DOCUMENT RESUME

ED 091 413 TM 003 625

AUTHOR Porter, Andrew C.; McSweeney, Maryellen
TITLE Comparison of Three Common Experimental Designs to Improve Statistical Power When Data Violate Parametric Assumptions.
PUB DATE [74]
NOTE 40p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974)

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Analysis of Covariance; Analysis of Variance; *Comparative Analysis; Goodness of Fit; *Hypothesis Testing; *Nonparametric Statistics; *Statistical Analysis; *Statistics; Tests of Significance

ABSTRACT

A Monte Carlo technique was used to investigate the small sample goodness of fit and statistical power of several nonparametric tests and their parametric analogues when applied to data which violate parametric assumptions. The motivation was to facilitate choice among three designs, simple random assignment with and without a concomitant variable and randomized blocks, and between nonparametric or parametric tests. The criteria for choice were power and robustness. The parameters of the Monte Carlo investigation were strength of relationship between the concomitant and dependent variables, number of levels of the independent variable, sample size, and location parameter. (Author)

# Comparison of Three Common Experimental Designs to Improve Statistical Power When Data Violate Parametric Assumptions

Andrew C. Porter and Maryellen McSweeney
National Institute of Education — Michigan State University



Researchers have long recognized that an increase in statistical precision may result from the judicious use of information on relevant antecedent variables. Both the analysis of covariance and the analysis of variance of an index of response of the form Y - KX have been advocated as precision gaining analyses for a simple random assignment design (Fisher, 1925, 1935). An increase in precision may also result from the use of the antecedent variable as a blocking variable in a randomized block design. Gourlay (1953) compared the effectiveness of the analysis of covariance (ANCOVA) and the analysis of variance (ANOVA) of indices of response relative to simple ANOVA on the dependent variable to determine the extent to which precision was improved by use of information on the antecedent variable. Cox (1957) examined the same question for ANCOVA, ANOVA of correct and incorrect indices of response, and ANOVA of a randomized block design. Porter and McSweeney (1970, 1971) examined all four parametric procedures--ANCOVA, ANOVA of correct and incorrect indices of response, ANOVA of a randomized block design, and simple ANOVA on the dependent variable--and their nonparametric analogues under conditions for which the parametric tests would be optimal. The findings of these simulation studies and the earlier studies by Gourlay and Cox apply to data in which:

(1) the dependent variable, Y, is conditionally normally distributed with equal variances for each of the t treatment groups;

(2) the regression equations of Y on X for each treatment group are linear with equal slopes;

(3) the errors are independent.

An analytic study of the robustness of ANCOVA to violation of its assumptions (Atiqullah, 1964) demonstrated that use of the ANCOVA model

$$y_{ij} = \mu + \alpha_j + \beta(X_{ij} - \bar{X}) + e_{ij}$$

in the presence of a quadratic component of regression of Y on X leads to serious bias in the estimation of the treatment effects and the mean squares for $t > 2$ treatments. Surprisingly, the presence of identical normal distributions on the concomitant variable X and a large number of treatment groups--remedies the reader might assume would mitigate the effects of bias--are inadequate to overcome the bias in the mean squares for ANCOVA. Cox (1957) observed, "...if the regression is non-linear but smooth, blocking methods will remain effective, while covariance methods will not unless the linear component accounts for most of the regression..." (p.157). Sukhatme and Sukhatme's (1954) comparison of two sampling procedures and their associated estimation techniques in a sample survey context has implications for the comparison among the previously mentioned experimental designs and their respective analyses. Sukhatme and Sukhatme examined the precision of the simple regression estimate of the mean in a simple random sampling design relative to the stratum-weighted mean estimate in a stratified sample. Such a comparison is analogous to examining the precision of ANOVA of an index of response for a simple random assignment design relative to ANOVA of a randomized block design. The authors observed:

> "...the stratified sample will, in general, furnish a more efficient estimate than the simple regression method. The relationship between Y and X is also not always found to be linear in practice, in which case the efficiency of the regression estimate is further reduced. For, while stratified sampling with suitably chosen strata can take care of any type of relationship, the regression estimate can eliminate only the effects of the linear component of the relationship." (p.210).

Various writers on experimental design have observed that when the magnitude of the linear correlation between the concomitant variable and the dependent variable is less than .3, the use of ANCOVA will gain little if any precision relative to ANOVA in a simple random assignment design. For linear relations such that $\rho < .4$, blocking is preferable to covariance analysis, while for values of $.4 < \rho < .6$, there is no clear preference between ANOVA of a randomized block design and ANCOVA of a simple random assignment design. When $\rho > .6$ ANCOVA is somewhat more precise than ANOVA of a randomized block design, and at $\rho > .8$, ANCOVA is substantially more precise than ANOVA on blocked data. (Cox, 1957; Bancroft, 1968; Elashoff, 1969). The findings of Atiqullah, Cox, and the Sukhatmes suggest that in the presence of a substantial non-linear relationship, ANOVA of a randomized block design may evidence equal or greater statistical precision than ANCOVA or ANOVA of an index of response.

Reliance on the questionable robustness to nonlinearity of ANCOVA and ANOVA of an index of response is one option available to the researcher seeking increased statistical precision through the use of a concomitant variable. Another option consists of selecting a statistical procedure which uses the concomitant information but does not require linearity. Randomization tests for the analysis of variance of a randomized block design (Baker and Collier, 1968) meet these criteria as does a newly developed randomization test for the analysis of covariance (Robinson, 1973). Rather than work with randomization procedures, we have chosen to retain the relative but not the actual magnitude of the observations through ranking and the use of nonparametric tests based on ranks. A nonparametric test statistic satisfying these restrictions is available as an analogue for each of the parametric procedures reviewed. The four nonparametric techniques considered are:

(1) The Kruskal-Wallis test of equality of mean ranks for a one way ANOVA design (KW);

(2) The Friedman test on ranks in a randomized block design (Fr);

(3) ANOVA of an index of response on mean deviated ranks, $d_y - \rho_s d_x$, (NI1). Indices with underestimated slopes of $.8\rho_s$ (NI2) and overestimated slopes of $1.2\,\rho_s$ (NI3) are also examined.

(4) ANCOVA on ranks (NC).

The first two nonparametric statistics are well-known (c.f. Conover, 1971); ANOVA of an index of response and ANCOVA were constructed by applying parametric procedures to the ranks. Monte Carlo studies by the authors (1971a, 1971b) verified that the small sample properties of the ANCOVA based on applying parametric procedures to ranks were comparable to those of Quade's (1967) nonparametric ANCOVA which adjusts the dependent variable ranks on the basis of a total sample regression estimate. Because the small sample properties of the two nonparametric ANCOVAs were comparable, the test based on a direct analogy to parametric procedures was chosen for its greater familiarity and ease of computation via standard parametric ANCOVA techniques. The index of response on the mean-deviated ranks was created to take advantage of the retained degree of freedom when the regression slope for Y given X can be specified a priori rather than requiring estimation from the sample. Our past Monte Carlo studies demonstrate that the randomized blocks design analyzed by Friedman's ANOVA is a successful method for improving power over that for a simple random assignment design analyzed by the Kruskal Wallis when the correlation between the blocking variable and the dependent variable is greater than .4. When the correlation is equal to .4, power is not a relevant dimension for choosing between the two designs and when correlation is zero the simple random assignment design analyzed by the Kruskal-Wallis test is the more powerful. Nonparametric ANCOVA on data in a simple random assignment design is more powerful than Friedman's ANOVA on data in a randomized blocks design for all values of the correlation between the concomitant variable and dependent variable. Moreover the nonparametric ANCOVA is equal in power to the Kruskal-Wallis when $\rho_{XY} = .0$, but becomes progressively more powerful than the Kruskal-Wallis for increasing values of $\rho_{XY}$. The use of ANOVA on a correctly determined index of response proved more powerful than either the analysis of covariance or the analysis of variance on indices of response with over-estimated or under-estimated slopes for nonparametric tests. Overestimation of the slope impaired the power of the test more seriously than did underestimation, and the effects of overestimation became more severe as $\rho_{XY}$ increased.

All of the parametric tests were slightly more powerful than their nonparametric counterparts for data which completely satisfied parametric assumptions. As the linear correlation $\rho_{XY}$ increased, the discrepancy between the power of the parametric and nonparametric tests increased, but the relationships among the analyses of covariance and analyses of variance on the indices of response for both parametric and nonparametric tests were unchanged by increasing $\rho_{XY}$. The fact that the relative advantage of the parametric tests was slight even when the assumptions necessary for their valid use were completely satisfied suggests that little loss, and possibly considerable gain, in power will result from the more general use of these nonparametric analogues. Since the nonparametric tests require only the identity of the marginal distributions of X and the monotonicity of the XY relationship, they may well be preferred to the parametric tests whenever there is doubt as to whether the conditional distributions of Y are normal, the regressions linear or the variances equal.

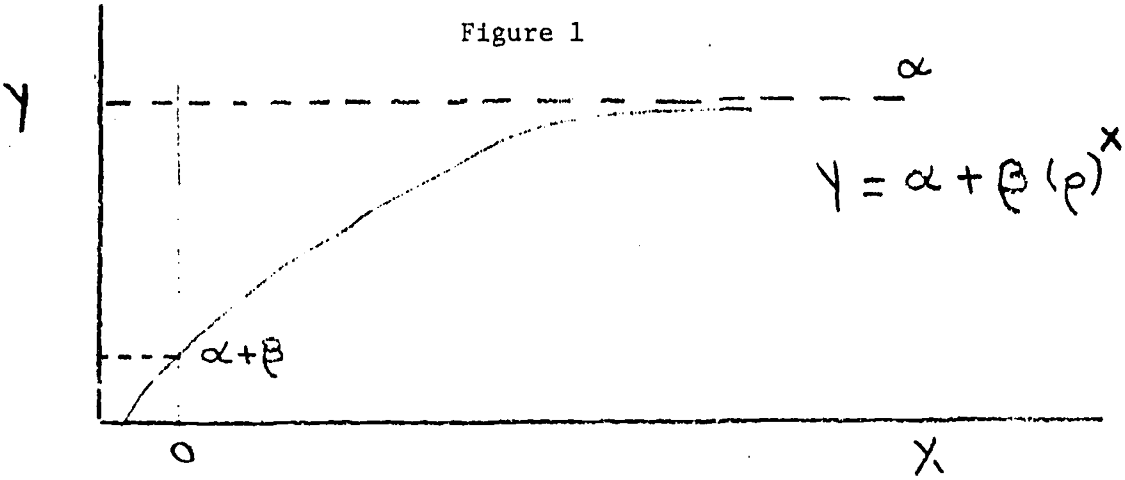
Our previous investigations of the small sample properties of the nonparametric statistics and their parametric counterparts were restricted to data in close agreement with the parametric assumptions. The single exception was our decision to use a normally distributed random concomitant variable rather than a fixed concomitant variable. The exception was motivated by our concern for the utility of the results since educational researchers rarely if ever have data available on a fixed concomitant variable measured on an ordinal or ratio scale. The purpose of the present study was to replicate our previous investigations of goodness-of-fit and small sample power of parametric and nonparametric tests with a data generation model that further violated the parametric assumptions while still satisfying the nonparametric assumptions. Thus the only restrictions on the selection of a data generation model were that it provide observations on independent experimental units for two monotonically related variables, that the observations on each variable allow rank ordering without ties, and that there be no treatment by concomitant variable interaction. Again the decision was made to have a random rather than fixed concomitant variable. Clearly a wide variety of models satisfy these less restrictive nonparametric assumptions.

The selection of a data generation model is one of the most important steps in conducting a Monte Carlo study since the utility of the subsequent results is in large part dependent upon that choice. Given the above restrictions, our selection of a general model was guided by the desire to simulate data that were consistent with those likely to be encountered by educational researchers. Our choice was the asymptotic regression curve or modified exponential defined by the equation

$$Y = \alpha + \beta(\rho)^{X} .$$

The asymptotic regression or modified exponential is a general form of growth curve with three parameters, $\alpha$, $\beta$, and $\rho$. In our use of this curve, the random variables X and Y represent the concomitant variable and the dependent variable respectively. When $\rho$ is restricted to values between zero and one and $\beta$ is negative, the curve rises from $\alpha + \beta$ at X=0 to the asymptote of α for large X. A member of the family of asymptotic regression curves satisfying these properties is presented in Figure 1.

Figure 1




When α=0, the curve simplifies to the exponential, $Y = \beta(\rho)^x$, which describes a series that changes by a constant ratio ρ. Exponential or modified exponential (asymptotic regression) growth seems more reasonable to investigate than second-degree polynomial growth if we want to simulate that type of curvilinearity resulting from the operation of a ceiling effect on the dependent variable measure. (Croxton and Cowden, 1960).

When the asymptotic regression curve is used to represent the underlying relationship between the concomitant and dependent variable, the full model for simulating data is

$$Y = \alpha + \beta(\rho)^X + \gamma Z$$

where α, β, ρ and γ are constant,

$$X \sim N(\mu_x, \sigma_x^2)$$

and

$$Z \sim N(0, 1).$$

The model represents a bivariate population with a normal marginal distribution of X, normal and homoscedastic conditional distributions of Y given X, and a negatively skewed marginal distribution of Y. The negatively skewed marginal distribution of Y violates the normality assumption of both one way and two way ANOVA, and the monotone increasing curvilinear relationship between X and Y violates the linearity assumption made by ANCOVA and ANOVA of an index of response. Educational researchers encounter similar data when there is a ceiling effect on the dependent variable either because of a test which was too easy, or because of a true ceiling effect on the latent dependent variable. One example of the latter situation is the use of mental age as a dependent variable and chronological age as a covariable.

Further specification of the data generation model required assigning specific values to each of the several constants in the equation. We wanted an asymptotic regression curve and a distribution of X such that the linear correlation between X and Y would be near .7 which we felt represented a substantial departure from one. A computer program was written to simulate data conforming to the asymptotic regression curve with X normally distributed. The output of the program was the Pearson correlation between X and Y($\rho_{XY}$) and the first four moments of the distribution of Y. Using samples of size 10,000 the values of $\rho$ were systematically varied for $\alpha=0$, $\beta=-1$, $\mu_x=0$ and $\sigma_x = 1$. Setting $\alpha=0$ simplified the curve to the exponential, $Y = \beta\rho^x$. As $\rho$ increased from .1 to .9, $\rho_{XY}$ increased from .29 to .95. Although $\rho = .3$ resulted in a desirable value of $\rho_{XY} = .72$, the standardized third and fourth moments for the marginal distribution of Y were -5.94 and 63.78 respectively. We decided to vary the other parameters in search of a similar Pearson correlation but a less substantial deviation from normal for the marginal distribution of Y. A value of $\rho = .5$ had resulted in $\rho_{XY} = .89$ and third and fourth moments of Y equal to - 2.53 and 12.62 respectively. Varying values of $\beta$ from - 1.0 to - 1.5 did not have a noticeable effect on reducing the value of $\rho_{XY}$ nor on the third and fourth moments. A lack of change in correlation and third and fourth moments was found when $\mu_x$ was varied from .0 to 2.0. Increasing $\sigma_x$ from 1.0 to 1.6 did reduce $\rho_{XY}$ to .72 but resulted in a concomitant increase in negative skew and kurtosis to approximately that of the value for $\rho = .3$. When we set $\rho = .4$, $\alpha = 0$, $\beta = -1$ and $\mu_x = 0$, an increase in $\sigma_x$ from 1 to 1.3 resulted in a decrease in $\rho_{XY}$ to .73 with third and fourth moments equal to −4.98 and 39.03. Therefore we used $\rho = .4$, $\alpha = 0$, $\beta = -1.0$, $\mu_x = 0$, and $\sigma_x = 1.3$ for our data generation model. Thus $Y = -.4^{x} + \gamma Z$ and $X \sim N(0, 1.3^2)$. Unfortunately, subsequent runs for the same set of parameters revealed that the third and fourth moments of Y were quite unstable for samples of size 10,000. The implication is that we could have left the standard deviation of X at one without markedly changing the third and fourth moments of Y.

The remaining constant to be defined in the data generation model was $\gamma$, the standard deviation of the conditional distribution of Y given X. The value of $\gamma$ was set according to the desired strength of relationship between X and Y as defined by

$$\eta^2 = 1 - \frac{\sigma^2_{y.x}}{\sigma_y^2} ,$$

where $\sigma_{y.x}^2 = \gamma^2$ is the variance of the conditional distribution of Y given X and $\sigma_y^2$ is the variance of the marginal distribution of Y. Since

$\sigma^2_y \geq \sigma^2_{y.x}$ , let $\sigma^2_y = \sigma^2_{y.x} + C$ where $C \geq 0$. Then

$$\eta^2 = 1 - \frac{\gamma^2}{\gamma^2 + C}$$

and

$$\gamma = \sqrt{\left(\frac{1 - \eta^2}{\eta^2}\right) C}$$

The standard deviation $\gamma$ can be evaluated in terms of the desired $\eta^2$ as soon as C is known. Analytic attempts to solve for $C = \sigma_y^2$ in the reduced model

$$Y = -.4^X \text{ where } X \sim N(0, 1.3^2)$$

failed. Consequently a simulation based on 100,000 observations from the reduced model was used to estimate C = 11.3906. Thus for a given value of $\eta^2$, $\gamma$ was set equal to

$$\gamma = \sqrt{\frac{1-\eta^2}{\eta^2} \; 11.3906}$$

Finally, a pseudo-random unit normal deviate generator for the IBM 370-165 computer was used to provide observations on the random variables X and Z in the data generation model. The unit normal deviate generation involved two stages. First, the multiplicative congruent method was used to generate sixteen pseudo-random numbers from a uniform distribution. Second, the sixteen numbers were summed and linearly rescaled to provide a pseudo-random unit normal deviate via the Central Limit Theorem.

## Design of the Monte Carlo Study

The parameters of the Monte Carlo study were t, the number of treatments; b the number of experimental units in each treatment; and $\eta^2$, the correlation ratio. Figure 2 presents a summary of the conditions under which the sampling distributions of the nonparametric and parametric test statistics were investigated. For each X in Figure 2, sampling distributions based on 1,000 samples of size tb were generated for the null case of no treatment effects and for a single noncentral case. Both central and noncentral sampling distributions were described by the frequencies of the various test statistics falling above their respective critical values for the .10, .05, and .01 levels of significance.

Figure 2

Conditions of the Designs Investigated

| | | | $\eta^2$ | |
|---|---|---|---|---|
| t | b | .4 | .6 | .8 |
| 3 | 5 | | | X |
| | 8 | X | X | X |
| | 10 | | | X |
| 5 | 5 | | | |
| | 8 | | | X |
| | 10 | | | |
| 8 | 5 | | | |
| | 8 | | | X |
| | 10 | | | |

t = number of treatments

b = number of blocks

$\eta^2$ = the correlation ratio between the blocking variable and the dependent variable

The population distributions indicated in Figure 2 were chosen to facilitate comparisons with the results from our earlier studies. Three values of t were investigated because of possible trends in the sampling distributions as the number of levels of the treatment independent variable increase. The smallest value of t was three since the Wilcoxon matched pairs test offers a more powerful alternative to the Friedman test when t equals two. Three values of b were investigated because of possible trends in the sampling distributions as the number of units under each level of the independent variable increases. The smallest value of b was five since exact tests would be more appropriate for smaller values. Three values of the correlation ratio ($\eta^2$ = .8, .6, .4) were investigated since the deviation from parametric assumptions increases with the size of the correlation ratio. The noncentral case was created by adding the value $\frac{1}{2}\sigma_y$ to the dependent variable value of each unit under one level of the treatment independent variable. The choice of $\frac{1}{2}\sigma_y$ was made because it seemed to represent a deviation from the null hypothesis that most educational researchers would wish to notice. Further, it produced intermediate values of power which facilitated comparisons of the various test statistics.

Defining the Indices of Response

Prior to the actual generation of the sampling distributions, a final set of parameters had to be defined. Both the parametric and the nonparametric indices of response require a priori knowledge of the slope of the regression line for predicting the dependent variable from the concomitant variable. The slope for the parametric index, $\beta_{y.x}$, is defined on the original observations while the slope for the nonparametric index, $\rho_s$, is the Spearman correlation. Both parameters were estimated for each of the three values of the correlation ratio by generating 4,000 samples with eight treatment groups and eight units per treatment group in each sample. The large samples (64) were used so that the estimates of $\rho_s$ would be minimally affected by sample size.

The three estimates of each parameter and their standard errors are presented in Table 1. The value of $\beta_{y.x}$ was approximately 1.78 for all three values of eta squared, but the value of $\rho_s$ increased with an increase in eta squared, i.e., .3595 for $\eta^2 = .4$ to .6027 for $\eta^2 = .8$.

Since nonparametric and parametric indices of response require a priori information about the parameters $\rho_s$ or $\beta_{y.x}$, it was also of interest to investigate the sampling distribution of the F test for each procedure when the a priori information was in error. Somewhat arbitrarily we decided to investigate the effect of a priori "guesses" about the slopes that were in error by twenty percent, either too high or too low. The slopes for the incorrect indices are also presented in Table 1.

The extensive simulations necessary for estimating $\rho_s$ and $\beta_{y.x}$ provided interesting additional descriptive information about the data generated by the model. The Pearson correlation between X and Y as well as the first four moments of the marginal distribution of Y for varying values of eta squared are contained in Table 2. The mean was approximately -2.0 for all three values of eta squared. The variances are in quite close agreement with what was predicted using the equation

$$\sigma_y^2 = 11.3906 + \gamma^2,$$

given earlier. Skewness, kurtosis and Pearson correlation increased with increases in eta squared as was expected. Skewness ranged from -1.8721 to -5.3158, kurtosis ranged from 10.4755 to 73.8704 and the Pearson Correlation ranged from .4436 to .6301.

TABLE 1

ESTIMATES OF THE SLOPES FOR DEFINING THE PARAMETRIC AND NONPARAMETRIC INDICES OF RESPONSE

| $\eta^2$ | $\beta_{y.x}$ | S.E. | $.8\beta_{y.x}$ | $1.2\beta_{y.x}$ | $\rho_s$ | S.E. | $.8\rho_s$ | $1.2\rho_s$ |
|---|---|---|---|---|---|---|---|---|
| .4 | 1.7837 | .0115 | 1.4270 | 2.1404 | .3595 | .0018 | .2876 | .4314 |
| .6 | 1.7807 | .0107 | 1.4246 | 2.1368 | .4666 | .0017 | .3733 | .5599 |
| .8 | 1.7864 | .0099 | 1.4291 | 2.1437 | .6027 | .0015 | .4822 | .7232 |

TABLE 2

A DESCRIPTION OF THE SIMULATED DATA FOR THE MARGINAL DISTRIBUTION OF Y FOR VARYING VALUES OF ETA SQUARED AND SAMPLES OF SIZE 256,000

| $\eta^2$ | .4 | .6 | .8 |
|---|---|---|---|
| Mean | - 1.9913 | - 1.9982 | - 2.0030 |
| Variance | 27.9597 | 18.8208 | 14.1016 |
| Skewness | - 1.8721 | - 3.9729 | - 5.3158 |
| Kurtosis | 19.4755 | 64.6519 | 73.8704 |
| Pearson Correlation | .4436 | .5414 | .6301 |

Correlations Within and Across the Treatment Groups for the Simulated Data

Table 3 reports the empirically generated correlations on the original data and on the ranks assigned to the data. Correlations are reported for the case of no treatment effects on the dependent variable (the central case) and for the case in which the dependent variable observations in one of the treatment groups have been increased by the addition of 1/2 $\sigma_y$ units (the noncentral case). The total sample correlation ratio, $E(\hat{\eta}^2)$, is an empirical estimate, based on 1000 data sets, of the expected value of $\hat{\eta}^2$ for the design defined by the specified values of t, b, and $\eta^2$. A comparison of each of the tabled values of $E(\hat{\eta}^2)$ with the respective value of $\eta^2$ shows that the simulation has been quite successful in achieving the desired amount of curvilinearity. The largest difference between the empirical estimate of the expected value of $\hat{\eta}^2$ and the desired parameter is .0257 for $\eta^2 = .6$, and the average size of the difference is only .0085. The total sample Pearson correlation for the underlying model, $E(r_{xy})$, is an empirical estimate of the expected value of $r_{xy}$ when the error-free model $Y = -.4^X$ with $X \sim N(0, 1.3^2)$ is used to describe the X Y relationship. The estimated expected values of $r_{xy}$ are very close to the desired measure of linear relationship, $\rho_{xy} = .7$, and are approximately equal across all designs studied. These data demonstrate that it was possible to keep the amount of linear relationship constant while varying the strength of curvilinear relation. Thus as $\eta^2$ increases, for constant $\rho_{xy}$, curvilinearity increases.

The remaining coefficients are also averages obtained over all 1000 data sets for their respective designs, and they represent empirical estimates of the corresponding mathematical expectations of those correlations in the population. The coefficient $E(r_W)$ is an average across 1000 data sets of the pooled within treatment group measures of association where

$$r_W = SS_{XY_W} \,/\, \sqrt{SS_{XX_W} \; SS_{YY_W}}$$

and $SS_{XY_W}$ represents a sum of squares within groups on the variables X and Y. Since it is a mea re of the linear relationship of X Y, we find, as expected, $E(r_W^2) \leq E(\hat{\eta}^2)$. The Spearman correlation, $E(r_{RW})$ is similarly defined as an average across all 1000 data sets of the pooled within group measures of association. However, the scores are ranked separately on both X and Y across the total sample prior to the computation of the sums of squares. The only comparable measures of association obtained from the application of Pearson and Spearman correlation coefficients to these data are the respective pooled within group estimates given by $r_W$ and $r_{RW}$. When empirical estimates of the expected values of these quantities are found from the summary results of the 1000 replications, $E(r_W)$ is consistently higher than $E(r_{RW})$ by .055 to .075 units, with no clear pattern of decrease in the discrepancy as a function of the total sample size, b t, alone or the value of $\eta^2$ alone. If the joint distribution of X and Y were bivariate normal, we would expect

$$\rho_S = 6\ (\sin^{-1}(\rho_{XY}\ /\ 2))/\pi$$

and $E(r_S) = 6\ (\ \sin^{-1}\rho_{XY} + (n - 2)\sin^{-1}(\rho_{XY}/2)\ )\ /\ (\pi(n + 1)\ )$ .

(Moran, 1948). Kendall (1949) demonstrated that these relationships may be substantially in error for samples from other than a bivariate normal distribution and that no simple modifications exist to express the exact relationship between $\rho_S$ and $\rho_{XY}$ for arbitrarily specified populations. Kendall's illustration of a particularly large discrepancy between predicted and actual values of $\rho_S$ occurs for a skewed, leptokurtic population not unlike the population defined by the dependent variable Y. For our data, use of the formula for $\rho_S$ in terms of $\rho_{XY}$ yields a value of $r_{RW} = .634$ when $E(r_W) = .6513$ ($b=8$, $t=3, \eta^2=.8$). This value is noticeably different from the obtained value $E(r_{RW}) = .5891$. The heavy concentration of Y values in the upper tail of the distribution may be responsible for the sizable differences between $E(r_W)$ and $E(r_{RW})$.

The remaining measure of correlation in the central case, $E(r_S)$, is the average across simulations of the Spearman correlation based on all b t ranks.

As expected, $E(r_S)$ increases with an increase in $\eta^2$ for fixed b t. The values of $r_S$ also increase slightly with an increase in b t, which may be reflective of the diminished effect of discreteness.

The average values of $r_S$ and $r_{RW}$ for the noncentral case can be compared with those for the central case. The values of $E(r_{RW})$ are relatively constant over the two cases with discrepancies of (.0012, .0002, .0018, .0034, .0005, .0009, and .0003) respectively. The total sample values $E(r_S)$ are smaller in the noncentral case than in the central case. The direction of this difference is expected since the addition of values of $1/2\ \sigma_y$ to the dependent variable in one group with no compensating change in the values of X, introduces noise into the X Y relationship.

<u>Goodness-of-Fit and Empirical Power for Varying Values of $\eta^2$ When t = 3 and b = 8</u>

Table 4 compares the empirical sampling distributions of the test statistics for the Kruskal-Wallis (KW), Friedman (Fr), parametric ANCOVA on ranks (NC) and nonparametric indices of response on mean deviated ranks with the correct slope (NI1) and with slopes that are systematically underestimated (NI2) and overestimated (NI3). The three rows labelled "central" depict the goodness-of-fit of the empirical distributions to their respective null distributions. Both the Kruskal-Wallis and Friedman testa are referred to $\chi^2_{t-1}$, the ANCOVA test refers to $F_{t-1,bt-t-1}$, and all of the ANOVAs on indices of response use $F_{t-1,t(b-1)}$. Since all of the sampling distributions were based on 1000 cases, standard errors for the estimated actual alphas can be determined from S.E. = $\sqrt{p(1-p)/1000}$ where p denotes nominal alpha level. For nominal alphas of .10, .05, and .01, the respective standard errors are .009, .007, and .003.

The nominal and empirically estimated actual alphas are in close agreement for the nonparametric ANCOVA and the ANOVAs of correct and incorrect indices of response. Without exception these empirical alphas are within two standard errors of the respective nominal $\alpha$ for all values of $\eta^2$. The Kruskal-Wallis test is noticeably conservative for nominal alpha values of .10 and .01 and slightly conservative for a nominal alpha of .05. The Friedman test is somewhat liberal at $\alpha$ = .10, but its empirical sampling distribution does not depart appreciably from good fit to the chi-square distribution at the other nominal alpha levels. On the basis of these data it can be concluded that the fit of the nonparametric analysis of covariance and analysis of variance of an index of response to their respective F distributions is quite good and is independent of the size of the correlation ratio, $\eta^2$. The fit of the Kruskal-Wallis and Friedman tests to $\chi^2_{t-1}$ is somewhat poorer, although it too seems independent of the size of $\eta^2$.

The goodness of fit of the corresponding parametric analogues:

ANOVA for a one way simple random assignment design (A1);

ANOVA for a two way randomized block design (A2);

ANCOVA for a one way simple random assignment design (PC);

ANOVA of an index of response for a correctly determined index (PI1), an underestimated slope (PI2), and an overestimated slope , (PI3);

to their respective F null distributions can be studied by examining the "central" case reported in the first three rows of Table 5. Fit to the respective null distributions is good throughout the table, with the only exceptions occurring for the index of response when $\eta^2 = .8$ and $\alpha = .01$. In that case all three ANOVAs of the indices yield slightly conservative tests; empirical alphas fall at least two standard errors below the nominal alpha = .01.

The fit of the test statistics to their corresponding null distributions is somewhat better for the parametric than for the nonparametric tests when the comparative goodness-of-fit is described by the relative incidence of values of estimated actual alpha that fall outside the 68 percent and 95 percent probability intervals defined on nominal alphas. For the nonparametric analyses, a total of 17 out of 54 values of estimated actual alphas fall outside their respective 68 percent probability intervals and, of these, 6 values are also outside their corresponding 95 percent probability intervals. For the parametric analyses, 13 of 54 values are not within the corresponding 68 percent probability intervals and, of these, one value falls outside the 95 percent probability interval. Although the parametric test statistics seem to exhibit slightly better fit than do their nonparametric analogues, the differences in absolute discrepancy between nominal $\alpha$ and estimated actual $\alpha$ ,

$$\varepsilon = |\text{nominal } \alpha - \text{estimated actual } \alpha| ,$$

are not large. For nominal $\alpha = .10$, the mean difference in fit between the parametric tests and their nonparametric analogues is $\varepsilon_P - \varepsilon_{NP} = -.005$, with the discrepancy almost entirely attributable to the poor fit of the Kruskal-Wallis and Friedman tests. The corresponding mean differences for $\alpha = .05$ and .01 are +.002 and +.001 respectively. The somewhat poorer fit of the Kruskal-Wallis and Friedman tests may be attributable to the use of the chi-square distribution, the traditional large sample approximation for each of these tests, rather than the F distribution, found by Wallace (1959) to give slightly better fit.

Tables 4 and 5 may also be used to compare the small sample power against a slippage alternative for the nonparametric and parametric tests included in this study. The values appearing in the three rows labelled "noncentral" represent the empirical power of the various tests at $\alpha = .10$, .05, and .01. The maximum standard error of the reported empirical powers, .016, occurs for power equal to .5. To facilitate power comparisons, we have averaged power differences between statistics across the nominal alpha levels. Although the power comparisons at the differing nominal alpha levels have slightly different precision, the differences in precision do not seem substantial enough to negate the use of these summary measures.

When the power of each of the nonparametric procedures which use information on the concomitant variable is compared with that of the Kruskal-Wallis test which makes no use of this information, the former is as high or higher for all values of $\eta^2$, all nominal alpha levels and all nonparametric tests using the antecedent information. The corresponding differences in power relative to the Kruskal-Wallis test for $\eta^2 = .4$ are .024 for the Friedman test, .049 for nonparametric ANCOVA, .055 for an ANOVA of a correct index of response, and .054 for both the underestimated and overestimated slope in ANOVA of an index of response. When $\eta^2 = .6$, the respective differences increase to .054, .062, .069, .067 and .064. Further increases are noted for $\eta^2 = .8$, where the corresponding differences are .088, .147, .159, .159 and .144 respectively.

As expected, the empirical power of nonparametric ANOVA of the correct and incorrect indices of response exceeds that of all other statistics when $\eta^2$ is large. This difference is already noted when $\eta^2 = .4$. The slight superiority of the indices with respect to ANCOVA was anticipated on the basis of the retained degree of freedom in the former technique and its loss in the latter. The superiority of the indices of response and ANCOVA to the Friedman test at $\eta^2 = .4$ is a little surprising if the designs alone are compared. Such a comparison overlooks the fact that the Friedman test does not exploit the existence of interblock differences to gain power. Hodges and Lehmann (1964) hypothesized that the use of intrablock ranking in the Friedman test, with the attendant disregard of interblock differences, resulted in a less than optimal nonparametric test for the randomized block design. Their test, based on intra- and interblock comparisons, is asymptotically more efficient than the Friedman test. We conjecture that because of the exclusive use of intrablock ranks in the Friedman test, the full advantage of blocking may not be gained relative to the use of direct antecedent variable adjustment in ANCOVA and ANOVA of indices of response. The power ANOVA of an index of response using an underestimate slope is comparable to that of an index using the correct slope, while that of an index using an overestimated slope is slightly less. The discrepancies in power between the under- and over-estimated slope increase with increasing $\eta^2$ from .000 to .003 and .012 respectively. We had found differences comparable to these in our earlier work with linear relations. Now, as then, it seems reasonable to ascribe increasing differences in the power of NI2 versus NI3 to underestimation in the average noncentral values of $r_s$. A comparison of the average noncentral values of the total sample Spearman correlation for $\eta^2 = .4$, .6, and .8 (Table 3) with the estimates of $\rho_s$ to define the incorrect slopes (Table 1) reveals that the underestimated slopes, $.8\rho_s$, are consistently closer to the average noncentral total sample $\rho_s$ than are the overestimated slopes, $1.2\ \rho_s$. The differences for the underestimated slopes relative to the average noncentral $\rho_s$ are -.0557, -.0591, -.0688 for $\eta^2 = .4$, .6, and .8 respectively. The corresponding differences for the overestimated slope are .0881, .1275 and .1722 respectively. The comparison of NI2 and NI3, although based on $.8\rho_s$ versus $1.2\rho_s$, employs sample correlations such that the magnitude of the error of underestimation is considerably smaller than that of overestimation. Consequently, the differences in empirical power are consistent with the differing size of the errors in estimating the slope.

A comparison of the empirical power of the nonparametric tests with that of their parametric analogues shows higher power for the nonparametric analysis of covariance and all nonparametric indices of response, irrespective of the value of $\eta^2$. Moreover, the advantage of the nonparametric tests relative to their parametric analogues increases with an increase in the correlation ratio. For the analysis of covariance the average differences in power for the nonparametric techniques versus the parametric techniques vary from .015 at $\eta^2 = .4$ to .011 at $\eta^2 = .6$ and to .071 at $\eta^2 = .8$. The differences for the correct and incorrect indices follow a similar pattern for increasing values of $\eta^2$. In the case of the correct index the values are .008, .014, and .095 respectively; for the index based on the underestimated slope, the respective quantities are .008, .011, and .077; and for the index based on the overestimated slope they are .006, .019, and .115. Since an increase in $\eta^2$ for constant $\rho$ in the underlying model implies an increase in curvilinearity in the model,* the results are consistent in demonstrating the relatively superior power of those nonparametric procedures which assume only monotonicity of the X Y relationship to parametric procedures assuming linearity of the relationship. The comparison of nonparametric to parametric procedures for the randomized block design tends to favor the parametric test slightly, despite the fact that the dependent variable distribution is negatively skewed and extremely peaked. The differences in power for the Friedman test relative to the F test for the randomized block design are - .007, .001, and - .003 for increasing values of $\eta^2$ when averaged over the nominal alpha levels. The relative weakness of the Friedman test in this context occurs primarily in the upper percentiles ($1 - \alpha \geqslant .95$) of the empirical sampling distribution and may be partly attributable to the discreteness of the sampling distribution of the Friedman test. As was noted earlier, the Friedman test does not use the interblock differences that are a source of increased precision for the F test in the randomized block design. This too may be responsible for the lower power relative to the parametric competitor. The Kruskal-Wallis statistic also exhibits somewhat lower power than the F test for one way ANOVA,especially at the .10 and .01 alpha levels. The conservativeness of the Kruskal-Wallis null distribution at these same alpha levels may explain the reduced power relative to the F test. When power differences are averaged over the nominal alpha levels, the results for KW - F are - .012, - .006 and .017 for increasing $\eta^2$. Only at $\eta^2 = .8$, is the power of the Kruskal-Wallis test superior to F's.

Although not of direct concern in this study, the empirical powers of the parametric tests can be examined to determine the extent to which they reflect the effects of curvilinearity. All of the parametric procedures which employ antecedent variable information are more powerful than the one way analysis of variance F test on a simple random assignment design. The following differences in power, averaged over the nominal alpha levels, are found when the power of each of the other parametric procedures is compared with that of the one way ANOVA F test. The differences are reported as three-tuples, with the elements ordered in terms of increasing $\eta^2$:

ANOVA in a randomized block design (.020, .048, .109)
ANCOVA for data in a simple one way design (.023, .046, .093)
ANOVA on a correct index of response (.035, .049, .093)
ANOVA on an index of response with an underestimated slope (.034, .050, .099)
ANOVA on an index of response with an overestimated slope (.031, .039, .049).

The results support the earlier conjecture that as $\eta^2$ becomes large, the analysis of variance of a randomized block design, which does not assume linearity, becomes more powerful than ANCOVA or ANOVA of response on correct or incorrect indices. The differences between the indices of response based on correct and incorrect slopes are slight at $\eta^2$ = .4 or .6 (.001 for the correct versus the underestimated slope and .004 and .010 for the correct versus the overestimated slope). However, at $\eta^2$ = .8, the index with the under-estimated slope is slightly more powerful than the index with the correct slope (.006) and noticeably more powerful than the index with the overestimated slope (.050).

Goodness-of-Fit and Empirical Power for Varying Values of t When b = 8 and $\eta^2$= .8

When the goodness-of-fit of the respective parametric and nonparametric tests is studied for increasing numbers of treatment groups (t = 3, 5, 8) and constant curvilinear relationship ($\eta^2$ = .8), no pronounced trends emerge. (Tables 6 and 7). The majority of the estimated actual alphas for these statistics fall within the 95% probability intervals based on the corresponding nominal alphas, with only five exceptions out of 54 cases for the nonparametric tests and fourteen out of 54 for the parametric tests. The fit is similar when 68% probability intervals are examined: 22 nonparametric estimated actual alphas and 31 parametric empirical alphas fall outside of the respective 68% intervals. Most of the instances of lack of good fit result in the over-estimation of the actual alpha through the use of the nominal alpha. Thus the associated tests are conservative. The fit of the nonparametric tests is somewhat better than that of the parametric procedures. The parametric tests are especially conservative for $t = 8$; however, it seems unlikely that the conservative behavior of almost all the parametric procedures for this design is a consequence of the change in design from $t = 3$ and 5 to $t = 8$. Unreported portions of this simulation study examined the central distribution of the parametric tests for varying t and $\eta^2 = .6$ and did not find comparably conservative behavior. Thus it seems more reasonable to assume that these results are an artifact of the particular 1000 data sets simulated for $t = 8$, $b = 8$, $\eta^2 = .8$ that are used for the analysis of all the parametric and non-parametric tests having these design dimensions.

When the empirical power of the nonparametric tests is examined as a function of an increasing number of treatment groups, the actual magnitude of the power decreases but the relative ordering among the tests remains unchanged. The decrease in empirical power with increased t merely reflects the fact that the single differing treatment group makes up a smaller proportion of the total sample size (1/3 versus 1/5 versus 1/8 for $t = 3, 5, 8$ respectively). The power of each of the other nonparametric tests using information on the concomitant variable is compared with that of the Kruskal-Wallis test and differences are averaged over the nominal alpha levels to yield the following results for increasing values of t:

| | | |
|---|---|---|
| Fr - KW | Power: | (.088, .114, .104) |
| NC - KW | Power: | (.147, .151, .153) |
| NI1 - KW | Power: | (.159, .159, .153) |
| NI2 - KW | Power: | (.159, .156, .154) |
| NI3 - KW | Power: | (.147, .148, .141) . |

The most powerful of the nonparametric techniques for $\eta^2 = .8$ are the analysis of variance on a correct index of response (NI1) and on an index of response with an underestimated slope (NI2). These are followed closely in power by the analysis of covariance (NC) and the analysis of variance on an index of response with an overestimated slope (NI3). The Friedman test (Fr) has substantially more power than the Kruskal-Wallis test in the presence of such a strong X Y relationship, but its power is considerably lower than that of the other procedures which incorporate information about the nature of the X Y relationship more directly into their respective test statistics.

A comparison of these nonparametric tests with their parametric analogues for the substantial curvilinear relationship implied by $\rho = .7$ and $\eta^2 = .8$ reveals that all of the nonparametric tests are more powerful than their parametric counterparts for all values of t. The gains in power for the Kruskal-Wallis and Friedman tests relative to the F tests for one way ANOVA and ANOVA of a randomized block design respectively are quite slight, while those for the other procedures versus the parametric tests demanding linearity are somewhat more substantial. For, increasing values of t

| | | |
|---|---|---|
| KW - A1 | Power: | (.017, .034, .020) |
| Fr - A2 | Power: | (.013, .013, .029) |
| NC - PC | Power: | (.073, .084, .107) |
| NI1 - PI1 | Power: | (.095, .102, .117) |
| NI2 - PI2 | Power: | (.077, .093, .085) |
| NI3 - PI3 | Power: | (.115, .117, .128) |

The strong showing of the nonparametric analysis of variance on an index of response with an overestimated slope is probably reflective of the very conservative behavior of the parametric analogue in the central case, and the reuse of the same data with 1/2 $\sigma_y$ added to one treatment group to create the noncentral case. A very slight increase in the advantage of the nonparametric procedures to the parametric tests occurs with increasing t. Such a finding is consistent with the results of many Monte Carlo studies which have established that nonparametric tests tend to make their best showing relative to the parametric analogues when differences between populations are small and power is low to moderate.

Goodness-of-Fit and Empirical Power for Varying Values of b When $t = 3$ and $\eta^2 = .8$

Tables 8 and 9 provide evidence with respect to the goodness-of-fit of the nonparametric and parametric tests for increasing numbers of observations per treatment group. Both tables indicate relatively good fit of the empirical sampling distributions of the test statistics to their corresponding null distributions. Only 6 of 54 estimated actual alphas for the nonparametric tests and 3 of 54 for the parametric tests fall outside the 95% probability intervals based on the corresponding nominal alphas. The fit is similar when 68% probability intervals are studied: 11 nonparametric and 18 parametric estimated actual alphas fall outside the respective 68% intervals. The quality of the fit does not vary appreciably with b, although there is slightly better fit for the nonparametric tests at larger values of b. The Kruskal-Wallis test and the Friedman test are once again the principal source of lack of fit among the set of nonparametric tests examined. All 5 estimated actual alphas that are outside the limits of the corresponding 95% probability intervals on the nominal alphas are identified with either the Kruskal-Wallis or the Friedman test.

The power of all of the parametric and nonparametric tests considered is a monotonically increasing function of the sample size per treatment group, b. Substantial increases in the power of the tests occur as b increases from 5 to 10, but the relative superiority of the various statistics is only moderately affected by changes in b.

Each of the nonparametric tests using information on the concomitant variable is more powerful than the Kruskal-Wallis test which ignores this information. Furthermore, the gain in power relative to the Kruskal-Wallis test increases with an increase in the number of observations. The analyses of variance on a correct index of response and on an index with an under-estimated slope are again slightly higher in power than the analysis of covariance or the analysis of variance on an index of response with an over-estimated slope and substantially more powerful than the Friedman test. The resulting differences in power relative to the Kruskal-Wallis test, averaged over the nominal alpha levels, are:

Fr - KW Power: (.032, .088, .109) For, increasing values of b
NC - KW Power: (.086, .147, .186)
NI1 - KW Power: (.097, .159, .192)
NI2 - KW Power: (.097, .159, .185)
NI3 - KW Power: (.085, .147, .183) .

All of the nonparametric tests except the Friedman test are more powerful than their parametric counterparts. The Friedman test, which exhibits a conservative null distribution, is slightly less powerful than the F test for a randomized block design. The Kruskal-Wallis test realizes only a slight advantage in power relative to the F test for one way analysis of variance, but this advantage increases with an increase in the number of observations per group. The nonparametric analysis of covariance and the analyses of variance of the indices of response are all substantially more powerful than their parametric analogues, and this gain in power increases with increasing b. These results are indicative of the effect of curvilinearity in decreasing the power of the parametric tests which assume linearity. Our earlier work dealing with similar design dimensions but a data model satisfying parametric assumptions provides an informative comparison with these data employing a curvilinear relationship.

Empirical Power of the Nonparametric Tests Minus That of Their Parametric Analogues Requiring Linearity, Averaged Over α Levels for t = 3, b = 5, 8, 10 and Differing Assumed X Y Relationships

| | Parametric Assumptions Met | | | Curvilinearity Present | | |
|---|---|---|---|---|---|---|
| | | $\rho = .6$ | | $\rho = .7$ | | $\eta^2 = .8$ |
| NC - PC | - .033 | - .043 | - .053 | .041 | .071 | .112 |
| NI1 - PI1 | - .050 | - .058 | - .058 | .056 | .095 | .128 |
| NI2 - PI2 | - .022 | - .050 | - .049 | .053 | .077 | .110 |
| NI3 - PI3 | - .068 | - .077 | - .075 | .055 | .115 | .154 |

As the results indicate, the advantage of the parametric tests increases with increasing b when parametric assumptions are met, but the disadvantage of the parametric tests increases when the X Y relationship is markedly curvilinear.

## Summary of the Empirical Results

The data for this study indicate some consequences, in terms of the estimated Type I Error and power, of the choice between a nonparametric test and a parametric test when there is a nonlinear relation between the dependent variable and the concomitant variable. The empirical sampling distributions of all of the parametric and nonparametric tests employing the F distribution showed relatively good fit to their respective null distributions. The Kruskal-Wallis and Friedman tests, both of which were referred to the chi-square distribution, occasionally demonstrated poor fit. The Kruskal-Wallis test was consistently somewhat conservative, while the Friedman test tended to be conservative in the extreme tail of the distribution but liberal at the α = .10 level.

All of the nonparametric tests were more powerful than the Kruskal-Wallis test at all levels of relationship studied and for all design variations of t and b. Among the nonparametric tests using information on a concomitant variable, the tests could be ranked in order of decreasing empirical power as:

ANOVA of correct index of response $\doteq$ ANOVA of index slope too low $>$ ANCOVA $\doteq$ ANOVA of index slope too high $>$ Friedman test

The differences between an index of response based on a correct slope and one based on an underestimated slope were typically very slight, as were the differences in power between analysis of covariance and analysis of variance of an index of response on an overestimated slope.

When the nonparametric tests were compared with those parametric tests which demand linearity of the X Y relation, the relative superiority of the nonparametric tests increased with increases in the size of $\eta^2$, the number of observations per group, b, and the number of treatments, t. The effect of skewed, leptokurtic distributions on the dependent variable was not sufficiently strong to yield nonparametric tests for the one way ANOVA (KW) and for the randomized block design (Fr) which were consistently more powerful than their parametric analogues. However, for $\eta^2 = .8$, the Kruskal-Wallis and Friedman tests tended to be slightly more powerful than their parametric counterparts.

When the parametric tests were compared among themselves, our conjecture that a sizable amount of nonlinearity might favor the F test for the randomized block design over its parametric counterparts which require a linear X Y relation was substantiated. Thus conclusions regarding the decreasing power of parametric procedures for data in which the X Y relation is linear,

ANOVA of correct index of response $>$ ANCOVA $>$ ANOVA of a randomized block design

are not supported in the nonlinear case. Instead,

ANOVA of a randomized block design $>$ ANOVA of a correct index of response $>$ ANCOVA

The study as a whole supports the viability of selecting a nonparametric test in preference to a parametric one if the relationship between the dependent variable and the concomitant variable is nonlinear. The declared chances of committing a Type I Error adequately describe the actual chances, and the power of the nonparametric analyses of variance of indices of response and of analysis of covariance is higher than that of their parametric analogues.

*We have asserted that, "Since an increase in $\eta^2$ for constant $\rho$ in the underlying model implies an increase in curvilinearity in the model,..." This statement does not imply that $\eta^2$ and $\rho$ are reported for comparable data. The value of $\rho$ pertains to the linear correlation in the error-free model,

$$Y = -\ .4^X \quad \text{with } X \quad N(0,\ 1.3^2)\ .$$

The value $(1 - \rho^2)$ is a measure of the lack of fit of a linear X Y relationship to the function specified by

$$Y = -\ .4^X.$$

The correlation ratio

$$\eta^2 = 1 - \frac{\sigma^2_{y.x}}{\sigma^2_y}$$

is based on the full model which includes normally distributed error,

$$Y = -\ .4^X + \gamma Z \quad \text{with } Z \quad N(0,\ 1)\ .$$

The value $(1 - \eta^2) = \sigma^2_{y.x} \ / \ \sigma^2_y$ is a measure of "pure error" in fitting empirical data to the nonlinear functional X Y relation. Unfortunately, no comparable data are reported for the total sample values of the Pearson correlations. These would measure both "pure error" and lack of fit (Draper and Smith, 1966). Of necessity, comparisons will be made between $E(\eta^2)$ and $E(r_W)$, both computed on the full model, but the former measuring association in the total sample and the latter measuring the pooled within groups linear correlation. Under the null hypothesis of no treatment effects and the assumption of identical distributions on the concomitant variable, needed by the nonparametric tests, $r_W$ and r for the total sample both serve as estimators of the corresponding population Pearson correlation. It can be noted that as $\eta^2$ increases, both $E(\hat{\eta}^2)$ and $E(r_W)$ increase, but at differing rates. Since $r^2 < \hat{\eta}^2$ in the presence of curvilinearity and $r^2 = \hat{\eta}^2$ for linearity, the increasing values of $E(\hat{\eta}^2) - (E(r_W))^2$ with an increase in $\eta^2$ imply an increase in curvilinearity.

References

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

Baker, F. B. and Collier, R. O. "An empirical study into factors affecting the F-test under permutation for the randomized block design," Journal of the American Statistical Association, 63 (1968), 902-911.

Bancroft, T.A. Topics in Intermediate Statistical Methods, Vol. I. Ames, Iowa: The Iowa State University Press, 1968.

Conover, W. J. Practical Nonparametric Statistics. New York: John Wiley,1971.

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

Croxton, F. E. and Cowden, D. J. Practical Business Statistics, 3rd. ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1960.

Draper, N. R. and Smith, H. Applied Regression Analysis. New York: John Wiley, 1966.

Elashoff, J. D. "Analysis of covariance: A delicate instrument," American Educational Research Journal, 6 (1969), 383-402.

Fisher, R. A. Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd, 1925.

Fisher, R. A. The Design of Experiments. Edinburgh: Oliver and Boyd, 1935.

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

Hodges, J. L.,Jr. and Lehmann, E. L. "Rank methods for combination of independent experiments in analysis of variance," The Annals of Mathematical Statistics, 33 (1962), 482-497.

Kendall, M. G. "Rank and product moment correlation," Biometrika, 36 (1949), 177.

McSweeney, M. and Porter, A. C. "Small sample properties of nonparametric index of response and rank analysis of covariance," Paper presented at 1971 AERA convention.

Moran, P. A. P. "Rank correlation and product moment correlation," Biometrika 35 (1948), 203.

Porter, A, C. and McSweeney, M. "Randomized blocks design and nonparametric statistics," Paper presented at 1970 AERA convention.

Porter, A. C. and McSweeney, M. "Comparison of rank analysis of covariance and nonparametric randomized blocks analysis," Paper presented at 1971 AERA convention.

Quade, D. "Rank analysis of covariance," Journal of the American Statistical Association, 62 (1967), 1187-1200.

Robinson, J. "The analysis of covariance under a randomization model," Journal of the Royal Statistical Society, B, 35 (1973), 368-376.

Sukhatme, P. V. and Sukhatme, B. V. Sampling Theory of Surveys With Applications, 2nd. ed., Ames, Iowa: Iowa State University Press, 1970.

Wallace, D. L. "Simplified beta-approximations to the Kruskal-Wallis H test," Journal of the American Statistical Association, 54 (1959), 225-230.

TABLE 3

TOTAL SAMPLE CORRELATION RATIO $E(\eta^2)$, TOTAL SAMPLE PEARSON CORRELATION FOR THE UNDERLYING MODEL $E(r_{xy})$, AND AVERAGE CORRELATIONS WITHIN TREATMENT GROUPS $E(r_W)$, ACROSS ALL RANKED OBSERVATIONS $E(r_S)$ AND WITHIN TREATMENT GROUPS ON RANKED DATA $E(r_{RW})$

| | | .4 | .6 | .8 | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\eta^2$ = | .4 | .6 | .8 | | | | |
| | t = | 3 | 3 | 3 | | | 5 | 8 |
| | b = | 8 | 8 | 5 | 8 | 10 | 8 | 8 |
| CENTRAL | $E(\eta^2)$ | .3920 | .5743 | .8029 | .8010 | .8135 | .8018 | .8068 |
| | $E(r_{xy})$ | .7109 | .7175 | .6978 | .6925 | .6887 | .6876 | .7048 |
| | $E(r_W)$ | .4187 | .5261 | .6445 | .6513 | .6516 | .6592 | .6553 |
| | $E(r_S)$ | .3539 | .4496 | .5698 | .5910 | .5893 | .6038 | .6044 |
| | $E(r_{RW})$ | .3505 | .4499 | .5670 | .5891 | .5893 | .6031 | .6019 |
| NON-CENTRAL | $E(r_S)$ | .3433 | .4324 | .5353 | .5510 | .5522 | .5772 | .5887 |
| | $E(r_{RW})$ | .3517 | .4501 | .5688 | .5857 | .5898 | .6022 | .6022 |

TABLE 4

EMPIRICAL PROPORTIONS OF KRUSKAL-WALLIS (KW), FRIEDMAN (Fr), NONPARAMETRIC ANCOVA (NC), AND NONPARAMETRIC INDICES OF RESPONSE BASED ON KNOWN SLOPE (NI1) AND SLOPES SYSTEMATICALLY IN ERROR (NI2,NI3) TEST STATISTICS FALLING IN THEIR RESPECTIVE REGIONS OF REJECTION FOR $\alpha$ = .10,.05,.01; t = 3; b = 8; $\eta^2$ = .4,.6,.8[@]

| $\eta^2$ = | | .4 | | | | | | .6 | | | | | | .8 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | KW | Fr | NC | NI1 | NI2 | NI3 | KW | Fr | NC | NI1 | NI2 | NI3 | KW | Fr | NC | NI1 | NI2 | NI3 |
| Central | 10 | 081 | 127 | 098 | 095 | 097 | 087 | 072 | 124 | 108 | 105 | 101 | 103 | 025 | 123 | 105 | [illegible] | [illegible] | [illegible] |
| | 05 | 042 | 051 | 051 | 043 | 042 | 040 | 048 | 048 | 057 | 054 | 057 | 058 | 049 | 059 | 054 | 048 | [illegible] | [illegible] |
| | 01 | 004 | 006 | 008 | 011 | 009 | 011 | 003 | 011 | 012 | 010 | 015 | 010 | 006 | 011 | 013 | 011 | 013 | [illegible] |
| Non-Central* | 10 | 224 | 282 | 302 | 308 | 309 | 309 | 230 | 378 | 357 | 366 | 360 | 365 | 373 | 525 | 533 | 563 | 561 | 550 |
| | 05 | 164 | 164 | 194 | 203 | 197 | 207 | 203 | 222 | 248 | 252 | 254 | 248 | 282 | 331 | 433 | 431 | 428 | 413 |
| | 01 | 029 | 044 | 068 | 071 | 072 | 063 | 043 | 078 | 108 | 114 | 113 | 105 | 068 | 131 | 198 | 207 | 211 | 196 |

[@]Decimals omitted for the proportion in this table

*The non-central case was $\mu_1 = \frac{1}{2}\sigma_y; \mu_2, \mu_3, \cdots \mu_t = 0$

TABLE 5

EMPIRICAL PROPORTIONS OF ONE AND TWO WAY ANOVA (A1,A2), PARAMETRIC ANCOVA (PC), AND PARAMETRIC INDICES OF RESPONSE BASED ON KNOWN SLOPE (PI1) AND SLOPES SYSTEMATICALLY IN ERROR (PI2,PI3) TEST STATISTICS FALLING IN THEIR RESPECTIVE REGIONS OF REJECTION FOR $\alpha$ = .10,.05,.01; t = 3, b = 8; $\eta^2$ = .4,.6,.8[©]

| $\eta^2$ = | | .4 | | | | | | .6 | | | | | | .8 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | A1 | A2 | PC | PI1 | PI2 | PI3 | A1 | A2 | PC | PI1 | PI2 | PI3 | A1 | A2 | PC | PI1 | PI2 | PI3 |
| Central | 10 | [illegible] | 094 | 098 | 092 | 100 | 099 | 092 | 099 | 096 | 096 | 093 | 098 | [illegible] | 092 | 104 | [illegible] | [illegible] | [illegible] |
| | 05 | 049 | 045 | 043 | 042 | 044 | 042 | 051 | 055 | 052 | 049 | 051 | 054 | 041 | 048 | [illegible] | [illegible] | [illegible] | [illegible] |
| | 01 | 005 | 010 | 005 | 007 | 006 | 006 | 004 | 012 | 010 | 012 | 012 | 012 | 007 | 009 | 007 | 003 | [illegible] | [illegible] |
| Non-Central* | 10 | 248 | 267 | 279 | 298 | 297 | 294 | 287 | 364 | 343 | 357 | 353 | 345 | 348 | 499 | 469 | 455 | 476 | 410 |
| | 05 | 157 | 182 | 186 | 203 | 198 | 196 | 189 | 238 | 243 | 242 | 245 | 232 | 241 | 355 | 333 | 327 | 342 | 293 |
| | 01 | 047 | 062 | 055 | 056 | 060 | 054 | 067 | 084 | 094 | 091 | 096 | 084 | 082 | 143 | 148 | 135 | [illegible] | [illegible] |

[©]Decimals omitted for the proportions in this table

*The non-central case was $\mu_1 = \frac{1}{2}\sigma_y; \mu_2, \mu_3, \cdots \mu_t = 0$

TABLE 6

EMPIRICAL PROPORTIONS OF KRUSKAL-WALLIS (KW), FRIEDMAN (Fr), NONPARAMETRIC ANCOVA (NC), AND NONPARAMETRIC INDICES OF RESPONSE BASED ON KNOWN SLOPE (NI1) AND SLOPES SYSTEMATICALLY IN ERROR (NI2, NI3) TEST STATISTICS FALLING IN THEIR RESPECTIVE REGIONS OF REJECTION FOR $\alpha = .10, .05, .01$; $\eta^2 = .8$; b = 8; t = 3,5,8[@]

| | t = | 3 | | | | | | 5 | | | | | | 8 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | KW | Fr | NC | NI1 | NI2 | NI3 | KW | Fr | NC | NI1 | NI2 | NI3 | KW | Fr | NC | NI1 | NI2 | NI3 |
| Central | 10 | 035 | 123 | 105 | 107 | 095 | 106 | [illegible] | 093 | 099 | 104 | 105 | [illegible] | 064 | 093 | 092 | [illegible] | [illegible] | [illegible] |
| | 05 | 049 | 059 | 054 | 048 | 052 | 052 | 037 | 046 | 054 | 053 | 062 | 052 | 028 | [illegible] | 042 | [illegible] | [illegible] | [illegible] |
| | 01 | 006 | 011 | 013 | 011 | 013 | 010 | 004 | [illegible] | 013 | 015 | 014 | 013 | [illegible] | 006 | 007 | [illegible] | [illegible] | [illegible] |
| Non-Central* | 10 | 373 | 525 | 533 | 563 | 561 | 550 | 341 | 489 | 508 | 520 | 511 | 503 | 277 | 438 | 469 | 465 | [illegible] | [illegible] |
| | 05 | 282 | 331 | 433 | 431 | 428 | 418 | 199 | 339 | 373 | 385 | 372 | 380 | 163 | 285 | 337 | 345 | 342 | [illegible] |
| | 01 | 068 | 131 | 198 | 207 | 211 | 196 | 066 | 119 | 179 | 177 | 191 | 166 | 049 | 077 | 141 | 139 | 145 | 135 |

@ Decimals omitted for the proportions in this table

* The non-central case was $\mu_1 = \frac{1}{2}\sigma_y$; $\mu_2, \mu_3, \cdots \mu_t = 0$

TABLE 7

EMPIRICAL PROPORTIONS OF ONE AND TWO WAY ANOVA (A1,A2), PARAMETRIC ANCOVA (PC), AND PARAMETRIC INDICES OF [illegible] BASED ON KNOWN SLOPE (PI1) AND SLOPES SYSTEMATICALLY IN ERROR (PI2,PI3) TEST STATISTICS FALLING IN THEIR RESPECTIVE REGIONS OF REJECTION FOR $\alpha$ = .10,.05,.01; $\rho^2$ = .8; b = 8; t = 3,5,8[@]

| | t = | 3 | | | | | | 5 | | | | | | 8 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | A1 | A2 | PC | PI1 | PI2 | PI3 | A1 | A2 | PC | PI1 | PI2 | PI3 | A1 | A2 | PC | PI1 | PI2 | PI3 |
| Central | 10 | 086 | 092 | 104 | 092 | 089 | 099 | 081 | 087 | 102 | 086 | 089 | 094 | 085 | 073 | [illegible] | [illegible] | [illegible] | [illegible] |
| | 05 | 041 | 048 | 044 | 044 | 043 | 043 | 044 | 033 | 048 | 038 | 038 | 043 | 043 | 033 | 035 | 032 | [illegible] | [illegible] |
| | 01 | 007 | 009 | 007 | 003 | 004 | 004 | 007 | 003 | 007 | 009 | 006 | 009 | 009 | 003 | 005 | [illegible] | [illegible] | [illegible] |
| Non-Central* | 10 | 348 | 499 | 469 | 455 | 476 | 410 | 279 | 474 | 397 | 382 | 391 | 357 | 239 | 391 | 335 | 328 | [illegible] | [illegible] |
| | 05 | 241 | 355 | 333 | 327 | 342 | 293 | 171 | 328 | 279 | 276 | 279 | 249 | 143 | 270 | 219 | 208 | 216 | [illegible] |
| | 01 | 082 | 143 | 148 | 135 | 150 | 115 | 054 | 133 | 132 | 117 | 126 | 093 | 046 | 101 | 071 | 061 | 069 | [illegible] |

[@] Decimals omitted for the proportions in this table

*The non-central case was $\mu_1 = \frac{1}{2}\sigma_y; \mu_2, \mu_3, \cdots \mu_t = 0$

TABLE 8

EMPIRICAL PROPORTIONS OF KRUSKAL-WALLIS (KW), FRIEDMAN (Fr), NONPARAMETRIC ANCOVA (NC), AND NONPARAMETRIC INDICES OF RESPONSE BASED ON KNOWN SLOPE (NI1) AND SLOPES SYSTEMATICALLY IN ERROR (NI2,NI3) TEST STATISTICS FALLING IN THEIR RESPECTIVE REGIONS OF REJECTION FOR $\alpha$ = .10,.05,.01; $\eta$ = .8; t = 3; b = 5,8,10[@]

| | b = | 5 | | | | | | 8 | | | | | | 10 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | KW | Fr | NC | NI1 | NI2 | NI3 | KW | Fr | NC | NI1 | NI2 | NI3 | KW | Fr | NC | NI1 | NI2 | NI3 |
| Central | 10 | 078 | 124 | 098 | 099 | 106 | 094 | 085 | 123 | 105 | 107 | 095 | 106 | 081 | 097 | 107 | 107 | [illegible] | [illegible] |
| | 05 | 046 | 037 | 051 | 052 | 056 | 044 | 049 | 059 | 054 | 048 | 052 | 052 | 052 | 046 | 048 | 053 | 056 | 052 |
| | 01 | 000 | 002 | 011 | 010 | 007 | 009 | 006 | 011 | 013 | 011 | 013 | 010 | 008 | 011 | 012 | 013 | 011 | 013 |
| Non-Central* | 10 | 245 | 353 | 355 | 369 | 372 | 354 | 373 | 525 | 533 | 563 | 561 | 550 | 446 | 590 | 654 | 656 | 637 | 651 |
| | 05 | 178 | 166 | 244 | 257 | 254 | 241 | 282 | 331 | 433 | 431 | 428 | 418 | 351 | 441 | 526 | 526 | 523 | 519 |
| | 01 | 007 | 008 | 088 | 096 | 096 | 089 | 068 | 131 | 198 | 207 | 211 | 196 | 107 | 199 | 283 | 298 | 298 | 282 |

[@]Decimals omitted for the proportions in this table

*The non-central case was $\mu_1 = \frac{1}{2}\sigma_y; \mu_2, \mu_3, \cdots \mu_t = 0$

TABLE 9

EMPIRICAL PROPORTIONS OF ONE AND TWO WAY ANOVA (A1,A2), PARAMETRIC ANCOVA (PC), AND PARAMETRIC INDICES OF RESPONSE BASED ON KNOWN [illegible] (PI1) AND SLOPES SYSTEMATICALLY IN ERROR (PI2,PI3) TEST STATISTICS FALLING IN THEIR RESPECTIVE REGIONS OF REJECTION FOR $\alpha$ = .10,.05,.01; $\rho^2$ = .8; t = 3; b = 5,8,10@

| | b = | 5 | | | | | | 8 | | | | | | 10 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | A1 | A2 | PC | PI1 | PI2 | PI3 | A1 | A2 | PC | PI1 | PI2 | PI3 | A1 | A2 | PC | PI1 | PI2 | PI3 |
| Central | 10 | 085 | 082 | 106 | 093 | 087 | 090 | 086 | 092 | 104 | 092 | 089 | 099 | 078 | 091 | 106 | 106 | 099 | 105 |
| | 05 | 035 | 043 | 052 | 040 | 042 | 041 | 041 | 048 | 044 | 044 | 043 | 043 | 043 | 041 | 052 | 053 | 051 | 055 |
| | 01 | 011 | 012 | 012 | 008 | 007 | 013 | 007 | 009 | 007 | 003 | 004 | 004 | 009 | 008 | 014 | 011 | 012 | 014 |
| Non-Central* | 10 | 229 | 309 | 296 | 295 | 313 | 283 | 348 | 499 | 469 | 455 | 476 | 410 | 392 | 586 | 544 | 526 | 531 | 493 |
| | 05 | 136 | 194 | 193 | 193 | 201 | 176 | 241 | 355 | 333 | 327 | 342 | 293 | 268 | 461 | 403 | 395 | 406 | 343 |
| | 01 | 041 | 057 | 070 | 067 | 059 | 059 | 082 | 143 | 148 | 135 | 150 | 115 | 104 | 242 | 180 | 174 | 190 | 150 |

@Decimals omitted for the proportion in this table

*The non-central case was $\mu_1 = \frac{1}{2}\sigma_y; \mu_2, \mu_3, \cdots \mu_t = 0$