc Blocking Methods, Non-Additive Condition, $\phi = 0.0$

						Fixed-Range Post Hoc Blocking Method Expected Rate (1000α)				Sampled-Range Post Hoc Blocking Method Expected Rate (1000α)			
ρ	t	b	n	df_1	df_2	250	100	50	10	250	100	50	10
0.2	2	4	5	1	3	159	51	30	7	156	51	28	4
	4	4	5	3	9	173	47	18	1	197	54	14	1
	Median					166	49	24	4	178	52	21	2
0.4	2	4	5	1	3	64	7	7	0	74	18	8	4
	4	4	5	3	9	42	7	2	0	45	9	2	0
	Median					53	7	4	0	60	14	5	2
0.6	4	2	5	3	3	12	2	1	0	15	2	2	2
	2	4	5	1	3	8	1	0	0	17	1	0	0
	4	4	5	3	9	2	1	0	0	4	0	0	0
	Median					8	1	0	0	15	1	0	0
0.8	2	4	5	1	3	1	0	0	0	2	0	0	0
	4	4	5	3	9	1	0	0	0	0	0	0	0
	Median					1	0	0	0	1	0	0	0
Median						12	2	1	0	17	2	2	1
Range						172	51	30	7	180	54	28	4

represents a slightly different set of b blocks, since block endpoints are defined by the b-1 quantiles of the pooled X observations. In small samples the block endpoints are relatively variable and the blocking variable behaves as though the population of blocks is larger than b. As sample sizes increase, the pooled sample quantiles become less variable and the blocking variable behaves as though the size of the block population approaches b. In very large samples the blocks would be fixed, for all practical purposes, and results should be similar to those of the fixed-range post hoc blocking method.

For the relatively small sample sizes in this investigation, as the degree of non-additivity increased, the average treatments mean square became larger than the average residual mean square, yet smaller than the average interaction mean square. Similarly the ratio MS_T/MS_R became increasingly liberal as the interaction increased, while the ratio MS_T/MS_{TB} became extremely conservative. Both ratios appeared to have central F distributions under the null hypothesis in the additive condition.

Block-Treatment Non-Additivity

Block-treatment additivity has been treated as one of the assumptions of the model and previous discussion has focused on what effect the violation of that assumption

had on the tests of the null treatments hypothesis. If, however, the primary motivation for post hoc blocking is to investigate non-additivity or as suggested earlier, there is a desire to use the sampled-range method when non-additivity is suspected, then additivity of block and treatment effects should be considered as an hypothesis to be tested, rather than an assumption which may have been violated. The remainder of this section refers to the null hypothesis of block-treatment additivity as simply the null hypothesis unless it is unclear which null hypothesis is involved. The two conditions in the empirical investigation were additive block and treatment effects, null treatments and non-additive block and treatment effects, null treatments.

Table 21 contains empirical Type-I error rates for the ratio MS_{TB}/MS_R of the fixed- and sampled-range post hoc blocking methods under the null hypothesis. The median observed error rates were close to theoretical values for both methods. The variability in observed error rates was consistent with that of the additive cases in Tables 14 - 18 and there were no consistent trends across variations in t, b, or ρ. Observed and nominal α levels were in close agreement when testing the additivity hypothesis with the ratio MS_{TB}/MS_R.

Table 22 contains the observed power of the ratio MS_{TB}/MS_R for the FRPB and SRPB methods under conditions of non-additivity. For both methods, power increased

Table 21. Observed Type-I Error Rates for the Ratio MS_{TB}/MS_R: Fixed- and Sampled-Range Post Hoc Blocking Methods, Additive Condition, $\phi = 0.0$

						Fixed-Range Method Expected Rate (1000α)				Sampled-Range Method Expected Rate (1000α)			
ρ	t	b	n	df_1	df_2	250	100	50	10	250	100	50	10
0.0	4	2	5	3	32	252	96	45	7	251	93	52	11
	2	4	5	3	32	260	112	57	10	272	97	44	8
	4	4	5	9	64	228	104	60	19	265	99	53	14
0.2	2	4	5	3	32	265	110	46	10	273	106	50	10
	4	4	5	9	64	239	94	51	9	240	106	53	9
0.4	2	4	5	3	32	253	117	65	17	252	98	54	12
	4	4	5	9	64	231	104	61	17	248	102	55	12
0.6	4	2	5	3	32	238	113	40	10	241	100	57	15
	2	4	5	3	32	243	91	45	10	256	96	50	11
	4	4	5	9	64	242	117	59	11	273	114	61	16
0.8	2	4	5	3	32	248	104	54	10	256	98	57	12
	4	4	5	9	64	246	103	55	15	264	108	61	12
Median						245	104	54	10	256	100	54	12
Range						37	26	25	12	33	21	17	8

Table 22. Observed Power of the Ratio MS_{TB}/MS_R: Fixed- and Sampled-Range Post Hoc Blocking Methods, Non-Additive Condition, $\phi = 0.0$

						Fixed-Range Method α Level (x1000)				Sampled-Range Method α Level (x1000)			
ρ	t	b	n	df_1	df_2	250	100	50	10	250	100	50	10
0.2	2	4	5	3	32	389	212	137	40	427	229	134	41
	4	4	5	9	64	426	205	129	33	417	224	137	36
	Median					408	208	133	36	422	226	136	38
0.4	2	4	5	3	32	756	550	426	201	784	594	466	216
	4	4	5	9	64	829	641	497	263	831	656	532	279
	Median					792	596	462	232	808	625	499	248
0.6	4	2	5	3	32	926	823	718	459	917	818	719	442
	2	4	5	3	32	982	930	864	654	977	934	870	686
	4	4	5	9	64	998	981	961	880	998	987	972	878
	Median					982	930	864	654	977	934	870	686
0.8	2	4	5	3	32	1000	1000	1000	995	1000	999	997	986
	4	4	5	9	64	1000	1000	1000	1000	1000	1000	1000	1000
	Median					1000	1000	1000	998	1000	1000	998	993
Median						926	823	718	459	917	818	719	442
Range						611	795	871	967	583	776	866	964

rapidly as ρ varied from 0.2 to 0.8, with the SRPB method slightly, but consistently, more powerful. At least for the specific cases investigated, the ratio MS_{TB}/MS_R was quite sensitive to non-additivity. For example, with $\rho = 0.6$ and $\alpha = 0.5$, the median observed power was 0.864 and 0.870 for the FRPB and SRPB methods, respectively. This is the same value of ρ for which the ratio MS_T/MS_R was unreasonably liberal in Table 19. It appears that the ratio MS_{TB}/MS_R provides a reasonably powerful prior test for non-additivity at the point where the non-additivity is contaminating the test for null treatment effects with the ratio MS_T/MS_R.

Precision

Precision, as shown in Chapter I, is one of the central topics in experimental design. For a fixed number of treatments and fixed numbers of experimental units per treatment group, the most precise design provides the most powerful test of the hypothesis of no treatment effects. The generalized index of apparent imprecision I_g, which is used in this investigation is the ratio of the squared non-centrality parameter for an ideal design--minimum residual variance and equal cell frequencies--to the squared non-centrality parameter for the design under consideration, modified to reflect the number of degrees of freedom

available to estimate residual variance. If the particular design of interest has equal cell frequencies, the generalized index is identical to the better-known index of apparent imprecision used by Cox (1957) and Feldt (1958), for example. For both indices, large values indicate a relatively imprecise design; the closer the index approaches the lower limit of 1.0, the more precise the design. Equivalently, the design with the numerically lowest index of imprecision provides the most powerful test of the null hypothesis.

Comparisons of precision between the one-factor design, one-factor analysis of covariance, and the fixed- and sampled-range post hoc blocking methods are most appropriately made within each combination of treatments, blocks, average cell size, and correlation between dependent and concomitant variables. Within each combination, I_g is a function of:

(a) variation in cell frequencies as reflected in $\tilde{n}$, the harmonic mean of the cell frequencies;

(b) the degrees of freedom available for estimating residual variance; and

(c) the magnitude of the residual variance, a function of the concomitant variable's within-block variance. It was expected, from prior research on a priori blocking methods, that within block variance of the concomitant variable would be smaller for the fixed-range method, contributing to greater precision.

Inspection of Tables 4 - 7 reveals that without exception the concomitant variable's residual variance was smaller for the fixed-range method than for the sampled-range method. However, the same tables show that the average harmonic mean of cell frequencies was always less for the fixed-range method, which contributes to lesser precision.

Table 23 presents indices of imprecision for the four methods for correlations of 0.0 and 0.6 and selected combinations of treatments, blocks, and average cell sizes. The table also presents empirical non-central F distributions (power) for the one-factor design and the fixed- and sampled-range methods where blocks and treatments are additive. They were included to provide verification of conclusions drawn from the generalized index of imprecision. For $\rho = 0.0$, the one-factor analysis of variance had the lowest value of I_g, followed by the one-factor analysis of covariance, the sampled-range method, and finally, the fixed-range method. In one case, the sampled-range and analysis of covariance indices exchanged places in the rank ordering. This was probably due to the fact that the indices were quite close in value and the sampled-range index was computed from empirical estimates while the analysis of covariance index was computed directly from the residual degrees of freedom. For $\rho = 0.6$, the one-factor analysis of covariance was most precise (lowest value of I_g), followed by the sampled-range

Table 23. Observed Power of the Ratio MS_T/MS_R and Empirical Indices of Imprecision: Additive Condition; $\phi = 1.0$; $\rho = 0.0, 0.6$

t	b	n	Method	df_1	df_2	$\rho = 0.0$ α Level (x1000) 250	100	50	10	I_g	$\rho = 0.6$ α Level (x1000) 250	100	50	10	I_g
2	2	5	One-Factor	1	18	605	187	96	15	1.053	594	201	87	19	1.727
			FRPB	1	16	530	184	90	14	1.254	680	226	120	18	1.498
			SRPB	1	16	575	181	82	16	1.177	695	236	114	22	1.425
			ANCOVA	1	17	-	-	-	-	1.181	-	-	-	-	1.181
4	2	5	One-Factor	3	36	714	337	169	35	1.054	746	332	168	45	1.647
			FRPB	3	32	655	296	149	28	1.201	789	388	208	47	1.488
			SRPB	3	32	679	313	148	32	1.162	817	401	223	55	1.404
			ANCOVA	3	35	-	-	-	-	1.087	-	-	-	-	1.087
2	4	5	One-Factor	1	38	449	243	151	50	1.052	449	246	141	49	1.643
			FRPB	1	32	422	246	147	40	1.273	476	287	181	52	1.365
			SRPB	1	32	448	240	146	48	1.163	501	292	197	56	1.261
			ANCOVA	1	37	-	-	-	-	1.082	-	-	-	-	1.082
4	4	5	One-Factor	3	76	594	369	249	97	1.026	597	369	255	98	1.603
			FRPB	3	64	551	326	213	75	1.250	655	439	314	124	1.348
			SRPB	3	64	551	325	218	80	1.192	650	449	331	136	1.286
			ANCOVA	3	75	-	-	-	-	1.040	-	-	-	-	1.040
4	4	10	One-Factor	3	156	602	382	265	97	1.013	603	382	263	100	1.582
			FRPB	3	144	558	361	242	85	1.109	687	488	335	157	1.195
			SRPB	3	144	584	369	252	89	1.082	698	486	345	161	1.167
			ANCOVA	3	155	-	-	-	-	1.019	-	-	-	-	1.019

method, then the fixed-range method, and finally, the one-factor analysis of variance. The ordering was consistent across all combinations of t, b, and n.

Table 24 presents the same indices of imprecision and power for five values of ρ(0.0, 0.2, 0.4, 0.6, 0.8), four treatments, four blocks, and an average cell size of five. In both Tables 23 and 24, the analysis of covariance was most precise for $\rho \geq 0.2$ and the sampled-range method was always more precise than the fixed-range method. Neither post hoc method was more precise than the one-factor analysis of variance until ρ reached 0.6 and then both methods were more precise. These results were confirmed by the non-central F distributions. In all cases lower values of I_g were associated with greater power.

The relative precision of the two post hoc blocking methods could be improved by increasing the number of blocks and/or the number of experimental units, but how much is unclear. Simply increasing the number of blocks without also increasing the number of experimental units quickly reaches a point of diminishing returns. Tables 4 - 10 include the number of additional iterations required to generate 1000 cases with non-zero cell frequencies for each design. Consider just the case of four treatments and four blocks: for n = 3, the fixed- and sampled-range methods required over 700 and 300 extra iterations, respectively; for n = 5, they required roughly 50 and 20 extra iterations, respectively; and for n = 10, no extra

Table 24. Observed Power of the Ratio MS_T/MS_R and Empirical Indices of Imprecision: Additive Case, $\phi = 1.0$, $t = 4$, $b = 4$, $n = 5$

ρ	df_1	df_2	Method	α Level (x1000) 250	100	50	10	I_g
	3	76	One-Factor	594	369	249	97	1.026
0.0	3	64	FRPB	551	326	213	75	1.250
	3	64	SRPB	551	325	218	80	1.192
	3	76	One-Factor	596	393	272	98	1.069
0.2	3	64	FRPB	572	335	227	85	1.262
	3	64	SRPB	590	367	245	99	1.198
	3	76	One-Factor	596	359	246	89	1.221
0.4	3	64	FRPB	594	368	238	81	1.282
	3	64	SRPB	602	373	248	92	1.223
	3	76	One-Factor	597	369	255	98	1.603
0.6	3	64	FRPB	655	439	314	124	1.348
	3	64	SRPB	650	449	331	136	1.286
	3	76	One-Factor	586	354	235	90	2.850
0.8	3	64	FRPB	771	588	449	211	1.551
	3	64	SRPB	792	617	459	214	1.486
Analysis of Covariance Index of Imprecision								1.040

FRPB = Fixed-Range Post Hoc Blocking Method
SRPB = Sampled-Range Post Hob Blocking Method
ANCOVA = Analysis of Covariance, One-Factor Design
Data for $\rho = 0.0$ and 0.6 are reproduced from Table 23.

iterations were required. Reducing either the number of blocks or treatments helped but clearly, small average cell sizes is likely to involve empty cells, in which case the post hoc methods are inapplicable.

Comparison to A Priori Block Formation

One final consideration: how much precision is lost by forming blocks after the fact rather than incorporating blocks directly into the experimental design? Pingel (1968) provides the estimates of concomitant variable within-block variance necessary to compute indices of imprecision for some of the specific combinations of design parameters in the present investigation. Indices of imprecision for the fixed- and sampled-range methods, both a priori and post hoc, are presented in Table 25. Indices of imprecision for the one-factor analyses of variance and covariance are repeated from Tables 23 and 24 for ease of comparison. Without exception, the a priori methods had smaller values of I_g. In fact, the a priori blocking methods were more precise than the one-factor analysis of variance for $\rho \geq 0.2$, while the post hoc blocking methods were more precise only for $\rho \geq 0.6$. The analysis of covariance was the most precise of all methods for $\rho \geq 0.4$. Within block variances were virtually identical for the respective a priori and post hoc methods,

Table 25. Generalized Indices of Apparent Imprecision for the Fixed- and Sampled-Range Methods when Used in the Design (A Priori) or Incorporated After Data Collection (Post Hoc), and the One-Factor Analysis of Variance and Covariance

ρ	t	b	n	Fixed-Range		Sampled-Range		ANOVA	ANCOVA
				Post Hoc	A Priori	Post Hoc	A Priori		
0.0	2	2	5	1.254	1.118	1.177	1.118	1.053	1.181
	4	2	5	1.201	1.061	1.162	1.061	1.054	1.087
	2	4	5	1.273	1.061	1.163	1.061	1.052	1.082
	4	4	5	1.250	1.031	1.192	1.031	1.026	1.040
0.2	4	4	5	1.262	1.037	1.198	1.037	1.069	1.040
0.4	4	4	5	1.282	1.058	1.223	1.058	1.221	1.040
0.6	2	2	5	1.498	1.346	1.425	1.359	1.727	1.181
	4	2	5	1.448	1.277	1.404	1.276	1.647	1.087
	2	4	5	1.365	1.144	1.261	1.147	1.643	1.082
	4	4	5	1.348	1.112	1.286	1.112	1.603	1.040
0.8	4	4	5	1.551	1.286	1.486	1.288	2.850	1.040

Indices for A Priori methods calculated from estimates of concomitant variable average within-block variance computed by Pingel (1968).

so the loss of precision was due to the unequal cell frequencies of the post hoc methods. It clearly pays to incorporate the blocks into the experimental design prior to assigning units to treatments or use the analysis of covariance if its assumptions are satisfied.

It should also be noted that the indices of imprecision for the post hoc blocking methods were computed by averaging across all values of the harmonic mean in the emprical sampling distributions. If, in a particular application, the cell frequencies should happen to be equal, the precision for equal cell frequencies would be essentially the same as for the a priori methods. If, on the other hand, the cell frequencies happen to be extremely disparate, then the precision is even less than that reported for this investigation.

CHAPTER IV

SUMMARY AND CONCLUSIONS

The major purpose of the investigation of post hoc block formation methods was to investigate the feasibility of imposing a generalized randomized blocks design on observations from a simple, one-factor design with fixed treatments when additional observations were available on some concomitant variable. The three post hoc blocking methods considered were post hoc blocking within treatments, fixed-range post hoc blocking, and sampled-range post hoc blocking.

For purposes of simplicity and comparability to earlier reseach, attention was restricted to the case of dependent and concomitant variables with homogeneous bivariate normal distributions across treatment populations. For each post hoc blocking method, the major topics were: testing the hypothesis of null treatment effects under conditions of block-treatment additivity; the effect of introducing non-additivity; testing the hypothesis of block-treatment additivity; and the precision of the method in the additive condition, relative to the one-factor analyses of variance and covariance.

The investigation relied heavily on Monte Carlo methods to generate empirical distributions of mean squares and

F ratios where analytic solutions were either unknown or unreasonably difficult. For each combination of treatments (t), blocks (b), average treatment-block cell frequency (n), and population correlation (ρ) between the dependent variable and the concomitant variable, 1000 samples of observations were generated on the dependent variable (Y) and the concomitant variable (X). Within each sample, equal numbers of observations were randomly assigned to treatments and then assigned to blocks within each treatment group according to the three post hoc blocking methods. A one-factor analysis of variance was computed on the unblocked observations, followed by unweighted means analyses for the three post hoc blocking methods. If any of the methods had zero cell frequencies, the sample was dropped for that method and additional samples were later generated to bring the total number of samples to 1000.

Empirical sampling distributions were generated for three separate conditions:

(a) true null treatment effects hypothesis and no block treatment interaction (central treatments, additive condition);

(b) false null treatment effects hypothesis and no block-treatment interaction (non-central treatments, additive condition);

(c) true null treatment effects hypothesis with block-treatment interaction introduced by generating

observations on X and Y such that one-half of the treatment populations had correlations equal to ρ and the remaining half had correlations equal to $-\rho$(central treatments, non-additive condition).

Results of the investigation are summarized separately for each post hoc blocking method.

Post Hoc Blocking Within Treatments

The post hoc blocking within treatments method rank orders observations on X and forms b blocks, all of size n, within each treatment group. The observations on Y are then analyzed by a two-factor analysis of variance. This method has no effect on the variability of treatment means and affords no gains in precision. While the treatments mean square is identical to the treatments mean square for the one-factor design, the residual and interaction mean squares are reduced for $\rho > 0.0$ and neither the ratio MS_T/MS_R, nor the ratio MS_T/MS_{TB} has a central F distribution under the hypothesis of null treatment effects. Even if reference distributions were derived, fewer degrees of freedom would be available and the method would have less power than the one-factor analysis of variance. This method is specifically excluded from future remarks about post hoc blocking methods.

Fixed Range Post Hoc Blocking (FRPB)

The fixed-range post hoc blocking method divides the population distribution of X into b ranges of equal probability. The b-1 quantiles of X determine the assignment of Y observations to blocks within treatment groups. An unweighted means analysis of variance is computed on the Y observations, subject to the restriction that all block-treatment combinations must contain at least one observation for the method to be applicable. Empirical results supported the conclusion that blocks are a fixed factor and under the hypothesis of null treatment effects, the ratio MS_T/MS_R has a central distribution, even under conditions of extreme non-additivity. The FRPB method had greater precision (and power) than the one-factor design when $\rho \geq 0.6$. Since the precision is partly a function of the harmonic mean of cell frequencies, the precision can be as great as for a priori blocking designs when cell freqeuncies happen to be equal or very poor if cell frequencies are extremely disparate. The ratio MS_{TB}/MS_R followed the central F distribution when block and treatment effects were additive and provided a reasonably powerful test for non-additivity, at least for the relatively extreme forms of non-additivity generated in this investigation.

The conclusions about the FRPB method can be generalized

with some caution to clearly fixed blocking variables with properties other than those considered in the empirical investigation. Sex of the respondent or some other nominal variable might be used to form blocks, provided that the usual analysis of variance assumptions are still met. Unless block effects are large or cell frequencies are very nearly equal, the FRPB does not provide a more powerful test for non-central treatments than the one-factor design, but it does provide a means for investigating block-treatment interaction.

Sampled-Range Post Hoc Blocking (SRPB)

The sampled-range post hoc blocking method ranks observations on X across all treatment groups, forming b blocks of size tn each. Blocks are formed within treatment groups according to the b-1 sample quantiles of X and an unweighted means analysis is performed on the Y observations, provided that no block-treatment combinations have zero frequencies. Empirical results indicated that blocks behave like a finite effect where the treatments mean square contains an interaction component, reduced by a factor of $(1-b/B)$ where $_B$ denotes the unknown, but finite, number of blocks in the population. Nonetheless, the ratio MS_T/MS_R had empirical Type-I error rates close to theoretical values for all but extreme cases of non-additivity, while the ratio MS_T/MS_{TB}

was clearly too conservative for all non-additive cases considered. When block and treatment effects were additive, the ratio MS_T/MS_R followed the central F distribution under the null treatment effects hypothesis and was consistently more powerful than the comparable test for the FRPB method when the null hypothesis was false. Like the FRPB method, the SRPB method was more precise than the one-factor design when $\rho \geq 0.6$.

The SRPB ratio MS_{TB}/MS_R had a central F distribution when block and treatment effects were additive and provided a reasonably powerful test for non-additivity when it was present. If the SRPB method is used to investigate non-additivity, e.g., forming blocks on a covariable when regression heterogeneity is suspected, the additivity hypothesis should be tested prior to testing for treatment effects. If strong interaction is present, little faith can be placed in the test of the null treatment effects hypothesis, but then such main effects would have to be interpreted in the light of the interaction anyway.

Like the FRPB method, the SRPB tests lose power as cell frequencies become more variable. When disparate cell sizes are encountered, precision may be increased by reducing the number of blocks until the arithmetic and harmonic means of cell frequencies are approximately equal. The increased residual variance should generally be offset by the larger harmonic mean and increased degrees of freedom for estimating the residual variance.

If a reasonably strong correlation ($\rho \geq 0.4$) exists between the dependent and concomitant variables and the assumptions are tenable, the analysis of covariance provides a more powerful test of treatment effects than any of the other methods considered here, including a priori blocking methods (blocks are incorporated into the experimental design prior to assigning experimental units to treatment conditions). If the analysis of covariance is not applicable, a priori blocking methods are preferable to the post hoc blocking methods. The post hoc methods are more precise than the one-factor design only when block effects are very strong, but if non-additivity is suspected after data are gathered, either post hoc method provides a viable means for investigating the interaction. Under such conditions of non-additivity the SRPB test for treatment effects tends to be too liberal and should be used with caution--and a prior test for non-additivity.

The a priori sampled range method, where blocks are formed on the basis of sample quantiles rather than population values, suffers from the same sensitivity to block-treatment interaction as the SRPB method. Contrary to Pingel's research (1968), the MS_R is an appropriate error term for testing the null treatment effects hypothesis in the additive condition. However, the test becomes too liberal under conditions of severe non-additivity and should be preceded by a test for non-additivity with the ratio MS_{TB}/MS_R.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, R.L. and T.A. Bancroft, Statistical Theory in Research. New York: McGraw-Hill, 1952.

Bancroft, T.A., Topics in Intermediate Statistical Methods. Ames, Iowa: Iowa State U. Press, 1968.

Cochran, W.G., Analysis of covariance: its nature and use. Biometrics, 1957, 13, 261-281.

Collier, R.O., The effects of subject-treatment interaction in a randomized block, one factor experiment. Educational and Psychological Measurement, 1961, 21, 819-829.

Cornfield, J. and J.W. Tukey, Average values of mean squares in factorials. Annals of Mathematical Statistics, 1956, 27, 907-949.

Cox, D.R., The use of a concomitant variable in selecting an experimental design. Biometrika, 1957, 44, 150-158.

_____, Planning of Experiments. New York: Wiley, 1958.

DeGracie, J.S., Analysis of covariance when the concomitant variable is measured with error. Dissertation Abstracts, 1969, 29, 3529B.

Elashoff, J.D., Analysis of Covariance: a delicate instrument. American Educational Research Journal, 1969, 6, 383-401.

Feldt, L.S., A comparison of the precision of three experimental designs employing a concomitant variable. Psychometrika, 1958, 23, 335-353.

_____ and M.W. Mahmoud, Power function charts for specification of sample size in analysis of variance. Psychometrika, 1958, 23, 201-210.

Fisher, R.A., The Design of Experiments. (8th ed.) Edinburgh: Oliver and Boyd, 1935.

Fox, M., Charts of the power of the F-test. Annals of Mathematical Statistics, 1956, 27, 484-497.

Glass, G.V. and J.C. Stanley, Statistical Methods in Education and Psychology. Englewood Cliffs: Prentice-Hall, 1970.

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

Lord, F.M., Large sample covariance analysis when the control variable is fallible. Journal of the American Statistical Association, 1960, 55, 307-321.

Millman, J. and G.V. Glass, Rules of thumb for writing the ANOVA table. Journal of Educational Measurement, 1967, 4, 41-51.

Pearson, E.S. and H.O. Hartley, Charts of the power function of the analysis of variance tests, derived from the non-central F-distribution. Biometrika, 38, 112-130.

Peckham, P.D., An investigation of the effects of non-homogeneity of regression slopes upon the F-test of analysis of covariance. Dissertation Abstracts, 1969, 29, 3531B.

Pingel, L.A., A comparison of the effects of selected methods of block formation on the analysis of variance. Unpublished Ph.D. Dissertation, University of Wisconsin, 1968.

Porter, A.C., The effects of using fallible covariables in the analysis of covariance. Unpublished Ph.D. Dissertation, University of Wisconsin, 1967.

_____, How errors of measurement affect anova, regression analyses, ancova and factor analyses. Occasional Paper 14, Office of Research Consultation, College of Education, Michigan State University.

_____ and M. McSweeney, Randomized blocks design and non-parametric statistics. Paper presented at the American Educational Research Association Annual Meeting, March, 1970.

Sarhan, A.E. and B.G. Greenberg (ed)., Contributions to Order Statistics. New York: Wiley, 1962.

Scheffe', H., The Analysis of Variance, New York: Wiley, 1959.

Silverman, W. (adapted by R.G. Wolfe), Random floating point numbers in normal distribution (RANSS). Madison, Wisconsin: University of Wisconsin Computing Center, 1963.

Snedecor, G.W. and W.G. Cochran, Statistical Methods. Ames, Iowa: Iowa State University Press, 1967.

Teichroew, D., Tables of expected value of order statistics for samples of size twenty and less from the normal distribution. Annals of Mathematical Statistics, 1956, 27, 410-426.

Wilk, M.B., The randomization analysis of a generalized randomized block design. Biometrika, 1955, 42, 70-79.

_____ and O. Kempthorne, Fixed, mixed, and random models. American Statistical Association Journal, 1955, 50, 1144-1167.

_____ and O. Kempthorne, Derived linear models and their use in the analysis of randomized experiments. WADC Technical Report No. 55-244, V. II, 1956. (1956a)

_____ and O. Kempthorne, Some aspects of the analysis of factorial experiments in a completely randomized design. Annals of Mathematical Statistics, 1956, 27, 950-985. (1956b)

APPENDIX A

Table A1. Summary Statistics for Ten Samples of One Thousand Pseudorandom Normal Deviates Generated by Subroutine RANSS

Variable	Mean	S.D.	Skew	Range Min	Range Max
1	.081	.998	.041	-2.983	3.036
2	-.010	.991	-.095	-3.571	2.734
3	.017	1.000	-.091	-2.977	2.677
4	.027	1.008	.001	-3.320	3.201
5	-.004	1.024	-.051	-3.087	2.779
6	.029	.986	-.040	-3.435	2.937
7	.029	1.019	.039	-3.101	2.892
8	-1.00	1.021	.018	-3.419	3.485
9	-.040	1.002	-.051	-3.382	2.639
10	-.026	1.014	.019	-3.192	3.446
Combined	.000	1.007	-	-3.571	3.485
Median	.007	1.005	-.012	-3.256	2.914

Correlations

	1	2	3	4	5	6	7	8	9	10
1	1.000									
2	-.048	1.000								
3	.004	.030	1.000							
4	-.070	.068	-.021	1.000						
5	.029	-.004	-.056	.033	1.000					
6	-.005	-.020	.021	-.026	-.015	1.000				
7	-.032	.022	-.033	.023	.023	.030	1.000			
8	.010	-.012	-.044	-.017	.055	-.039	.007	1.000		
9	.012	-.032	-.026	.016	.045	.009	-.016	.008	1.000	
10	-.002	.009	-.048	-.024	.009	-.009	.001	.041	.021	1.000

Median r = .001

APPENDIX B

Table B1. Empirical Mean Squares for the Sampled-Range A Priori Blocking Method with Both Additive and Non-Additive Block and Treatment Effects: $\rho = 0.6$; $t = 2$; $b = 4$; $n = 5$

	Condition	
Source	Additive	Non-Additive
Treatments	.732	1.010
Blocks	4.661	.753
TxB Interaction	.731	4.751
Residual	.688	.693

Table B2. Observed Type-I Error Rates for the Sampled-Range A Priori Blocking Method with both Additive and Non-Additive Block and Treatment Effects: $\rho = 0.6$; $t = 2$; $b = 4$; $n = 5$

Ratio	Df_1	Df_2	Error Rate (1000α) 750	500	250	100	50	10
MS_T/MS_R								
(Additive)	1	32	731	502	258	112	61	17
(Non-Additive)	1	32	782	574	338	157	98	40
MS_T/MS_{TB}								
(Additive)	1	3	743	476	253	102	50	13
(Non-Additive)	1	3	444	132	14	0	0	0
MS_{TB}/MS_R								
(Additive)	3	32	759	535	285	110	58	15
(Non-Additive)	3	32	999	991	987	955	916	775

Non-Additive Case constructed by:

$$Y_i = p*X_i + \sqrt{1-p^2}*Z_i \text{ for treatment 1}$$

$$Y_i = -p*X_i + \sqrt{1-p^2}*Z_i \text{ for treatment 2}$$