DOCUMENT RESUME

ED 365 712 TM 020 906

AUTHOR Wu, Yi-Cheng; McLean, James E.
TITLE To Block or Covary a Concomitant Variable: Which Is Better?
PUB DATE Nov 93
NOTE 26p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 10-12, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Covariance; Analysis of Variance; Computer Simulation; Correlation; Monte Carlo Methods; *Research Design; Research Needs; Statistical Distributions
IDENTIFIERS *Blocking; *Concomitant Variables; Power (Statistics)

ABSTRACT

By employing a concomitant variable, researchers can reduce the error, increase the precision, and maximize the power of an experimental design. Blocking and analysis of covariance (ANCOVA) are most often used to harness the power of a concomitant variable. Whether to block or covary and how many blocks to be used if a block design is chosen become important. This paper provides an historical review of the problem and recommends future research to examine the problem based on how subjects are assigned, how data are analyzed, and the distributions of the variables. In this study, subjects were randomly assigned to treatments ignoring the concomitant variable, and data were analyzed by one-way analysis of variance (ANOVA), post-hoc two-block, four-block, and eight-block ANOVA and ANCOVA. Distributions of the concomitant and dependent variables were normal. The Monte Carlo method was used to generate 20,000 data sets for 8 experimental conditions (2 levels of subject and 4 levels of correlation between concomitant and dependent variables. The five analysis procedures were examined under each experimental condition. Results show that ANCOVA is more powerful than post-hoc rank blocking. Eight tables present analysis results. (Contains 36 references.) (Author/SLD)

# TO BLOCK OR COVARY A CONCOMITANT VARIABLE:

# WHICH IS BETTER?



by

Yi-Cheng Wu and James E. McLean

The University of Alabama

Paper to be presented at the annual meeting of the
Mid-South Educational Research Association
New Orleans, LA, November 10-12, 1993

## Abstract

By employing a concomitant variable, researchers can reduce the error, increase the precision, and maximize the power of an experimental design. Blocking and ANCOVA are most often used to harness the power of a concomitant variable. The questions of whether to block or covary and how many blocks to be used if a block design is chosen become important. This paper provides an historical review of the problem and recommends future research to examine the problem based on three dimensions: (1) how subjects are assigned, (2) how data are analyzed, and (3) the distributions of the variables. In this study, (1) subjects were randomly assigned to treatments ignoring the concomitant variable, (2) data were analyzed by one-way ANOVA; post-hoc two-block, four-block, and eight-block ANOVA; and ANCOVA, and (3) the distributions of the concomitant and dependent variables were normal. The Monte Carlo method was used to generate 20,000 sets of data for 8 experimental conditions (two levels of the number of subjects and four levels of correlation between the concomitant and the dependent variables). The five analysis procedures were examined under each experimental conditions. The results showed that ANCOVA was more powerful than post-hoc rank blocking.

## Introduction

Most educational experiments involve assigning students to treatments. Traditional one-way analysis of variance can be used to analyze the differences among treatments. However, differences among subjects, such as, sex, socioeconomic status, or level of ability, often mask or obscure the effects of a treatment (Kennedy and Bush, 1985; Kirk, 1982). Nuisance variation due to such differences can be extracted from the error variance. By controlling the concomitant (nuisance) variable, researchers often reduce the background noise, increase the precision, and enhance the statistical power of a design (Bonett, 1982; Keppel, 1991; Maxwell & Delaney, 1984). The most widely used procedures to harness the power of a concomitant variable are the block design and the analysis of covariance. The decisions on whether to block or covary and how many blocks to be used if a block design is selected are often based on rules of thumb with no empirical support. An empirical study that can offer the scientific foundation on which to base such decisions is desirable.

**Historical Review of the Problem**

In two classic design books, The Design of Experiment and Statistical Methods for Research Workers, Fisher (1937; 1973) developed the analysis of variance of block design and the analysis of covariance. He demonstrated that the precision of an experimental design could be improved by controlling a concomitant variable in the two analysis procedures. Lindquist (1953) used the term, treatments-by-levels design, which consists more than one observation in a cell, to differentiate it from the randomized complete block design, which consists only one observation in a cell. The treatments-by-levels design is also called the treatments-by-blocks design (Kennedy & Bush 1985). Lindquist recommended that the treatments-by-blocks design be used over the analysis of covariance because: (1) the treatments-by-blocks design required much less restrictive assumptions than the analysis of covariance, (2) the computational procedure were considerably simpler with the treatments-by-blocks design, and (3) the use of treatments-by-blocks design permitted a study on the simple effects of the treatments at any given block.

Gourlay (1953) compared the analysis of covariance with the randomized complete block design in which blocks were formed by matching subjects on the concomitant variable. He recommended that the analysis of covariance be used in preference to the matching block technique; this view was shared by Greenberg (1953) in a similar study.

Federer (1955) favored the block design over the analysis of covariance. He offered the following rule of thumb: "if the experimental variation cannot be controlled by stratification (blocking), then measure related variates and use covariance" (p. 483-484). However, he also pointed out that "it may be more advantageous to use covariance than to use stratification, since fewer degrees of freedom are usually required to control the variation" (p. 484).

Cox (1957) developed the Apparent Imprecision measure and used it to compare the analysis of covariance with the randomized complete block design in which blocks were formed by ranking subjects on the concomitant variable. Based on this measure, he found that the randomized complete block design was somewhat better than the analysis of covariance if the correlation coefficient was less than .6 while the analysis of covariance became appreciably better than the randomized complete block design when the correlation coefficient was .8 or more. He suggested that the analysis of covariance be preferable to the block design only if the correlation coefficient between the concomitant and the dependent variable was at least .6.

The most rigorous research on this topic was conducted by Feldt (1958). He used Cox's Apparent Imprecision measure to compare three experimental designs. The three experimental designs were: (1) stratification (blocking), (2) the analysis of covariance, and (3) the analysis of variance of difference scores. Feldt found the analysis of variance of difference scores was the least precise procedure; "for $\rho < .4$ the factorial (blocking) approach results in approximately equal or greater precision than covariance; for $\rho >= .6$ the advantage is in favor of covariance"; and "for $\rho < .2$ and small values of N neither covariance nor the factorial design yields appreciably greater precision than a completely randomized design" (p. 347). Feldt also provided a table for the optimal number of blocks to be used if the block design was selected. He summarized that the optimal number of blocks tended to be larger for (1) larger values of correlation coefficients, (2) lager numbers of subjects, and (3) smaller numbers of treatments. This study should be considered the classic study comparing the block design and the analysis of covariance; its findings have been quoted most often by textbooks in the area of experimental design (e.g., Cook & Campbell, 1979; Dayton, 1970; Kennedy and Bush, 1985; Keppel, 1991; Kirk, 1982; Myers, 1979).

In a block design, subjects are usually grouped into blocks before the experiment according to the value of the concomitant variable. However, there are times that the value of the concomitant variable is not available before the experiment. When blocks are formed after the experiment, the block design is defined as post-hoc block design. Keppel (1973) gave advantages of the post-hoc block design over the analysis of covariance: (1) reduction in computational effort, (2) free from the more strict assumptions of the analysis of covariance, and (3) possibility of testing the treatment X block interaction. However, he also pointed out two disadvantages of post-hoc blocking: (1) impossible to calculate the within-groups mean square when cells had fewer than 2 subjects, (2) unable to adjust the treatment means for differences on the concomitant variable.

Post-hoc blocking is popular because the value of the concomitant variable can be unknown before the experiment. Nevertheless, Myers (1979) pointed out the danger of abusing the post-hoc block design by demonstrating that reordering scores within each treatment would not change the treatment means but generally would reduce the error variance, which resulted in significant Fs which "merely reflect the reduction in error variance due to blocking rather than any variability due to treatments" (p. 155). However, he did not consider the loss of degrees of freedom with the block design.

Bonett (1982) compared the post-hoc block design with the analysis of covariance and offered the following rule: "if the assumptions for each method can be satisfied and if the probability of a Type II error is of concern, the analysis of covariance will be preferred when the form of the regression equation is known but the magnitude of the correlation is known. Post-hoc blocking, on the other hand, will be preferred when the magnitude of the correlation is known" (p. 38).

The only study found using the Monte Carlo method and using statistical power as the criterion variable to compare the block design and the analysis of covariance was performed by Maxwell and Delaney (1984). Their study was limited to two treatments. The procedures they compared were based on the following two dimensions: (1) the method of assignment and (2) the method of data analysis. Each of the two dimension had three levels: (1) the concomitant variable was ignored, (2) the concomitant variable was categorized, and (3) the concomitant variable was continuous. This resulted in nine procedures being compared. Maxwell and Delaney (1984) preferred the analysis of covariance over the block design. They argued that "the recommendation of most experimental design texts to consider the correlation between the dependent and concomitant variables in choosing the best technique for utilizing a concomitant variable is incorrect. Instead, the two factors that should be considered are whether scores on the concomitant variable are available for all subjects prior to assigning any subjects to treatment conditions and whether the relationship of the dependent and concomitant variables is linear" (p. 136). They also illustrated that the Apparent Imprecision measure, which was used in Cox's (1957) and Feldt's (1958) study, might provide a different perspective from statistical power, but, the Apparent Imprecision measure and statistical power are not independent.

**Summary of the Review**

While some research favored the block design, other research preferred the analysis of covariance. Based on the historical review of the problem, we agree with Maxwell and Delaney that "the relative merits of blocking and ANCOVA are more complicated, because neither is uniformly superior to the other" (1984, p. 136). It is likely that different procedures may be preferable to others under different experimental conditions. One significant consequence of applying the analysis of variance of block design and the analysis of covariance, which has been neglected often in early research but frequently stressed in recent research, is the decrease of the probability of the Type II error, i.e., the increase of the statistical power.

Based on the review of the relative literature, it is suggested that future research examine the problem based on three dimensions: (1) how subjects are assigned, (2) how data are analyzed, and (3) the distributions of and the relationship between the concomitant and the dependent variables (i.e., considering the assumptions of the block design and the analysis of covariance); that the experimental conditions include three factors: (1) the number of treatments, (2) the number of subjects per treatment, and (3) the magnitude of the relationship between the concomitant and the dependent variables;

and that the criterion variable on which to base the comparison be the statistical power, the Type I error ($\alpha$), and the Apparent Imprecision measure.

### Justification of the Study

This section provides the rationale for selecting statistical power as the criterion variable and using computer generated data to simulate the experiment.

#### Statistical Power as the Criterion Variable

The expressions; "reduce error", "increase precision", "enhance efficiency", and "maximize statistical power"; have been used frequently and interchangeably to describe the objective of employing a concomitant variable in the block design and the analysis of variance (e.g., Bonett, 1985; Kennedy & Bush, 1985; Maxwell & Delaney, 1984). Among those expressions, the term, statistical power, is unambiguously understood and operationally defined by every researcher and every book.

The neglect of statistical power in research, textbooks, and curricula has been constantly reported. As Cohen (1962; 1977; 1988; 1992) has stressed, one of the most pervasive threats to the validity of the statistical conclusions reached by behavioral research is the low statistical power. The investigation of statistical power in experiment designs has gain more and more significance (Chase & Tucker, 1976; Sedlmeier & Gigerenzer, 1989). Furthermore, the optimal number of blocks Maxwell and Delaney (1984) used to compare the statistical power of the block design and the analysis of covariance was based on the Apparent Imprecision measure—which may not be the optimal number of blocks to achieve statistical power. Therefore, examining the optimal number of blocks to achieve statistical power is desirable.

Computer Simulation

This is an empirical study using the Monte Carlo method to simulate the experiment. The Monte Carlo method has been used effectively in examining many sensitive properties of statistics (Harwell, Rubinstein, Hayes, & Olds, 1992; Shapiro, Wilk, & Chen, 1968; Wilcox, Charlin, & Thompson, 1986). Computer simulations have many advantages. "We can often simulate situations more readily on the computer than perform the corresponding experiments in real life"; "one can also easily vary parameters in computer experiments"; and "furthermore, the simulations tend to be very flexible in that a whole multitude of differing models can be simulated with relative ease with essentially the same computer code" (Jain, 1992, p. 2). Therefore, using a high speed computer to calculate the statistical power based on empirical sampling is the most direct and effective way to answer the research questions of this study.

**Procedures**

This study compared five analysis procedures under eight experimental conditions using empirical power as the criterion (dependent) variable. The five analysis procedures were: one-way analysis of variance; two-block, four-block, and eight-block analysis of variance; and analysis of covariance. The eight experimental conditions were the combinations of two levels of the number of subjects per treatment (8 and 40), and four levels of the correlation coefficient (.0, .5, .7, and .9). For each experimental condition, 2,500 sets of data were generated using the computer. Each set of data was analyzed using all five analysis procedures at the .05 significant level. The percentage of the significant analyses was the empirical power, for example, if 600 out of the 1,000 analyses were significant, empirical power would be .6.

Statistical power is a function of three major factors: (1) the significance level, (2) the sample size, and (3) the effect size (Dayton, Schafer, & Rogers 1973; Hinkle, Wiersma, & Jurs, 1988; Lipsey, 1990; Sawyer & Ball, 1981). Statistical power increases as the significance level, the sample size, or the effect size increases. In order to make statistical power comparable among the five analysis procedures, the power of the one-way analysis of variance was controlled at .5. Therefore, the effect sizes to achieve a power of .5 for one-way analysis of variance were calculated before the experiment. The calculation was based on tables provided by Cohen (1988). The effect sizes were 1.057 and 0.444 for n=8 and n=40 respectively with two treatments.

**Generation and Analyses of the Data**

The generation and analyses of the data were accomplished by a computer simulation system running on the IBM 3090/400E mainframe computer at The University of Alabama. Data were generated using the SAS commands provided by Clark and Woodward (1992). These commands generate random data from a bivariate normal distribution (the concomitant and the dependent variables) with a mean of 0, a variance of 1, and the user-specified correlation coefficient. Random samples were generated separately for each treatment. Only the means of the dependent variable of the second treatment was transformed based on the calculated effect sizes, while the other parameters were not changed. Data in each treatment were grouped into 2, 4, and 8 blocks by their ranks on the concomitant variable. For example, to group 40 subjects into 4 blocks, the top 10 ranked subjects were in the first block, the 11-20 ranked subjects were in the second block, the 21-30 ranked subjects were in the third block, and the 31-40 ranked subjects were in the fourth block.

The computer simulation system included one executable file and two SAS programs (International Business Machines, 1988a; International Business Machines, 1988b; SAS Institute Inc., 1990a; SAS Institute Inc., 1990b). For each of the eight experimental conditions, the executable file ran the first SAS program 2,500 times, then ran the second SAS program. The computer programs under the condition of $n=40$ and $\rho=.7$ is provided in the appendix. The first SAS program generated a set of data, analyzed that set of data with the five analysis procedures being compared, and output the results of the analyses to a data file. After the first SAS program had run for 2,500 times, the data file would contain 2,500 records of the results of the analyses. The second SAS program calculated the empirical power based on the 2,500 records. Totally, there were 20,000 (2,500 X 8) sets of data generated and 100,000 analyses conducted.

## RESULTS

The resulting empirical power under each experimental condition is shown in Table 1. Each value represents the percentage of the significant analyses out of 2,500 analyses.

Since there is only one observation per cell, we can only test the results using the randomized complete block analysis. The outcomes of the analysis are shown in Table 2. The overall F, the main effects, and the two-way interactions are all significant.

Table 1

Empirical Power (All Combinations)

| Sample Size | Correlation Coefficient | Analysis Procedures | | | | |
|---|---|---|---|---|---|---|
| | | ANOVA | Two-Block | Four-Block | Eight-Block | ANCOVA |
| 8 | .0 | 51.4 | 50.8 | 49.7 | 46.8 | 48.2 |
| | .5 | 50.4 | 56.0 | 57.5 | 54.9 | 57.8 |
| | .7 | 49.6 | 61.9 | 66.4 | 63.9 | 74.0 |
| | .9 | 49.4 | 71.8 | 79.2 | 80.0 | 98.8 |
| 40 | .0 | 49.3 | 49.3 | 49.0 | 49.0 | 49.1 |
| | .5 | 50.2 | 56.8 | 59.0 | 59.4 | 61.3 |
| | .7 | 50.5 | 63.5 | 68.7 | 70.4 | 77.7 |
| | .9 | 51.2 | 73.3 | 81.6 | 84.9 | 99.2 |

(HSD: P@NXC = 2.9 and NXC@P = 3.3)

Table 2

Summary for Randomized Complete Block Analysis

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 27 | 7654.98325000 | 283.51789815 | 663.65 | 0.0001 |
| Error | 12 | 5.12650000 | 0.42720833 | | |
| Total | 39 | 7660.10975000 | | | |

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| N | 1 | 30.45025000 | 30.45025000 | 71.28 | 0.0001 |
| C | 3 | 4245.71675000 | 1415.23891667 | 3312.76 | 0.0001 |
| P | 4 | 1787.56850000 | 446.89212500 | 1046.08 | 0.0001 |
| N*C | 3 | 14.49475000 | 4.83158333 | 11.31 | 0.0008 |
| N*P | 4 | 24.05350000 | 6.01337500 | 14.08 | 0.0002 |
| C*P | 12 | 1552.69950000 | 129.39162500 | 302.88 | 0.0001 |

The randomized complete block design assumes no treatment-by-block interaction, which is not likely true in this analysis. If a significant interaction exits, the analysis will be conservative because of the inability to exclude interaction from the error term. Even if they may be conservative, all sources are significant with low p-values (0.0001 - 0.0008). Another evidence supporting the precision of this analysis is the accuracy of the resulting power value. The power of the one-way analysis of variance was controlled at .5 before the experiment. The resulting power values of the one-way analysis of variance have a mean of .5 and a standard deviation of .008, which indicates that the empirical power, outcome of the analyses of 2,500 sets of data, is accurate. Therefore, the true mean square error should be small.

The following tables provide the means for the main factors and their two-way combinations. The means of the main factors are in the last row and the last column; and the grand mean is at the bottom right corner.

Table 3

Empirical Power (Sample Size X Procedure)

| | Analysis Procedure | | | | | |
|---|---|---|---|---|---|---|
| Sample Size | ANOVA | Two-Block | Four-Block | Eight-Block | ANCOVA | |
| 8 | 50.2 | 60.1 | 63.2 | 61.4 | 69.7 | 60.9 |
| 40 | 50.3 | 60.7 | 64.6 | 65.9 | 71.8 | 62.7 |
| | 50.3 | 60.4 | 63.9 | 63.7 | 70.8 | 61.8 |

(HSD: N = .5, P = 1.0, P@N = 1.5, and N@P = 1.0)

Table 4

Empirical Power (Correlation Coefficient X Procedure)

| | Analysis Procedure | | | | | |
|---|---|---|---|---|---|---|
| Correlation Coefficient | ANOVA | Two-Block | Four-Block | Eight-Block | ANCOVA | |
| .0 | 50.4 | 50.1 | 49.4 | 47.9 | 48.7 | 49.3 |
| .5 | 50.3 | 56.4 | 58.3 | 57.2 | 59.6 | 56.3 |
| .7 | 50.1 | 62.7 | 67.6 | 67.2 | 75.9 | 64.7 |
| .9 | 50.3 | 72.6 | 80.4 | 82.5 | 99.0 | 76.9 |
| | 50.3 | 60.4 | 63.9 | 63.7 | 70.8 | 61.8 |

(HSD: C = .9, P = 1.0, P@C = 2.1, and C@P = 1.9)

Table 5

Empirical Power (Sample Size X Correlation Coefficient)

| | Correlation Coefficient | | | | |
|---|---|---|---|---|---|
| Sample Size | .0 | .5 | .7 | .9 | |
| 8 | 49.4 | 55.3 | 63.2 | 75.8 | 60.9 |
| 40 | 49.1 | 57.3 | 66.2 | 78.0 | 62.7 |
| | 49.3 | 56.3 | 64.7 | 76.9 | 61.8 |

(HSD: N = .5, C = .9, C@N = 1.2, and N@C = .9)

Tukey's Honest Significant Difference (HSD) was used for multiple comparisons. The HSDs for the respective main effect and simple effect comparisons were reported at the bottom of the mean tables. The following tables provide the results of multiple comparisons for main effects.

Table 6

Multiple Comparisons (Sample Size)

Alpha = .05    df = 12    MSE = .4272

Critical Value of Studentized Range = 3.081

HSD = .5

Means with different letters are significantly different.

| | Means | N | Sample Size |
|---|---|---|---|
| A | 62.7 | 20 | 40 |
| B | 60.9 | 20 | 8 |

Table 7

Multiple Comparisons (Correlation Coefficient)

Alpha = .05    df = 12    MSE = .4272

Critical Value of Studentized Range = 4.199

HSD = .9

Means with different letters are significantly different.

| | Means | N | Corr. Coeff. |
|---|---|---|---|
| A | 76.9 | 10 | .9 |
| B | 64.7 | 10 | .7 |
| C | 56.3 | 10 | .5 |
| D | 49.3 | 10 | .0 |

Table 8

Multiple Comparisons (Procedure)

Alpha = .05 df = 12 MSE = .4272

Critical Value of Studentized Range = 4.508

HSD = 1.0

Means with different letters are significantly different.

| | Means | N | Procedure |
|---|---|---|---|
| A | 70.8 | 8 | ANCOVA |
| B | 63.9 | 8 | Four-Block |
| B | 63.7 | 8 | Eight-Block |
| C | 60.4 | 8 | Two-Block |
| D | 50.3 | 8 | ANOVA |

The results show that all pair-wide main effect comparisons are significant except for that between the four-block and eight-block procedures. The multiple comparison for simple effects can be done by simply examining whether or not the difference between two means exceeds the corresponding HSD value offered at the bottom of the mean tables--if it does, the comparison is significant.

## Conclusions and Implications

Based on the results, we summary:

A. The power increases as the number of subjects increases or the correlation coefficient increases.

B. For $\rho = 0$ and n=8, neither the block design nor the ANCOVA is more powerful than the one-way analysis of variance.

C. For $\rho = 0$ and n=40, the five procedures yield approximately the same power.

D. The optimal number of blocks increases as the number of subjects increases or the correlation coefficient increases.

E. The ANCOVA is the most powerful design when $\rho > .5$.

This study does not include the treatment-by-block interaction in the block design since the interaction does not exist in the population. Future study can examine the effects of including the interaction using the same computer simulation system, or, by varying the parameters of the population, examine the effects of including and excluding the interaction when the interaction does exist in the population. The greatest contribution of this study might not be the specific results reported here, but the potential for examining many other situations. This computer simulation system can be used to simulate a whole multitude of relevant studies with minor modifications; these include investigating other criteria such as the Type I error, examining other levels of the experimental conditions, and testing other blocking methods in addition to the post-hoc blocking used in this study.

## References


Bonett, D. G. (1982). On post-hoc blocking. Educational and Psychological Measurement, 42, 35-39.

Chase, L. J., & Tucker, R. K. (1976). Statistical Power: derivation, development, and data-analytic implications. The Psychological Record, 26, 473-486.

Clark, M. R. & Woodward, D. E. (1992). Generating random numbers with base SAS software. Observations, 1(4), 12-19.

Cohen, J. (1962). The statistical power of abnormal-social psychological research. Journal of Abnormal and Social Psychology, 65(3), 145-153.

Cohen, J. (1977). Statistical power analysis for the behavioral Science (rev. ed.). New York: Academic Press.

Cohen, J. (1988). Statistical power analysis for the behavioral Science (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155-159.

Cook, T. D. & Campbell, D. T. (1979). Quasi-experimentation: design & analysis Issues for field settings. Chicago: Rand McNally College Publishing Company.

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

Dayton, C. M. (1970). The Design of Educational Experiments. New York: Mcgraw-Hill.

Dayton, C. M., Schafer, W. D., & Rogers, B. G. (1973). On appropriate uses and interpretations of power analysis: a comment. American Educational Research Journal, 10(3), 231-234.

Federer, W. T. (1955). Experimental design. New York: Macmillan Company.

Feldt, L. S. (1958). A comparison of the precision of three experimental designs employing a concomitant variable. Psychometrika, 23(4), 335-353.

Fisher, R. A. (1937). The design of Experiments (2nd ed.). Edinburgh: Oliver and Boyd.

Fisher, R. A. (1973). Statistical Methods for Research Workers (14th ed.). New York: Hafner Publishing Company.

Greenberg, B. G. (1953). The use of analysis of covariance and balancing in analytical surveys. American Journal of Public Health, 43, 692-699.

Gourlay, N. (1953). Covariance analysis and its applications in psychological research. British Journal of Statistical Psychology, 6, 25-34.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. Journal of Educational Statistics, 17(4), 315-339.

Hinkle, D. E., Wiersma, W. & Jurs, S. G. (1988). Applied Statistics for the Behavioral Sciences (2nd. ed.). Boston: Houghton Mifflin Company.

International Business Machines. (1988a). VM/XA System Product interpreter: Reference Manual, First Edition, March 1988: Author.

International Business Machines. (1988b). VM/XA System Product interpreter: User's Guide, First Edition, March 1988: Author.

Jain, S. (1992). Monte Carlo simulations of disordered system. Singapore: World Scientific Publishing Company.

Kennedy, J. J. & Bush, A. J. (1985). An introduction to the design and analysis of experiments in behavioral research. Lanham, M.D.: University Press of America.

Keppel, G. (1973). Design and analysis: a researcher's handbook. Englewood Cliff, N.J.: Prentice Hall.

Keppel, G. (1991). Design and analysis: a researcher's handbook (3rd ed.). Englewood Cliff, N.J.: Prentice Hall.

Kirk, R. E. (1982). Experimental design (2nd ed.). Monterey, CA: Brooks/Cole Publishing Company.

Lindquist, E. F. (1953). Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin Company.

Lipsey, M. W. (1990). Design sensitivity: statistical power for experimental research. Newbury Park, CA: SAGE Publications.

Maxwell, S. E. & Delaney, H. D. (1984). Another look at ANCOVA versus blocking. Psychological Bulletin, 95(1), 136-147.

Myers, J. L. (1979). Fundamentals of experimental design. Boston: Allyn and Bacon.

SAS Institute Inc. (1990a). SAS Procedures Guide, Version 6, Third Edition. Cary, N.C.: Author.

SAS Institute Inc. (1990b). SAS/STAT User's Guide: Volume 2, GLM-VARCOMP, Version 6, Third Edition. Cary, N.C.: Author.

Sawyer, A. G. & Ball, A. D. (1981). Statistical power and effect size in marketing research. Journal of Marketing Research, 18, 275-290.

Sedlmeier, P. & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? Psychological Bulletin, 105(2), 309-316.

Shapiro, S. S., Wilk, M. B., & Chen, M. J. (1968). A comparative study of various tests for normality. Journal of the American Statistical Association, 63, 1343-1372.

Wilcox, R. R., Charlin, V. L., & Thompson, K. L. (1986). New Monte Carlo results on the robustness of the ANOVA F, W and F* statistics. Communications in Statistics—Simulation and Computation, 15, 933-943.

## APPENDIX

Exec File

```
/*    */
ADDRESS COMMAND
"ERASE PVALUE DATA A"
SEED = 123456789
TIME = 1
DO WHILE TIME < 2501
SEED = SEED + 99999
"EXECIO 1 DISKW" NEWSEED DATA A "(STRING" SEED
"EXEC SAS G2N40"
"ERASE NEWSEED DATA A"
TIME=TIME+1
END
"EXEC SAS G2N40P"
```

First SAS Program

```
CMS FILEDEF INDATA DISK NEWSEED DATA A;
CMS FILEDEF PVALUE DISK PVALUE DATA A (LRECL 133 BLKSIZE 133
RECFM FBS;
CMS FILEDEF SASLIST DISK G2N40 LISTING A;
DATA BIVNORM (DROP=I);
   INFILE INDATA;
   INPUT SEED 1-9;
   RETAIN SEED;
   DO I=1 TO 40;
      GROUP=1;
      X=RANNOR(SEED);
      Y=.0*X+SQRT(1-.0**2)*RANNOR(SEED);
      OUTPUT;
   END;
   DO I=1 TO 40;
      GROUP=2;
      X=RANNOR(SEED);
      Y=.0*X+SQRT(1-.0**2)*RANNOR(SEED);
      Y=0.444+1*Y;
      OUTPUT;
   END;
PROC SORT;
   BY GROUP X;
DATA BIVNORM;
   SET BIVNORM;
   IF _N_<=20 OR (_N_>=41 AND _N_<=60) THEN B2=1;
   ELSE B2=2;
   IF _N_<=10 OR (_N_>=41 AND _N_<=50) THEN B4=1;
   ELSE IF _N_<=20 OR (_N_>=51 AND _N_<=60) THEN B4=2;
   ELSE IF _N_<=30 OR (_N_>=61 AND _N_<=70) THEN B4=3;
```

```
   ELSE B4=4;
   IF _N_<=5 OR (_N_>=41 AND _N_<=45) THEN B8=1;
   ELSE IF _N_<=10 OR (_N_>=46 AND _N_<=50) THEN B8=2;
   ELSE IF _N_<=15 OR (_N_>=51 AND _N_<=55) THEN B8=3;
   ELSE IF _N_<=20 OR (_N_>=56 AND _N_<=60) THEN B8=4;
   ELSE IF _N_<=25 OR (_N_>=61 AND _N_<=65) THEN B8=5;
   ELSE IF _N_<=30 OR (_N_>=66 AND _N_<=70) THEN B8=6;
   ELSE IF _N_<=35 OR (_N_>=71 AND _N_<=75) THEN B8=7;
   ELSE B8=8;
PROC PRINT;
PROC CORR DATA=BIVNORM;
   VAR X Y;
   BY GROUP;
PROC GLM;
   CLASS GROUP;
   MODEL Y=GROUP/SS3;
PROC GLM;
   CLASS GROUP B2;
   MODEL Y=GROUP B2/SS3;
PROC GLM;
   CLASS GROUP B4;
   MODEL Y=GROUP B4/SS3;
PROC GLM;
   CLASS GROUP B8;
   MODEL Y=GROUP B8/SS3;
PROC GLM;
   CLASS GROUP;
   MODEL Y=GROUP X/SS3;
DATA;
   INFILE SASLIST;
   INPUT WORD1 $ WORD2 $ @;
   FILE PVALUE MOD;
   IF WORD1 = 'X' AND WORD2 ='40' THEN DO;
      INPUT MEAN STDDEV;
      PUT MEAN 6.4 STDDEV 6.4 @;
      INPUT Y $ N MEAN STDDEV;
      PUT MEAN 6.4 STDDEV 6.4 @;
      END;
   ELSE  IF WORD1="X" AND WORD2 = '1.00000' THEN DO;
      INPUT CORR;
      PUT CORR 6.4 @;
      END;
   ELSE  IF WORD1="GROUP" AND WORD2 = '1' THEN DO;
      INPUT SS MS F PR;
      PUT PR 6.4 @;
      INPUT BLOCK $ DF SS MS F PR;
      PUT PR 6.4 @;
      END;
```

Second SAS Program

```
CMS FILEDEF INDATA DISK PVALUE DATA A;
DATA PVALUE;
INFILE INDATA;
INPUT (G1XMEAN G1XSD G1YMEAN G1YSD G1CORR G2XMENA G2XSD G2YMEAN
       G2YSD G2CORR GROUP1B BLOCK1B GROUP2B BLOCK2B GROUP4B
       BLOCK4B GROUP8B BLOCK8B GROUPANC BLOCKANC) (20* 6.4);
     TOTAL=0;
     G1BSG=0;
     B1BSG=0;
     G2BSG=0;
     B2BSG=0;
     G4BSG=0;
     B4BSG=0;
     G8BSG=0;
     B8BSG=0;
     GANCSG=0;
     BANCSG=0;
     TOTAL=1;
     IF GROUP1B <= 0.05 THEN G1BSG=1;
     IF BLOCK1B <= 0.05 THEN B1BSG=1;
     IF GROUP2B <= 0.05 THEN G2BSG=1;
     IF BLOCK2B <= 0.05 THEN B2BSG=1;
     IF GROUP4B <= 0.05 THEN G4BSG=1;
     IF BLOCK4B <= 0.05 THEN B4BSG=1;
     IF GROUP8B <= 0.05 THEN G8BSG=1;
     IF BLOCK8B <= 0.05 THEN B8BSG=1;
     IF GROUPANC <= 0.05 THEN GANCSG=1;
     IF BLOCKANC <= 0.05 THEN BANCSG=1;
PROC FREQ;
     TABLE G1BSG -- BANCSG;
PROC SUMMARY DATA=PVALUE;
     VAR G1XMEAN -- BANCSG;
     OUTPUT OUT = DESCRIPT;
PROC PRINT DATA=DESCRIPT;
PROC UNIVARIATE DATA=PVALUE PLOT NORMAL;
     VAR G1XMEAN -- BANCSG;
```