ED 365 720 TM 020 936

AUTHOR	Johnson, Colleen Cook
TITLE	The Effects of Violation of Data Set Assumptions when Using the Oneway, Fixed Effects Analysis of Variance and the One Concomitant Analysis of Covariance Statistical Procedures.
PUB DATE	Nov 93
NOTE	15p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 10-12, 1993).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Analysis of Covariance; *Analysis of Variance; Comparative Analysis; Computer Simulation; Monte Carlo Methods; Reliability; *Research Design; *Sample Size; *Sampling; Statistical Distributions
IDENTIFIERS	*Balanced Designs; *Violation of Assumptions

ABSTRACT

The purpose of this study is to help define the precise nature and limits of the tolerable range in which a researcher may be relatively confident about the statistical validity of his or her research findings, focusing specifically on the statistical validity of results when violating the assumptions associated with the one-way, fixed-effects analysis of variance (ANOVA) and one concomitant analysis of covariance (ANCOVA) statistical procedure. Methodological and data set statistical assumptions that must be met for ANOVA and ANCOVA are discussed. Research results were obtained from an exploratory study of the effects of single and compound violations of the mathematical conditions (assumptions) underlying use of ANOVA and ANCOVA through Monte Carlo methods using randomly created data for a mathematical simulation (665 data set conditions). For all of the analyses, comparisons were made between the empirical F sampling distributions and the theoretical (i.e., nominal) F distributions expected when one uses normal theory. For balanced designs, the ANOVA and ANCOVA F tests were found to be remarkably robust when faced with most of the violations in the simulation. Research does reaffirm that ANOVA and ANCOVA should be avoided when group sizes are not equal. Specific recommendations are made for checking the ratio of largest to smallest group variances. (Contains 29 references.) (SLD)

THE EFFECTS OF VIOLATION OF DATA SET ASSUMPTIONS WHEN USING THE ONEWAY, FIXED EFFECTS ANALYSIS OF VARIANCE AND THE ONE CONCOMITANT ANALYSIS OF COVARIANCE STATISTICAL PROCEDURES

Colleen Cook Johnson, Ed.D.

Paper presented at the Mid-South Educational Research Association Annual Meeting, November 1993 in New Orleans, LA

INTRODUCTION

As educational researchers, we frequently use statistical procedures (ie., statistical tools) to aid in the interpretation of our data. There are a number of such statistical tools available to researchers. One of the most important tasks in the design of research is choosing the correct statistical procedure to use in the interpretation of the data yielded by research.

There are two general types of statistical procedures: parametric and non-parametric. Parametric procedures are used to test hypotheses about specific population parameters (e.g., the population mean) when only sample statistics are available. The oneway analysis of variance and the one concomitant analysis of covariance are two among the many parametric statistical tools available, but their use in both educational research and the social sciences is widespread: indeed, Halpin and Halpin (1988) argue that the analysis of variance is the most widely used statistical procedure by practitioners in both disciplines. *Used appropriately*, parametric statistical procedures are both very powerful (that is, they should be sensitive to change in the specific factors being tested by the researcher) and robust as well (that is, they should not be sensitive to changes in extraneous factors of a magnitude likely to occur in real life situations) (Box and Anderson, 1955). This, of course, contributes to their popularity among researchers.

Often overlooked by researchers, however, is the fact that statistical procedures are like tools used for any other purpose: they are designed to perform a specific function *under the appropriate set of conditions.* We would not choose to use a chisel if our goal was to cut a 2 by 4 in half. Nor would we ask a dc type battery to power a household applicance that is designed to operate from ac current. Yet parametric statistical procedures may be used by researchers in situations that the procedure was not designed to handle; situations where the alternative non-parametric procedure would yield a truer picture of the relationship between variables in one's research.

When they are initially developed by mathematicians, parametric statistical procedures are designed to be used only when specific conditions (ie., "assumptions") exist. The reason for this stems back to two conflicting sets of needs that developers of the mathematical procedures had to balance as they developed these statistical procedures. On the one hand, they had to develop procedures that would be able to process data in a form that useful to researchers. But on the other hand, they also had to develop these procedures in a manner that would simplify many mathematical derivations and operations (Glass, Peckham and Sanders, 1972). The resulting parametric statistical procedures do balance the two sets of needs: however, they are able to do so only when the researcher's data set meets the specific assumptions appropriate for his or her statistical procedure of choice.

Seldom, however, do data sets adhere perfectly to the assumptions a statistical procedure was designed to handle. Therefore the question that the researcher must ask in reference to the data that he or she has collected is not *whether* the assumptions have been satisfied, but instead, *are the violations that do occur extreme enough to compromise the validity of the results?* Put another way, the crucial question is how much difference is there between the conditions that the model was designed to handle and the actual conditions that exist in a particular research situation? If that difference is within a "tolerable range," then use of the chosen parametric procedure should produce information that is statistically robust in its interpretation of the relationship between variables. It is only when the differences between the data collected and the ideal data set exceeds that "tolerable range" that the non-parametric alternatives must be considered.

One methodology for estimating the limits of that "tolerable range" is through the use of Monte Carlo simulation techniques. Simulation studies such as this project are designed to determine how much difference can exist between a researcher's data set and the conditions that the procedure was designed to operate under. If this difference is within the "tolerable range," then the results produced by the parametric procedure should produce statistically valid results. If, however, the differences between the ideal data set and the actual observed data exceeds that "tolerable range," then parametric statistical procedures should be abandoned in favor of their non-parametric alternatives.

The purpose of this study is to help define the precise nature and limits of this "tolerable range" within which a researcher may be relatively confident about the statistical validity of his or her research findings. This study focuses specifically on the statistical validity of results when violating the assumptions associated with the oneway, fixed-effects analysis of variance (ANOVA) and one concomitant analysis of covariance (ANCOVA) statistical procedures. Widespread use of these statistical procedures by educational researchers and social scientists demands that we understand as precisely as possible when ANOVA and ANCOVA results can and cannot be trusted.

ASSUMPTIONS OF THE ANOVA AND ANCOVA PROCEDURES

Two Types of Assumptions

The statistical assumptions which must be met when using the ANOVA or ANCOVA procedures can be classified as falling into one of two categories: methodological assumptions or data set assumptions (Johnson, 1992). **Methodological assumptions** are concerned with the design of the research, the mathematical methodology and/or the sampling procedures. **Data set assumptions** are concerned with the mathematical characteristics of the observed data set and the population from which the observed data was drawn. Both the methodological and data set assumptions for these two statistical procedures will be discussed below.

ANOVA Assumptions

In 1972, Glass et al. identified three assumptions of concern for the ANOVA. The first is additivity - that is, each observation must be the simple sum of three components: the grand mean, the treatment effects, and the error associated with each individual observation. This, Cochran (1947) argued, is important because the least amount of information is lost in an additive model. The second assumption - more technically, a mathematical restriction adopted to allow for a unique solution to the least squares equation - is that the sum of the treatment effects equal zero. Finally, the third assumption is that errors made while using the model should be normally distributed with a population mean of zero and a variance of ^{2}. The third assumption involves the nature of the errors in the population that the sample data comes from, and takes three distinctive forms: (a) normality of the error distribution, (b) homogeneity of group variances, and (c) the independence of errors. Independence of errors is, of course, a methodological concern. Therefore it is forms (a) and (b) of the third assumption that are the subject of most theoretical and empirical research into ANOVA.

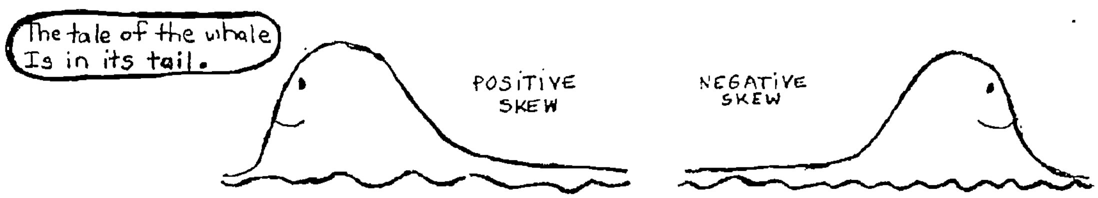
Homogeneity of Group Variances

The term homogeneity of variances refers to the assumption that the degree of variance (i.e., the spread of the scores from the group mean) within each of the groups be very similar. In 1972, Glass et al., following an extensive review of empirical research into the assumptions of the ANOVA and ANCOVA procedures, suggested that when there are an equal number of subjects in each of a researcher's groups (in other words, when there is a "balanced design"), F test results should be sufficiently robust, provided the ratio of largest to smallest group variance does not exceed three. This has become a standard for judging the validity of ANOVA test results in the two decades following their work. In 1990, however, this conclusion was questioned by Harwell, Hayes, Olds and Rubinstein. Following a meta-analytic study of empirical research, they suggested that even when sample sizes are equal, inflated type I errors are possible when the ratio of largest to smallest variance is as small as two.

When sample sizes are unequal (in other words, in an "unbalanced design"), empirical research conducted throughout the decades suggest that the validity of the F ratio is suspect. When group sizes are unequal and only two groups are involved, research suggests that inflated type I error rates occur when the larger group size is paired with the smaller group variances (e.g., Scheffe', 1959).

Normality of the Distribution of Errors

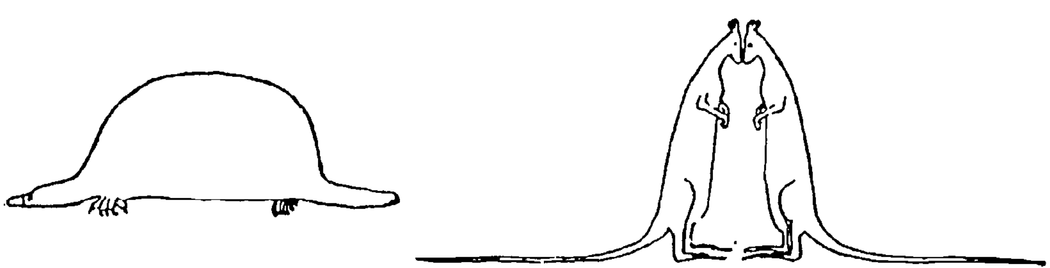
Another ANOVA assumption is that the errors (that is, the differences between the individual scores and the mean) be normally distributed. Distributions containing skewed errors, when graphed in frequency polygon form, have a shape similar to a whale in water:



where the extended tail (i.e., the extreme scores) determine whether a distribution is negative (i.e., contains extreme scores below the mean) or positive (i.e., contains extreme scores above the mean).

A distribution having either a leptokurtic or platykurtic shape, on the other hand, might have a shape similar to one of these zoological figures:

The important property which follows from this is that platykurtic curves have shorter "tails" than the



normal curve of error and leptokurtic longer "tails." I myself bear in mind the meaning of the words by the above *memoria technica*, where the first figure represents platypus, and the second kangaroos, noted for "lapping," though, perhaps, with equal reason they should be hares!

(from "Errors of Routine Analysis" by Student t, 1929)

Games and Lucas (1966) suggest that skewed distributions are a greater threat to the robustness of ANOVA than are either leptokurtic or platykurtic distributions. These researchers also suggest that results may actually improve when ANOVA is conducted on data that has highly leptokurtic error distributions, although ANOVA results when used with platykurtic error distributions are adversely affected.

Extension of ANOVA Asssumptions to ANCOVA

The simplest form of the analysis of covariance (which consists of one independent, one concomitant and one dependent variable) is merely an extension of the oneway, fixed-effects ANOVA. Therefore, researchers generally accept the assumptions of ANOVA as applying to the ANCOVA as well, provided that the concomitant variable is normally distributed. (e.g., Cochran, 1957; Winer, 1962).

The Seven Assumptions of the Analysis of Covariance

Elashoff (1969) and McLean (1979, 1989) report the following seven assumptions associated with ANCOVA: (1) the cases are assigned at random to treatment conditions; (2) the covariate is measured error-free (that is, there is a perfect reliability in the measurement of the covariate); (3) the covariate is independent of the treatment effect; (4) the covariate has a high correlation with the dependent variable; (5) the regression of the dependent variable on the covariate is the same for each of the treatment groups; (6) for each level of the covariate, the dependent variable is normally distributed; and (7) the variance of the dependent variable at each given value of the covariate is constant across treatment groups. Again, these assumptions can be classified as either methodological or data set assumptions.

Methodological Assumptions

Two of the ANCOVA assumptions deal with the research design and sampling methods: (1) the cases are assigned to random treatments (randomization) and (2) the covariate has perfect reliability. Concerning the issue of randomization, Evans and Anastasio (1968) distinguish between three separate situations: (1) individuals are assigned to groups at random after which the treatments are randomly assigned to the groups; (2) intact groups are used, however treatments are randomly assigned to the groups; and (3) intact groups are used where treatments occur naturally rather than being randomly assigned by the researcher. These researchers maintain that ANCOVA is appropriate for the first situation, can be used with caution in the second, but should be abandoned altogether (perhaps in favor of the less restraining factorial block ANOVA design) in the third. They provide two reasons for their recommendations: first, it is never quite clear whether the covariance adjustment has removed all of the bias when proper randomization has not taken place, and second, when there are real differences among the groups, covariate adjustments may involve computational extrapolation.

Raajimakers and Pieters (1987) and also McLean (1974) have addressed the issue of an unreliable covariate. Raajimakers and Pieters note that there are two ways that a researcher can conceptualize covariate reliability. If one assumes that the dependent variable is linearly related to the observed value of the covariate, then the ANCOVA results will retain their statistical validity. If, however, it is assumed that the dependent variable is linearly related to the underlying true score on the covariate (rather than the sample of scores that were actually observed), then the resulting F ratio will produce biased results. McLean's research, however, suggests that the issue of perfect reliability becomes less of a threat to the validity of the F ratio if there is an independence of the covariate measure and the treatment groups.

The Covariate's Relationship with the Independent and Dependent Variables

The covariate should have no significant correlation with the independent variable, but should be highly correlated with the dependent variable. Feldt (1958) recommends the use of a covariate only when the zero-order correlation between the covariate and the independent variable is greater than 0.6. McLean (1979,1989) sees the relationship between the covariate and the independent variable to be the most fundamental of all of the assumptions, and suggests that ANCOVA not be performed until after the data has been tested to see if it meets this assumption. If this assumption is not met, then the F test results are not invalidated as such, however it reduces the ANCOVA's efficiency to slightly below that of doing a simple oneway ANOVA on the same data.

Homogeneity of Group Regression Slopes

This assumption requires that the slope of the regression line between the concomitant and dependent variables be the same for all levels of the grouping variable. The problem, if this assumption is violated, is analogous to trying to interpret main effects in the presence of significant interactions in an n-way factorial ANOVA. If heterogeneous regression slopes are suspect, the researcher would be wiser to use the randomized block ANOVA instead of ANCOVA.

Empirical research using balanced ANCOVA designs suggests that small differences in the actual vs. expected significance levels may occur when regression slopes differ between groups (Peckham, 1968; McClaren, 1972). Peckham also found that as the degree of heterogeneity in the regression slopes created in his simulations increased, the heterogeneity of group vairances likewise increased - this, in turn, decreasing the rate of Type I errors that would otherwise be expected.

With unbalanced designs, empirical research (e.g., Box, 1954; McClaren, 1972; Scheffe', 1959) suggests that when the smallest regression coefficient and the largest variance are combined with the smallest sample size, the empirical significance levels will be biased in a non-conservative direction. When the pairings are reversed, however, the test results become conservative.

Homogeneity of Group Variances and Non-Normal Error Distributions in ANCOVA

As has been discussed previously, most researchers accept the claim by Cochran (1957) and Winer (1962) that the effects of the simple ANOVA violations are equally viable when the model is extended to include one or more concomitants.

RESEARCH METHODOLOGY

The research results which will be summarized below were obtained as a result of an exploratory study of the effects of both single and compound violations of the mathematical conditions (i.e., assumptions) underlying use of the analysis of variance (ANOVA) and analysis of covariance (ANCOVA) statistical procedures. A Monte Carlo methodology was used, which allowed for the empirical investigation of problems identified by theoretical mathematicians as potential threats to the robustness of the ANOVA and/or ANCOVA results under conditions common to research practitioners in the behavioral sciences, the social sciences and education. Because of advances both in methodological techniques and computing technology, the capability has emerged to study this topic in depth, yet with a global perspective not possible just a few years ago. Capitalizing on these advances, this study has integrated into one comprehensive laboratory experiment a vast array of previously defined and substantively interrelated research avenues that have spanned across seven decades of statistical inquiry.

For this mathematical simulation, a mainframe computer randomly created sets of data which were checked to assure that they violated no data set assumptions. These data sets were then perturbed algebraically to simulate the following mathematical conditions: skewness within the dependent variable, kurtosis, heterogeneity of group variances, and (for the ANCOVA analyses only) heterogeneity of group regression slopes and a skewed covariate. Specifically, three degrees of skewness were imposed on the dependent variable data (no skew, moderate skew and extreme skew); while three degrees of kurtosis were also imposed on the data (platykurtic, mesokurtic and leptokurtic). All skewness and kurtosis conditions were simulated both singly and in combination except for two: extremely skewed and platykurtic distributions and extremely skewed and mesokurtic distributions. These two were not possible to create mathematically for technical reasons (see Johnson, 1993; Fleishman, 1978).

Four different group variance ratios were imposed on the dependent variable data representing four different degrees of differences in group variances: homogeneity of group variances (group variance ratio of 1:1:1), a slight degree of heterogeneity of variances (group variance ratio of 1:1/2:3), a moderate degree of heterogeneity of variances (group variance ratio of 1:2:3), and extreme heterogeneity of variances (group variance ratio of 1:3:5). For the ANCOVA simulations, vectors of data were created for the covariate vector as well, simulating both homogeneity and heterogeneity of group regression slopes and a normally distributed and moderately skewed covariate. In addition, four experimental conditions were simulated: one balanced design using three groups of size 15, one balanced design using three groups of size 30, one balancd design using three groups of size 45, and one unbalanced design using three groups with unequal sizes (15, 30 and 45 per group).

In the end, every single and compound violation of each of these combinations were simulated in the data sets created by the computer for each of the three balanced designs and the one unbalanced design. FORTRAN subroutines from the International Mathematical and Statistical Libraries (IMSL) were then used to run ANOVA and ANCOVA on each of the simulated data sets. This procedure was run 4000 times, allowing the creation of F sampling distributions - most containing 4000 F ratios each. The sampling distributions created in the presence of the 665 different data set conditions resulting from this process were then compared against the F sampling distributions for the appropriate degrees of freedom derived using normal theory. In the end, this allowed for the direct comparison of what actually occurs in the presence of known violations of the data set assumptions with what would have happened if the data sets violated no assumptions. (Note: a complete, detailed description of the simulation methodology can be found in "The Effects of Single and Compound Violations of the Data Set Assumptions When Using the Oneway, Fixed-Effects Analysis of Variance and the One Concomitant Analysis of Covariance Statistical Procedures;" Johnson, 1993).

FINDINGS AND CONCLUSIONS

For all of the analyses, comparisons were made between the empirical F sampling distributions and the theoretical (i.e., "nominal") F distributions expected using normal theory. Specific results (including tables containing the differences between the theoretical and nominal F distributions for each of the 665 F sampling distributions) can be found in the complete paper by Johnson (1993). The discussion below will be limited to tying together the specific results of this simulation with the existing theory.

About Balanced Designs

Previous research (Glass, Peckham and Sanders, 1972; Harwell, Hays, Olds, and Rubinstein, 1990; etc.) suggest that heterogeneity of variances is the greatest single threat to robustness. Conventional thought suggests that when a balanced ANOVA or ANCOVA design is used, problems arise only when the ratio of largest to smallest group variances exceeds three. Meta-analytic findings by Harwell et al., however, suggested differently: they argued that balanced designs may suffer from inflated type I error rates when the ratio between the largest and smallest group variances is as small as two.

The group variance ratios used in this simulation were chosen to directly compare Harwell et al.'s claim against the standard set by Glass et al. two decades ago. No support was found for Harwell's claim; quite the contrary, there were almost no significant differences found in any of the balanced designs, even when the ratio between the largest and smallest group variance was as high as five.

The results of this simulation when using balanced designs suggests a robustness far beyond that proposed by Glass et al. The unique methodology employed in this study may help to explain why. As part of the data generating process, the base vectors (which were later used to create the various data set perturbations) were tested to see if they were significantly different from zero skew and kurtosis. If they were significantly different, then they were disgarded and new vectors created in their place - vectors which again were checked to assure that they were not significantly skewed or kurtotic. This procedure increased the probability that the algebraic perturbations imposed on the base vectors were truly what they are proported to be. Following removal of this sampling noise, the causes for the differences that remained were easier to isolate and interpret.

Most of the studies that Glass et al. reviewed, on the other hand, used a methodology whereby parent populations with the desired characteristics were created and repeated random samples were drawn. No check was made to insure that the samples drawn possessed the mathematical properties being tested. Therefore, when significant differences emerged between the empirical and theoretical F distributions, it was unclear to what degree the differences were the result of known mathematical characteristics and at what point they became the product of selected samples that, by the luck of the draw, possess mathematical properties far different from their parent populations.

Within the balanced design simulations, the only significant difference between the empirical and nominal F distributions in this simulation that did occur was found with the smallest group n's (group n's of 15 for each of the three groups). Using the ANOVA, there were no significant differences at all, even with this small group size, however one data set condition almost achieved significant difference: specifically the extremely skewed and leptokurtic data distributions coupled with extreme heterogeneity of group variances. In the ANCOVA simulations, however, statistically significant differences did occur when the extremely skewed leptokurtic data distributions were coupled with extreme heterogeneity of group variances, a normally distributed covariate and either homogeneous or heterogeneous regression slopes. When balanced designs with larger groups (group n's of 30 and 45) were simulated, no significant differences emerged.

The fact that the only significant differences that did arise in the balanced designs did so among the small group size is worth noting. As has been mentioned previously, the data vectors originally created by the IMSL subroutine were tested to see if they were signifiantly different from zero skew and kurtosis. This testing procedure was done by calculating the 95% confidence intervals for zero skew and zero kurtosis for the appropriate sample size. If the original vectors created by IMSL had skew and/or kurtotic values that fell outside of these confidence bands, then they were disgarded and new ones created in their stead. This screening procedure was, of course, used to screen out samples that had mathematical characteristics different from those that what they were supposed to be. However, because of the mechanics of the process, confidence bands are widest when the sample size is small. It is possible that some samples that should have been disgarded were not because of the wide confidence bands. If this is the case, then the origin of the significant differences that emerged in the small sample size simulations remains unclear: are they the result of violations of the assumptions under test, or are they the result of the inclusion of extreme samples with mathematical characteristics far different from those being tested?

Games and Lucas (1966) suggested that a skewed dependent variable is a greater threat to robustness than either a leptokurtic or platykurtic dependent variable. Additionally, they have suggested that the validity of the F test improves for leptokurtic distributions but suffers for platykurtic distributions. Distributional shape, however, did not prove to be a major factor in influencing type I error rates in this simulation.

Potthoff (1965) suggests that a non-normal concomitant increases the sensitivity of F to departures from normality in the dependent variable. Surprisingly, this research found the opposite: the small (but statistically insignificant) differences that did emerge found analyses using the normal covariate - not the skewed - to be most sensitive to distortions in the dependent variable.

Unbalanced Designs

Although balance designs turned out to be very robust, unbalanced designs did not prove to be very robust at all. Statistically significant differences emerged in the face of almost all combinations except a few that involved only perturbations of shape. In both the ANOVA and ANCOVA simulations, significant differences emerged (at the $p<.01$ level) even when the heterogeneity of group variances was minimal (group variance ratio of 1: 1/2 2). Previous research (e.g., Scheffe', 1959; Shields, 1976) have suggested that when heterogeneity ov variance is coupled with unequal n's, the effect of the violation of equal variances will differ in nature depending on whether the larger group is paired with the larger group variance, or the larger group is paired with the smaller group sample. This trend did, in fact, emerge in this simulation. For the ANOVA analysis, when the largest variance was paired with the smallest group size, all sampling distributions were significantly less than the theoretical F distribution, most at the $p<.01$ level. When the smallest group contained the smallest variance, however, the opposite trend developed: sampling distributions having heterogeneity of variances were found to be significantly greater than theoretical F at the $p < .01$ level. This trend emerged in the ANCOVA analyses involving equal group regression slopes as well.

In the ANCOVA situation involving unequal group regression slopes, the effect of additivity gets more complex, however. For instance, in the unequal n simulations, the smallest regression slope is always paired with the smallest group size for all of the analyses. When this combination (which should increase the number of type I errors made) occurs jointly with heterogeneous variances where the smallest variance is found in the smallest group (which should decrease the number of type I errors), the net effect is a wash out; that is, no significant differences remain. Conversely, when the combination of the smallest slope and group size is paired with the largest variance, the number of type I errors increasd dramatically - higher than either one of the violating conditions alone could have produced.

Concluding Remarks

In summary, for balanced designs the ANOVA and ANCOVA F test was found to be remarkably robust when faced with most of the violations included in this simulation. The degree to which the F test was robust, however, was surprising. The procedure remained robust even when the ratio of largest to smallest group variance was as high as five. After the systematic removal of sampling noise due to the chance creation of skewed and/or kurtotic base vectors, F was found to be far more robust than previously believed. This research, however, reaffirms once again the findings of many previous studies that suggest that ANOVA and ANCOVA be avoided when group sizes are not equal.

In terms of specific recommendations to researchers using balanced designs, the ratio of the largest to smallest group variances should continue to be checked. If the ratio is less than three, then there is no need to fear statistically invalid results due to any of the data set violations included here. If the ratio is between 3 and 5, however, the researcher should test to see if his or her dependent data is within the 95% confidence bands surrounding zero skew and kurtosis. If the dependent's skew and kurtosis values are within this range, then the F statistic should still be sufficiently robust. If, however, either the skew or kurtotic values fall outside of the 95% confidence bands, then the researcher should consider the use of a statistical procedure that has less stringent assumptions.

BIBLIOGRAPHY

Atiquallah, M. (1964). The robustness of the covariance analysis of a one-way classification. Biometrika, 51, 365-372.

Bliss, C.I. (1967). Statistics in Biology. New York: McGraw - Hill.

Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I: Effect of inequality of variance in the oneway classification. Annals of Mathematical Statistics, 25, 290 - 302.

Box, G.E.P. & Anderson, S.L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. Journal of the Royal Statistical Society, 17, 1 - 26, June, 1955.

Cochran, W.G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics, 3, 22 - 38.

Cochran, W.G. (1957). Analysis of covariance: Its nature and uses. Biometrics, 13, 261 - 281.

Elashoff, J.D. (1969). Analysis of covariance: A delicate instrument. American Educational Research Journal, 6, 383 - 401.

Evans, S.H. & Anastasio, E.J. (1968). Misuse of analysis of covariance when treatment effects and covariate are confounded. Psychological Bulletin, 69, 225 - 234.

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

Fleishman, A.I. (1978). A method for simulating non-normal distributions. Psychometrika, 43, 521-532.

Games, P.A. & Lucas, P.A. (1966). Power of the analysis of variance of independent groups on nonnormal and normally transformed data.

Gibbons, J.D. & Chakrabort, S. (1992). Nonparametric Statistical Inference, 3'rd Ed. New York: Marcel Dekker.

Glass, G.V., Peckham, P.D. & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. Review of Educational Research, 42 (3), 237 - 288.

Halpin, G. & Halpin, G. (1988). Evaluation of research and statistical methodologies. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY, November.

Hammersley, J.M. & Handscomb, D.C. (1967). Monte Carlo methods. London: Methuen & Co.

Harwell, M.R., Hayes, W.S., Olds, C.C. & Rubinstein, E.N. (1990). Summarizing Monte Carlo results in methodological research: the oneway, fixed-effects ANOVA case. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Lord, F. M. & Novick, M.R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison - Wesley.

McClaren, V.R. (1972). An investigation of the effects of violating the assumption of homogeneity of regression slopes in the analysis of covariance model upon the F-statistic. Unpublished doctoral dissertation, North State Texas University, Denton, TX.

McLean, J.E. (1974). An empirical examination of analysis of covariance with and without Porter's adjustment for a fallible covariate. Unpublished doctoral dissertation, University of Florida.

McLean, J.E. (1979). The Care and Feeding of ANCOVA. Paper presented at the annual meeting of the Mid-South Educational Research Association in Little Rock, AR, November.

McLean, J.E. (1989). ANCOVA: A Review, Update and Extension. Paper presented at the annual meeting of the Mid-South Educational Research Association in LIttle Rock, November.

Pearson, E.S. (1929). The distribution of frequency constants in small samples from nonnormal symmetrical and skew populations. Biometrika, 19, 151 - 164.

Peckham, P.D. (1968). An investigation of the effects of non-homogeneity of regression slopes upon analysis of covariance. Unpublished doctoral dissertation, University of Colorado.

Potthoff, R.F. (1965). Some Scheffe´ type tests for some Behrens-Fisher type regression problems. Journal of the American Statistical Association, 60, 1163 - 1190.

Raaijmakkers, J.G. & Pieters, J.P.M. (1987). Measurement error and ANCOVA: Functional and structural relationships. Psychometrika, 52, 4, 521 - 538.

Scheffe´, H. (1959). The analysis of variance. New York: John Wiley & Sons.

Shields, J.L. (1978). An empirical investigation of the effect of heteroscedascity and heterogeneity of variance on te analysis of covariance and the Johnson-Neyman technique. Army Technical Paper 292, Project no. 2Q762722A777, July.

Student, (W.S. Gossett). (1908). The probable error of the mean. Biometrika, 6, 1 - 25.

Winer, B.J. (1962). Statistical principles in experimental design. New York: McGraw-Hill.

15A