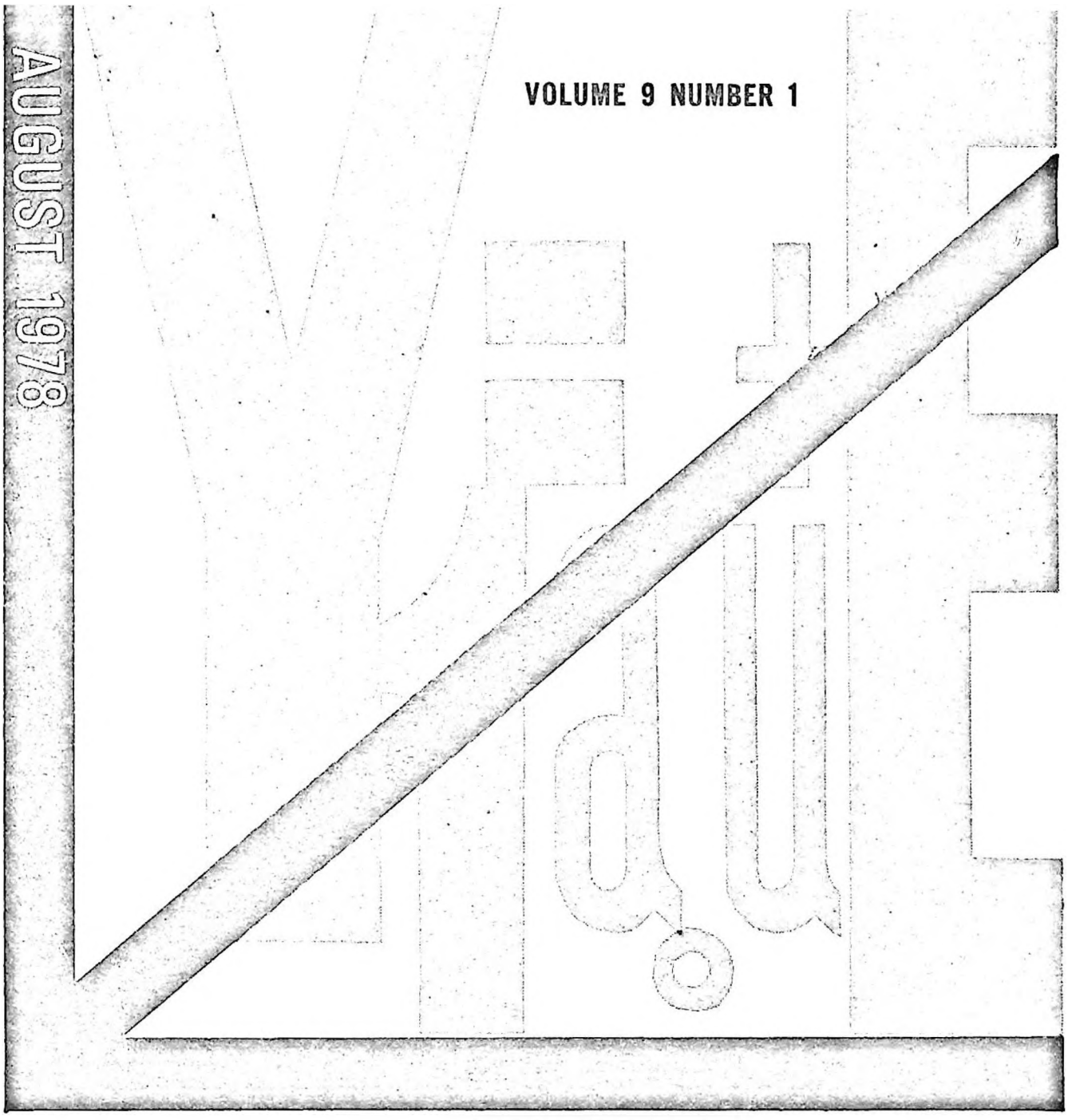
AUGUST 1978

VOLUME 9 NUMBER 1



MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group
on Multiple Linear Regression

MULTIPLE LINEAR REGRESSION VIEWPOINTS

EDITORIAL BOARD

TABLE OF CONTENTS

Multiple Linear Regression Viewpoints
Vol. 9 No. 1 1978

THE MALPRACTICE OF STATISTICAL INTERPRETATION

John W. Fraas
Ashland College

Isadore Newman
The University of Akron

Abstract

This paper examines problems that researchers may confront when interpreting statistical research results. The first section of the paper examines the problems associated with the use of gain scores. The second portion of the paper examines why the use of analysis of covariance is superior to the analysis of gain scores in aiding the researcher to avoid misinterpreting the data. The third section of the paper discusses the problem of disproportionality as it produces multicollinearity. The fourth section of the paper examines the difference between the interpretation of research results analyzed by part correlation as opposed to partial correlation. The final section presents a brief discussion of the effect of violating the assumption of rectilinearity in the regression effect.

PROBLEMS ASSOCIATED WITH GAIN SCORES

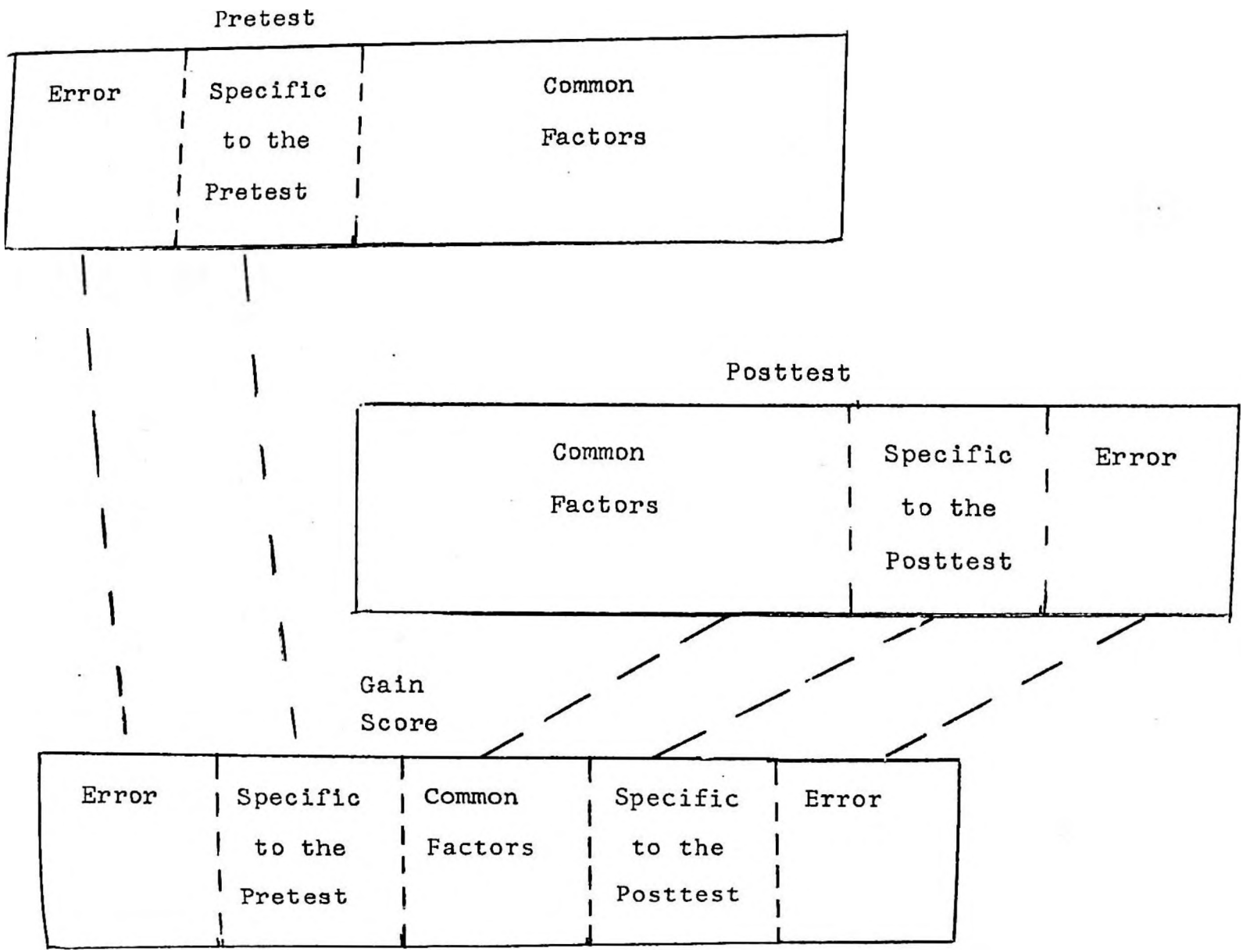
In the education research literature one often finds that gain scores, which are equal to posttest scores minus pretest scores, are used as a dependent variable. Unfortunately, gain scores tend to be unreliable (Gulliksen, 1950). This unreliability has two sources: (a) the unreliability of the pretest and the posttest, and (b) the correlation between the pretest and posttest.

The reasons why gain scores tend to be unreliable can better be understood with the aid of Figure 1. The first bar in Figure 1 represents the variance in performance on an economic knowledge pretest. The segments of this bar represent the different factors that account for the variance in student performance on the pretest.

Presented at the 87th Annual Meeting of The Ohio Academy of Science, Wright State Univeristy, Dayton, Ohio, April, 1978.

Figure 1

Variance in Gain Scores



The second bar represents the variance in an economic knowledge post-test. Similar to the segments of the first bar, the segments of the second bar represent the different factors that account for the variance in student performance on the posttest. Both the pretest and post-test bars have been divided into three factors. The factors entitled "Error" are the chance error of the measurements of the tests. The factors entitled "Specific Factors" are the factors that are measured only on that given test. And the factors entitled "Common Factors" are the factors measured by both tests.

The variance in the gain scores are represented by the third bar. Only the specific factors and the random error factors of each test remain in the gain scores. As indicated in Figure 1, the proportion of random error variance to the total variance of gain scores is larger than is the proportion of random error to the total variance of either the pretest or the posttest. One can also deduce from Figure 1 that only the error variance remains in the difference scores when the pretest and posttest measures exactly the same factors. Such a result would produce totally unreliable gain scores.

The reliability of gain scores can be improved, however, by changing two factors. First, an increase in the reliability of the pretest and/or the posttest would increase the reliability of the gain scores. Second, a decrease in the correlation between the pretest and posttest would increase the reliability of the gain scores. The importance of these two factors in determining the reliability of gain scores can be seen by examining the equation used to calculate the reliability coefficient for gain scores (Thorndike and Hagen, 1969).

The equation used to calculate the reliability coefficient for gain scores is as follows:

$$\text{Diff.} = \frac{\left(\frac{r_{11} + r_{22}}{2}\right) - r_{12}}{1 - r_{12}}$$

where r_{11} is the reliability of pretest.

r_{22} is the reliability of the posttest.

r_{12} is the correlation between the pretest and posttest.

As indicated by this equation, the reliability estimate for the gain scores would only be .40 when both the pretest and posttest had reliability estimates of .70 and the correlation between the pretest and posttest was .50. The reliability of the gain scores would increase if the reliabilities of the pretest and posttest were increased. However, the correlation between the tests would also increase if the reliability of each test was increased. Thus, the reliability of the gain scores would not increase by the degree that one might expect when the reliabilities of the pretest and posttest were increased. For example, if the reliabilities of the tests were increased from .70 to .80 and the corresponding correlation between the pretest and posttest increased from .50 to .60, the reliability of the gain scores would only increase to .50.

Gain scores of individuals are not of major interest in many research studies. The researcher is often interested in mean change of the groups. Again scores may be used as the dependent variable when dealing with mean changes of groups. However, the researcher

may encounter a problem when using gain scores in this manner in addition to the problem that gain scores are generally unreliable. An analysis of gain scores contains a built-in-bias in favor of the group that contained the students with the lower pretest scores. In general, the students with the lowest pretest scores will record the largest gain scores. Although this bias is not likely to be large unless the sample size is small, it would be preferable to avoid this bias altogether.

Both of the problems connected with the use of gain scores can be avoided if the researcher used analysis of covariance to analyze the difference between the students posttest scores. The following section of this paper disscuses the advantage of using analysis of covariance.

The Advantages of Using Analysis of Covariance

As discussed in the previous section, analysis of covariance offers an option to the use of gain scores. There are at least two advantages connected with the use of analysis of covariance as opposed to the use of gain scores in addition to avoiding the problems associated with the use of gain scores discussed in the previous section. These two advantages are: (a) analysis of covariance tends to be more powerful and (b) it is relatively simple to determine if the underlying assumptions of analysis of covariance have been violated.

Huck and McLean (1975) noted that covariance analysis will normally result in a more sensitive test of possible differences

among treatments. The use of the pretest scores will generally reduce the within-group variability and, therefore, increase the power of the test. The only time where covariance analysis will not be more powerful than an analysis of gain scores occurs when the correlation between the pretest scores and posttest scores is equal to a value of one. The following equation can be used to understand the impact that a perfect relationship between the pretest and posttest scores has on the relative power level of analysis of covariance and analysis of gain scores (Huck & McLean, 1975):

$ss'_{wg} = ss_{wg}$ (gain) $- (b_w - 1)^2$ ss (pretest) where

ss'_{wg} represents the adjusted within-groups sums of squares of the gain scores.

ss_{wg} (gain) represents the within-groups sums of squares of the gain scores.

b_w represents the within-groups regression coefficient for predicting the posttest scores from pretest scores.

ss_{wg} (pretest) represents the within-groups sum of squares of the pretest scores.

If $b_w = 1$, the product $(b_w - 1)^2 ss_{wg}$ (pretest) is equal to zero and the adjusted within-group sums of squares for the analysis of covariance is equal to the within-group sums of squares for analysis of gain scores $[ss'_{wg} = ss_{wg}$ (gain)]. Therefore, the sums of squares value in the denominator of the F ratio would be the same for analysis of covariance and gain scores. Since the analysis of covariance has one fewer degrees of freedom in the denominator of the F-ratio, the analysis of gain scores would be slightly more powerful than the co-

variance analysis in this case. However, the value for b_w will probably never be equal to one in an applied setting. Therefore, the adjusted within-groups sums of squares of the posttest scores would be smaller than the within-groups sums of scores of the gain scores by a factor related to the relationship between the pretest and posttest scores $[ss_{wg} < ss_{wg}\ (gain) - (b_w - 1)\ ^2ss_{wg}\ (pretest)]$. Analysis of covariance will, therefore, be more powerful in an applied setting than would an analysis of gain scores.

Just as important as the additional power provided by the analysis of covariance, the use of analysis of covariance allows the researcher to more easily test the underlying assumption of covariance. The failure to determine if the underlying assumptions of analysis of covariance have been satisfied may provide the setting in which the researcher may misinterpret the data. A researcher may be tempted to use gain scores to avoid the need to worry about satisfying the assumptions of analysis of covariance. However, Huck and McLean (1975) noted that gain scores and analysis of covariance are based on the same set of assumptions. It is, therefore, important to determine if the underlying assumptions are met regardless if the analysis of covariance or analysis of gain scores is used.

The analysis of covariance, however, provides an easier and more interpretable method of testing the underlying assumptions than does the analysis of gain scores. In a study conducted by Fraas (1978) the importance of testing for the homogeneity of regression slopes and the ease with which these tests can be handled by using analysis of covariance was demonstrated. In the study by Fraas two methods

of instruction were evaluated. The following procedure was used to test for the homogeneity of the regression slopes of the groups across the range of the students' pretest scores when evaluating the students' posttest scores:

1. The research hypothesis was established.

H_1: There was a statistically significant interaction between the methods of instruction and the students' pre-course economic knowledge when accounting for the variation in the students' post-course economic knowledge over and above the differences due to the students' pre-course economic knowledge and methods of instruction.

2. The following regression models were constructed:

Full Model: $X_4 = a_oU + a_1X_1 + a_2X_2 + a_5X_5 + a_6X_6 + E_1$

Restriction: $a_5 = a_6$

Restricted Model: $X_4 = a_oU + a_1X_1 + a_2X_2 + a_3X_3 + E_2$

where X_1 = the control group (1 if the control group; o otherwise)

X_2 = the experimental group (1 if in the experimental group; o otherwise)

X_3 = pretest scores

X_4 = posttest scores

$X_5 = X_1 * X_3$

$X_6 = X_2 * X_3$

$a_o, a_1 \ldots a_6$ = the regression coefficients of their respective variables

E_1, E_2 = the error vectors

3. The research hypothesis could be tested by placing the restriction $a_5 = a_6$ into the full model to obtain the restricted model

and calculating the F value by using the following formula:

$$F = \frac{(R_F^2 - R_R^2)/(m_1 - m_2)}{(1 - R_F^2)/(N - m_1)}$$

where R_F^2 = the total variance in the criterion vector that was accounted for by the predictor variables in the full model

R_R^2 = the total variance in the criterion vector that was accounted for by the predictor variables in the restricted model

m_1 = the number of linearly independent vectors in the full model

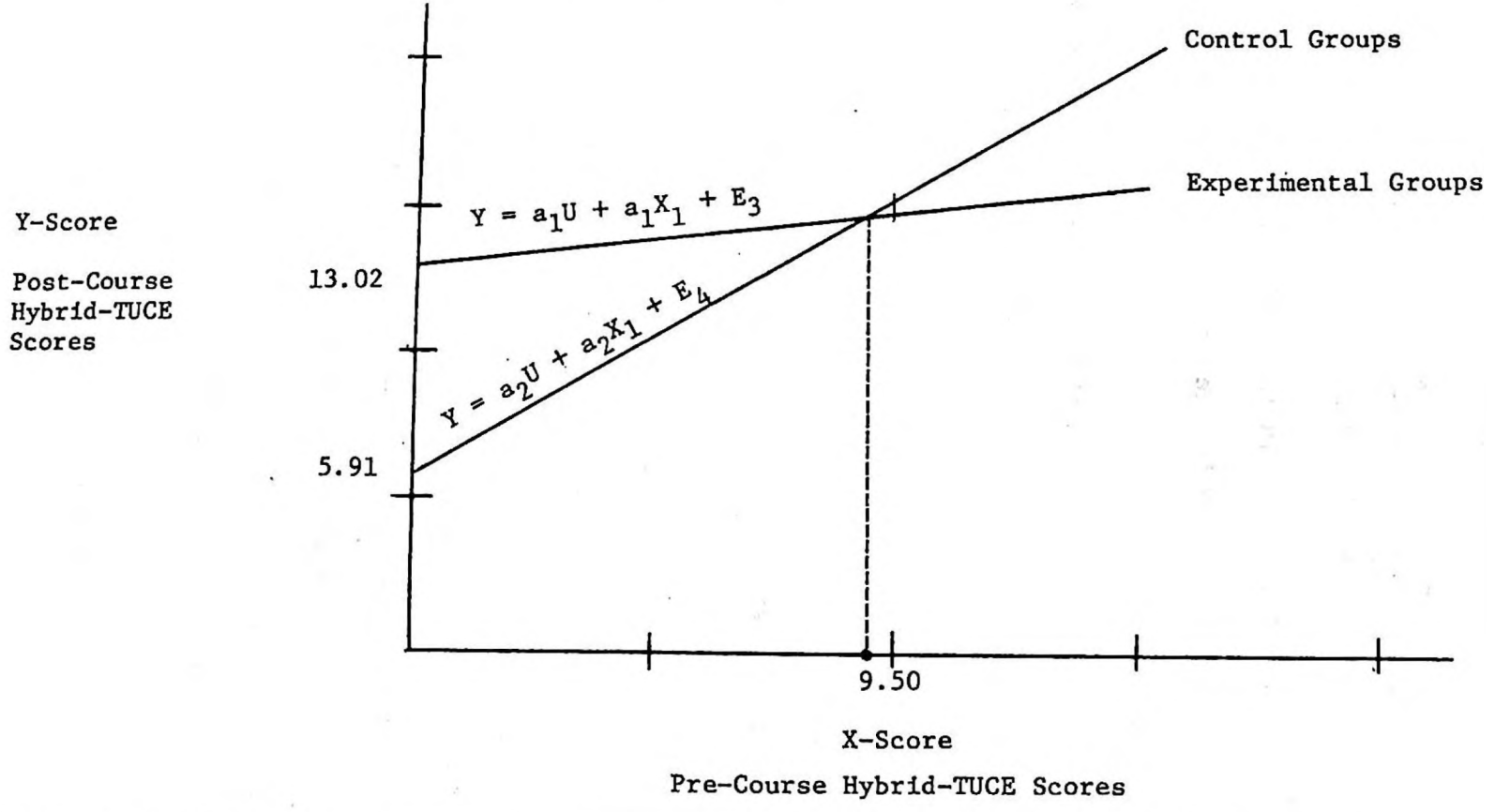
m_2 = the number of linearly independent vectors in the restricted model

N = the number of students (replication)

4. A significat F-value would have indicated that the slopes of the regression lines were not homogeneous. A diagram of the test results, which indicated that the research hypothesis was true, would give additional insight into the nature of the data. The graph contained in Figure 2 was obtained by plotting the values of the regression coefficients a_1, a_2, a_5, and a_6. The values of the coefficients a_1 and a_2 would have been equal to the y-intercepts of the regression lines for the control and experimental groups, respectively. The values of the coefficients a_5 and a_6 would have been equal to the slopes of the regression lines for the control and experimental groups. As can be seen from the graph in Figure 2, the interaction of the treatments and the pretest scores was disordinal. The effectiveness of each

FIGURE 2

THE INTERACTION EFFECT BETWEEN THE STUDENTS' PRE-COURSE HYBRID-TUCE SCORES AND THE METHODS OF INSTRUCTION WHEN ACCOUNTING FOR THE VARIATION IN THE STUDENTS' POST-COURSE HYBRID-TUCE SCORES



Note: The point were the two lines cross can be determined by solving the sum

Equation: $$\left\{\begin{array}{l} Y = a_1U + a_1X_1 \\ -Y = a_2U + a_2X_2 \\ \hline 0 = a_1 - a_2 + (a_1 - a_2)X \end{array}\right\} = X = \frac{a_2 - a_1}{a_1 - a_2};$$

X will be the point where the lines cross.

method of instruction was, therefore, dependent on the pretest scores of the students. A misinterpretation of the data could have easily been made if the researcher would not have tested for the homogeneity of regression slopes.

Summary: The use of analysis of covariance with the posttest scores as the dependent variable is preferable to analysis of gain scores for a number of reasons. First, gain scores are less reliable than posttest scores. Second, gain scores tend to "favor" the group with the lowest pretest scores. Third, the analysis of gain scores is less powerful than analysis of covariance. Finally, it is easier to test the underlying assumptions of analysis of covariance and interpret these results than it would be to test the underlying assumption of the analysis of gain scores.

PROBLEMS ASSOCIATED WITH DISPROPORTAIONAL CELL FREQUENCIES

Researchers generally believe that they must correct for unequal N's. One method for correction is to randomly throw out S's from the larger cells to obtain proportionality. A second method is to run the analysis by calculating the means of each group and acting as if there was an N of 1 in each cell, but using all S's to estimate the mean square within.

Methods 1 and 2 are generally referred to as approximate solutions to the correction of disproportionality and are generally not considered as desirable as the solutions that are called "exact."

When using one of the so-called "exact" solutions, one tends to believe the corrections are more "accurate." This quite often is not

the case since the "exact solution" may not reflect the research question of interest. (i.e., The statistical model does not reflect the research question. Type VI Error, Newman, et. al., 1976.)

The following are examples of three frequently-used corrections and the different questions they reflect. (Newman and Oravecz, 1977)

Three prominent least sum of squares solutions for disproportionality will be defined.

Solution I is the use of the general linear model to simultaneously adjust for the correlations between the main effects and the main effects with interaction. A symbolic example of this procedure is presented below for a two factorial design.

Model 1 $Y_{kab} = \delta + b_1\alpha_a + b_2\beta_b + b_3\alpha\beta_{ab} + \varepsilon_{kab}$

Model 2 $Y_{kab} = \delta + b_4\beta_b + b_5\alpha\beta_{ab} + \varepsilon_{kab}$

Model 3 $Y_{kab} = \delta + b_6\alpha_a + b_7\alpha\beta_{ab} + \varepsilon_{kab}$

Model 4 $Y_{kab} = \delta + b_8\alpha_a + b_9\beta_b + \varepsilon_{kab}$

Y_{kab} = is the score for subject k in row a and column b

δ = is the grand $\overline{X}$

α_a = is the effect for row "a"

β_b = is the effect for column "b"

$\alpha\beta_{ab}$ = is the interaction effect for the row "a" and column "b"

ε_{kab} = is the error term for each subject

$b, \ldots b_n$ are partial regression coefficients

Adjustment for Solution #1

Adjustment for A main effects test Model 1 against Model 2

Adjustment for B main effects test Model 1 against Model 3

Adjustment for A*B effects test Model 1 against Model 4

Solution 2 adjusts each main effect in terms of the other main effects. The interaction was adjusted for by all main effects. (This is the same as in Solution 1). The following is a symbolic representation of this solution.

Adjustment for Solution #2

Model 4 $Y_{kab} = \delta + b_{10}\alpha_a + b_{11}\beta_b + \varepsilon_{kab}$

Model 5 $Y_{kab} = \delta + b_{12}\beta_b + \varepsilon_{kab}$

Model 6 $Y_{kab} = \delta + b_{13}\alpha_a + \varepsilon_{kab}$

Adjustment for A main effects test Model 4 against Model 5

Adjustment for B main effects test Model 4 against Model 6

Adjustment for AB interaction effects test Model 4 against Model 1

Solution 3 assumes an apriori ordering of the importance of the variables under investigation. The apriori ordering decides which variables one will allow to account for as much variance as possible by themselves. The following is a symbolic representation of Solution 3, assuming the researcher considers the A main effects most important, B main effects second, and the interaction leaat important. (Many researchers feel that it is unlikely that most investigators will be able to order the importance of their variable. However, we believe this judgment can be made by a competent researcher who is aware of the underlying constructs and theories he is dealing with.)

Adjustment for Solution #3

Model 7 $Y_{kab} = \delta + b_{14}\alpha_a + \varepsilon_{kab}$

Model 8 $Y_{kab} = \delta + \varepsilon_{kab}$

Model 9 $Y_{kab} = \delta + b_{15}\alpha_a + b_{16}\beta_b + \varepsilon_{kab}$

Adjustment for A main effects test Model 7 against Model 8

Adjustment for B main effects test Model 9 against Model 7

Adjustment for AB interaction test Model 1 against Model 9

Marks (1974) presents a mathematical proof that one has to have A main effects and B main effects in the full and restricted models to test for AB interaction. It seems that this is true when dealing with traditional analysis of variance and catagorical variables. That is, the A main effects and B main effects must be fitted first before interaction can be tested. However, this does not seem to be necessarily true when dealing with continuous variables. To the best of our knowledge, this has never been investigated. At least testing the AB interaction without holding the A and B main effects constant is asking a different question. Then when one is partialing out the main effects, we feel it's appropriateness depends on the research questions being asked.

What is even more confusing is the argument presented by Timm and Carlson (1975) and Williams (1977) for the full-rank solution, (i.e., a correction that does not assume a zero interaction) rather than the non full-rank solutions. Solution 1, presented above, for the main effects is one type of full-rank solution. Part of the problem with the concept of full-rank solution arises when different data coding methods are used. The test of effects of a traditional ANOV is most accurately reflected by a full-rank solution when contrast coding of data is used. A different outcome occurs when binary coding is used. Even though the R^2s of the full models using both coding systems would be the same, the contrast coding procedures have a more direct interpretation in terms of the traditional ANOV; Kerlinger (1973), Cohn & Cohn (1975), Williams (1977). However, the binary coding (dummy coding) has greater heuristic value and increases ease of hypotheses testing, Williams (1977), Kerlinger (1973), McNeil, Kelly, McNeil (1975).

The advantage of the full-rank solution compared to the non full-rank solutions (such as Solutions 2 and 3 presented above) is that the full-rank solution does <u>not</u> assume a zero interaction when testing for main effects; however, traditionally, if one has interaction in ANOV, one generally does not desire to interpret main effects. Therefore, what seems to be an obvious advantage of the full-rank solution may not be such an advantage.

Newman and Oravecz (1977), using a Monte Carlo Research Design, found that if the χ^2 for testing significantly different proportionality of S's in cells was not significant at $\alpha = .25$. There did not seem

to be a strong need to correct for disproportionality (at least in a 2 x 2 ANOV) only when the χ^2 was found to be significant at $\rho \leq .05$ was there a strong need to use a correction.

The researcher should remember that the corrections for disproportionality are appropriate only for fixed effects designs and there is no appropriate correction for randomized designs, to the best of my knowledge.

Newman and Oravecz (1977) state:

"When the researcher feels disproportionality is severe enough to be of concern, there are a variety of procedures that he can utilize to attempt to correct for the potential problems. However, before any corrections are applied, one should be sensitive to the underlying assumptions that they are making about the population from which their data is drawn, and the investigator must also be very clear about the research question he is interested in asking.

If one had a research project in which the data and variables came from groups that already exist, such as age, intelligence, socio-economic status, etc., and if one was interested in generalizing and predicting back to the group from which the sample came, there is a good possibility that there would be a correlation that would not be spurious between such variables as I.Q. and socio-economic status. In other words, there may be significantly more above average socio-economic status people who have above average I.Q.'s than one would expect by chance. If the data

were forced to correspond to a balanced design in which there are an equal number of high and low I.Q. people for an equal number of high and low socio-economic status positions, the result of the study and the statistical analysis may allow one to say something that may only be true for that artificially forced relationship and one could not properly generalize to the population in which this proportionality did not actually exist.

The other side of the coin is if the disproportionality in a research design is an artifact, (it really does not exist in the population) and the disproportionality of vectors is causing a spurious correlation between the variables, then one would have to adjust for this disproportionality and would have to decide which solution of disproportionality would best adjust the data so that it would better reflect the question(s) of interest and the true state of affairs.

To be able to begin to decide upon the correct solution, one has to:

a. know something about the theoretical and/or empirical relationship between the variables being studied;

b. know some of the descriptive data about the population one wishes to generalize to in relation to the specific variables being studied;

c. know the specific research question under investigation if one decides an adjustment for disproportionality is needed, then,

d. know the underlying assumptions and implications for different adjustment procedures, and

e. know the consequences for using the selected adjustment procedure on the interpretation and generalization of the data."

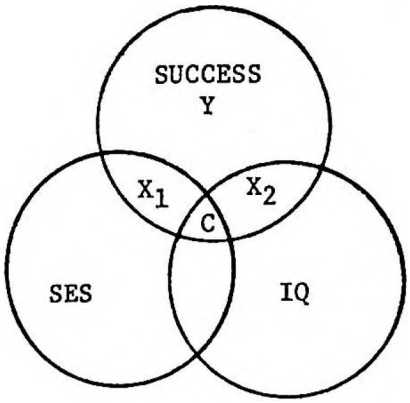
In addition, still assuming that one is dealing only with a fixed effect design, one alternative is to consider that the correlation between the variables, due to the disproportionality, actually exists in the population one wishes to generalize to. Then, the researcher may choose not to correct for the disproportionality. The problem that arises with this is the problem of multicollinearity, that is, one cannot attribute the variance accounted for by a particular variable. (This is the dilemma of much ex post facto research.)

PART VS. PARTIAL CORRELATION

Quite often a researcher may use part correlation and interpret it as a partial correlation or visa versus. This confusion can lead to severe misinterpretation. For this purpose, part and partial correlations will be identified and explained. Nunnally (1967) has called part correlation "semi partial" and McNemar (1962) refers to it as "part." The part correlation squared is the total variance that can be accounted for in the criterion (Y) by a predictor variable (X_1) after the relationship of (X_2) has been taken out (partialed out) of (X_1). For example:

$$Y_{(success)} = a_0 U + a.(SES) + a_2(IQ) + E$$

It is the relationship between how well socio-economic status (SES) can predict success after I.Q. is removed from SES. In this example, I.Q. would be the covariate (control variable). Graphically, it may look like:



X_1 = is the unique variance that is being accounted for in Y by SES independent of IQ.

C = is the common variance that IQ and SES have with Y (SUCCESS)

$$C = \begin{cases} R^2_{y.12} \\ \text{- part correlation of } r_{y1.2} \\ \text{- part correlation of } r_{y2.1} \end{cases}$$

FIGURE 3

Partial Correlation is the variance in (Y) which is not associated with IQ that can be accounted for by socio-economic status (SES) after IQ is removed. An easier way of saying this is that it is the correlation between two residuals, the criterion residual which is made up of the criterion after IQ is removed from it, (assuming IQ is the variable one wishes to control) correlated with the residual that results after IQ is removed from socio-economic status. That is, it is the percentage of unaccounted for variance in (Y) that can be accounted for by unique variance of SES. (SES-IQ) The reason why part correlation is sometimes called semi-partial is because it only partials out the variable one wishes to control for one side of the equation. For example, it subtracts IQ from SES but

not from the criterion while partial correlation subtracts the control variable from both sides of the equation and correlates the residuals; for example, success minus IQ correlated with SES minus IQ.

Since partial correlation is dealing with percentage of unaccounted variance that is being accounted for, it will always be larger than part correlation which accounts for total criterion variance. This is true except when the control variable, in this case IQ, is zero correlated with the predictor variable (SES); then they are equal.

The partial correlation calculated by predicting SUCCESS from IQ partitioned out from SES and SUCCESS can be more graphically seen by looking at Figure 3.

$$\text{Partial Correlation } (P_{r_1})^2 = \frac{X_1}{X_1 + Y} = \frac{R^2_{Y.12} - r^2_{Y2}}{1 - r^2_{Y_2}}$$

(See FIGURE 3)

If one interprets part correlation when they should have interpreted partial, they are most likely making a severe misinterpretation of their data.

VIOLATIONS OF THE ASSUMPTION OF RECTILINEARITY ON THE REGRESSION EFFECT

The Regression Effect states that extreme scores on a pre-test will regress toward the mean on a posttest. Pohlman & Newman (1978) state:

"If a researcher finds that extreme scores on a pretest becomes even more extreme on a posttest, i.e., regressed away from the mean, this means that the assumption of rectilinearity (straight line is the line of best fit) has been violated and one should look at a curved relationship (ETA, for example) to more appropriately represent or interpret the data. If this is not done, one is underestimating the relationship between the variable and increasing the probability of misinterpreting the data.

In summary, we identified the problems of using gain scores. These problems include increasing the unreliability of the dependent variable; a bias of increasing gains for the low group, and that at the very best gain scores estimate the true relationship that can be more accurately measured by ANCOV, which covaries the pretest scores. Gain scores approach the accuracy of ANCOV as the correlation between the pretest and posttest scores approaches 1.

What is most often not understood is that the underlying assumptions for gain scores is the same as the underlying assumptions for ANCOV. Since this is not as widely known, this is less likely to be checked and therefore increases the likelihood of misuse and misinterpretation.

As far as disproportionality is concerned, there are a variety of solutions. These solutions can be broken down into "Approximate" versus "Exact" of which the "Exact" solutions are generally preferred.

The "Exact" solutions can be divided into two major categories, "full-rank solution," which does not assume the interactions are zero and has all the variance components accounted for in the full model, and the "non full-rank solutions" which does not have all the variance components present in the full model. It would appear that the "full-ranked" solution is obviously preferable; however, there are problems with different results related to differing coding procedures. There are also problems with the accuracy of the model in reflecting the research problem of interest and the assumptions of how the variables are truly related in the population. It appears to the authors that there is no one preferable solution for all situations, and believing there is a preferable solution is more likely to lead to a malpractice of statistical use and interpretation.

In dealing with "part" versus "partial" correlation, one must remember that partial correlation is the percentage of unaccounted for variance being accounted for, and, therefore, the r tends to be higher than part (semi-partial) which relates to the proportion of total variance being accounted for.

Finally, the paper deals with the problems when one violates the assumption of rectilinearity (straight line is the line of best fit) without being aware of it. This tends to result in an underestimate of the relationship between the independent and dependent variable.

Obviously, there are many problems which can lead to misuse and misinterpretation of statistical procedures. We hope that this paper will have the desirable effect of increasing the awareness of researchers to some of these problems and thereby aid researchers in avoiding many of the problems that cause the malpractice and misinterpretation of statistics.

REFERENCES

Cohen, J. and Cohen P. Applied multiple regression/correlation analysis for the behavioral sciences. New York: Lawrence Erlbaum Associates, 1975.

Cronbach, J. and Ferby, L. "How should we measure "change" -- or should we?" Psychological Bulletin, 1970, 75, 68-80.

Edwards, L. Experimental Design in Psychological Research. New York: Holt, Rinehart & Winston, 1960.

Fraas, J. The effectiveness of a simulation-gaming method of instruction for teaching college-level introductory economics (Doctoral dissertation, The University of Akron, 1978). Dissertation Abstracts International, 1978.

Gulliksen, H. Theory of Mental Tests. New York: John Wiley & Sons, 1950.

Huck, S. The analysis of covariance: increased power through reduced variability. The Journal of Experimental Education, 1972, 41 (1), 42-46.

Huck, S. and McLean, R. Using a repeated measures ANOVA to analyze data from a pretest-posttest design: A potentially confusing task. Psychological Bulletin, 1975, 82, 511-518.

Kerlinger, F. N. Foundation of Behavioral Research,(2nd ed.). New York: Holt, Rinehart & Winston, Inc., 1973.

Kerlinger, F. N. and Pedhazur, E. J. Multiple Regression in Behavioral Research. New York: Holt, Rinehart & Winston, Inc., 1973.

Lord, F. Elementary models for measuring change, In Harris, C. (ed). Problems in Measuring Change, University of Wisconsin Press, Milwaukee: 1967.

Marston, P. and Borich G. Analysis of Covariance: Is it the Appropriate Model to Study Change? Paper presented at the American Educational Research Association, Annual Meeting, New York, April 1977.

Marks, Edmond. Analysis of variance techniques for unbalanced data. Review of Educational Research. Vol. 44, 3, 1974.

McNeil, K. A., Kelly, F. J. and McNeil, J. T. Testing research hypotheses using multiple linear regression. Carbondale, Ill.: Southern Illinois University Press, 1975.

McNemar, A. Psychological Statistics. New York: John Wiley and Sons, 1962.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R. "Type VI Error: Inconsistency between the Statistical Procedure and the Research Question." Paper presented to The World Population Society, Western Regional Meeting; aboard the Queen Mary, Long Beach, California, January 1976.

Newman, I. and Newman, C. 38-22-36 Conceptual Statistics for Beginners (4th edition). Akron, Ohio. Printed at The University of Akron, 1977. ISBN # 0-917180-06-Z.

Newman, I. Oravecz, M. T. "Solutions to the Problems of Disproportionality: A Discussion of the Models." Multiple Linear Regression Viewpoints, Vol. 7, No. 3, 1977.

Nunnally, J. Psychometric Theory. New York: McGraw-Hill, 1967.

Pohlman, J. T. and Newman, I. "Regression Effects when the Assumption of Rectilinearity is not Tenable." Ohio Journal of Science, Vol. 78, No. 2, 1978.

Searle, S. R. Linear Models. New York: Wiley, 1971.

Thorndike, R. and Hagen, E. Measurement and Evaluation in Psychology and Education (3rd edition). New York: John Wiley & Sons, 1967.

Timm, N. H., & Carlson, J. E. Analysis of Variance through Full-rank Models. Monograph issued by Multivariate Behavioral Research, 1975.

Werts, C. & Linn, R. A general Linear Model for Studying Growth. Psychological Bulletin, 1970, 73, 17-32.

Williams, J. D. "Full-rank and Non Full-Rank Models with Contrast and Binary Coding Systems for a Two-way Dispropotionalte Cell Frequency Analysis." Multiple Linear Regression Viewpoints, Vol. 8, No. 1, 1978.

Multiple Linear Regression Viewpoints

Vol. 9 No. 1 1978

A THREE-YEAR EX POST FACTO STUDY OF ARITHMETIC ACHIEVEMENT FOR ELEMENTARY PUPILS ELIGIBLE FOR A REMEDIAL ARITHMETIC PROGRAM

Gary D. House, St. Louis Public Schools

This study traced the three-year impact of a remedial arithmetic program on eligible St. Louis Public School pupils. Hypotheses were tested through multiple linear regression models for analyses of covariance. No treatment effects were found. The study reveals that changes in future program evaluation designs are needed.

Background

During the 1974-75 school year, a remedial arithmetic program funded by ESEA Title I was implemented in 41 low SES schools within the St. Louis Public Schools System. The program continued in 38 of those schools throughout the 1975-76 and 1976-77 school years.

Guidelines imposed on schools by the Missouri State Department of Education (1975-1977) required that pupils be selected for the remedial arithmetic program through annual administrations of standardized tests of arithmetic achievement. In the St. Louis Public Schools that test was the Iowa Tests of Basic Skills. Qualifying grade equivalent scores from the Arithmetic Total norm tables of the tests were required, and a general policy was that pupils with IQ scores less than 76 were ineligible for selection because they were eligible for special education programs.

The remedial arithmetic program received a process evaluation during its first year of operation, (House, 1976) and very little analysis of pupil achievement was undertaken during the program's first three years because of the strong regression effects created by local measurement practices. One test was given per year, and that test was used as the major criterion for student selection, as a pretest for those selected students, and as a post test for students who had been served during the preceeding year by the program. These regression effects have been widely discussed in educational literature by Campbell and Stanley (1967), Bracht and Glass (1968),

Linn and Slinde (1977) and Tallmadge (1976). Locally they have been documented empirically by Powers (1976) and have been discussed at some length by House and Powers (1977).

Program administrators requested in 1977 that the achievement of pupils served by the program be examined across the first 3 years of program operation. This ex post facto study reflects the writer's attempt to provide that information within the constraints of the real world.

The Design

Scope: This study was limited to a three year period, 1974-75, 1975-76, 1976-77 and included data from those pupils who were eligible for selection into the program at the end of their third-grade year, 1974, were promoted annually, remained enrolled during the three year period in schools offering the remedial arithmetic program, and had complete data in their history files. The decision to limit the study to these pupils was made because the program has tended to enroll pupils in the middle grades more than others (House, et.al 1975, 1976, 1977), and the study had to be limited in order to accomodate limited resources.

Data Elements: Data elements used were pupil intelligence as measured by the Lorge-Thorndike Intelligence Test Total IQ Score (obtained during 1974-75); pupil sex; 1974 ITBS Total G.E. (selection) score; 1975 ITBS Arithmetic Total G.E. score; 1976 ITBS Arithmetic Total G.E. score; 1977 ITBS Arithmetic Total G.E. score; school attended each year.

Data Analysis: Achievement was examined for all possible combinations of past and current treatment after each year through applications of multiple linear regression approaches to the analysis of covariance. The criterion measures were end-of-year ITBS Arithmetic Tests.

Covariates for each year were selected both logically and empirically. They were: Arithmetic Achievement at time of selection; Intelligence; and, Sex (Mobility was added in the third-year analysis only, and was defined as number of different schools attended during the 3 year period.). Mobility was used as a logical covariate in the final year

only because local program administrators and teachers hypothesized that pupil mobility had a negative effect on achievement. The empirical basis for the inclusion of Selection Achievement Score, Intelligence, and Sex, came from the school system's Management and Instructional Information System which annually derives models for the prediction of pupil achievement in several areas (DeBlauw, 1977). The 1977 model reported that Previous Achievement, Intelligence, and Sex accounted together for 56% of the variance in fourth-grade arithmetic achievement for St. Louis Public School Pupils. Further, each of the three variables contributed a statistically significant increase in variance accounted for ($p < .01$). Table 1 summarizes these data.

Table 1

MIIS DATA ON CONTRIBUTION OF PRE ACHIEVEMENT, POST-IQ, AND SEX TO VARIANCE IN FOURTH GRADE ARITHEMTIC ACHIEVEMENT FOR ST. LOUIS PUBLIC SCHOOL PUPILS

CRITERION VARIABLE: 1977 GRADE 4 ARITHMETIC TOTAL G.E. SCORE

PREDICTOR VARIABLES	R	R^2	R^2 Change	F	df	Alpha Level
1976 GRADE 3 ARITHMETIC TOTAL G.E. SCORE:	.69685	.48560	.48560	3009.47552	1/3188	$p < .01$
IQ:	.74945	.56167	.07607	2041.88975	2/3187	$p < .01$
SEX:	.74947	.56171	.00004	1361.03907	3/3186	$p < .01$

For each year the program operated there were different configurations of treatment in which pupils participated. In 1974-75, pupils were either selected or not selected for the program. Subsequently, they were again either selected or not selected in 1975-76 and 1976-77. Thus, the different treatment configurations to be studied doubled with each successive year. Letting 0= no treatment, and 1= treatment, Table 2 summarizes the naturally occurring configurations of treatment included in each of the points of analysis in this study.

Table 2

CONFIGURATIONS OF TREATMENT GROUPS FOR EACH ACHIEVEMENT CRITERION YEAR

Achievement Criterion Year:	1975	1976		1977		
Treatment Year:	1975	1975	1976	1975	1976	1977
Configurations of Treatment	0	0	0	0	0	0
	1	1	0	1	0	0
		0	1	0	1	0
		1	1	0	0	1
				1	1	0
				1	0	1
				0	1	1
				1	1	1

Research Hypotheses

Among each of the points of analysis, the research hypotheses were similar. They are shown separately below along with full and restricted linear models used to test their null counterparts.

Research Hypothesis No.1:

Knowledge of participation in the remedial arithmetic program during 1974-75 adds to the prediction of 1975 arithmetic achievement, over and above the effects of intelligence, sex, and arithmetic achievement at time of selection.

FULL MODEL

$$Y_{75\ Ach} = a_1 U + b_1 T_1 + b_2 Q + b_3 S + b_4 P + E_1$$

RESTRICTED MODEL

$$Y_{75\ Ach} = a_o U + b_5 Q + b_6 S + b_7 P + E_2$$

Where, $Y_{75\ Ach}$ = *Criterion (ITBS Arithmetic Total G.E. Score in Grade 4, 1975.)*

T_1 = *Pupils participating in the during 1974-75.*

Q = *Intelligence*

S = *Sex*

P = *Achievement at time of selection*

E = *Residuals*

Research Hypothesis No. 2:

Knowledge of participation in the remedial arithmetic program during 1974-75 and/or 1975-76 adds to the prediction of 1976 arithmetic achievement, over and above the effects of intelligence, sex and arithmetic achievement at time of selection.

FULL MODEL

$$Y_{76\ Ach} = a_1 U + b_1 T_1 + b_2 T_2 + b_3 T_3 + b_4 Q + b_5 S + b_6 P + E_1$$

RESTRICTED MODEL

$$Y_{76\ Ach} = a_o U + b_7 Q + b_8 S + b_9 P + E_2$$

Where, $Y_{76\ Ach}$= Criterion (ITBS Arithmetic Total G.E. Score in Grade 5, 1976.)

T_1= Pupils served in 1974-75, but not in 1975-76

T_2= Pupils not served in 1974-75, but served in 1975-76

T_3= Pupils served in both 1974-75 and 1975-76

Research Hypothesis No. 3:

Knowledge of participation in the remedial arithmetic program during 1974-75, and/or 1975-76, and/or 1976-77, adds to the prediction of 1977 arithmetic achievement over and above the effects of intelligence, sex, arithmetic achievement at time of selection, and pupil mobility.

FULL MODEL

$$Y_{77\ Ach} = a_1 U + b_1 T_1 + b_2 T_2 + b_3 T_3 + b_4 T_4 + b_5 T_5 + b_6 T_6 + b_7 T_7 \ldots + b_8 Q + b_9 S + b_{10} P + b_{11} M + E_1$$

RESTRICTED MODEL

$$Y_{77\ Ach} = a_o U + b_{12} Q + b_{13} S + b_{14} P + b_{15} M + E_2$$

Where, $Y_{77\ Ach}$= Criterion (ITBS Arithmetic Total G.E. Score in Grade 6, 1977.)

T_1= Pupils served in 1974-75 but not in 1975-76 or 1976-77

T_2= Pupils not served in 1974-75, served in 1975-76 but not served in 1976-77

T_3= Pupils not served in 1974-75, not served in 1975-76, but served in 1976-77

T_4= Pupils served in 1974-75 and 1975-76 but not served in 1976-77

T_5= Pupils served in 1974-75, not served in 1975-76 but served in 1976-77

T_6= Pupils not served in 1974-75 but served in 1975-76 and 1976-77

T_7= Pupils served all 3 years.

M= Mobility

In addition to each research hypothesis, tests for homogeneity of regression for each covariate were made in order to aid in the interpretation of results.

One example is given below in the interest of brevity.

EXAMPLE HOMOGENEITY OF REGRESSION HYPOTHESIS (H_o)

The interaction of the covariate, intelligence, with each treatment group will not add to the prediction of 1977 achievement over and above the effects of treatment groups, intelligence, sex, arithmetic achievement at time of selection, and mobility.

FULL MODEL

$$Y_{77\ Ach} = a_1 U + b_1 (T_1 * Q) + b_2 (T_2 * Q) + b_3 (T_3 * Q) \ldots + b_7 (T_7 * Q) \ldots + b_8 T_1 + b_9 T_2 \ldots + b_{14} T_7 + b_{15} Q + b_{16} S + b_{17} P + b_{18} M + E_1$$

RESTRICTED MODEL

$$Y_{77\ Ach} = a_o U + b_1 T_1 + b_2 T_2 \ldots + b_7 T_7 + b_8 Q + b_9 S + b_{10} P + b_{11} M + E_2$$

All statistical tests will utilize an alpha level, $p \leq .05$.

Data files were searched for qualified pupils enrolled during the time period of the study in grades 4, 5 and 6 successively, within schools offering the remedial arithmetic program. Those with incomplete data on either the covariate or criterion variables were excluded; of the original 1172 identified pupils, 400 were includable in the analyses. These pupils were identified with treatment conditions according to their program histories. The Missouri State Department of Education guideline for program participation (20 or more days of program enrollment) was used.

Descriptive statistics an all variables were obtained, and tests for interaction effects among treatment groups and criteria were undertaken. Following these tests, null counterparts of each research hypothesis were tested and both observed and adjusted criteria means and standard deviations were computed.

After completeing the statistical tests stipulated in the design of the study, the relative contribution of each predictor variable was obtained and reported for use in future studies of this type.

Results

An examination of correlation matrices for all of the three covariance analyses revealed virtually no relationship among treatment groups and criterion measures (Tables 3, 4 and 5). IQ and selection score correlated highest with the criterion each year, but those coefficients diminished in magnitude with each successive year. In no case was any coefficient greater than .50 (IQ x 1975 Criterion), and there appeared to be neither redundant nor suppressor variables in any of the three matrices.

As shown in Table 6, tests for homogeneity of regression among the treatment groups and each of the covariates revealed one significant interaction, selection score with 1976 treatment.

More descriptive information for pupils identified by treatment groups appears in Tables 7, 8 and 9. Table 7 shows that both 1975 groups had equal mean IQ's and virtually equal IQ variances. Likewise, they were nearly equal in numbers of males and females, and their selection score means were virtually equivalent. On the criterion, both groups had nearly equal observed means and variances and their adjusted variances were equal.

For the 1976 groups, Table 8 reports a slightly higher observed mean IQ for pupils served for the first time in 1976 than for other groups, and a lower observed standard deviation for pupils served for the second year in a row. The proportions of females in the T_{11} groups was higher than in any other group. The general trend for the T_{11} group to be somewhat different was continued through slightly lower selection score means as well as both observed and predicted standard deviations on the criterion.

The 1977 data summarized in Table 9 again supported the trend for pupils in continued treatment to have lower mean IQ, selection score and criterion levels, while containing higher proportions of females. Mobility means for all treatment groups were less than one, but in each case, standard deviations were

Results (Cont'd)

greater than corresponding means.

Results of analyses of covariance F tests are reported in Table 10. In all three instances, knowledge of group membership added statistically insignificant amounts to R^2 values. The maximum variance accounted for by full models was about 30% in 1975; the lowest was about 19% in 1977.

Standard regression (Beta) weights reported in Table 11 A, B and C show negative values for each treatment group during each criterion year, while IQ, and selection score were the variables weighting positively at all points. Sex weighted positively in 1975, negatively in 1976 and 0 in 1977. Mobility weighted negatively when it was added in 1977.

The final presentation of data, in Table 12, is that of the contribution to R^2 by the listwise inclusion of variables into each regression equation. It shows IQ and selection score contributing significant amounts of variance accounted for in each analysis, and sex adding a significant amount to the explanation of 1976 criterion variance. The contributions of all other variables are not statistically significant.

TABLE 3

CORRELATION MATRIX FOR VARIABLES IN 1975 COVARIANCE ANALYSIS

	1975 Arith G.E.	Control Group (T_0)	Treatment Group (T_1)	IQ	SEX	SELECTION SCORE
1975 Arith G.E.	—	.01	—.01	.50	.14	.38
Control Group	—	—	—	.00	—.04	—.04
Treatment Group	—	—	—	.00	.04	.04
IQ	—	—	—	—	.21	.37
SEX	—	—	—	—	—	.13
SELECTION SCORE	—	—	—	—	—	—

TABLE 4

CORRELATION MATRIX FOR VARIABLES IN 1976 COVARIANCE ANALYSIS

	1976 Arith G.E.	T_{oo}	T_{10}	T_{01}	T_{11}	IQ	SEX	SELECTION SCORE
1976 Arith G.E.	—	.02	—.03	.05	—.02	.42	—.02	.31
T_{oo}	—	—	—	—	—	—.07	—.01	—.05
T_{10}	—	—	—	—	—	.01	—.05	.05
T_{01}	—	—	—	—	—	.10	—.04	.00
T_{11}	—	—	—	—	—	.00	.10	.00
IQ	—	—	—	—	—		.21	.37
SEX	—	—	—	—	—	—	—	.13
SELECTION GROUP	—	—	—	—	—	—	—	—

TABLE 5

CORRELATION MATRIX FOR VARIABLES IN 1977 COVARIANCE ANALYSIS

	1976 Arith G.E.	T_{000}	T_{100}	T_{010}	T_{001}	T_{110}	T_{101}	T_{011}	T_{111}	IQ	SEX	SELECTION SCORE	MOBILITY
1976 Arith G.E.	—	.07	—.03	.08	—.05	—.06	—.01	.02	—.04	.38	.08	.27	—.09
T_{000}	—	—	—	—	—	—	—	—	—	—.08	—.03	—.02	.06
T_{100}	—	—	—	—	—	—	—	—	—	—.03	—.07	.01	.02
T_{010}	—	—	—	—	—	—	—	—	—	.08	—.06	—.01	—.05
T_{001}	—	—	—	—	—	—	—	—	—	.01	.04	—.05	.00
T_{110}	—	—	—	—	—	—	—	—	—	—.01	.03	—.01	—.02
T_{101}	—	—	—	—	—	—	—	—	—	.06	.02	.07	.01
T_{011}	—	—	—	—	—	—	—	—	—	.05	.01	.03	.00
T_{111}	—	—	—	—	—	—	—	—	—	.00	.12	.00	—.05
IQ	—	—	—	—	—	—	—	—	—	—	.21	.37	—.09
SEX	—	—	—	—	—	—	—	—	—	—	—	.13	—.05
SELECTION SCORE	—	—	—	—	—	—	—	—	—	—	—	—	—.04
MOBILITY	—	—	—	—	—	—	—	—	—	—	—	—	—

TABLE 6

TESTS FOR HOMOGENEITY OF REGRESSION ON 1975, 1976 AND 1977 CRITERION MEASURES

INTERACTION OF TREATMENT WITH...	R^2 RESTRICTED MODEL	R^2 FULL MODEL	R^2 diff	F	dfN / dfD	ALPHA	CRITERION
...IQ	.296	.296	.000	0.0	1/394	1.000	1975 ARIH ACH
...SEX	.296	.301	.005	2.6894	1/394	0.102	
...SELECTION SCORE	.296	.302	.006	3.1390	1/394	0.077	
...IQ	.224	.225	.001	0.1923	3/390	0.902	1976 ARITH ACH
...SEX	.224	.230	.006	1.1540	3/390	0.327	
...SELECTION SCORE	.224	.246	.022	3.9679	3/390	0.008	
...IQ	.188	.195	.007	0.4405	7/381	0.875	1977 ARITH ACH
...SEX	.188	.197	.009	0.6089	7/381	0.748	
...SELECTION SCORE	.188	.190	.002	0.0923	7/381	1.001	
...MOBILITY	.188	.210	.022	1.5201	3/381	0.158	

TABLE 7

DESCRIPTIVE COVARIATE AND CRITERION STATISTICS FOR PUPILS IN GROUPS 1975

	COVARIATE STATISTICS							CRITERION STATISTICS (1975 ARITH ACH)					
	IQ			SEX (N)		SELECTION SCORE		OBSERVED			ADJUSTED		
GROUPS	N	$\bar{x}$	σ	M	F	$\bar{x}$	σ	$\bar{x}$	σ		$\bar{x}$	σ	
T_o	200	84.2	10.7	104	96	2.62	.43	3.73	.73		3.73	.39	
T_1	200	84.2	10.8	97	103	2.66	.41	3.72	.70		3.72	.39	

TABLE 8

DESCRIPTIVE COVARIATE AND CRITERION STATISTICS FOR PUPILS IN 1976 GROUPS

Groups	COVARIATE STATISTICS							CRITERION STATISTICS (1976 ARITH ACH)			
		IQ		SEX (M)		SELECTION SCORE		OBSERVED		ADJUSTED	
	N	$\bar{x}$	σ	M	F	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
T_{∞}	148	83.2	10.5	75	73	2.61	.42	4.73	.86	4.72	.40
T_{10}	118	84.3	12.3	64	54	2.67	.42	4.66	.87	4.66	.44
T_{01}	52	86.8	10.9	29	23	2.65	.47	4.81	1.10	4.81	.47
T_{11}	82	84.1	8.3	33	49	2.64	.41	4.65	.74	4.66	.33

TABLE 9

DESCRIPTIVE COVARIATE AND CRITERION STATISTICS FOR PUPILS IN 1977 GROUPS

	COVARIATE STATISTICS									CRITERION STATISTICS (1977 ARITH ACH)			
	IQ			SEX (N)		SELECTION SCORE		MOBILITY		OBSERVED		ADJUSTED	
	N	$\bar{x}$	σ	M	F	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
T_{000}	108	82.7	11.4	57	51	2.63	.43	.65	.85	5.56	.98	5.56	.42
T_{100}	88	83.7	12.9	50	38	2.65	.43	.60	.77	5.39	1.13	5.39	.49
T_{010}	35	86.9	12.2	21	14	2.63	.50	.46	.66	5.69	1.03	5.69	.51
T_{001}	40	84.7	7.3	18	22	2.58	.41	.58	.78	5.30	8.2	5.30	.34
T_{110}	56	84.0	9.2	26	30	2.63	.44	.54	.79	5.30	9.3	5.30	.39
T_{101}	30	86.3	10.1	14	16	2.75	.36	.60	.68	5.41	1.02	5.41	.36
T_{011}	17	86.6	8.0	8	9	2.69	.39	.59	.71	5.56	1.01	5.56	.34
T_{111}	26	84.3	6.1	7	19	2.64	.34	.42	.50	5.27	.98	5.28	.25

TABLE 10

F TESTS FOR TREATMENT EFFECTS OVER AND ABOVE COVARIATES FOR EACH CRITERON YEAR

CRITERION YEAR	R^2 RESTRICTED MODEL	R^2 FULL MODEL	R^2 DIFF	F	dfN / dfD	ALPHA
1975	.296	.296	.000	1.268	7/388	0.265
1976	.220	.224	.004	0.6664	3/393	0.573
1977	.170	188	.018	0.3232	1/395	0.570

TABLE 11 A,B,C

STANDARD REGRESSION WEIGHTS FOR FULL MODEL PREDICTOR VARIABLES FOR EACH CRITERION YEAR

A. 1975 CRITERION

Predictors	Standard Weights	Error
T_1	−.024	.000
IQ	.409	.001
SEX	.025	−.001
SELECTION SCORE	.230	−.001

B. 1976 CRITERION

Predictors	Standard Weights	Error
T_{10}	−.068	.000
T_{01}	−.018	.001
T_{11}	−.036	.002
IQ	.376	.000
SEX	−.130	.000
SELECTION SCORE	.195	.001

C. 1977 CRITERION

Predictors	Standard Weights	Error
T_{100}	−.084	.002
T_{010}	−.004	.000
T_{001}	−.090	.000
T_{110}	−.107	.000
T_{101}	−.079	.001
T_{011}	−.029	.001
T_{111}	−.086	.001
IQ	.323	.000
SEX	0.0	.001
SELECTION SCORE	.158	.002
MOBILITY	−.059	.000

TABLE 12

R^2 VALUES FOR EACH PREDICTOR VARIABLE ADDED TO EQUATIONS FOR EACH CRITERION YEAR

VARIABLE NAME	R^2 BEFORE INCLUSION	R^2 AFTER INCLUSION	R^2 DIFF	dfN / dfD	F	ALPHA	CRITERION YEAR
IQ	0.000	.248	.248	1/398	131.59	0.000	1975
SEX	.248	.250	.002	1/397	0.7182	0.397	
SELECTION SCORE	.250	.296	.046	1/396	25.8705	0.000	
TREATMENT	.296	.296	.000	1/395	0.3232	0.570	
IQ	0.000	.175	.175	1/398	84.4026	0.000	1976
SEX	.175	.188	.013	1/397	6.5682	0.011	
SELECTION SCORE	.188	.220	.032	1/396	15.8268	0.000	
TREATMENT	.220	.224	.004	3/393	0.6664	0.573	
IQ	0.000	.146	.146	1/398	67.966	0.000	1977
SEX	.146	.146	.000	1/397	0.0032	0.954	
SELECTION SCORE	.146	.167	.021	1/396	9.9549	0.954	
MOBILITY	.167	.170	.003	1/395	1.4484	0.229	
TREATMENT	.170	.188	.018	7/388	1.268	0.265	

Discussion

It is recognized that the majority of pupils originally identified for this study were eliminated from the analyses because of incomplete data, and that the remaining group may not have been representative of the total group of identified pupils. The effects of subject elimination are not known, but it is the writer's expectation that any biasing would have been in favor of finding significant treatment effects since remaining subjects progressed in school normally, were present in school to a sufficient extent to have complete data histories on all measures, and remained enrolled in Title I schools for the entire 3 year term of the study.

The limitations of an ex post facto design are also recognized. There were never any controls placed upon the delivery of program services to pupils, pupil selection or any other confounding variable. The design is the writer's best effort at testing the effectiveness of treatment. While admitting confounding factors, it is not recommended that the findings be discounted as meaningless, but considered as partial evidence regarding program effectiveness and suggestive of the need for more controlled studies to be implemented.

The one significant interaction indicating a lack of homogeneity of regression for 1976 treatment groups on selection score is likely to be spurious given the multiple F tests obtained in this study, and is therefore discounted by the writer in interpreting the findings.

The data did not support any of the research hypotheses. Further, the negative Beta weights for each treatment condition suggested a trend toward the remedial arithmetic program negatively influencing arithmetic achievement, although to a statistically insignificant degree. Since each level of treatment was included in this study, and since their combined knowledge accounted for no significant amounts of explained criterion variance, it can be concluded that treatment made no difference for identified pupils remaining enrolled in Title I eligible schools offering the remedial arithmetic program for a three-

Discussion (Cont'd)

year period in sequential grades 4, 5 and 6, and having complete data histories.

Mean IQ scores were well below the national average as were all mean grade equivalent values for these pupils, and the amounts of variance accounted for (19% to 30%) were lower than for the St. Louis Public Schools in general given similar predictor variables. These facts suggest either that norm referenced achievement objectives may be inappropriate for these pupils, or that the program is ineffective in overcoming academic deficiencies. Certainly the ITBS and Lorge Thorndike tests are very difficult for these Title I eligible pupils.

The results of this study show, as did a similar study of a remedial reading program (House and Powers, 1977), that continued enrollment in the remedial arithmetic program is a function of continued low achievement, which is in turn related to lower intelligence. In order to study either cumulative or single year effects of a program such as this, it would be necessary to randomly assign pupils to treatments or control groups and then follow their achievement across time. Such a design would not necessarily bias results in favor of control groups because the influences identified as significant contributors to criterion achievement would be controlled through random assignment, and any random influences could be statistically controlled.

It is not currently known what factors others than State guidelines are used by teachers and administrators in selecting students. It is likely that many variables are considered. Random assignment should limit both known and unknown selection effects.

Only one type of criterion measure was included in this study because only one was obtained. Program objectives refer to only that measure (ITBS). It is possible that program effects other than arithmetic achievement as measured by ITBS have occurred or will occur in the future. It is the writer's

Discussion *(Cont'd)*

recommendation that other criteria including criterion tests identified as representative of program curricular content be applied, with the random assignment of pupils, to any further evaluation design for this program, and that future studies be conducted to utilize on-going rather than past implementation of the program. It is also recommended that any further studies span at least 3 year periods and include pupils in primary and upper as well as middle grades.

References

Bracht, Glenn H., and Glass, Gene U., The External Validity of Experiments, AERJ, Vol. 5, No. 4, November, 1968, pp 437-467

Campbell, Donald T., and Stanley, Julian C., Experimental and Quasi-Experimental Designs for Research, 1963, The American Educational Research Association.

DeBlauw, Robert A., St. Louis Public Schools, Division of Evaluation MIIS Achievement Prediction Models, 1975, 1976, 1977.

House, Gary, Final Evaluation Report on the Mathemtics Improvement Teams Program, St. Louis Public Schools, Division of Evaluation, ESEA Title I State Report, 1976.

House, Gary, et.al, St. Louis Public Schools, Division of Evaluation, ESEA Title I State Report, 1975, 1976, 1977

House, Gary, and Powers, Jerry, An Ex Post Facto Study of Reading Achievement With The Reading Improvement Teams, St. Louis Public Schools, Division of Evaluation, ESEA Title I State Report, 1977

Linn, Robert L. and Slinde, Jeffrey A., The Determination of The Significance of Change Between Pre- and Posttesting Periods, Review of Educational Research, Winter 1977, Vol. 47, No. 1, pp 121-150

Missouri State Department of Education, ESEA Title I Manual of Operational Policies and Guidelines, 1975, 1976, 1977

Powers, Jerry M., An Empirical Study On The Use of Gain Scores In Title I Programs, St. Louis Public Schools, Division of Evaluation, ESEA Title I State Report, 1976

Tallmadge, G. Kasten, The Regression Effect, ESEA Title I Evaluation Reporting System, Department of Elementary and Seconday Education, Jefferson City, Mo., October, 1976

Multiple Linear Regression Viewpoints
Vol. 9 No. 1 1978

AN APPROXIMATION TECHNIQUE FOR VARIABLE SELECTION USING COST CRITERIA

Thomas P. Ryan
Old Dominion University

ABSTRACT

The problem of selecting regression variables using cost criteria is considered. A method is presented which approximates the global minimum of one of several criterion functions which might be employed. Examples are given and the results are compared with the results of other methods. The outcome of a simulation study is also discussed, and suggestions are made as to the practical use of the method.

INTRODUCTION

A considerable amount of attention has been given to the problem of selecting variables in multiple regression, but little attention has been devoted to this problem when acquisition costs for the variables are present and are to be utilized. The problem has been previously considered by Lindley (1968), Pohlmann (1973), and McCabe and Ross (1973). The objective of the present paper is essentially the same as that given by McCabe and Ross (1973); namely, to obtain the best or nearly best subset ("best" in accordance with an appropriate criterion) without performing a complete enumeration.

METHOD

We agree with McCabe and Ross (1973) that essentially a minimization of a linear function of the residual sum of squares

and the costs of the included variables is required, and that for a large number of variables an optimal solution is computationally impractical. However, we prefer a different approach to the problem. Consider the function

$$L(S_k) = SSE(S_k)/SST + M\left[\sum_{i\epsilon S_k} c_i \Big/ \sum_{i=1}^{p} c_i\right] \qquad (1)$$

$$0 < M < \infty$$

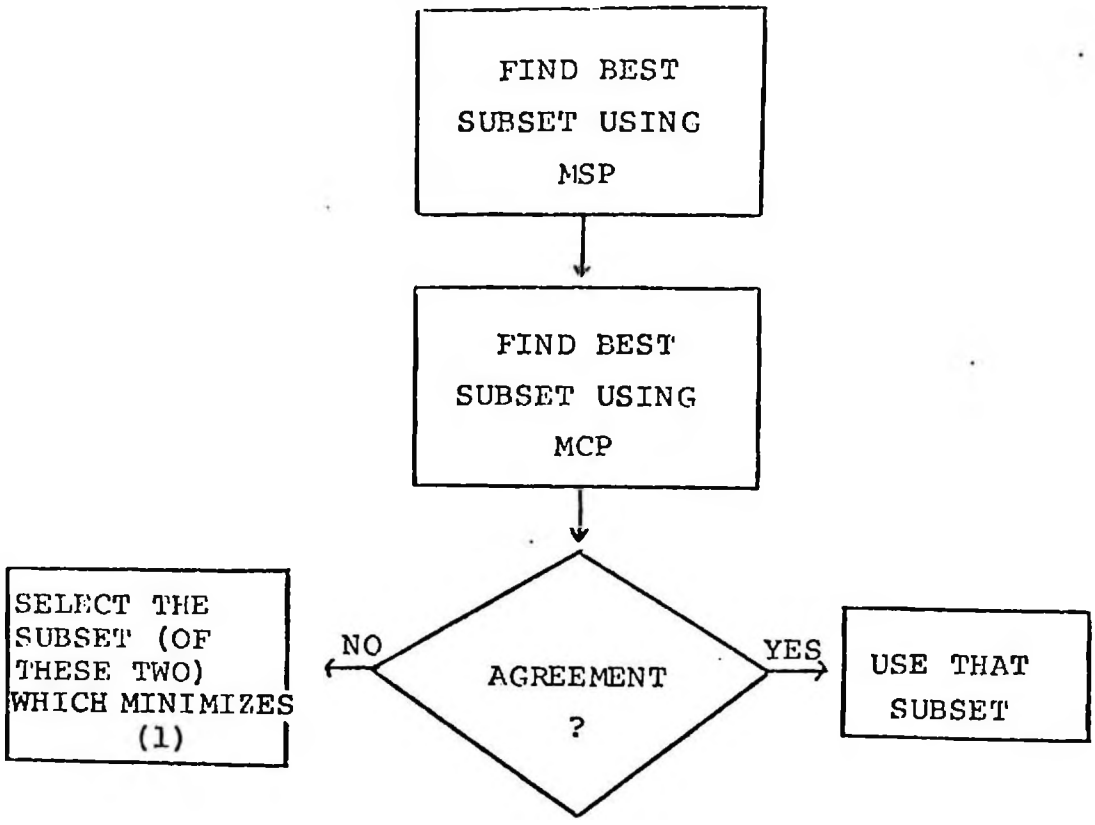
where S_k denotes an arbitrary subset of size k, c_i is the cost of the i^{th} variable, p is the number of available variables, M is a constant, $SSE(S_k)$ is the error sum of squares for that subset, and SST is the total sum of squares. Thus, if M = 1 we attach equal importance to cost and error sum of squares, if M < 1 cost is given secondary consideration, and if M > 1 cost is given primary consideration. The rationale behind the selection of equation (1) as the loss function is as follows. First, each of the two components is a unit-free number; and, second, with M = 1, when k = 0, $L(S_k) = 1$, and when k=p, $L(s_k) = 1.0 + SSE(S_k)/SST$, where the second term will usually be close to zero. Thus, if the loss function is minimized for k < p, this implies that the scaled improvement in the residual sum of squares exceeds the scaled increase in cost. Thus, for small values of p(say, $4 \leq p \leq 6$) we would expect the loss function to be somewhat parabolic (for most problems) with optimal k less than or equal to p/2. This seems intuitively appealing since the point of diminishing returns

is reached rather quickly in regard to the residual sum of squares (even when costs are ignored).

Similar supportive arguments can probably be made for the loss functions proposed by Pohlmann (1973) and McCabe and Ross (1973), each of which utilizes at least one scale factor. Subsequently, the objective here is not necessarily to determine the most appropriate function but rather to approximate the minimum of whatever function is chosen. The minimum of equation (1) may be approximated as follows. Let

$$L_1(S_k) = \frac{SSE(S_k) - SSE(S_{k+1})}{SST} \Bigg/ M\left[\frac{C(S_{k+1}) - C(S_k)}{\sum_{i=1}^{p} c_i}\right] \qquad (2)$$

where S_k now denotes the subset of size k with the smallest SSE, and $C(S_k)$ is the associated cost. Starting with $k = 1$, we increase k until $L_1(S_k) \leq 1$; i.e., when the scaled reduction in SSE no longer exceeds the scaled increase in costs. We call this the minimum SSE procedure (MSP). It then seems reasonable to redefine the terms in equation (2) so that $C(S_k)$ is now the cost of the least expensive subset of size k, and $SSE(S_k)$ is the associated residual sum of squares, and to apply the same stopping rule. We call this the minimum cost procedure (MSP). This leads to the following flowchart.



This appears to be a reasonable sequential procedure because we are attempting to "entrap" the best subset by moving toward it from two directions - one which emphasizes residual sum of squares and the other which emphasizes cost. It can be easily shown that (with M = 1) this sequential procedure will always select the subset that minimizes the lost function (over all values of k), if the subset that minimizes the loss function for each value of k is either the least expensive subset or the subset with smallest SSE. Thus, for small values of p (e.g., $4 \leq p \leq 6$) the procedure should have high efficiency since we are looking at two subsets of each subset size. A modification of the sequential procedure

is used for larger values of p, but discussion of the modified procedure is deferred to the last section.

EXAMPLES

We applied this procedure to the same four-variable examples considered by Lindley (1968) and McCabe and Ross (1973), and for each example the sequential procedure selected the same subset that would have been selected with a complete enumeration - but this should not be surprising in light of the comments made above. These were also the same subsets selected by Lindley (1968) and McCabe and Ross (1973) - the latter using a stepwise algorithm with a single scale factor, while the former performed a complete enumeration without scale factors.

The use of the sequential procedure can be illustrated with one of the four-variable examples used by Lindley (1968) and McCabe and Ross (1973). The results are shown in the following table.

TABLE 1

Application of the Sequential Procedure

A. MSP		B. MCP	
Variables In Subset	Value of Equation (2)	Variables In Subset	Value of Equation (2)
4*		4	
	0.4055		3.12813
12		34*	
			0.10692
124		134	

* indicates the chosen subset

The subset selected by the minimum cost procedure, (3, 4), was the overall best subset whereas the subset selected by the minimum SSE procedure was rated second. Thus, the combined procedure selected the best subset - as was the case with the other examples.

To further assess the worth of our procedure we simulated 330 data and cost structures for four-variable problems. The data structures were generated using the simulation procedure described in Ryan (1977) in which intercorrelations from 0 to 1 are systematically generated. Most of the common statistical distributions were used to provide the cost structures - normal, inverted normal, exponential increasing, exponential decreasing, etc. Even though a complete enumeration can easily be performed with only four predictors, we nevertheless restricted our attention to this number so that the necessary simulation would not be unmanageable, and the efficiency of the procedure could be easily assessed. The results are shown in Table 2.

TABLE 2

Summary of Results

Procedure	Number of Times Best* Subset Selected	Efficiency (%)	Rank of Seven Sub-Optimal Subsets Selected By CP
MSP	301	91.21	
Combined Procedure (MSP and MCP)	323	97.88	2,2,2,2,3,3,2

* as determined by the minimum value of (1), M = 1

primary interest is centered upon the efficiency of the combined procedure, but the figures are given for MSP to indicate the high efficiency of that separate procedure.

DISCUSSION

Although the combined procedure was obviously quite efficient, the results should be viewed in light of the following considerations. First, the effectiveness of the procedure will certainly depend upon the total number of variables from which the selection is made. Thus, for large problems (say, $p > 10$) we would expect the efficiency to fall of somewhat. Second, we would also expect the efficiency to be lower when the data exhibits a high degree of multicollinearity and the differences in costs are small. This is especially true when the ranking of the variables in terms of predictive ability is virtually the opposite of the ranking of the variables in terms of cost; but under such circumstances whichever subset is selected will be highly competitive with the best subset. Nevertheless, if optimality is deemed an important objective, the following modification of the sequential procedure might be used for large problems: (A) use the procedure to initially determine the subset size (for the simulation study, the correct subset size was selected for all but 1 of the 330 structures), (B) apply the procedure to 6-10 subsets of that size (of course, this begins to approach a complete enumeration, but we are working with only one subset size so the extra effort would be minimal).

The input for this procedure can be efficiently generated using the latest version of Furnival and Wilson (1974) "Leaps and Bounds" algorithm, although one of the modifications of their program might have to be employed if the least expensive subset of each size is not among the best ten subsets in terms of SSE.

REFERENCES

Furnival, G. M., and Wilson, R. W. "Regression by leaps and bounds." Technometrics. 16, 499 - 511, 1974.

Lindley, D. V. "The choice of variables in multiple regression." Journal of the Royal Statistical Society. Series B, 30, 31 - 53, 1968.

McCabe, George P., Jr., and Ross, Mary Ann. "A stepwise algorithm for selecting regression variables using cost criteria." Mineograph Series #349, Dept. of Statistics, Purdue Univ., Dec., 1973.

Pohlmann, J. T. "Incorporating cost information into the selection of variables in multiple regression analysis." Multiple Linear Regression Viewpoints. 4, 18 - 26, 1973.

Ryan, T. P. Selection of regression variables using cost criteria. Doctoral Dissertation, Univ. of Georgia, 1977.

Vol. 9 No. 1 1978

UNIVARIATE NONPARAMETRIC ANALYSIS OF VARIANCE THROUGH MULTIPLE LINEAR REGRESSION

Bradley E. Huitema
Western Michigan University

Abstract

Many methodologists are aware that parametric tests associated with the analysis of variance and the analysis of covariance can be computed using regression procedures. It is shown that multiple linear regression can also be employed to compute the Kruskal-Wallis nonparametric analysis of variance.

Introduction

Researchers frequently encounter data analysis situations in which it is reasonable to believe that the population distributions associated with J treatments deviate far from normality and/or homogeneity of variance. In these extreme cases the conservative data analyzer may decide to compute the Kruskal-Wallis test rather than a conventional parametric analysis of variance or regression analog - especially if the sample sizes are unequal. The purpose of this paper is to show how multiple linear regression can be employed to compute the Kruskal-Wallis test statistic.

The computation of the test statistic H using the conventional Kruskal-Wallis formula (Siegel, 1956) will be presented first; then the regression approach will be described. Both procedures will be applied to the data of Table 1.

Table 1. Raw Data and Ranks for Kruskal-Wallis Test

Treatment							
I		II		III		IV	
RAW	RANK	RAW	RANK	RAW	RANK	RAW	RANK
95	12	67	5	91	11	103	14
40	1	48	4	47	3	99	13
88	10	79	8	43	2	124	16
82	$\frac{9}{32}$	76	$\frac{7}{24}$	70	$\frac{6}{22}$	105	$\frac{15}{58}$

Conventional Computation

The conventional Kruskal-Wallis formula is

$$\frac{12}{N(N+1)}\left[\sum \frac{(\sum^{n} R_i)^2}{n_i}\right] -3(N+1) = H$$

where

H is the test statistic which is distributed (approximately) as chi square with J-1 degrees of freedom,

N is the total number of subjects,

$\sum^{n} R_i$ is the sum of the ranks for the j^{th} group and

n_i is the number of subjects in the j^{th} group.

The application of this formula to the data of Table 1 yields

$$\frac{12}{16(17)}\left[\frac{(32)2 + (24)^2 + (22)^2 + (58)^2}{4}\right] -3(17)$$

$$= .044117\,[1362] - 51$$

$$= 9.09$$

Regression Approach

The computation of H using multiple linear regression involves the following steps:

A. Arrange ranked data using dummy variables to identify groups

D_1	D_2	D_3	Y
1	0	0	12
1	0	0	1
1	0	0	10
1	0	0	9
0	1	0	5
0	1	0	4
0	1	0	8
0	1	0	7
0	0	1	11
0	0	1	3
0	0	1	2
0	0	1	6
0	0	0	14
0	0	0	13
0	0	0	16
0	0	0	15

B. Regress Y on the dummy variables using an ordinary multiple regression program.

C. Multiply the squared multiple correlation coefficient by N-1. For the example data $R^2 = .60588$ and $N - 1 = 15$. The test statistic H is therefore $15(.60588) = 9.09$ which agrees with the answer obtained using the conventional procedure.

Our data include only four subjects per treatment in order to keep the computation simple. It should be kept in mind, however, that regardless of the method of computation, the H statistic is distributed approximately as chi square only if the number of subjects per group is reasonably large. The definition of "reasonably large" for this test is generally considered to be six or more.

Multiple Comparisons

Since the obtained H is significant (using alpha = .05) we may be interested in identifying the contrasts that have led to the significant result. Nemenyi's rank order analog of Tukey's test (Miller, 1966) is recommended in this situation. Before proceeding with this test it will be necessary to compute the mean rank for each treatment group. The regression equation may be employed in obtaining each mean rank. The computation of the mean ranks and Nemenyi's tests for the example data are shown below:

A. First compute the regression equation parameter estimates; the mean ranks are based on these estimates.

$\hat{\beta}_0 = 14.5$

$\hat{\beta}_1 = -6.5$

$\hat{\beta}_2 = -8.5$

$\hat{\beta}_3 = -9.0$

Mean rank for group I = $\hat{\beta}_0 + \hat{\beta}_1 = 8.0$

Mean rank for group II = $\hat{\beta}_0 + \hat{\beta}_2 = 6.0$

Mean rank for group III = $\hat{\beta}_0 + \hat{\beta}_3 = 5.5$

Mean rank for group IV = $\hat{\beta}_0 = 14.5$

B. The next step is to determine the absolute mean rank difference for all pairwise differences.

Groups	Mean Rank	Absolute Mean Rank Difference $\lvert \bar{R}_i - \bar{R}_j \rvert = d$
I - II	8-6	2.0
I - III	8-5.5	2.5
I - IV	8-14.5	6.5
II - III	6-5.5	.5
II - IV	6-14.5	8.5
III - IV	5.5-14.5	9.0

C. The value that the absolute mean rank difference must equal or exceed in order to be declared significant is obtained using the following formula:

$$\underline{q}_{\alpha, J, \infty} \sqrt{\frac{J(Jn + 1)}{12}}$$

where $\underline{q}_{\alpha, J, \infty}$ is the studentized range statistic with J and infinite degrees of freedom,

J is the number of groups in the experiment and

n is the common number of subjects found in each group.

Four groups are involved in the example; $\underline{q}_{.05, 4, \infty}$ is 3.633 and the critical mean rank difference is

$$3.633 \sqrt{\frac{4(17)}{12}} = 8.65.$$

We conclude that the difference between the mean ranks for groups III vs. IV is significant at the .05 level.

A Note on the Interpretation of the Kruskal-Wallis Test

There appears to be some confusion in the methodological literature concerning the assumptions and interpretations associated with the Kruskal-Wallis test. It turns out that, as might be expected, the interpretation depends upon what is assumed.

If it can be assumed that (1) the population distributions are continuous, (2) the population distributions are of the same shape (that is, they have the same variance, skewness and kurtosis), and (3) the samples from the populations are independent random ones, then we can conclude, given a significant test statistic, that there are differences amoung the population means. In this case this "nonparametric" test is, in fact, an acceptable test of differences amoung population means which are, of course, parameters.

If we cannot make all of the assumptions listed above but can assume that (1) the population distributions are continuous and (2) the samples are independent random ones, the test is still useful but the interpretation is different.

In this case a significant test statistic allows us to conclude only that the populations are not identical. This means that the interpretation of the significant effect is ambiguous. It is possible that the population means are equal but the variances differ, or the skewness differs or several of these distribution characteristics differ. It is not clear what the "effect" is in this case.

References

Miller, R. G. Jr. Simultaneous statistical interference. New York: McGraw-Hill, 1966.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Multiple Linear Regression Viewpoints
Vol. 9 No. 1 1978

UNIVARIATE NONPARAMETRIC ANALYSIS OF VARIANCE: A COMMENT

Lee M. Wolfle
Virginia Polytechnic Institute and State University

Huitema (1978) correctly pointed out that there is an association between the Kruskal-Wallis H statistic and the multiple R^2 based on regressing ranks on k-1 dummy variables used to identify the k groups. Yet he failed in a number of ways to be convincing. Specifically, first, he demonstrated by means of example that $H = (N - 1)R^2$, but did not offer a proof. Second, he failed to consider the more general case of tied ranks. Third, he failed to discuss the utility of the regression approach over the traditional computation. Fourth, he failed to note that there is an associated relationship between the H statistic and an F-test performed on ranks. Finally, he failed to demonstrate the advantages, if any, of using the nonparametric test on ranks, regardless of how computed, over that of the more traditional one-way analysis of variance. It is my purpose in appending this comment to Huitema's paper to address these five points. I hope by offering these comments that readers will be made sensitive to considerations underlying the choice of these statistics.

First, note the mean of a rectangular distribution of ranks, $\bar{R} = (N + 1)/2$, the variance, $Var(R) = (N^2 - 1)/12$, and the total sum of squares, $SS_t = N(N^2 - 1)/12$.

Then,

$$R^2 = SS_{reg}/SS_t$$

$$= \Sigma n_j (\bar{R}_j - \bar{R})^2 / N(N^2 - 1)/12$$

$$= \left[\Sigma \frac{R_j^2}{n_j} - 2N\bar{R}^2 + N\bar{R}^2\right] / N(N^2 - 1)/12$$

$$= \left[\Sigma \frac{R_j^2}{n_j} - \frac{N(N + 1)^2}{4}\right] / N(N^2 - 1)/12$$

$$= \frac{12}{N(N^2 - 1)} \left[\Sigma \frac{R_j^2}{n_j}\right] - \frac{3(N + 1)^2}{(N^2 - 1)}$$

where n_j = number of cases in the j-th sample, for j = 1...k,

$N = \Sigma n_j$, the number of cases in all samples combined,

R_j = sum of ranks in j-th sample, and

$\bar{R}_j$ = mean of ranks in j-th sample.

Proving the equality Huitema demonstrated by example is accomplished by multiplying R^2 by (N-1):

$$H = (N - 1)R^2$$

$$= \frac{12(N - 1)}{N(N^2 - 1)} \left[\Sigma \frac{R_j^2}{n_j}\right] - \frac{3(N + 1)^2 (N - 1)}{(N^2 - 1)}$$

$$= \frac{12}{N(N + 1)} \left[\Sigma \frac{R_j^2}{n_j}\right] - 3(N + 1); \qquad \text{Q.E.D.}$$

Second, consider the more general case of tied ranks. The consequence is that the distribution is no longer rectangular, and while the mean is unchanged the variance is reduced below $(N^2 - 1)/12$. The new total sum of squares is:

$$SS_t = \left[\frac{N(N^2 - 1)}{12}\right] \left[1 - \frac{\Sigma T}{N^3 - N}\right]$$

where $\Sigma T = t^3 - t$, and t is the number of ties in each group of tied

ranks. Applying the correction in the computation of the multiple R^2, and multiplying the result by (N - 1) results in the usual Kruskal-Wallis formula (e.g., 1952, p. 587) for H with tied ranks. Thus, $H = (N - 1)R^2$ with or without tied ranks.

My third point is that regardless of the equality demonstrated by Huitema, and proved above, I can imagine only the rarest of instances in which it makes sense to employ dummy variable regression rather than the Kruskal-Wallis formula. Indeed, Kruskal and Wallis listed as the first advantage of their (then) new statistic, "The calculations are simplified" (1952, p. 585). Personally, I can easily and quickly compute H on my hand-held calculator, but to obtain the multiple R^2 by dummy variable regression requires me to use the main-frame computer to which I have access, and I must still use my calculator to obtain H. Thus, the relationship Huitema discussed has some formal interest for me, but has no practical utility.

Fourth, recalling that F and R^2 are related brings to mind the possibility that H and F are also related. If an F-test is performed using ranks as the dependent variable and the k groups as the independent, then:

$$H = \frac{(N - 1)\ (k - 1)F}{(N - k) + (k - 1)F} .$$

To prove this, recall the proof that $H = (N - 1)R^2$, and substitute the well-known equality $R^2 = v_1\ F/(v_2 + v_1\ F)$, where v_1 = the number of independent variables, and $v_2 = N - v_1 - 1$. I find this formal relationship equal in interest to the relationship of H to R^2, but of more practical utility because I can use my hand-held calculator to compute the F statistic.

This leads to my final point. I am not convinced that there exists a practical situation in which I would use the Kruskal-Wallis statistic in place of parametric, one-way analysis of variance. The Kruskal-Wallis test is unusually recommended in place of one-way ANOVA when the researcher is unwilling to assume the populations are distributed normally within groups. However, employing the Kruskal-Wallis test in lieu thereof does not excuse the researcher from assumptions I find equally limiting. Specifically, the Kruskal-Wallis test assumes the populations within groups are distributed with approximately the same form. If I must make that assumption in order to use the Kruskal-Wallis test on ordinal ranks, I would just as soon use an analysis of variance on interval raw scores. The F-test is sufficiently robust, and has the advantage of following an exact probability distribution, whereas the H statistic only approximates the chi square distribution. By way of example, I calculated a one-way ANOVA on the data Huitema reported in his paper: $F = 4.5956$, and $P\ (F_{3,12} = 4.5956) = .023$. In contrast, for the same data, $H = 9.0882$, and $P\ (\chi^2_3 = 9.0882) = .028$. There would have been no difference in substantive interpretation, regardless of approach. Another example produced a similar result, this one taken from Kruskal and Wallis (1952, p. 589): $F = 2.9913$ with $P = .011$, and $H = 18.5654$ with $P = .010$. The choice of ANOVA over Kruskal-Wallis is clearly one of style, and I expect opinions will vary.

In conclusion, Huitema has pointed out that there exists an alternative way to compute the Kruskal-Wallis H statistic. While I have proved the relationship, I can foresee no instance in which I would use dummy variable regression rather than computing H directly.

Indeed, I can imagine only rare instances in which I would use H at all. I wish Huitema had addressed these issues more thoroughly, because I think they are much more important than the formal relationship he did discuss.

REFERENCES

Huitema, Bradley E.
1978 "Univariate Nonparametric Analysis of Variance Through Multiple Linear Regression." *Multiple Linear Regression Viewpoints, supra.*

Kruskal, William H., and W. Allen Wallis.
1952 "Use of Ranks in One-Criterion Variance Analysis." *Journal of the American Statistical Association,* 47 (December): 583-621.

ultiple Linear Regression Viewpoints
'ol. 9 No. 1 1978

A DEFENSE OF INFERENTIAL STATISTICS IN EDUCATIONAL RESEARCH

Paula L. Woehlke, Dennis W. Leitner, and Ernes L. Lewis
Southern Illinois University, Carbondale

Several recent criticisms have focused on the perceived overuse or abuse of inferential statistics in educational research. Brown (1975) criticized the use of multiple regression, while Derrick (1976) decried the used of inferential statistics in general, particularly techniques which are based on the general linear model. Although Spaner (1977) has responded to Brown's technical criticisms regarding multiple linear regression, to date there has been no reply to the general points made by these critics. However, the charges they make appear to be easily refuted, and it is the purpose of the present paper to do so, and to maintain that the use of general linear model techniques and inferential statistical techniques in general is not overemphasized in educational research.

The Criticisms

The criticisms made by Brown (1975) and Derrick (1976) can be summarized as follows:

(a) Inferential statistics in general, and the general linear model in particular, create a restricted view of educational reality by forcing research into a correlational mold

(Derrick, p. 36) or by placing it in strictly "independent variable-dependent variable terms" (Brown, p. 492).

(b) Use of statistics places undue emphasis on the test of statistical significance as an indication of the quality of research (Derrick, p. 36; Brown, p. 493.

(c) Inferential statistics require data to meet certain assumptions which are frequently ignored or overlooked by researchers (Derrick, pp. 36-37; Brown, p. 493).

(d) Studies using inferential statistics produce results which are not practically useful or meaningful. Brown (1975) phrases this as the presence of "a discussion which is strictly empirical and often of no theoretical importance" (p. 493) while Derrick (1976) says inferential statistics may "declare results to be encouraging in a statistical sense, whilst the practical sense is not considered" (p. 37).

Refutations

In order to refute these criticisms of inferential statistical procedures, it is necessary to first consider the goal of scientific inquiry in education. As with other sciences, basic or applied, this goal is to discover the cause or causes of observed phenomena. Educational research differs from other sciences (e.g., astronomy, physics, biology) only in that experimentation in a controlled setting is much more difficult and often impossible. Thus, scientific research in education often consists of the observation of natural changes in variables. Once these changes

have been observed and recorded, it is hoped that the reasons for these changes can be investigated in order to isolate and explain them. In looking for correlations among variables, the behavioral researcher is interested not only in whether the correlations exist, but also in using them to make predictions about what might happen in similar circumstances, and in isolating those factors which appear to explain the predicted outcome (see Kerlinger, 1977, for an extended discussion of this topic).

Given this view of scientific research in education, we can make the following replies to the critics of inferential statistics and the general linear model, point by point:

(a) Inferential statistics and the general linear model restrict the views of educational reality only to the extent that researchers allow their research topics to be dictated by the tools available to them. Nowhere is it written that one must do correlational research or that experimental research is superior to descriptive research, or inferior to it. Each type of research has a unique role to play in the conduct of educational research.

(b) Many theoreticians have discussed the problem of overemphasis on the test of statistical significance (e.g., Binder, 1963; Skipper, Guenther, & Nass, 1967; Edwards, 1965). However, theoretical discussions of the process of hypothesis testing presented in even elementary statistics texts (e.g., Hays, 1963, pp. 299-300; Games & Klare, 1967, pp. 422-426; Welkowitz, Ewen & Cohen, 1971, pp. 159-160) emphasize that statistical

significance relates only to the numbers gathered in the experiment and does not directly indicate the quality of the theory under investigation. Thus, the emphasis on a statement of statistical significance is not a function of the statistical technique but rather of its users.

(c) While it is true that most inferential statistical techniques including analysis of variance, multiple linear regression, and t-tests, require that data meet certain distributional and procedural assumptions and that many users of the techniques do not check the fit of their data to these assumptions, this does not invalidate the usefulness of the techniques. Rather, it indicts the users of the techniques as careless, unknowledgeable, or perhaps incompetent. One must not confuse deficiencies of statistical techniques with the deficiencies of those who use them.

(d) The criticism that the results of studies using inferential statistics are not immediately applicable, or do not definitively resolve crucial aspects of theory, appears to be based on a somewhat naive conception of science. Science does not consist of an accumulation of facts (Kuhn, 1962). Nor does science, at least behavioral science, consist of a series of "crucial experiments" which leads to definite answers about the truth of alternative theories, as described by Platt (1964). One cannot necessarily expect a single study in education to provide either practically useful results or final answers about basic

educational issues. Single studies which are not part of a program of educational research will give results which are at best "encouraging" (Derrick, 1976, p. 37), and at worst useless. The purpose of inferential statistics is to indicate when the outcomes of research are so unusual as to be rarely observed under certain hypothesized conditions; thus statistics cannot prove anything, and can never substitute for carefully designed studies which test aspects of theory. Those who require statistical inference to prove theories ask the impossible, in that theories are accepted on the basis of accumulations of positive evidence and, perhaps, esthetic considerations, which cannot be provided by any objective, mathematical process.

Summary

Critics of inferential statistics in educational research, particularly statistics based on the general linear model, have tended to attack the use of these techniques in current research rather than the techniques themselves. Specifically, statistical techniques cannot create restricted views of research, cannot be blamed for undue emphasis on the test of statistical significance, and are not to blame if researchers ignore their limitations and assumptions. Statistical techniques were never designed to provide answers about the ultimate truth or falsity of a substantive theory or about the utility of a procedure in any and all situations. Those who criticize inferential statistics on such grounds are blaming statistics for the shortcomings of researchers who misuse or do not understand the techniques.

References

Binder, A. Further considerations on testing the null hypothesis and the the strategies and tactics of investigating theoretical models. Psychological Review, 1963, 70, 107-115.

Brown, D. J. Mirror, mirror ...Down with the linear model. American Educational Research Journal, 1975, 12, 491-505.

Derrick, T. The criticism of inferential statistics. Educational Research, 1976, 19, 35-40.

Edwards, W. A tactical note on the relation between scientific and statistical hypotheses. Psychological Bulletin, 1965, 63, 400-402.

Games, P. A. & Klare, G. R. Elementary Statistics. New York: McGraw-Hill, 1967.

Hays, W. L. Statistics for Psychologists. New York: Holt, Rinehart & Winston, 1963.

Kerlinger, F. N. The influence of research on education practice. Educational Researcher, 1977, 6, 5-12.

Kuhn, T. S. The Structure of Scientific Revolutions. Chicago, Ill.: University of Chicago Press, 1962.

Platt, J. K. Strong inference. Science, 1964, 146, 3642.

Skipper, J. K., Guenther, A. L. & Nass, G. The sacredness of .05: A note concerning the uses of statistical levels of significance in social science. The American Sociologist, 1967, 2, 16-18.

Spaner, S. D. What inferences are allowable with a significant F in regression analysis? Multiple Linear Regression Viewpoints, 1977, 7, 62-74.

Biographical Resumes

Paula L. Woehlke is assistant professor in the Department of Guidance and Educational Psychology, Southern Illinois University, Carbondale, Illinois 62901. She received her B.A.Ed., M.A.Ed., and Ph.D. at Arizona State University, in educational psychology with a subspecialization in measurement and research design. She is a member of AERA Divisions C and D.

Dennis W. Leitner is assistant professor in the Department of Guidance and Educational Psychology, Southern Illinois University-Carbondale. He received his B.A. at San Diego State College and his M.A. and Ph.D. degrees at the University of Maryland, College Park. His specialization is educational statistics, and he is a member of AERA Divisions D and H.

Ernest L. Lewis is associate professor and chairman in the Department of Guidance and Educational Psychology at Southern Illinois University-Carbondale. He received his B.A. from Illinois State University and M.A. and Ph.D. from Southern Illinois University-Carbondale, specializing in educational statistics and measurement. He is a member of AERA Divisions D and H.

tiple Linear Regression Viewpoints
9 No. 1 1978

A CLOSER LOOK AT STATISTICAL INDEPENDENCE, ANALYSIS OF COVARIANCE AND DIRECTIONAL HYPOTHESES

Bradley E. Huitema

Western Michigan University

Spaner (1977) has summarized some findings and offered some opinions on the inferences that are allowable with a significant F in regression analysis. The purpose of this note is to clarify some of the points that appear to be misleading or incorrect. My concern is primarily with his statements concerning the independence assumption, the analysis of covariance and directional hypotheses. The paper is organized around these three topics.

Statistical Independence of Observations

After listing three assumptions associated with the F statistic (random-independent measures, homogeneity of variance and normality), Spaner states that he will "summarily dismiss these F ratio assumptions with an impressive list of citations." (p. 63)

It appears that he (appropriately) backs down on his dismissal of the independence assumption two pages later where he refers to Glass and Stanley (1970) who warn that violations of this assumption can be serious. Spaner then, on page 69, goes on to state that there are suprisingly no investigations of this "tenet" and suggests that Snedecor (1956) or Vasu and Elmore (1975) claim that " ... dependence of observations ($r \geq .95$) can cause disruption of accurate calculations. Snedecor (1956) has recommended the elimination of one of the pair of Xs with correlation greater than .95 (based on a redundancy interpretation)." Two aspects of his statements on the independence assumption are incorrect. First, investigations of this assumption have been conducted;

second, Snedecor's suggestion and the paper by Vasu and Elmore have nothing to do with the independence assumption.

Some of the investigations of dependence of observations are with reference to F and some are with reference to other test statistics, but one conclusion of these studies is very clear: nonindependence of the residuals produces serious distortions in probability levels of the test statistics. One might question what is meant by "serious distortion." Scheffé (1959) carried out some early work on the robustness of t and ANOVA F tests with respect to the independence assumption. He showed that when the observations (X_i) are autocorrelated with the lag 1 autocorrelation ρ, the variance error of the mean, to order 1/n is

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}\left[1 + 2\rho\right]$$

and the true probability of Type I error with nominal α can be determined through the use of a function that will not be described here. Padia (1976) extended Scheffé's method to the lag k case for large n. A small portion of Padia's results are summarized below.

Table 1 Probability of Type I Error for an Autoregressive (lag 1) Process for Nominal α = .01 and .05

	Actual Probability of a Type I Error	
ρ Autocorrelation	Nominal α = .01	Nominal α = .05
.50	.1362	.2584
.30	.0474	.1498
.10	.0204	.0768
.00	.0100	.0500
-.10	.0044	.0300
-.30	.0004	.0074
-.50	.0000	.0006

Notice that for nominal $\alpha = .01$ and $\rho = .50$ the true probability of Type 1 error is almost 14 times the stated value! Certainly we should not "summarily dismiss" the independence assumption.

The reference to Snedecor and to Vasu and Elmore is irrelevant with respect to the independence assumption because these authors deal with the problem of multicollinearity - which, of course, has to do with parameter estimation issues associated with high correlation among predictors. The problems of multicollinearity must not be confused with the independence assumption issue. In the former case the problem is dependence among predictor variables and in the latter case the problem is dependence among the observations or residuals.

If the reason for dependence among the observations is that the observations have been collected from a unit repeatedly across time, there are well established methods (e.g., Box and Jenkins, 1970) of transforming the data to yield uncorrelated residuals. Other causes of dependence among observations are much more difficult to handle because it is usually impossible to know whether or not independence is present. There is no such thing as a definitive test of the independence of observations unless the design is one in which time-series data are collected. In this case serial correlation or autocorrelation, which can easily be computed, is relevant.

In the typical non-time-series design, however, we simply don't know if the response one subject makes has an effect on the responses other subjects make or if some event common to all subjects in the sample has caused some dependence among all responses. This concern with dependence of observations is at the heart of much current research on the choice of the correct experimental unit to be employed in any statistical analysis, including regression.

Analysis of Covariance

Spaner touches on the analysis of covariance in discussing assumptions,

models and interpretations. He suggests (p. 63) that the only additional assumption (beyond independence, homogeneity of variance and normality) that is associated with ANCOVA is the homogeneity of variance assumption. I agree that this is a very important assumption but it is not the only additional assumption that should be considered. It is essential that researchers be aware that the covariate is assumed to be error-free and that the covariate and treatment effects are assumed to be statistically independent - especially if ANCOVA is employed with the ex post facto studies Spaner seems to favor. If ANCOVA is employed with this design the inferences allowable are very likely to be much different from those the researcher has in mind. And I am not referring to slight inaccuracies in probability values. I am referring to complete misinterpretations of the results: bias in the degree and even direction of the differences between adjusted means as well as gross errors in probability statements.

As I previously mentioned, Spaner does appropriately emphasize that the homogeneity of regression assumption is important. I must differ, however, with his statement on page 68 that "inferences based on significant F tests of covariance analyses will be rendered inaccurate if not invalid without the homogeneity of regression test." We cannot render the F test on adjusted means inaccurate or invalid by simply failing to compute the homogeneity of regression test. It is true, however, that we will be in the dark about whether or not the adjusted mean difference is constant across the various levels of the covariate included in the study if we do not carry out this test. This, of course, is very important information.

Directional Hypotheses

In his discussion of three aspects of curve fitting Spaner lists the intercept, the slope and the number of inflection points and states that "of these three factors only one is available for nondirectional hypotheses: the comparison of intercept points " (p. 69). Since he refers to the slope rather

than a plane or hyperplane he is dealing with the one predictor variable case and since he refers to the "comparison of intercept points" he is dealing with at least two samples. Why is the comparison between intercepts the only possible nondirectional test? If we once again employ Spaner's one predictor variable - two group example, we have two sample slopes as well as two sample intercepts from which population parameters are inferred. We may be interested in testing the equality of the two within group regression slopes. This test is, of course, the homogeneity of regression test. The associated null hypothesis is

$$H_0: \beta_1^{(1)} = \beta_1^{(2)}$$

where the superscript refers to populations one and two. This is a nondirectional hypothesis.

Spaner's description of the directional nature of the test on the slope may not, however, deal with the hypothesis stated above. Perhaps it deals with the case in which there are two (or more) groups and the predictor variable is simply a dummy variable. In this case the hypothesis associated with the F test is

$$H_0: \beta_1 = 0.0$$

which is, once again, nondirectional. If we prefer to deal with the population correlation coefficient rather than the population slope the hypothesis is

$$H_0: \rho = 0.0$$

which is also nondirectional. Hence, Spaner's conclusion that "Only in the statement of hypotheses about intercepts (group mean differences) is there the potential for a nondirectional F test ... " is not correct.

References

Box, G.E.P. and Jenkins, G. M. Time-series analysis: forcasting and control. San Francisco: Holden Day, 1970.

Glass, G. V. and Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Padia, William L. Effect of autocorrelation on probability statements about the mean. Paper presented at the annual meeting of the American Educational Research Association, 1976.

Scheffé, H. The analysis of variance. New York: John Wiley and Sons, 1959.

Snedecor, G. W. Statistical methods (5th ed.), Ames, Iowa: Iowa State University Press, 1956.

Spaner, Steven D. What inferences are allowable with a significant F in regression analysis? Multiple Linear Regression Viewpoints, 1977, 7, 62-74.

Vasu, E. J. and Elmore, P. B. The effect of multicollineanity and the violation of the assumption of normality on the testing of hypotheses in regression analysis. Multiple Linear Regression Viewpoints, 1975, 6, 21-50.

Multiple Linear Regression Viewpoints
Vol. 9 No. 1 1978

ON THE COMPARABILITY OF MULTIPLE LINEAR (MULR-05) AND INTERACTION (AID-4) REGRESSION TECHNIQUES

THOMAS E. Jordan
University of Missouri at St. Louis

INTRODUCTION

Creative use of regression techniques requires that use of linear and interaction terms be approached carefully. Given even a simple set of predictors the linear model may not be the most effective configuration, and the proportion of criterion variance assigned may be less than optimal. The obvious response is to hypothesize nonlinear forms of vectors, interactions of vectors, and conceivably, interactions involving nonlinear terms. For some areas of research cubed terms seem as high as one need go (Joossens and Brems-Heyns, 1975), and interactions beyond triple terms resist interpretation anyway. However, the conceivable range of complex vectors is enormous, even when theory eliminates some of the more forebidding combinations.

Koplyay's (1971) paper on interaction regression has done much to help the compulsive if perplexed investigator by pointing out that interaction terms in regression are manageable by computer programming. Jordan (1975) has shown that parsimonious models derived from use of the MULR-05 program can be explicated by use of the AID-4 program, and Spaner and Jordan (1973) have shown that models generated by AID-4 can be recaptured through MULR-05 with the R^2 value remaining unchanged. However, the two regression programs are different in their mathematical techniques and in their assumptions. Of the two approaches interaction regression makes fewer rigorous assumptions

and, in the view of some, may require large data sets because of loss of degrees of freedom (Sonquist, Baker, and Morgan, 1973).

The interaction regression algorithm explains the variance of the criterion variable by sequential splitting of the original group into subgroups. The splitting is done to minimize the within groups (error) sum of squares. This is accomplished by the examination of each possible split of every predictor variable of the current candidate group to be split and by the selection of the split giving the smallest within groups sum of squares. Three basic statistics are reported at each split in the branching process. An R squared value indicates the percentage of criterion variance explained through the current splitting. An F value indicating the significance of the reduction in the error sum of squares due to the current split. A one-way analysis of variance considering all groups at this stage leads to an F-value.

Multiple linear regression is a technique in which a full regression model for a criterion score is compared with a second model which is restricted by deletion of a critical variable. Evaluation is in the form of a probability level for the F test resulting from comparison of R^2_{full} and $R^2_{restricted}$. This technique is well suited to long predictor series and permits examination of data in both continuous and discrete form.

PROBLEM

This investigation addresses the question of whether interaction regression provides results concordant with those of multiple linear regression. It does so by asking the empirical questions:

1. Are predictors identified as sources of variance from both techniques comparable?

2. Are R^2 values from both regression techniques comparable?

METHOD

Data from the St. Louis Baby Study were assembled on just under two hundred children (N=196) using variables gathered from birth to age three years by prospective longitudinal study. These data plus others are reported more fully elsewhere (Jordan, 1975, 1979). The children whose traits are reported in Table I are mostly white and middle class, and have a mean PPVT IQ of 97 at age three, according to published norms. The data set was analyzed by use of interaction regression and multiple linear regression.

RESULTS

The multiple linear regression analysis summarized in Table 2 employed a predictive full model composed of five variables plus briefer models each deleting one of the five variables listed. The full model created an R^2 value of .14 (p=.0002), and the least powerful model, that deleting maternal education, created an R^2=.09 (p=.0005); all models were statistically significant. The most significant predictor was maternal education, accounting for over one-third of the full model's variance (R^2=.05). It was followed by race which accounted for two percent of the variance.

In the case of the interaction regression analysis, summarized in Table 3, the same full model composed of five predictors was used. The model used two predictor variables the first of which, and the prime source of variance, was maternal education. The second variable of the five was perinatal SES score. The interaction regression model in Table 3 may be written as:

$$Y = (\text{Mom's Ed.}) + (\text{SES}) + (\text{Mom's Ed.} * \text{SES}) + (\text{Mom's Ed}^2 * \text{SES}) + e$$

TABLE 1

DESCRIPTION OF THE SUBJECTS

(N=196)

	Sex[1] (M)	Race[1] (B)	SES[1]	Pop's[2] Occupation	Mom's[2] Work	Mom's[2] Education	PPVT(A)[2]
Range			16-78	11-70	1-3	1-5	6-54
Mean or %	.51	.06	48.04	40.09	2.53	3.12	27.32[3]
Sigma			15.29	17.72	.77	.95	11.14

[1]Gathered at delivery.

[2]Gathered at child age three years.

[3]=IQ 97.

TABLE 2

MULTIPLE REGRESSION ANALYSIS OF INFLUENCES ON 36-MONTH PPVT(A) RAW SCORES (N=196)

Predictor Variable	Models Compared	R^2	F	P
	Full Model 1	.14		.0002*
			2.01	.15
SES	Model 2	.13		.0002*
	Full Model 1	.14		.0002*
			3.66	.06
Race (B)	Model 3	.12		.0004*
	Full Model 1	.14		.0002*
			.86	.35
Paternal Occupation	Model 4	.13		.0001*
	Full Model 1	.14		.0002*
			.03	.84
Maternal Occupation	Model 5	.13		.0001*
	Full Model 1	.14		.0002*
			9.29	.002
Maternal Education	Model 6	.09		.0005*

*Significance of the difference from zero.

TABLE 3

INTERACTION REGRESSION ANALYSIS OF INFLUENCES

ON 36 MONTH PPVT(A) RAW SCORES

Antecedent Predictor	Predictors in Order of Contribution to Increased R^2	R^2 Contribution (-Cumul.)	p. of R^2 Increment	Cell Means		Cell N		Predictor Levels	
				High Cell	Low Cell	High Cell	Low Cell	High Cell	Low Cell
	Mom's	.08	<.001	29.05		154		2-4	
	Education				21.00		42		0,1
Mom's	SES	.12	.007	30.10		128		0-5	
Education					23.85		26		6-8
SES	Mom's	.13	.10	31.82		55		3-4	
	Education				28.81		73		2
Mom's	SES	.16	.01	28.00		11		3-4	
Education					18.52		31		5-8

This model explained sixteen percent of the criterion variance (R^2=.16).

Our first question asked if the two regression techniques provide concordant results, in terms of the variables identified by the two techniques. The answer may be given in three parts. First, both analyses show that no significance need be attached to paternal occupation, SES, or race. Second, both techniques identify maternal education as the variable of prime interest. In short, there is agreement on positive findings, agreement on negative findings, and a region in which divergent findings emerge.

The second question dealt with the matter of R^2 values. In fact, both regression systems give similar R^2 values; the lower R^2 value of .14 was generated by the simple multiple linear regression model. On the other hand, the more complex model composed of nonlinear and interaction terms generated a higher R^2 value, R^2=.16. This model is parsimonious, in terms of the number of developmental variables, while creating several terms within the regression equation. Comparability of the two models exists because both generated similar R^2 values using a sample of modest size (N=196).

DISCUSSION

We note the original problem faced by regression analysts: namely, how to cope with the potential permutations of variables implicit when relationships can be complex interactions and when nonlinearity of regression cannot be ruled out. We conclude that interaction regression may be a guide for the perplexed since it creates interactions as a part of the program introduced to regression analysts by Koplyay (1971).

It is evident from our analysis that compatible results can be created by interaction regression, both in terms of proportion of variance explained

and in terms of variables found useful and those not found useful. In addition, we note that this degree of compatibility emerges from a sample of moderate size. We conclude that interaction regression has much to offer based on the outcomes when multiple linear regression findings are used as the criteria of validity and utility. We confine these observations to premises of modest sample size and to the domain of developmental data on children. Larger data sets, longer predictor aggregates, and data from other domains may well provide a different degree of compatibility.

BIBLIOGRAPHY

Hollingshead, A. B. The Two-Factor Index of Social Position, 1957.

Joossens, J. F. and Brems-Heyns, E. High Power Polynominal Regression for the Study of Distance Velocity and Acceleration of Growth. Growth, 1975, 39, 535-551.

Jordan, T. E. Influences on Preschool Cognitive Attainment. Multiple Linear Regression Monograph, 1975, 1. (Separate).

Jordan, T. E. Old Man River's Children. Academic Press, 1979.

Koplyay, J. B., Gott, C. D., and Elton, J. H. Automatic Interaction Detection, AID-4. Technical Report, Air Force Systems Command, Brook A.F.B., 1971.

McGuire, C. M. and White, G. C. The Measurement of Social Status: Research Paper in Human Development #3. University of Texas, 1955.

McNeil, K. A., Kelly, F. J. and McNeil, J. T. Testing Research Hypotheses Using Multiple Linear Regression. Southern Illinois University Press, 1975.

Sonquist, J. A., Baker, E. L., and Morgan, J. N. Searching for Structure, Revised Edition, University of Michigan, 1973.

Spaner, S. D. and Jordan, T. E. An AID-4 Analysis of Maternal Antecedents To Locus of Control - Internal (LoC-1) At Age 60 Months, paper presented to the American Psychological Association, 1973.

The University of North Dakota
DEPARTMENT OF EDUCATION
GRAND FORKS 58201

DIVISION OF MEASUREMENT & STATISTICS
TELEPHONE: (701) 777-2511

TO: Members of the Multiple Linear Regression SIG

RE: A Planned Meeting of the SIG in Orlando, Florida in June, 1979

After several years on the back burner, we are going ahead with plans to have a special meeting of the SIG in or near Disney World (Orlando, Florida) in the early part of June, 1979. This meeting is in addition to the AERA Convention in San Francisco.

The tentative plans for the June meeting is to meet approximately four hours a day, perhaps 9:30-11:30 a.m. and 12:30-2:30 p.m. Tentatively, most sessions would be in a workshop format, with two hours devoted to each workshop. We could also have one or more paper sessions if there is sufficient interest. We would meet for three days. Obviously, it would be a great time to bring along families to sightsee and go to Disney World.

If you would be interested in either presenting a workshop or a paper, send your idea to Isadore Newman, University of Akron. Also, if there is a particular topic you'd like to see as a workshop topic, send that idea to Isadore.

Cordially,

John D. Williams
Chairperson, Multiple Linear Regression SIG

JDW:jcf

If you are submitting a research article other than notes or comments, I would like to suggest that you use the following format, as much as possible:

Title
Author and affiliation
Indented abstract (entire manuscript should be single spaced)
Introduction (purpose—short review of literature, etc.)
Method
Results
Discussion (conclusion)
References

All manuscripts should be sent to the editor at the above address. (All manuscripts should be camera-ready copy.)

It is the policy of the sig=multiple linear regression and of *Viewpoints* to consider for publication articles dealing with the theory and the application of multiple linear regression. Manuscripts should be submitted to the editor as an original, single-spaced typed copy. A cost of $1 per page should be sent with the submitted paper. Reprints are available to the authors from the editor. Reprints should be ordered at the time the paper is submitted and 20 reprints will cost $.50 per page of manuscript. Prices may be adjusted as necessary in the future.

"A publication of the *Multiple Linear Regression Special Interest Group* of the American Educational Research Association, published primarily to facilitate communication, authorship, creativity, and exchange of ideas among the members of the group and others in the field. As such it is not sponsored by the American Educational Research Association nor necessarily bound by the Association's regulations.

"Membership in the *Multiple Linear Regression Special Interest Group* is renewed yearly at the time of the American Educational Research Association Convention. Membership dues pay for a subscription to the *Viewpoints* and are divided into two categories: individual=$3.00; and institutional (libraries and other agencies)=$12.50. Membership dues and subscription requests should be sent to the Executive Secretary of the MLRSIG."

THE UNIVERSITY OF AKRON
AKRON, OHIO 44325

TABLE OF CONTENTS