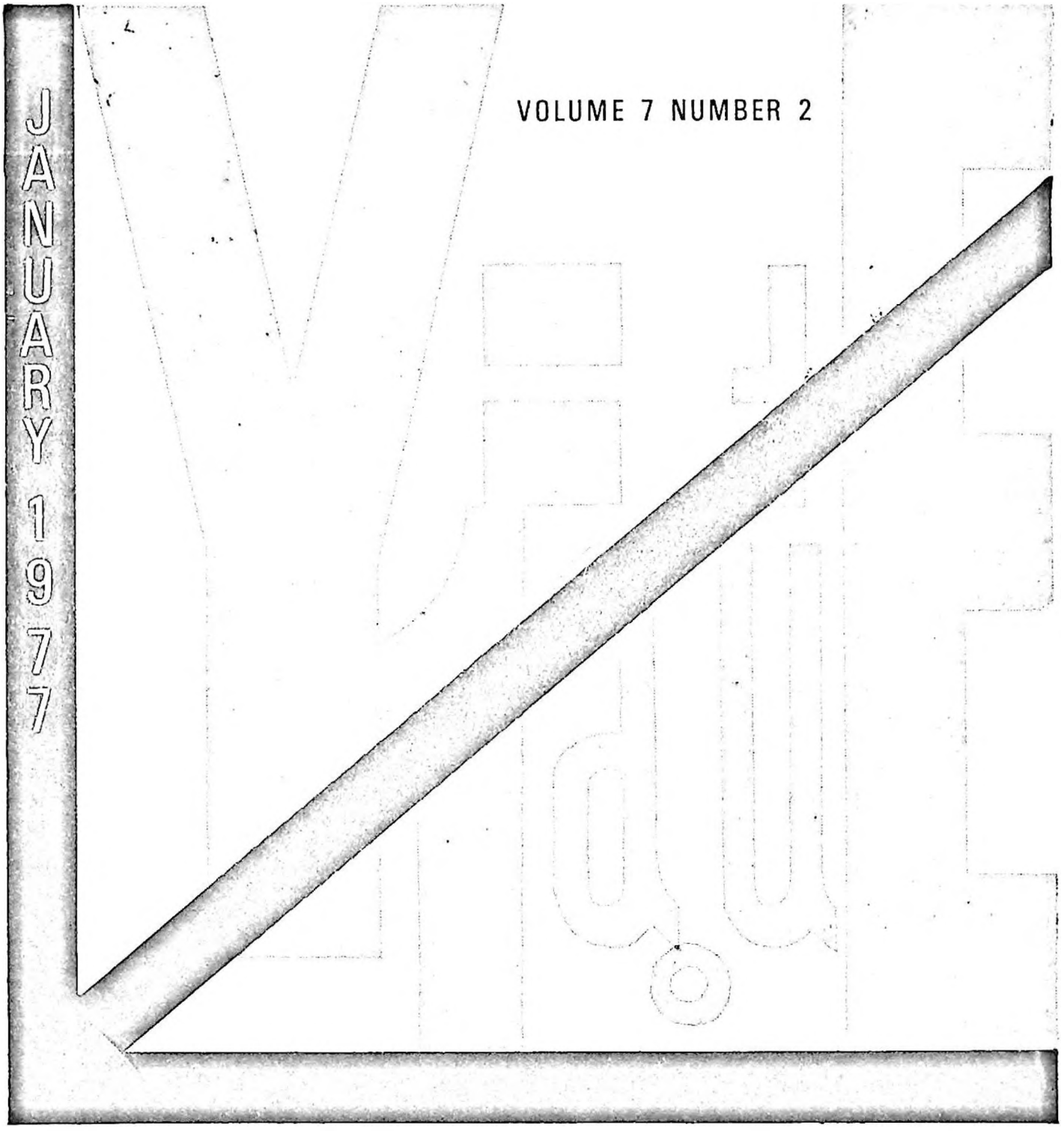
VOLUME 7 NUMBER 2

JANUARY 1977



MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group
on Multiple Linear Regression

TABLE OF CONTENTS

Multiple Linear Regression Viewpoints
Vol. 7 No. 2 1977

SPURIOUS AGGREGATION AND THE UNITS OF ANALYSIS*

Hugh Poynor

Kirschner Associates, Inc.

Washington, D.C.

In educational research studies the principal investigator takes steps to insure the validity of reported findings by designing the study properly and analyzing the data thoroughly. It is only when failure of design and analysis efforts occurs that spurious findings are reported. Beyond the reporting stage may lie the scrutiny of collegues, an audience of policy makers with their own urgent perspectives, a need for replication, and a hope of further understanding of reported effects.

Our symposium today is concerned with the research implications of data aggregation and unit of analysis issues. These topics many times are obscured by overly detailed statistical and mathematical methodologies. Yet it is meaningful to characterize their intention with the familiar story of the man seeking to locate his house keys two blocks from where they were lost because of better light where he was looking. In this same vein let me say that those with information needs concerning state-level policy will not find appropriate answers using the pupil as the analysis unit even though that is many times the most conventional place to look. Similarly, one cannot expect to

---

*Paper presented at annual AERA convention held in San Francisco, April, 1976. Author's current address is: Suite 1137, 733 15th Street, N.W., Washington, D.C. 20005

shed light on the effects of a particular instructional technique on the pupil by analyzing district-level aggregates. The choice of a unit of analysis in all cases must be directed by the research goals, or by the audience for the findings and their special needs, and by methodological considerations such as independence of the units.

It will now be noted how aggregation and unit of analysis phenomena have potential for creating spurious analysis outcomes. By way of proceeding to these problems, it is necessary to consider several areas that are often thought to be unrelated, and then to re-introduce analysis techniques from a new point of view.

Subsample Boundary (the G variable)

Drawing subsamples of the units of analysis in a data base may be done in several ways and for several reasons. Subsamples are often separated according to identifying features such as school grade level, instructional treatment and so on. Once such samples are identified it is a simple task to devise a mapping variable called G that will serve to define the boundaries of the subsample. This G variable is nothing more than a rule for one of two actions: (1) forming aggregates (or averages, as is usually the case), and (2) identifying interactions among elements of X and Y. Definition of interactions through the use of G is accomplished by simply using the rules to expand the analysis model, or, in general linear model terminology, by expanding the number of predictor variables. The present paper shows

several effects of sample boundary variables in routine analysis efforts in light of hypothetical parent-sample homogeneity or heterogenity.

Traditional Suppressor Phenomena

The action of independent variables adding to prediction of a dependent variable through apparent relationships with other independent variables but not through direct relationships with the dependent variable has been termed the suppressor effect. Two conditions must be met before the independent variable is thought to be operating as a classical or traditional (Cohen & Cohen, 1975; Conger, 1974) suppressor in multiple regression settings: (1) the X variable must be only slightly or not at all related to the dependent variable (Y) and (2) this same X variable must be strongly related to at least one other X variable. Under these conditions multiple prediction will be increased by a factor that usually exceeds the bivariate relation between Y and X (suppressor) variable. For this reason, suppressors are welcome additions to multiple regression models. Beyond this highly specific pattern of bivariate correlations, a traditional suppressor is identified by its negative regression weight. Thus, the regression weight which a suppressor acquires is positive when it is negatively correlated with the dependent variable. In other language, the raw weight and the standardized (or beta) weight have opposite signs.

Suppressor phenomena occur only when an inconsistency exists between the X variable set and the Y variable. Inconsistency arises from the manner in which subsamples of Y are related differentially to X.

The variance in Y may be visualized as being composed of a subset of Y scores bearing a positive correlation with selected X_1 values and another subset of Y bearing no correlation with the remainder of X_1 values in the equation $Y = X_1 + X_2$. This latter subset of Y is highly predictible from the corresponding X_2 values even though on the whole Y and X_2 are not related. In this example, X_2 is the suppressor variable. When all Y values are used to compute r (Y, X_2), the correlation is near zero, but it is possible to find a smaller set of Y which is highly correlated with a corresponding X_2 set. In order for the suppressor effect to operate, then, X_1 and X_2 must share these subsample boundaries.

Guilford (1964) and others do not appeal to this type of explanation of suppressor effects, although a discussion of variance borders on the subsample boundary explanation proposed in the present paper. It is explained that X_1, in spite of a correlation with Y, has some variance that correlates near zero with Y. It is because of this X_1 variance that the correlation between Y and X_1 is prevented from being even larger. Now, variable X_1 correlates highly with X_2 (the suppressor) because they have in common that variance not shared by Y. Thus, including X_2 in the multiple regression model permits this portion of the Y variance to find a capable predictor. A boundary explanation is more generally useful than this traditional variance explanation of suppressors since it leads to the understanding of X-Y relationships in a broader number of analysis situations. Usefulness of this explanation will next be demonstrated with commonality analysis interpretations.

Negative Commonalities

More recently than Guilford, Veldman (1974) and Kerlinger & Pedhazur (1973) have called attention to "negative contributions" in commonality analysis results. The commonality procedure examines each X variable as if it is the last variable in the dependent variable set to be added to the regression equation. The net value of X in the equation is the percent increase in explained Y variance as computed by subtracting the R^2 figures of the before-and-after X models. The commonality name for the procedure arises because pairs of variables, triplets, and so on are similarly treated as the last additions to the model so that their common net value may be determined. Since the output of commonality analysis consists of percents of Y variance explained by both single and joint contributions X's make toward explaining Y variance, then the negative contribution outcome is a signal that <u>less than nothing</u> in Y has been explained. Negative contributions are possible only for commonality values, that is for pairs of X variables while they are impossible for unique (single X) contributions.

Besides being a sad state of affairs after research funds have been expended for the study, explanation of less than nothing can also be a clear signal that suppressors are operating in the data set. Extending the present interpretation of suppressors to negative contributions, then, suggests that they are products of the action of a boundary variable. Just as there are inconsistencies in X-Y relationships which will produce suppressor phenomena, commonality analysis will also register the

inconsistencies. By inconsistency is meant the irregular quality of X-Y correlations when subsamples of X and Y are considered. In other words the sample is not homogeneous with regard to the manner in which X and Y relate statistically. For example one portion of the sample may be high and positively related on X and Y, while another may be negatively related.

Analysis of Variance

The problem of sample heterogenity can also be considered in the context of analysis of variance situations, where effects are sought as indicators of experimental treatment outcomes. Traditional use of the term "relationships" is made for observational-correlational research studies while use of the term "effects" is usually reserved for experimental studies. In either case the G variable is a useful heuristic for the topic of spurious aggregation and the units of analysis.

Attention to parent sample heterogenity in experimental school effect studies comes from the popular textbook by Glass & Stanley (1970). Their verbal analysis of independence and the sampling units of analysis led them to conclude that degrees of freedom must suffer if the research setting is the intact classroom. A setting such as this is quite frequent in our resesrch, and unfortunately it is many times mistreated in the statistical analysis work. Glass and Stanley (1970) recommend that one form classroom means and adjust degrees of freedom accordingly. That is, the researcher should aggregate the pupil scores, form classroom means, and adjust df to reflect the number of classrooms rather than the number of pupils.

In terms of the G variable approach to this analysis of variance situation, it may be clear already that the G variable is the classroom identification of each pupil. If Glass & Stanley are empirically correct, we must expect there to be considerable subsample heterogenity with respect to G, for if the sample is found to be homogeneous with respect to G then aggregation and loss of df is a questionable procedure.

Because aggregation may be questionable and because it is often done--sometimes automatically--when analyzing both effects and relationships the recommendation is best subjected routinely to empirical test. Poynor (1974) has provided evidence that the untested use of aggregate units of analysis of classroom (and sometimes, individual units of analysis) can lead to grievous Type I and II errors. Either unit of analysis can lead to errors, so exclusive use of either unit accomplishes nothing. A pre-analysis step is required to identify the proper unit of analysis. In an investigation of the independence of analysis units, Glendening (1976) discusses the use of this pre-analysis test, but finds it to be too conservative or too liberal a test in selected situations of independent and dependent units. These two studies show the empirical test of a unit of analysis with respect to sample homogeneity, and the importance of management control in effect studies for protection of the selected units. Because of the great loss of information that occurs with aggregation (Poynor, 1975) it should not be done until after it is proven necessary.

Correlation and Regression

Thorndike (1939) called attention to the operation of G in correlational-relationship studies by illustrating the spurious development of a relationship across higher order aggregations. His early example calls attention to real-world manifestations of G. Twelve school districts were each asked to provide two data points on each pupil in the district: the pupil's IQ score (X) and the number of rooms available in the school building where each pupil was taught (Y). Each individual district was quite heterogeneous with respect to these variables. That is, within each district there was a broad range of IQs and rooms yet there was no correlation between these X and Y variables within each district. The problem in the example comes from heterogenity on these X-Y variables across districts. When the 12 districts were combined and a correlation coefficient was computed, it was .45, not .00 as it has been in each district individually. No aggregation had yet taken place. Once school district aggregates (means) were used, the resulting correlation was .90.

Extensive research work with simulations and literature reviews by the team of Hannan and Burstein have been done in the correlation and regression analysis areas (Hannan, 1971; Burstein, 1975). These studies provided the impetus for the present synthesis of the effects and relationships areas using the G variable concept. This author has sought to apply their detailed statistical research findings to popular analysis models, using nontechnical language. Where the above authors refer to bias and inefficiency (or inaccuracy and inconsistency) of

regression weights, these terms have been collected here under the label "spurious." Using their findings to predict spurious correlations, it is necessary to employ the G variable again. Briefly, they conclude that spurious relationships will be products of situations where strong relationships exist among G-X or G-Y. Recall that both these situations were present in the Thorndike paper, which is somewhat classic as an example of confounded or proxy relationships.

Measurement of G

Before ending this paper, let me say that it is very little comfort to know that suppressors, negative commonalities, Type I and Type II errors and inflated correlation coefficients can be explained in terms of a G variable used to establish the unit of analysis. Research which is carefully planned and conducted will rarely be affected by this as a nuisance at the time of data analysis. Still, our understanding of effects and relationships among measures of interest is many times insufficient to control all the potential contaminants of our findings.

An ultimate solution is believed to be the measurement of G itself, for if G is viewed as an abstract rule for selecting or forcing observations into groups prior to analysis, then G is truly a potent treatment variable. This solution does not refer to the practice of simply including dummy variables in a regression equation as indicators of school district, and school building location of the observation.

The importance of fully specified models, or true starting models as they are sometimes called, is well known to data analysts. While the

practice of using such dummy variables often increases the percent of explained criterion variance, it does nothing more than acknowledge G as potent.

It is the measurement of the underlying classroom, school building or district differences that will promote our understanding of G and the effects and relationships associated with our criterion variables. Instead of making sterile statements such as "Twenty percent of the criterion variance was explained by school building differences," the researcher may someday be able to offer richer, more meaningful statements relating criterion variance to specific features or conditions, such as teacher feedback, time spent on material, amount of direct questioning, presence of open classrooms or other substantive learning variables.

Burstein, Leigh. Data Aggregation in Educational Research. Paper presented at the Annual meeting of AERA, Washington, D.C., 1975.

Cohen, Jacob and Patricia Cohen. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. John Wiley, New York, 1975.

Conger, A.J. A revised definition for suppressor variables: a guide to their identification and interpretation. Educ. Psychol. Meas., 1974, 34, 35-46.

Glass, G.R. and Stanley, J.C. Statistics in Education and Psychology, Prentice-Hall, Englewood Cliffs, N.J. 1970.

Glendening, Linda. The Effects of Correlated Units of Analysis: Choosing the Appropriate Unit. Paper presented at the Annual Meeting of AERA, San Francisco, 1976.

Guilford, J.P. Fundamental Statistics in Psychology and Education, McGraw-Hill, New York, 1964.

Hannan, Michael T. and Burstein, Leigh, Estimation from grouped observations. Am. Soc. Rev., 1974, 39, 374-392.

Hannan, Michael T. Problems of Aggregation. In Blalock, H.M. (ed.) Causal Models in the Social Science. Aldine Press, Chicago, 1971.

Hannan, Michael T., Young, A.A. and Finielsen. Specification Bias Analysis of the Effects of Grouping of Observations in Multiple Regression Models. Paper presented at the Annual Meeting of AERA, Washington, D.C., 1975.

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

Poynor, Hugh. To Aggregate is to Aggravate. Paper presented at the Annual Meeting of AERA, Washington, D.C., 1975.

Poynor, Hugh. Selecting units of analysis. In Gary Borich (ed.) Evaluating Educational Programs and Products, Englewood Cliffs, N.J., Educational Technology Press, 1974.

Thorndike, E.L. On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. Am. J. Psychol., 52, 1939, 122-124.

Veldman, D.J. Negative Contributions in Commonality Analysis. Project PRIME working paper, The University of Texas at Austin, 1974.

tiple Linear Regression Viewpoints
.7 No. 2 1977

COMMENTS ON POYNOR'S PAPER

Earl Jennings

The University of Texas at Austin

The paper is an inadequate vehicle for communicating clearly anything in particular. In common with many speeches and papers, an effort is being made to cover too much ground in too little space or time. The net result is probably that only those persons who already understand the point Poynor is trying to make will understand the paper.

I have grave reservations about the utility of drawing inferences from "suppressor" effects. Consider the claim made at the end of the third page, "The variance in Y may be visualized as being composed of a subset of Y scores bearing a positive correlation with selected X_1 values and another subset of Y bearing no correlation with the remainder of X_1 values in the equation $Y = X_1 + X_2$." I am not sure that such a situation leads to "a suppressor effect," but if all that is being claimed is that the difference in Y per unit difference in X_1 at fixed levels of X_2 is non-zero in some range of X_1, but zero in some other range of X_1, it would seem appropriate to reflect that fact in the model.

Likewise, I have reservations about the utility of drawing inferences in the commonality procedure from negative "joint contributions." In general, even the use of "unique contributions" has some drawbacks because it is not an unbiased estimate of anything.

The mathematics of this procedure guarantees that the unique contribution value will be greater than or equal to zero even if X is taken from a table of random numbers. Use of the language "less than nothing has been explained" is equivalent to the concept of negative variance. If you subtract one variance from another it is possible to get a negative number, but it is not a negative variance. Naming the number is not sufficient to make it interesting.

Finally, I regard two points as technically inaccurate. Near the end of the third page, contrary to the claim, raw weights and standardized weights cannot have opposite signs. At the top of the fifth page, the word should be independent rather than "dependent."

As to the disposition of the paper, a number of possibilities exist. One of the original goals in founding Viewpoints was to provide a vehicle for ventilating viewpoints among a relatively homogeneous group of individuals without "publication lag." The recently adopted review process may very well sabotage that goal if it leads to rewrites, re-reviews, etc. A possible mechanism for having our cake and eating it too, would be to publish an unedited paper along with the review and perhaps a rebuttal of the review. In order to speed publication, these items would not necessarily have to appear in the same issue. If Poynor does not object, why don't we try it.

tiple Linear Regression Viewpoints
.7 No. 2 1977

REJOINER TO JENNINGS

Hugh Poynor

Kirschner Associates, Inc.

If a clear distinction can be made between conceptual organizers that promote efficient storage of large quantities of facts and the outright glutting of facts, then the distinction should be employed in contrasting my paper and Jennings' comments. The suggestion of employing a subsample boundary (G variable) explanation for diverse regression phenomena is an appeal to a convenient conceptual organizer. If G doesn't provide new statistical knowledge to the reader, it is because it is not a real cause of regression outcomes. Rather, it is a framework (or filing system) for users of linear regression. Continuing in this same vein, let me say it is handy and important to have pants pockets even though they don't explain how or why I wear pants and they certainly don't explain why I've chosen the contents of the pockets.

G is handy because it gets together several different aspects of regression analysis, and it is useful to collect thoughts when generating models for analysis. Beyond the research questions that serve to stimulate analysis models, the internal statistical features of data sets should not be omitted from analysis models. G gets together several of these features, that is, it is a helper in collecting thoughts about suppressors, negative commonalities, aggregate levels in classroom experiments, spurious correlation, and aggregation artifacts. In time, other topics are certain to be added to this list.

Turning to Jennings' comments on my suppressor effects, I definitely agree that they are entirely appropriate (and valuable)

aspects of data sets to be reflected in models. By restating my "subsample of Y" conception in terms of the range of X_1, it seems Jennings has come to this same conclusion. Nevertheless, after thinking about my statement made in the original paper it does seem trivial to locate a subsample of data points without regard to the range of either X or Y, so I would restate the subsample boundary method to indicate a "selected range of X, (or Y) values."

I share the concern about inferences from negative commonalities. I don't make such inferences. In the paper negative commonality was pointed to as provacative and not as an interpretation of anything.

Two technical inaccuracies are acknowledged. Raw and standardized weights cannot have opposite sign. When these values have signs different from the validity (zero-order correlation), however, then the suppressor relationship is definitely signified. At the top of the fifth page, the word should be "independent."

Nearly anyone benefits from directed interaction with their colleagues, certainly I do. Convention papers are too brief for adequate consideration by an audience. This review has provided consideration of a convention paper which needed careful thought. In one instance an idea has proven to be poorly stated or wrong and can be improved. I welcome the opportunity for directed interaction and plan to employ Viewpoints as a quick turn-around forum on some more recent thoughts concerning marker variables for classroom assignments in research studies of schooling effects.

* Multiple Linear Regression Viewpoints
Vol. 7 No. 2 1977

REGRESSION APPROACHES AND APPROXIMATE SOLUTIONS TO ANALYSIS OF VARIANCE WITH DISPROPORTIONALITY VARIED

STARRETT DALTON
University of California at Riverside

The degree of nonorthogonality in a factorial design was systematically increased. Five methods of dealing with nonorthogonality were selected and applied: two were least squares solutions (Method 1 and Method 2); two were approximate solutions (the unweighted means analysis and the method of expected frequencies); and the fifth was the alternative of data elimination. Under extreme nonorthogonality all methods converged in yielding conclusions which while erroneous were similar across methods. Under moderate nonorthogonality, however, the unweighted means analysis and Method 1 were superior. Overall, the data elimination alternative was inferior in that it led to more type II errors than any of the other four methods.

Authors of current textbooks in design and statistics (Keppel, 1973; Dayton, 1970) acknowledge the superiority of the least squares methods yet refer researchers whose data evidences only slight departures from proportionality to the more easily understood and computationally simpler approximate solutions. The most commonly recommended ones are the unweighted means analysis and the method of expected frequencies. Data elimination to achieve equal n is also frequently recommended.

Several investigators, (particularly Carlson & Timm, 1974; Timm & Carlson, 1975; Overall, Spiegel, & Cohen, 1975; Cohen & Cohen, 1975; & Overall & Spiegel, 1969) have compared the various regression solutions and clarified the hypotheses tested by each. Yet, despite this clarification no one has empirically compared the best known regression solutions to the more popular approximate ones.

Three regression procedures for analyzing disproportional data will be defined. The reader who is not familiar with these methods is referred to the seminal article by Overall and Spiegel and a more recent investigation by Carlson and Timm. These solutions are conceptually and algebraically similar in the univariate and the multivariate cases (Woodward & Overall, 1975), but the present discussion will be limited to univariate analyses since multivariate extensions of the approximate solutions have not been developed.

Method 1 requires mutual adjustment of all main effects and interactions. The structural model for a two-way analysis of variance is: $X_{ijm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijm}$, where μ is the grand mean, α_i is the treatment effect for level i of the first factor, β_j is the treatment effect for level j of the second factor, $\alpha\beta_{ij}$ is the interaction term for cell ij, and e_{ijm} is the error for individual m in cell ij. Carlson and Timm have argued cogently that Method 1 is the best extension of traditional analysis of variance because the same parameters are estimated and the same hypotheses are tested in the orthogonal and the nonorthogonal cases. The reader is referred to their writing for an extensive presentation concerning the hypotheses tested by Method 1. No rigorous proof is known, but empirical tests carried out by the present author indicate that results with Method 1 coincide with results found by using the weighted squares of means method as described by Winer (1971) and by Schuessler (1971).

Method 2 is similar to Method 1. The structural model for Method 2 is $X_{ijm} = \mu + \alpha_i + \beta_j + e_{ijm}$ where the terms are defined as they were for Method 1. The essential difference between the two methods is that Method 2 requires the assumption that no true interaction exists and consequently the interaction is not taken into consideration when estimating main effects. An empirical test shows that the results with Method 2 are identical to those with the method of fitting constants as described by Winer. In fact, the congruence of these two methods is so obvious it seems odd that they are still referred to by separate names. Based primarily on tradition and popular acceptance, Overall and Spiegel initially suggested that Method 2 was the best generalization of analysis of variance to the nonorthogonal case. However, they later reversed their stance (Overall, Spiegel, & Cohen) and argued that Method 1 is the proper generalization of analysis of variance to the nonorthogonal case. Methods 1 and 2 are identical when the interaction is exactly zero.

The structural model for Method 3 is identical to the model for Method 1. Method 3, however, assumes a priori evidence to justify an ordered entry of vector sets representing α, β, $\alpha\beta$ into the regression equation. The hypotheses tested with Method 3 are not the same as those tested with analysis of variance. While this method has the advantage that the component sums of squares equal the total sum of squares, the requirement of establishing a logical a priori ordering of variables is

typically impossible for a researcher to establish, thus it is of very limited usefulness. Consequently it will not be considered in this study.

The approximate solutions for the analysis of nonorthogonal data are adequately described in a number of textbooks (Myers, 1967; Winer, 1971; Kirk, 1968; & Glass & Stanley, 1970). The unweighted means analysis uses cell means to estimate main effects and interaction, and adjusts the error term by a factor which reflects the unequal cell sizes. The method of expected frequencies involves multiplication of cell means by the expected cell frequency to obtain a sum for each cell. Sums obtained in this manner are used in estimating main effects and interactions. The third and final commonly recommended non least squares procedure to be considered here is not really a solution to nonorthogonal data analysis: it consists of randomly eliminating data to achieve equal n.

Some general, but largely unsubstantiated, guidelines are available concerning the appropriate selection of a least squares or an approximate solution. Winer and Dayton indicate that the unweighted means analysis is applicable if the experimental design called for equal n and if subject loss was essentially random. Myers (1967) offers a similar caution. In addition, he warns the experimenter not to use the unweighted means analysis when the "n's are very disparate." Myers also views the method of expected frequencies as appropriate when proportionality can be assumed and when departure from proportionality is not too great. When the departure is large, he indicates that a least squares method should

be undertaken. If one follows this recommendation, a least squares solution would be selected when disproportionality is relatively large.

Williams (1972) has compared certain selected least squares and non least squares methods. However, he omitted the least squares solution (Method 1) which in view of the subsequent research of Carlson and Timm may well be the most acceptable generalization of traditional analysis of variance to the nonorthogonal case. In addition, Williams included a discussion of the hierarchical method (Method 3), the unadjusted main effects method, and the method of weighted means. These three methods are of limited utility. The difficulty inherent in the hierarchical method has been discussed above. The method of weighted means is seldom recommended when there are two or more missing scores per cell. The unadjusted main effects method places the variance shared by main effects into all main effect estimates thereby unjustifiably inflating the F values and leading to type I errors.

No one seems to have indicated when departures from proportionality should be considered large or to have established a criterion for judging when a least squares solution becomes more appropriate than a non least squares procedure. The main purpose of this study is determine if the least squares solutions and approximate solutions diverge as the degree of nonorthogonality increases. Based on the literature review it is hypothesized that they will diverge when nonorthogonality is extreme.

Other possible methods were also considered but omitted from detailed examination due to characteristics which mark them as clearly unacceptable. These were (1) a modification of Method 2 which entails pooling insignificant interactions and the error term, and (2) substitution of the cell mean for the missing values in that cell. The mean-substitution method tends to yield F values much higher than those in the basic data analysis which is conducted before any subject loss is incurred. A clearly unacceptable rate of type I errors ensues. The pooling procedure in conjunction with Method 2, even when the interaction is insignificant, can have effects which are very difficult to predict because the pooled interaction will itself change with increasing degrees of nonorthogonality. As a result it was not included in the detailed comparisons which follow.

Method

Two 3 X 3 factorial designs were used as the initial data sets (Table 1). Each basic data set contains 81 scores, nine per cell, but only three different scores. Such score duplication, which allows frequency reduction and the introduction of nonorthogonality without modifying cell means or the parameter estimated was used by Overall, Spiegel, and Cohen. Analysis of Basic Data I results in a significant interaction as well as two significant main effects. Basic Data II differs in that it produces no significant interaction. Analyses of Basic Data I and II are assumed to represent the correct results.

Table 1

Two sets of basic data and four cases in which certain cell frequencies ( ) were reduced to simulate random data loss or to achieve equal n.

Basic Data I
Interaction Present

3 6(9) 9	6 4(9) 7	9 12(9) 14
5 6(9) 8	6 5(9) 7	3 5(9) 8
10 12(9) 9	11 13(9) 15	12 8(9) 10

Basic Data II
No Interaction Present

3 5(9) 8	3 4(9) 9	10 6(9) 9
6 6(9) 7	5 6(9) 8	9 12(9) 8
5 12(9) 7	4 14(9) 10	11 13(9) 15

(9)	(6)	(6)
(6)	(9)	(6)
(6)	(9)	(9)

1. Case 1, moderate nonorthogonality. $X^2 = 2.2$, N = 66

(6)	(6)	(6)
(6)	(6)	(6)
(6)	(6)	(6)

2. Case 2. Frequencies reduced from 66 found in Case 1 to 54 to achieve equal n.

(9)	(3)	(3)
(3)	(9)	(3)
(3)	(9)	(9)

3. Case 3, extreme nonorthogonality $X^2 = 11.99$, N = 51

(3)	(3)	(3)
(3)	(3)	(3)
(3)	(3)	(3)

4. Case 4. Frequencies reduced from 51 found in Case 3 to 27 to achieve equal n.

be undertaken. If one follows this recommendation, a least squares solution would be selected when disproportionality is relatively large.

Williams (1972) has compared certain selected least squares and non least squares methods. However, he omitted the least squares solution (Method 1) which in view of the subsequent research of Carlson and Timm may well be the most acceptable generalization of traditional analysis of variance to the nonorthogonal case. In addition, Williams included a discussion of the hierarchical method (Method 3), the unadjusted main effects method, and the method of weighted means. These three methods are of limited utility. The difficulty inherent in the hierarchical method has been discussed above. The method of weighted means is seldom recommended when there are two or more missing scores per cell. The unadjusted main effects method places the variance shared by main effects into all main effect estimates thereby unjustifiably inflating the F values and leading to type I errors.

No one seems to have indicated when departures from proportionality should be considered large or to have established a criterion for judging when a least squares solution becomes more appropriate than a non least squares procedure. The main purpose of this study is determine if the least squares solutions and approximate solutions diverge as the degree of nonorthogonality increases. Based on the literature review it is hypothesized that they will diverge when nonorthogonality is extreme.

Case 1 and Case 3 represent reductions in frequencies that lead to moderate (Case 1) and extreme (Case 3) nonorthogonality. Further reduction of the frequencies found in Case 1 and in Case 3 leads to the equal, although reduced, frequencies found in Case 2 and Case 4. The four solutions, i.e., Method 1, Method 2, the method of expected frequencies, and the unweighted means analysis, will be applied to both sets of data under both conditions of nonorthogonality. Based on these results the four solutions will be compared to each other and to the fifth alternative of further eliminating subjects to obtain equal n in all cells. Such elimination produced Case 2 and Case 4.

An overview of the various analyses and comparisons can be offered in terms of Table 2. The data in cells G and H will be analyzed using the four solutins to nonorthogonal data analysis and these results will be compared to the results from analysis of data in cells A and B. The four solutions to the nonorthogonal data in cells G and H will also be compared to the analyses of the data configurations in cells C and D. The data in cells I and J, representing extreme nonorthogonality, will be analyzed using all four solutions and the results will be compared to the results from the analysis of data in cells A and B. Results from cells I and J will also be compared to results from cells E and F. As can be seen in Table 2, it is possible to compare the four solutions to each other and to the alternative of eliminating scores to achieve equal n both when nonorthogonality is moderate and when it is extreme.

Table 2

Schematic representations of all data configurations and two sets of data.

Data Configurations	Basic Data I	Basic Data II
Equal n (n = 9)	A	B
Case 2 (n = 6)	C	D
Case 4 (n = 3)	E	F
Case 1, moderate nonorthogonality (X^2 = 2.20)	G	H
Case 3, extreme nonorthogonality (X^2 = 11.99)	I	J

Results and Discussion

When nonorthogonality was moderate and when interaction was present (Table 3), Method 1 and the unweighted means analysis both yielded essentially the same conclusions (α = .05) that were reached by analysis of Basic Data I prior to any frequency reduction whatsoever, namely,

Table 3

Data analyses of Basic Data I (Interaction present) with varying degrees of nonorthogonality.

Degree of Orthogonality		Basic Data I Results used as a standard for comparison			Case 2			Case 4		
		df	F	P	df	F	P	df	F	P
	A	2	60.12	.0001	2	37.58	.0001	2	15.04	.0001
Orthogonal	B	2	4.49	.0145	2	2.81	.0708	2	1.12	.3480
Data	AB	4	18.30	.0001	4	11.44	.0001	4	4.58	.0100
(N=81, 54, 27)	WG	72			45			18		

Four Solutions to Nonorthogonal Data Analysis

		Method 1		Method 2		Expected Frequencies Method		Unweighted Means Analysis	
		df	F	P	F	P	F	P	F	P
Case 1	A	2	46.66	.0001	48.97	.0001	50.56	.0001	45.12	.0001
Moderate	B	2	3.23	.0469	3.06	.0546	3.01	.0572	3.38	.0408
Nonorthogonality	AB	4	13.31	.0001	13.31	.0001	14.01	.0001	13.73	.0001
(X^2=2.2, N=66)	WG	57								
		df	F	P	F	P	F	P	F	P
Case 3	A	2	27.79	.0001	37.49	.0001	40.36	.0001	25.50	.0001
Extreme	B	2	1.72	.1914	2.57	.0885	1.77	.1828	1.91	.1601
Nonorthogonality	AB	4	7.64	.0001	7.64	.0001	9.85	.0001	7.76	.0001
(X^2=11.99, N=51)	WG	42								

rejection of all three null hypotheses: $\Sigma\alpha_i^2 = 0$, and $\Sigma\beta_j^2 = 0$, and $\Sigma(\alpha_i\beta_j)^2 = 0$. Method 2, the method of expected frequencies and further data reduction to achieve equal n (Case 2) would all lead to the same error, the failure to reject $\Sigma\beta_j^2 = 0$.

When nonorthogonality was extreme and interaction was present (Case 3) all four solutions and the alternative of further data elimination (Case 4) led to essentially the same results. Thus, contrary to what the recommendations of Winer and Myers might lead one to anticipate, Method 1 did not become more preferable as deviations from orthogonality increased. To the contrary, when nonorthogonality was extreme the F values resulting from application of Method 1 deviate more from the standard than do the F values resulting from either of the approximate solutions. Method 2 was not an appropriate analysis because a significant interaction was present in Basic Data I.

The analysis of Basic Data II was used as a standard for comparing the various methods of dealing with nonorthogonality when no interaction is present (Table 4). When nonorthogonality was moderate all four solutions led to the same results as those found in the standard, namely rejection of $\Sigma\alpha_i^2 = 0$, and $\Sigma\beta_j^2 = 0$, but failure to reject $\Sigma(\alpha_i\beta_j)^2 = 0$. There was a type II error in Case 2, the data elimination alternative. It consisted of the failure to reject $\Sigma\beta_j^2 = 0$.

When no interaction was present and nonorthogonality was extreme (Case 3) all four solutions and the data elimination alternative led

Table 4

Data analyses of Basic Data II (no interaction present) with varying degrees of nonorthogonality.

Degree of Orthogonality		Basic Data II Results used as a standard for comparison			Case 2			Case 4		
		df	F	P	df	F	P	df	F	P
	A	2	79.01	.0001	2	49.38	.0001	2	19.75	.0001
Orthogonal	B	2	4.58	.0131	2	2.87	.0671	2	1.15	.3388
Data	AB	4	.68	.6081	2	.43	.7862	4	.17	.9509
(N=81, 54, 27)	WG	72			45			18		

Four Solutions to Nonorthogonal Data Analysis

		Method 1			Method 2		Expected Frequencies Method		Unweighted Means Analysis	
		df	F	P	F	P	F	P	F	P
Case 1	A	2	56.75	.0001	59.18	.0001	59.60	.0001	58.47	.0001
Moderate	B	2	3.50	.0368	3.63	.0328	3.59	.0339	3.40	.0403
Nonorthogonality	AB	4	.50	.7358	.50	.7358	.56	.6930	.50	.7358
(X^2=2.2, N=66)	WG	57								
		df	F	P	F	P	F	P	F	P
Case 3	A	2	30.26	.0001	38.06	.0001	40.64	.0001	32.37	.0001
Extreme	B	2	2.03	.1440	2.38	.1049	2.59	.0873	1.88	.1652
Nonorthogonality	AB	4	.28	.8893	.28	.8893	.34	.8498	.28	.8893
(X^2=11.99, N=51	WG	42								

to the same conclusions and resulted in one type II error. Again it is found that the conclusions reached by least squares and those reached by approximate methods do not diverge with increased nonorthogonality.

Summary

In comparing the four solutions, (a) to the analyses of the initial factorials, (b) to each other, and (c) to the data elimination alternative, it was obvious that, regardless of the degree of nonorthogonality and regardless of the presence or absence of an interaction, data elimination to achieve equal n was the worst alternative due to the strong tendency to yield type II errors.

When nonorthogonality was extreme, contrary to the hypothesis put forth in this paper, the four solutions tended to converge in yielding the same results. A slight divergence of results was found when a moderate degree of nonorthogonality was present, but not along the dimension of regression solutions versus nonregression solutions. Rather Method 1 and the unweighted means analysis appear to be best when results differ.

Method 1 is probably the most defensible based on criteria external to this study (Carson and Timm); yet the present results indicate that the more traditional solutions yield F values very close to the F values obtained by Method 1. In general, it appears that the traditional solutions may serve as satisfactory substitutes for Method 1 when interaction is present or for Method 1 or Method 2 when no interact

is present. However, Monte Carlo work is needed to establish the generality of the present results.

These conclusions are important to researchers who are accustomed to the unweighted means analysis or the method of expected frequencies but feel the more computationally complex regression solutions should be adopted. In some cases this fear may be unfounded. The present results are also important to those who use the regression approaches to nonorthogonal data analyses. It offers some assurance that results obtained via regression analyses are comparable to analyses from previous studies conducted by researchers who used the method of expected frequencies or the unweighted means analysis when they encountered disproportional data.

References

1. Carlson, J. E., & Timm, N. H. Analysis of nonorthogonal fixed-effects designs. Psychological Bulletin, 1974, 81 (9), 563-570.
2. Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.
3. Dayton, D. M. The design of educational experiments. New York: McGraw-Hill, 1970.
4. Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.
5. Keppel, G. Design and analysis: a researcher's handbook. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.
6. Kirk, R. E. Experimental design: procedures for the behavioral sciences. Belmont, California: Brooks/Cole, 1968.
7. Myers, J. L. Fundamentals of experimental design. Boston, Massachusetts: Allyn and Bacon, 1972.
8. Overall, J. E., & Spiegel, D. K. Concerning least squares analysis of experimental data. Psychological Bulletin, 1969, 72, 311-322.
9. Overall, J. E., Spiegel, D. K., & Cohen, J. Equivalence of orthogona and nonorthogonal analysis of variance. Psychological Bulletin, 1975, 82, 182-186.
10. Schussler, K. Analyzing social data: a statistical orientation. Boston: Houghton Mifflin, 1971.

11. Timm, N. H., & Carlson, J. E. Analysis of variance through full rank models. Monograph issued by Multivariate Behavioral Research, 1975.

12. Williams, J. D. Two way fixed effects analysis of variance with disproportionate cell frequencies. Multivariate Behavioral Research, 1972, 67-83.

13. Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

14. Woodward, J. A., & Overall, J. E. Multivariate analysis of variance by multiple regression methods. Psychological Bulletin, 1975, 82, 21-32.

tiple Linear Regression Viewpoints
7 No. 2 1977

"PATH ANALYSIS AND CAUSAL MODELS AS REGRESSION TECHNIQUES": A COMMENT

LEE M. WOLFLE
Virginia Polytechnic Institute and State University

Williams and Klimpel's (1975) paper, "Path Analysis and Causal Models as Regression Techniques," replicates a misunderstanding of the measurement of indirect effects in path analysis. Duncan (1966) initiated the misunderstanding by ambiguously specifying the total indirect effect (TIE) of a predictor on a dependent variable to be $r_{12} - p_{21}$, where r_{12} is the correlation of the predictor variable, X_1, and the dependent variable, X_2, and p_{21} is the path coefficient. Land (1969) and Kerlinger and Pedhazur (1973: 316-17) replicated the ambiguity, but the matter was resolved by Finney (1972).

A more informed understanding of indirect effects will help one to employ the heuristic value of path analysis to best advantage. By way of example, the model used by Williams and Klimpel might be illustrative. It is a simple recursive model relating the variables:

X_1: ownership in acreage in the past;

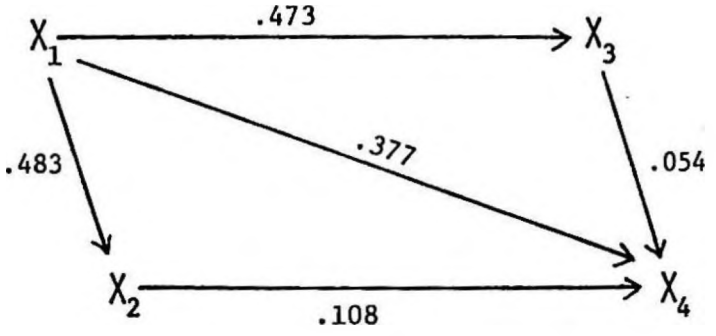
X_2: number of mentionings for leadership in the past;

X_3: ownership in acreage (present);

X_4: number of mentionings for leadership in the present.

The model allows for continuity over time by letting X_1 be a direct cause of X_3, and X_2 to be a direct cause of X_4. The

assumption for North Dakotan respondents was that leadership mentionings were caused by land ownership; the model therefore specifies effects of X_1 on X_2 and X_3 on X_4, and also a lagged effect of previous land ownership on current leadership mentionings, X_1 on X_4. The model, with path coefficients (or beta weights), may be diagramed as:



Williams and Klimpel then compute total indirect effects by taking the difference between the correlation coefficient and the path coefficient. Let's look at an example in their model where it makes logical sense to consider the indirect causal effect of one variable on another. The total correlation of variable X_1 and X_4 may be expressed from the fundamental theorem of path analysis (Duncan, 1966: 5) as

$$r_{14} = p_{41} + p_{43}p_{31} + p_{42}p_{21}$$

$$r_{14} = .454 = .377 + .025 + .052$$

The direct causal effect of X_1 on X_4 is measured by the path coefficient, p_{41} (= .377). In addition, X_1 has indirect causal effects separately through X_2 ($p_{42}p_{21}$ = .052) and X_3 ($p_{43}p_{31}$ = .025). By far the most important cause of present leadership mentionings was the lagged effect of past land ownership.

Williams and Klimpel also considered the total indirect effect of, among others, X_3 on X_4. But what variables intervene between X_3 and X_4? The answer is, none. How then can one conceptually think of an indirect causal effect of X_3 on X_4? It makes more sense to consider the total correlation of X_3 and X_4 to be the result of a direct causal effect of X_3 on X_4, and an effect due to common antecedent causes of X_3 and X_4. In the usual way of thinking about such effects, we may call these effects spurious. Thus, the total correlation of X_3 and X_4 may be expressed in terms of a direct effect, and a spurious effect resulting from both these variables being caused by previous land ownership, X_1. Thus,

$$r_{34} = p_{43} + p_{41}p_{31} + p_{42}p_{21}p_{31}$$

$$r_{34} = .262 = .054 + .178 + .025$$

(These do not sum exactly to $r_{34} = .262$ since Williams and Klimpel assume the path p_{32} to be zero, when in fact it is nonzero and positive, although small. Note also that this equation is somewhat different than equation (35) on page 18 of their article; applying their assumption that $p_{32} = 0$ to their equation resolves the discrepancy.) Thus, one sees that most of the covariation of X_3 and X_4 is due in a causal sense to the fact that these variables are both caused by X_1.

What Williams and Klimpel called "total indirect effect" was really two (actually, three) different kinds of effects. First, there was an indirect causal effect occurring through intervening variables. Second, there was a spurious association due to joint dependence on prior variables. There is a third source of

covariation which was not exemplified in the Williams and Klimpel model, namely, an association due to the correlation between predetermined variables. Since only one variable, X_1, in this model was predetermined, there could be no such association. If, however, we consider the relationship between X_1 and X_2 to be unspecified in a causal sense, we would merely consider the correlation between the two predetermined variables. The correlation of X_1 and X_2 is equivalent to the path coefficient in this case. The correlation of X_3 and X_4 would thus be divisible into four components: a direct causal effect ($p_{43} = .054$), an indirect causal effect (which is undefined in the absense of any intervening variables), a spurious effect due to X_3 and X_4 being jointly dependent on X_1 ($p_{41}p_{31} = .178$), and a joint association due to the correlation of X_1 and X_2 which separately effect X_3 and X_4 ($p_{42}r_{12}p_{31} = .025$).

Given Williams and Klimpel's model, a more explicit presentation of their Table 1 would look something like the following:

Variables	Total Association	Direct Effects	Indirect Effects	Spurious Effects	Joint Associations
X_1 on X_2	.483	.483	--	--	--
X_1 on X_3	.473	.473	--	--	--
X_1 on X_4	.454	.377	.077	--	--
X_2 on X_4	.305	.108	--	.195	--
X_3 on X_4	.262	.054	--	.203	--

If more convenient, these may be expressed as percentages of total association. The reader may recognize that the ratio of, say,

$$\frac{p_{41}}{r_{14}} = \frac{.377}{.454} = 83.0\%$$

is equivalent to the ratio of

$$\frac{b_{41.23}}{b_{41}},$$

where $b_{41.23}$ is the partial regression coefficient from the equation:

$$X_4 = a + b_{41.23}X_1 + b_{42.13}X_2 + b_{43.12}X_3 + e;$$

and b_{41} is the simple slope from:

$$X_4 = a' + b_{41}X_1 + e'.$$

I find the presentation of causal models more convenient in path analytic format. However, the same analytic result may be achieved without resorting to the vocabulary of path analysis, although the same issues must be confronted regardless of language.

I would also like to raise a matter of style. Here I exposit a personal preference. Williams and Klimpel present on page 11 of their article a model they call "more parsimonious," from which they have eliminated all paths except those eminating from X_1.

It seems to me that the model should remain in its final form as reproduced above. If one assumes that land ownership causes leadership mentionings in the past, one should also allow the relationship in the present. If the latter relationship is small, then one should show it. It is a substantive finding worth presenting to readers, regardless of its numeric value. For similar reasons the path from X_2 to X_4 should remain, for it gives the over-time continuity of leadership mentionings. My point here is clearly one of style, and goes beyond any statistical decision rule for retaining paths with small, or even statistically insignificant, values. I find my preference based on the substantive argument no less arbitrary than Williams and Klimpel's statistical decision

rule. They eliminate paths p_{43} and p_{42} because they do not pass a test for statistical significance. According to that test, however, retaining paths depends to a large extent on the sample size. With no change in the numeric values of p_{43} and p_{42}, Williams and Klimpel would have retained the paths given more respondents. An alternative, no less arbitrary, is to simply eliminate every path smaller in absolute value than some subjectively chosen value (see Land, 1969: 35), but I would prefer to leave such paths in the model when *a priori* theory predicted nonzero effects.

It has not been the purpose of this comment to advocate or defend the use of path analysis. The techniques of path analysis have been useful for a few sociologists studying status achievement, but elsewhere seem to have had limited usefulness. Either the models are so contrived that they appear utterly implausible, or the models are so complex that the heuristic value usually associated with path analysis becomes rather slight.

The advantage of path analysis is to force one to explicitly formulate the structural relationship of one's variables. There is no statistical difference between estimating path coefficients and regression coefficients in recursive models. The difference lies in the fact that path analysis is a method for thinking about cause. On the one hand, although causal or temporal ordering provides the basis for the interpretation of regression statistics, there are no such assumptions required for the computation of the same statistics. As a result, variables are frequently entered into regression equations that at best can be considered implausible

causes of the dependent variable. On the other hand, path analysis forces one to be explicit about the causal connection of variables, and if no causal relationship can be formulated from theory or one's other *a priori* notions of the real world, the variable should be excluded from the regression.

A good example of the explicitness advocated here may be found in Blau and Duncan (1967: 166-68). In constructing a model of the process of occupational placement, they had to order within their model the two variables, U: Respondent's educational attainment, and W: Status of respondent's first job. Although education is usually thought of as being temporally antecedent to occupational status, for an appreciable minority of people a first job is entered before the completion of one's education. Blau and Duncan postulated that for most respondents the sequence of events follows the usual U-W ordering, and obtained quantitative estimates for the U-W sequence model. The point illustrated is that these authors were explicit in their ordering of variables. One may disagree with their postulate, but what they did was explicitly stated. The presentation of assumptions may be wrong, but should not be misunderstood. Path analysis thus becomes a tool for confronting the structural relationship of variables.

Although path analysis is a method for considering cause, neither it, nor any other method, can be used for inferring causality from non-experimental data. Path analysis and other methods merely allow quantitative estimates for models from which

a number of possible causal relations have been excluded on *a priori* bases.

Despite its limitations, path analysis does present some heuristic advantages. I hope that practitioners of path analytic techniques in educational research will adopt the practice of considering the difference between the correlation coefficient and the path coefficient as being composed of three parts. Indirect effects may then consistently be considered indirect *causal* effects.

REFERENCES

Blau, Peter M. and Duncan, Otis Dudley. *The American Occupational Structure*. New York: John Wiley & Sons, 1967.

Duncan, Otis Dudley. Path analysis: sociological examples. *The American Journal of Sociology*, 1966, *72*, 1-16.

Finney, John M. Indirect effects in path analysis. *Sociological Methods and Research*, 1972, *1*, 175-86.

Kerlinger, Fred N. and Pedhazur, Elazar J. *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart and Winston, 1973.

Land, Kenneth C. Principles of path analysis. In Edgar F. Borgatta (Ed.) *Sociological Methodology 1969*. San Francisco: Jossey-Bass, 1969.

Williams, John D. and Klimpel, Ronald J. Path analysis and causal models as regression techniques. *Multiple Linear Regression Viewpoints*, 1975, *5*, 1-20.

ultiple Linear Regression Viewpoints
ol. 7 No. 2 1977

PATH ANALYSIS: A COMMENT ON WOLFLE'S COMMENT

John D. Williams and Ronald M. Klimpel
The University of North Dakota and North Dakota State Social Service Board

Wolfle's (1976) comment on our paper (Williams and Klimpel, 1975) is, on the whole, a useful contribution to those who use regression techniques but who may also have little information regarding path analysis. Surely, Wolfle's breaking down the contribution of path coefficients into four components (direct effects, indirect effects, spurious effects, and joint associations) makes more sense than our lumping all effects except the direct effects into being called indirect effects. As Wolfle points out, we joined a rather illustrious list of transgressors (Duncan, 1966; Land, 1969; and Kerlinger and Pedhazur, 1973).

So that we might give at least a feeble excuse for our use of the term "indirect effect", we would point out the raison d'etre of our original article was to put into a more readable form (at least to users of regression) the concepts of path analysis. At the time we wrote the article, Duncan's approach was probably more prevalent than the correction offered by Finney (1972). However, it does appear that the suggestions offered by Wolfle are in fact becoming part of the path analysis literature, if the paper by Spaeth (1975) can be taken as being representative. In any event, it does make good sense to consider that the portion of the effect of a variable later in a causal sequence which is due to earlier variables to be spurious rather than indirect.

In regard to Wolfle's preference for keeping in the model the two non-significant paths (p_{42} and p_{43}), we don't concur with his reasoning. To be sure, it is necessary to carry sufficient information in an article so that those who have a different orientation can analyze the data to their satisfaction. Wolfle's reanalysis of our data illustrates this point; sufficient information was given so that he could reanalyze the data. The issue seems to us to be in the interpretation. We opted for the "parsimonious" interpretation that variables X_2, X_3, X_4, were caused by X_1, and that other paths were unnecessary (and non-significant) to explain the relationships among the variables. Not only were the dropped paths non-significant, they also were accompanied with small path coefficients; only p_{42} exceeded .10 (p_{42} = .108).

The issue of sample size as brought out by Wolfle is valid; while researchers will differ, our point of view is this: the sample size should be such that meaningful (interpretable) relationships should be detectable; on the other hand sample sizes that are so large that even the smallest departure from zero is significant is an obvious case of wasteful collection of data. A useful approach to choosing sample size is to a priori select a value such that if the value is exceeded, statistical significance will also be reached; if the sample yields a number less than the posited value, non-significance will occur. (Use of this rule is not always accomplishable.) As many researchers opt for the largest possible sample under whatever monetary constraints exist for them, Wolfle's point is not entirely countermanded by our viewpoint.

Finally, the use of path analysis as a method to require researchers to think about the causal links in their data is commendable; all too often researchers put a bunch of variables into a stepwise regression solution to see what kind of sausage comes out. If they interpret such data without replication, the possiblity of spurious findings is enormous.

REFERENCES

Duncan, O.D. Path Analysis: Sociological examples. The American Journal of Sociology, 1966, 72, 1-16.

Finney, J.M. Indirect effects in path analysis. Sociological Methods and Research, 1972, 1, 175-186.

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

Land, K.C. Principles of path analysis. In Borgatta, E. F. (ed.) Sociolog ical Methodology 1969. San Francisco: Jossey-Bass, 1969.

Spaeth, J.L. Path analysis. In Amick, D.J. and Walberg, H.J. (eds.) Introductory Multivariate Analysis. Berkely, California: McCutchan Publishing Corp., 1975.

Williams, J.D. and Klimpel, R.J. Path analysis and causal models as regress ion techniques. Multiple Linear Regression Viewpoints. 1975, 5, No.3, 1-20.

Wolfle, L.M. Path analysis and causal models as regression techniques: a comment. Multiple Linear Regression Viewpoints. 1976, this issue.

near Regression Viewpoints
2 1977

MULTIPLE REGRESSION TECHNIQUES APPLIED TO TEST THE EFFECTS OF THREE TYPES OF SPECIAL CLASS PLACEMENT ON THE ARITHMETIC ACHIEVEMENT OF EDUCABLE MENTALLY RETARDED PUPILS

Joseph D. George*
Columbus College
Columbus, Georgia

ABSTRACT

Multiple regression analysis was used to examine the different effects of special class placement on the arithmetic achievement of Educable Mentally Retarded (EMR) pupils. Self-contained classes, selected academic placement programs and learning resource centers were the types of placement studied. A significant interaction between sex and type of placement was observed with respect to arithmetic achievement. Girls in self-contained classes gained more than boys in the same classes. Boys gained more in selected academic placement programs than in the other two types of placement; girls did best in selected academic placement programs.

LITERATURE REVIEW

The educational value of self-contained classes has been discussed primarily from two viewpoints. One stresses the social growth resulting from such placement. The other recognizes the importance of social growth but accentuates the need for academic growth.

In a study designed to determine the social position of mentally handicapped children in regular grades, G. Orville Johnson (October, 1962) found the chronological age of the mentally handicapped pupils to be greater than that of the non-handicapped pupils at every grade level.

---

*Date for this study was acquired as a result of this researcher's participation in the ESEA Title III Project, number 45-72-207-2. Appreciation is extended to Dr. Thomas L. Noffsinger, director of the project, for granting permission to use the data.

Mean differences in academic achievement between the handicapped and non-handicapped were found to increase as the grade level of the pupils increased.

After studying a school system noted for its progressive methods, G. Orville Johnson and Samuel A. Kirk (October, 1950-May, 1951) observed that, ". . .teachers stressed social adjustment and that academic achievement of both typical and mentally handicapped children was not overly stressed. This point was emphasized when the school board had a number of complaints from parents that the teachers were not stressing the three R's and that they feared their children were becoming deficient in the basic academic skills." Concluded was that mentally retarded children attending regular grades could be segregated but not necessarily separated from other children in the same grades. Such segregation resulted from a lack of social acceptance experienced by mentally retarded pupils.

Willie Kate Baldwin (November, 1958), in her investigation of the social position of educable mentally retarded children in regular public school grades, concluded that as mentally retarded children grew older and progressed through the elementary grades they became less socially accepted by non-retarded peers.

Research results imply that educable mentally retarded pupils are segregated from non-retarded ones whether they attend regular classes or self-contained special classes. Those attending regular classes are segregated because they are not accepted by their peers. Educable mentally retarded pupils in self-contained classes are segregated because they are separated from their peers. Those in regular classes experience frustration due to their being rejected and exhibit deviant behavior. Mentally handicapped pupils in self-contained classes experience less rejection and, as indicated by Farson (July, 1945), more

opportunities for success. Hence, they are less frustrated and exhibit less deviant behavior. However socially beneficial to the child, some Special Educators question the academic value of self-contained class placement.

Mentally retarded children, by definition achieve less than non-retarded children. With respect to the educational value of special class placement, the question arises: Who achieves more, mentally retarded children in regular classes or mentally retarded children in special education classes?

R. W. Edmiston and J. G. Benfer (1949) report that when slow learning students were removed from regular classes reading achievement scores of children whose I.Q.'s ranged 40 points apart were higher than those of children whose I.Q.'s ranged 60 points apart. Implied here is the notion that the presence of mentally retarded children in regular classes deters the overall progress of non-retarded children in the same class. Two questions appear obvious. First, how is achievement deterred—in terms of the group's <u>average</u> score or in terms of each child's individual score? What is the specific effect of the wide range placement on the achievement of the mentally retarded child? The answer to the first question is self evident. Because the retarded child has a limited growth potential, and because the most commonly used statistical procedures in education are dependent upon a comparison of group means, a group's average achievement score will be affected by the presence of scores of mentally retarded children, but the scores of individual children will not necessarily be affected by the presence of such children. The second question is not so easily disposed of.

In a study comparing the general academic achievement of mentally retarded children participating in regular classes in New York City's

Public Schools and mentally retarded children attending self-contained special classes in the same schools, Philip A. Cowen (July, 1938) found ". . .that the general academic achievement of the non-segregated group is somewhat superior to that of the equivalent segregated groups; that the unsegregated boys make greater gains in mechanical ability and in personality traits than do the special class boys, but that the special class girls make greater gains in mechanical ability and personality tha do the unsegregated groups."

Cowen's findings reveal a sex difference. Mentally retarded girls i regular classes did not show as much gain in mechanical ability and in personality traits as boys in the same classes. Moreover, girls in special classes did better in developing mechanical ability and personal traits than girls in regular classes. Interestingly, mentally handicapp boys and mentally handicapped girls in regular classes profited more in the academic areas of reading and arithmetic than those attending specia classes. With respect to the issue of special class placement the implication is that--academically--mentally retarded children will profi more from regular classes; but that--socially--girls will benefit more from special class placement and boys will do better in regular classes. Hence, the Special Educator's dilemma: How can the best of two worlds--regular class placement and self-contained special class placement--be combined to form an improved educational program for educable mentally retarded children?

Robert Fruininks and John Rynders (September, 1971) reported that among educable mentally retarded children, those with lower I.Q.'s profited more academically from special class placement, while those having higher I.Q.'s achieved greater academic gains in regular classroom settin

Implicit here is the notion that special class settings will provide different educational benefits for different students.

It appears that mentally retarded children benefit socially from special class placement but that such placement slows their academic growth. Therefore, the real issue involved in the question of the efficacy of special class placement seems to be: Which aspect of a mentally retarde pupil's educational growth--the social or the acedemic--will most benefit him in later life? Obviously, the answer to that question will vary with the individual child. Therefore, Special Educators must be concerned with developing programs which will provide the mentally handicapped youngster with the best possible combination of social and academic growth--for him.

The emphasis of the present study is to examine the effects of three different types of special class placement upon the arithmetic achievement of educable mentally retarded pupils.

PROCEDURES*

For purposes of this study, three types of special class placements were identified. They were (1) Self-Contained Classes; (2) Selected Academic Placement Programs; and (3) EMR-Learning Resources Centers. Self-contained classes were considered to be ones in which educable mentally retarded pupils:

1. Received all of their instruction in the basic subjects like language arts, math, and social studies from an appropriately certificated teacher.

---

*Procedures used in the conduct of this study are identical to the ones utilized by George (1975) to examine the impact of three types of special class placement on the reading achievement of EMR pupils. George's study appeared in the March 1975 issue of Viewpoints.

2. Received instruction in subjects like physical education, art, and music from specialists in each of those fields or from an EMR certificated teacher.

3. Remained with other EMR pupils throughout the entire school day, with the exception of recess, lunch periods or study halls.

Selected academic placement programs were identified as ones in which educable mentally retarded children:

1. Received all of their instruction in the basic subjects like language arts, math and social studies from a teacher who has been appropriately certificated to work with such pupils.

2. Remained with instruction in subjects like physical education, art, and music from specialists in each of those fields.

3. Remained with other EMR pupils when receiving instruction in the basic subjects like language arts, math and social studies, but are permitted to join non-retarded pupils when receiving instruction in subjects like physical education, art and music. They also join non-retarded students during recess and lunch.

EMR-learning resource centers were defined as ones in which pupils:

1. Received instruction in the basic subjects like language arts, math, and social studies from teachers who have been appropriately certificated to work with pupils in those subject areas and who work closely with an EMR certificated teacher to develop a course of instruction for the retarded pupil.

2. Received instruction in subjects like physical education, art, and music from specialists in each of those fields.

3. Remained with non-retarded pupils except for short periods of time when they went to a room equipped with specialized materials and/or equipment to receive highly individualized instruction in a subject area in which they exhibit a weakness.

THE EXPERIMENTAL DESIGN

The design of this study was a nonequivalent control group design as noted in Campbell and Stanley's (1972) monograph in experimental design.

Groups of educable mentally retarded children examined by this study were placed into special education classrooms as prescribed by standards set forth by the Ohio Department of Education, Division of Special Education. Hence, assignments of subjects into groups could not be random.

However, Campbell and Stanley (1972) have indicated that the greater the number of characteristics in which the control and experimental groups are similar, the more the prior assigned groups can be considered to be equivalent. Because a pre-test is required by the nonequivalent control group design, and because statistical techniques could be used to examine equivalency as related to pre-test scores, the groups were assumed equivalent.

THE SAMPLE

Because EMR pupils attending schools participating in the ESEA Title III Project Number 45-72-207-2 during the 1972-1973 school year were the subjects of this study, the population sampled was limited to groups of educable mentally retarded children attending 385 public schools in 8 project areas in the state of Ohio (Noffsinger 1972-1973). Further, the population was limited to educable mentally retarded pupils attending special classes which met the state of Ohio's requirements for one or another of the three types of special class placements being studied. Ninety of the 385 schools met those requirements. All of the groups of educable mentally retarded pupils attending those 90 schools were included in this study.

SOURCES OF DATA

The Ohio Special Achievement Inventory, a test that measures the achievement of educable mentally retarded pupils, was used to gather data pertaining to the reading achievement of the subjects. The pre-tests were administered during the months of September and October 1972, by test administrators who were trained by Title III personnel. Post-tests were given during April and May of 1973.

The arithmetic test was administered individually to children in the primary and intermidiate grades. Pupils in the junior high grades were administered the test individually. Senior high pupils received the test as a group.

The Metropolitan Achievement Test of arithmetic was administered to all of the children in the sample. This testing was done only once and occurred as closely after the pupil received the Ohio Special Achievement Inventory as possible.

Groups failing to meet the following criteria were eliminated from the sample population:

1. O.S.A.I. pre- and post-test score of 75% or more of the pupils in each group.
2. Two or more boys and two or more girls in each group.

Of the 90 groups in the sample, 30 met the requirements stated above. Ten of those were Self-Contained Classes; eight were Selected Academic Placement Programs; and twelve were EMR-Learning Resource Centers. Utilizing a table of random numbers, eight Self-Contained Classes and eight EMR-Learning Resource Centers were selected. All eight Selected Academic Placement Programs were used.

There were 43 boys and 37 girls attending Self-Contained Classes; 50 boys and 36 girls attending Selected Academic Placement Programs; and 43 boys and 38 girls participating in Learning Resource Centers. The total number of boys included in this study was 136. The total number of girls was 111. Because of absenses and for other reasons, complete sets of data were not obtained for all of the subjects.

STATISTICAL TREATMENT OF DATA

To establish a degree of equivalency between the prior assigned groups,

a one way analysis of variance using Multiple Regression Techniques was computed from the arithmetic pre-test scores. The hypotheses tested, the regression models used to test them, and the findings are reported in Table 1. For a description of the variables see Table 5.

Reliability and validity data for the Ohio Special Achievement Inventory were unavailable. The Kuder-Richardson procedure, Formula 21, as reported by Thorndike and Hagen (1969), was used to compute reliability. Scores for each level of the test were obtained and are reported in Table 2.

To compute concurrent validity, the Pearson Product Moment Correlation procedure was used (Newman, Frye and Newman, 1973). The arithmetic sections of the Ohio Special Achievement Inventory were correlated with the corresponding sections of the Metropolitan Achievement Tests. Results of those computations are reported in Table 3.

MODELS AND EXPLANATIONS	R^2	df	ALPHA	F	P
Null Hypothesis 1: There will be a significant difference between arithmetic achievement pre-test scores, as measured by the O.S.A.I., of EMR pupils attending Self-Contained Classes, Selected Academic Placement Programs, and Learning Resource Centers.		2/45	.05	2.6672*	.1094
Full: $Y_9 = a_0U + a_3X_3 + a_4X_4 + a_5X_5 + E$	.0016				
Restricted: $Y_9 = a_0U + E$	.0000				
Null Hypothesis 2: There will be a significant difference between the arithmetic pre-test scores, as measured by the O.S.A.I., of EMR boys attending Self-Contained Classes, Selected Academic Placement Programs, and Learning Resource Centers and the arithmetic pre-test scores of girls attending the same classes.		2/46	.05	0.038*	.9648
Full: $Y_9 = a_0U + a_6X_6 + a_7X_7 + E$	.0559				
Restricted: $Y = a_0U + E$	.0000				

*Not significant at the .05 alpha level.

**The format in which the above data appears was suggested by Isadore Newman in a recent issue of Viewpoints.

NOTE: See Table 5 for a description of the variables used.

Table 2
O.S.A.I. ARITHMETIC RELIABILITY SCORES

Test Level	r_{11}	SD	n	M_t
Primary	.69	2.56	14	10.97
Intermediate	.84	4.63	18	9.09
Junior High	.83	4.05	18	12.95
Senior High	.76	3.98	18	8.50

gathered for each level of the reading test and are reported in Table 4.

Table 3
PRODUCT-MOMENT CORRELATION COEFFICIENTS
COMPUTED BETWEEN THE O.S.A.I. ARITHMETIC TEST
AND THE MAT ARITHMETIC TEST

Test Level	r	n
Primary	.67*	24
Intermediate	.60*	44
Junior High	.48*	26
Senior High	.52*	31

*Significant beyond the .01 alpha level for a one-tailed test.

Finally, using Multiple Regression Techniques, analysis of covariance was used to test the research hypotheses. Data decks used were sorted. The research hypotheses, regression models used to test them, and findings are reported in Table 4.

Table 4
MODELS, F-RATIOS, AND R^2 FOR PREDICTING ARITHMETIC SCORES OF SUBJECTS

MODELS AND EXPLANATIONS	R^2	df	ALPHA	F	P
Null Hypothesis 3: There will be a significant difference in arithmetic achievement post-test scores as measured by the O.S.A.I. between EMR pupils attending Self-Contained Classes, Selected Academic Placement Programs, and Learning Resource Centers when co-varying O.S.A.I. arithmetic pre-test scores.		2/44	.05	0.0915*	.9127
Full: $Y_2 = a_0U + a_3X_3 + a_4X_4 + a_5X_5 + a_9X_9 + E$	.6204				
Restricted: $Y_2 = a_0U + a_9X_9$	.6188				
Null Hypothesis 4: Male EMR pupils attending Self-Contained Classes, Selected Academic Placement Programs, and Learning Resource Centers will score significantly different on the O.S.A.I. arithmetic post-test than female EMR pupils attending the same classes when co-varying O.S.A.I. arithmetic pre-test scores.		1/45	.05	0.2550*	.6159
Full: $Y_2 = a_0U + a_6X_6 + a_7X_7 + a_9X_9 + E$	.6209				
Restricted: $Y_2 = a_0U + a_9X_9 + E$	.6187				

*Not significant at the .05 alpha level

Table 4 (Continued)

MODELS AND EXPLANATIONS	R^2	df	ALPHA	F	P
Null Hypothesis 5: There will be a significant difference between O.S.A.I. arithmetic post-test scores or EMR boys attending Self-Contained Classes and O.S.A.I. arithmetic post-test scores of EMR girls attending the same classes when co-varying O.S.A.I. arithmetic pre-test scores.		1/13	.05	0.3830*	.5467
Full: $Y_2 = a_0U + a_6X_6 + a_7X_7 + a_9X_9 + E$	.6128				
Restricted: $Y_2 = a_0U + a_9X_9 + E$	.6014				
Null Hypothesis 6: There will be a significant difference between O.S.A.I. arithmetic post-test scores of EMR boys attending Selected Academic Placement Programs and O.S.A.I. arithmetic post-test scores of EMR girls attending the same classes when co-varying O.S.A.I. arithmetic pre-test scores.		1/13	.05	0.1197*	.7349
Full: $Y_2 = a_0U + a_6X_6 + a_7X_7 + a_9X_9 + E$	.5634				
Restricted: $Y_2 = a_0U + a_9X_9 + E$	.5593				

*Not significant at the .05 alpha level.
NOTE: See Table 5 for a description of the variables used.

Table 4 (Continued)

MODELS AND EXPLANATIONS	R^2	df	ALPHA	F	P
Null Hypothesis 7: There will be a significant difference between O.S.A.I. arithmetic post-test scores of EMR boys attending Learning Resource Centers and O.S.A.I. arithmetic post-test scores of EMR girls attending the same classes when co-varying O.S.A.I. pre-test scores.		1/13	.05	0.1426*	.7118
Full: $Y_2 = a_0U + a_6X_6 + a_7X_7 + a_9X_9 + E$	.7155				
Restricted: $Y_2 = a_0U + a_9X_9 + E$	.7124				
Null Hypothesis 8: The interaction between sex and type of class will not account for a significantly greater amount or EMR student variance on the O.S.A.I. arithmetic post-test scores than will sex and type of class separately when co-varying O.S.A.I. arithmetic pre-test scores.		2/39	.05	8.5613**	.0001
Full: $Y_2 = a_0U + a_9X_9 + a_{10}X_{10} + a_{11}X_{11} + a_{12}X_{12} + a_{13}X_{13} + a_{14}X_{14} + a_{15}X_{15} + E$	.6284				
Restricted: $Y_2 = a_0U + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 + a_7X_7 + a_9X_9 + E$	.0575				

*Not significant at the .05 alpha level.

Table 5

A DESCRIPTION OF THE AVRIABLES USED

Where the Full Models are:

$$Y_2 = a_0U + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 + a_7X_7 + a_9X_9 + E$$

The Variables are:

Y_2 = A criterion, O.S.A.I. arithmetic post-test scores

a_0 = a_3 through a_9 = Partial regression weights

U = The unit vector (a "1" for each sample)

aX_3 = The Self-Contained Class, 1 if in the class, zero otherwise

aX_4 = The Selected Academic Placement Program, 1 if in the program, zero otherwise

aX_5 = The EMRLearning Resource Center, 1 if in the center, zero otherwise

aX_6 = Sex, 1 if male, zero otherwise

aX_7 = Sex, 1 if female, zero otherwise

aX_9 = O.S.A.I. arithmetic pre-test scores

E = Error vecor, difference between predicted score and actual score

aX_{10} = X(10) = X(3) * X(6) = males in Self-Contained Classrooms

aX_{11} = X(11) = X(4) * X(6) = males in Selected Academic Placement Programs

aX_{12} = X(12) = X(5) * X(6) = males in Learning Resource Centers

aX_{13} = X913) = X(3) * x(7) = females in Self-Contained Classrooms

aX_{14} = X(14) = X(4) * X(7) = females in Selected Academic Placement Programs

aX_{15} = X(15) = X(5) * X(7) = females in Learning Resource Centers

Hupotheses 1 and 2, as shown in Table 1, were not rejected providing evidence that the arithmetic pre-test scores of the subjects were not differe enough to be considered significant at the .05 level. This provides a degree of support for the internal validity of the Non-equivalent Control Group design used in this study (Campbell and Stanley, 1972). Further supporting the internal design of this study is the notion that 11 of the subjects were identified and placed into special class programs pursuant to standards set forth by Ohio's Department of Education, Division of Special Education.

It must be pointed out that, although the F-ration calculated for hypotheses 1 and 2 were not significant at the .05 level, they both approached the .1 level showing the possibility that there were differences between the groups prior to the pre-test. Additionally, it should be noted that the interpretation of standards for identifying and placing EMR pupils vary depending upon the interpreter and upon the situation of the school system by whom the interpreter is employed.

The O.S.A.I. arithmetic reliability scores, shown in Table 2, ranged from .69 at the primary level to .84 at the intermediate level. Thorndike and Hagen (1969) suggest that the Kuder-Richardson procedure (Formula 21) is a conservative estimate of reliability. Hence, it can be assumed that the ability, and--for purposes of the present research--the O.S.A.I. arithmetic test can be considered reliable.

The correlations reported in Table 3, give evidence that--for purposes of the present research only--concurrent validity for the O.S.A.I. test arithmetic exists with the M.A.T. test of arithmetic. With respect to arithmetic achievement null hypothesis 8 was the only null rejected. Knowledge of sex and type of special class placement was a predictor of success in arithmeti achievement post-test scores.

Table 6

A COMPARISON OF THE MEAN GAIN SCORES BETWEEN CHILDREN PARTICIPATING IN SELF-CONTAINED CLASSES, SELECTED ACADEMIC PLACEMENT PROGRAMS AND LEARNING RESOURCE CENTERS

	S.C.	SAPP	LRC
Males	-6.60	+4.90	+3.40
Females	+2.20	+7.43	-1.90

Examination of mean gain scores (Table 6) of subjects in each placement model shows that EMR girls in Self-Contained Classes and in Selected Programs achieved more in arithmetic than EMR boys in the same classes. Moreover, EMR girls in Selected Academic Placement Programs did better than EMR girls in Self-Contained Classes. However, EMR girls in Learning Resource Centers scored less in arithmetic achievement than EMR girls in either Self-Contained Classes or Selected Academic Placement Programs. EMR boys in Selected Academic Placement Programs scored higher in arithmetic achievement than EMR boys in Learning Resource Centers. Moreover, EMR boys in Selected Academic Placement Programs achieved higher scores than EMR boys in Self-Contained Classes, suggesting that Selected Academic Placement Programs are more beneficial for boys than for girls with respect to arithmetic achievement. For the population examined, girls in Selected Academic Placement Programs scored higher in arithmetic achievement than girls in the other two placements studied. Null hypotheses 3, 4, 5, 6, and 7 did not approach significance. This may have been due to the low in which may have caused a Type II error. Because no distrinction was made between pupils in primary, intermediate, junior high, and senior high grades it was assumed that they would score similarly in arithmetic achievement. This may not have been the case. Students

scoring high at one level could have cancelled out the effects of students scoring low at other levels. The possibility that differences exist should not be overlooked.

CONCLUSIONS

For the population studied, the data supports the conclusion that boys in Learning Resource Centers achieved more in arithmetic than girls in Learning Resource Centers. However, girls in Selected Academic Placement score better in arithmetic than boys and girls in any of the other placement models. The evidence indicates that for this sample special class placement had a different effect upon boys than upon girls.

Additional and more careful examination of the interaction between sex and type of special class placement is recommended.

REFERENCES

Baldwin, Willie Kate. "The Social Position of The Educable Mentally Retarded Child in the Regular Grades in the Public Schools," Journal of Exceptional Children, 25: (November, 1958), pp. 106-112.

Bruininks, Robert H. and John E. Rynders. "Alternatives to Special Class Placement for Educable Mentally Retarded Children," Focus on Exceptional Children, Vol.3: No. 4 (September, 1971), pp. 1-12.

Campbell, Donald J. and Julian C. Stanley. Experimental and Quasi-Experimental Designs for Research (Chicago: Rand McNally and Company,

Cowen, Phillip A. "Special Class Vs. Grade Groups for Subnormal Pupils," School and Society, 48:1227 (July, 1938), pp. 27-28.

Edmiston, R.W. and J.G. Benfer. "The Relationship of Group Achievement and Range of Abilities Within the Groups," Journal of Educational Research, 42:7 (1949), pp. 547-548.

Farson, Mable R. "A Program for Low Ability Children in the Regular Grade American Journal of Mental Deficiency, 50:1 (July, 1945), pp. 107-114.

Johnson, Orville G. "Special Education for the Mentally Handicapped--A Paradox," Exceptional Children, 29 (October, 1962), pp. 62-69.

Johnson, Orville G. "A Study of the Social Position of Mentally Handicappe Children in the Regular Grades," American Journal of Mental Deficienc 55: (July, 1950-April, 1951), pp. 60-89.

Johnson, Orville G. and Samuel A. Kirk. "Are Mentally Handicapped Children Segregated in the Regular Grades?" Journal of Exceptional Children, 17:2 (October, 1950-May, 1951), pp. 65-68 and 87-88.

Noffsinger, Thomas L., editor, Program Models for EMR Students: Second Year Evaluation, ESEA Title III Project Number 45-72-207-2, February 1, 1972-January 31, 1973.

Isadore, Isadore. "Editorial Comment," Multiple Linear Regression Viewpoints, January, 1973.

ltiple Linear Regression Viewpoints
.7 No. 2 1977

WHAT INFERENCES ARE ALLOWABLE WITH A SIGNIFICANT F IN REGRESSION ANALYSIS?*

Steven D. Spaner

University of Missouri - St. Louis

Multiple linear regression (MLR) procedures (Wainer, 1976) and the linear model in general (Brown, 1975) have come under attack in the most recent months. And, sad to say many of the points raised are cogent criticisms indeed. However, none of the remarks are foreign to the members of SIG-MLR or the readers of Multiple Linear Regression Viewpoints. But, perhaps these admonitions from our brethren would be less audible if the topics to be discussed here today were more universally heard. What appears to be needed is a better understanding of MLR: its foundations, its applications, and its ramifications. My presentation addresses some of these misunderstandings.

We calculate regression coefficients, we determine R^2s, we compute F statistics, but do we know what it's all about? If the audience is composed of Ward-Kelly-McNeil protégés then the word "assumptions" elicits a discounting laugh; but there are those who read a list of assumptions in a statistics text and live by them or die by them. I believe an examination of the foundations of MLR would do both the jester and the jouster some good.

Drawing from a variety of sources (Ward & Jennings, 1973; Lindquist, 1953; McNeil, Kelly, & McNeil, 1975; Snedecor, 1956; Young & Veldman, 1972; and Glass & Stanley, 1970) the following list of assumptions have been identified for the F statistic:

* Not reviewed

1. The F statistic must have been generated from randomly selected and independent entities or criterion measures.
2. The variance of the criterion measures within each population subgroup must be equal (homogeneity of variance).
3. The distribution of the criterion measures in each population must be normal.

Specific authors have added to or restated in other words the above list. The only meaningful change would occur in the case where a covariance analysis had been performed and the F statistic generated from these data. In such an instance the above assumptions apply to the adjusted criterion measures and a fourth assumption is added:

4. The regressions of the criterion measures onto the covariate(s) are equal for each population subgroup (homogeneity of regression).

Before I summarily dismiss these F ratio assumptions with an impressive list of citations, it should be pointed out that another list of assumptions are of concern. While the F distribution is the theoretical sampling distribution, the overall calculation technique is regression which is synonymous with correlation (in point of fact, Galton postulated his "law of universal regression" before Pearson developed the index of correlation/ship/). Hence we must also acknowledge the set of assumptions underlying regression. These assumptions are essentially the characteristics of the bivariate normal distribution:

1. X scores, disregarding Y scores, are normally distributed.
2. Y scores, disregarding X scores, are normally distributed.
3. The Y scores for each X score are normally distributed with a common variance ($\sigma^2_{y.x}$).
4. The X scores for each Y score are normally distributed with a common variance ($\sigma^2_{x.y}$).

5. The means of the Y score distributions for each X score fall on a straight line.

Since we are here to discuss multiple regression the above assumptions must be expanded to address the multivariate normal distribution; but I will leave that task to the reader, the features are essentially the same. However, the expansion brings up my first point with regard to allowable inferences.

Snedecor (1956) has classified multiple regression into two basic models: Model I - the values of X are considered fixed, that is, chosen by the investigator, only the Y values or the criterion is a random, normally distributed variable, and Model II - the values of X are not selected, individuals are randomly selected leaving the values of every variable measured on the individuals available to chance; that is, a random sample is drawn from a multivariate normal population. The first model is exemplified by the ex post facto research design and designs in which treatments are not randomly assigned. The second model is more in the tradition of the experimental design. Investigations utilizing model I often take the form of covariance analyses which call for the 4th assumption to the F statistic and call for a limitation on the generalizations; i.e., the population of adjusted Y scores. However, statistical control should not be viewed as limiting since it allows one to study the actual situation instead of one that has been artificially produced by experimental control.

Lindquist (1953) has observed that in educational research the application of model two is often amended to random assignment of treatments instead of random assignment of subjects to treatments. The reasons

are obvious and the amendment ingeniously adaptive; however, it must be understood by the researcher that his population has changed. The population is no longer individuals who are potentially available for random selection but instead intact groups who are available for random assignment of treatments.

Lest you begin thinking that assumptions are all this paper is going to address let me put your thoughts at rest. A number of investigations have dealt with the F distribution assumptions and their violation (Norton, 1952; Bonneau, 1960, 1963; Young & Veldman, 1963; Pearson, 1931; Box & Anderson, 1955). The summarized conclusion from these investigations is that there is no appreciable effect on the accuracy of the F test from nonnormality and if sample sizes are equal, heterogeneity of variance has a negligible effect. The only apparent serious violation that can be committed is failure to randomly select independent entities or measures (Glass & Stanley, 1970). However, surprisingly, there are no impirical investigations of this tenet. Furthermore, there is a dearth of inquiry into the effects of violation of the homogeneity of regression assumption and the list of postulates for regression and correlation. From the few statements offered on these topics (Snedecor, 1956; Vasu & Elmore, 1975) it appears that once again normality is a mute issue but 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). However, McNeil and Spaner (1971) have made a case for a more judicious examination of such a recommendation, especially as it would apply to nonlinear problems.

Having fulfilled my obligation to the conservatives; having acknowledged the assumptions we are operating under in MLR; I will now turn to the topic sentence: "What inferences are allowable with a significant F in regression analysis?" The first and foremost limitation on our inferences is the research tool; i.e., the type of regression technique, stepwise or hypothesis testing regression. My colleague (McNeil, 1976) has addressed himself to this matter so I will make only a brief remark. The calculation of an F statistic carries with it the implication that a comparison is desired, a decision is to be made, and that an inference will follow. All three of these activities suggest that a comparison hypothesis has been adopted (a null hypothesis) and an alternative hypothesis will be accepted should the comparison hypothesis be found untenable. Hypothesis testing regression fits this research format precisely: a null hypothesis is formulated - the restricted model as Bottenburg and Ward (1963) tagged it, and an alternative hypothesis is proposed - the full model in Texas terminology (this form of labeling has been principally associated with University of Texas faculty and graduates). A significant F statistic, in the "grand tradition", calls for rejection of the null hypothesis (restricted model) and acceptance of the alternative hypothesis (full model).

Let us look now at stepwise regression. Stepwise comes in two forms: ascending and descending. Ascending stepwise regression adds variables to a null set until a new set is created which has maximum predictive efficiency (according to some criterion). Descending stepwise regression operates in just the reverse; from a defined set of variables, variables are removed which least add to efficient prediction (until some "stop" criterion is obtained). Examination of these two techniques in

relation to an F test leaves the researcher with a "loose end". In the case of ascending stepwise, the null hypothesis is known, it's a model of the criterion grand mean. But, the alternative hypothesis is unknown; hence, no decision can be contemplated and no inferences entertained that are not sample generated. Likewise, in descending stepwise the alternative hypothesis is known, but we must wait for the "marvelous toy" to tell us what our comparison (null) hypothesis is. Therefore, if inference and generalizations are to be allowed upon significant F tests we must conduct hypothesis testing regression analyses.

A second limitation on our allowable inferences relates to the research hypothesis. Many of the more interesting and pertinent questions in education call for the statistical control of variables that are practically or explicitly beyond experimental control. Indeed, one of the attractive features of MLR is the ease with which covariance analysis can be conceptually as well as operationally understood (Williams, 1976). And while we have a number of admonitions against causal interpretations in ex post facto (basically correlational) studies (Campbell & Stanley, 1969; McNeil, Kelly, & McNeil, 1975; Newman, et. al., 1976); we would do well to remember ". . . that in many cases statistical control is more to be desired; the actual situation is studied instead of one artificially produced, the observations are extended over a greater range, thus broadening the foundation for inference, and in the end one has knowledge of the variation of two [or more] quantities instead of one, together with the relationship between them." (Snedecor, 1956, p. 146) However, covariance analysis it will be remembered imposes a 4th assumption on the F distribution - homogeneity of regression. Ironically, McNeil and company (1975, p. 131) have shown us how easy it is to make a test of this assumption with MLR

(i.e., interact the membership variables with the covariates and test the equality of the coefficients); yet, users and teachers (Williams, 1976) seem to be unfamiliar with this easy but necessary test. Hence, inferences based on significant F tests of covariance analyses will be rendered inaccurate if not invalid without the homogeneity of regression test.

Indeed the value of and need for interaction tests has been grossly underemphasized in MLR studies. I suspect that this phenomenon arises out of a misunderstanding, perhaps even fear, of a significant interaction finding. True, a significant interaction hampers the interpretation of main effects, but the positive view is that a significant F test of interaction tells us how to appropriately limit our generalizations (Glass & Stanley, 1970).

Another issue related to hypotheses and having a bearing on our inferences with a significant F statistic is that of directionality. Without going into the rudiments of sampling theory, I will summarize the directionality vs. nondirectionality decision as one which doubles your chances of rejecting the null hypothesis, if you hypothesize in the right direction. It's that last phrase that seems to be overlooked by many researchers. There seems to be a preponderance of cautious research hypotheses (nondirectional) and bold research conclusions (directional). Let me give some guidance as to this choice of hypotheses as it relates to MLR. It must be remembered that regression, especially least squares regression, is basically a curve fitting technique (Lewis, 1960; Snedecor, 1956). This being the case there are only three aspects of a curve that are manipulatable: 1) the point where the curve intersects with some reference axis (intercept point), 2) the rate of rise of the curve (the slope, which is an indicator of relationship), and 3) the number of inflection points

in the curve (this is governed by the exponentiation factor). Of these three factors only one is available for nondirectional hypotheses: the comparison of intercept points. The weightings or regression coefficients that identify level of intercept are indicative of group means. And direction of differences between group means can be unknown such that a nondirectional hypothesis is conceivable. Of course, each researcher is characterized by his or her own risk-taking-behavior, but there are strong arguments for stating all research hypotheses as directional hypotheses (McNeil, Kelly, McNeil, 1975).

The other two aspects of a curve call for directional hypotheses by the very nature of their source. The source of rate or slope changes is the addition of information into a model. The worse that can happen is no change; hopefully the added information will enhance the relationship of the predicted scores with the actual scores (i.e. increase r_{yy}). And, since our yardstick of measurement is error sum of squares we cannot generate more prediction error by including information which is related to the criterion. We can, however, possibly reduce the error by the inclusion of new information (Ward & Jennings, 1973; Snedecor, 1956; McNeil, Kelly, & McNeil, 1975).

The same mathematical truths hold for hypotheses about inflection points. That is, in hypothesizing models that are to fit nonlinear data, the only possible hypothesis is a directional hypothesis. Inclusion of exponentiated variables allows the best fit line to bend and turn with the data thus reducing the sum of squared deviations from the line. However if the exponentiated variables do not create a better fit, they do not create more error, they simply take up degrees of freedom without error reduction and receive zero or low weightings.

Hence, the point to be made is that only directional hypotheses are being tested with a significant F in two of the types of regression

hypotheses, whether the researcher has stated them as such or not. Only in the statement of hypotheses about intercepts (group mean differences) is there the potential for a nondirectional F test and there is great suspect of a researcher who would not have some expectation of the direction of mean differences.

Returning momentarily to a point made about nonlinear fits of data; there are recurring calls for the development of nonlinear hypotheses in educational research, the most recent coming from Brown (1975). Being myself one of the more ancient heralds of this idea (McNeil & Spaner, 1971) and knowing that I was not the first to send smoke, I find it a puzzlement as to why there are not more nonlinear hypotheses in educational research. I come up with two possible reasons: one, the "Pandora's box" fear, and two the fear of violation of assumptions. With regard to the first reason, Pandora's box, it is true that there is an infinite set of nonlinear terms. However, bivariate plots of sample data can narrow down the field of fruitful hypotheses quite well. Additionally, there are computer aids to "snoop around" in sample data and arrive at some tenable hypotheses (Automatic Interaction Detection-Version 4 by Koplyay, Gott, & Elton, 1973).

The second fear, violation of assumptions, must refer to the linearity of regression assumption since the first portion of this paper has indicated that normality and homoscedasticity are not critical. So let's examine the meaning of this assumption: it states that the means of all sampled populations lie on a straight line. But what is the effect of this assumption? It allows us to infer for populations not sampled. We assume the population mean for unsampled populations falls on the sampled regression line. And, this is a valuable principle, indeed, for without

it we have no prediction. However, does this regression line have to be straight? If it does, then there are a number of logarithmic transformations of curve line functions that will bring the curved best fit line back into line (Lewis, 1960). But all this manipulation is unnecessary; it is simply a rescaling process. It's a recognition that the measuring devices we use are not necessarily monotonic. It's a recognition, as Pohlman & Newman (1973) have suggested, that the assumption of rectilinearity has not been met. But more than that, it's a recognition that the assumption can be met if a curved best fitting line is used. McNeil & Kelly (1970) and Ward & Jennings (1973) have addressed themselves to this issue and have suggested that the investigator attempt to express functional relationships in data first and not worry about whether conditions or assumptions are met. In other words, if it works, use it.

As a concluding remark, it should be pointed out that "what inferences are allowable with a significant F in regression analysis "are zero if the R^2 of the regression model is not practically significant." Ward & Jennings (1973) have indicated that very large samples can produce statistically significant F statistics with very little practical consequences. Scientific rigor and the scientific process produce significant results, statistics simply apply a probability level to your potential error in judgment. Focus should be placed on 1) random and independent sampling, 2) clear and precise statements of research hypotheses, 3) construction of appropriate models to test the hypotheses, 4) examination of R^2 values, and 5) allowable inferences from significant findings.

References

Boneau, C.A. The effects of violations of assumptions underlying the t-test. Psychological Bulletin, 1960, 57, 49-64.

Boneau, C.A. A comparison of the power of the U and t-test. Psychological Review, 1963, 69, 246-256.

Bottenberg, R.A. & Ward, J.H. Applied multiple linear regression (AD 413128). Lackland Air Force Base, Texas: Aerospace Medical Division, 1963.

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

Campbell, D.T. & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago, Illinois: Rand McNally, 1969.

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

Koplyay, J.B., Gott, C.D. & Elton, J.H. Automatic interaction detection-version 4 (AID-4) (AFARL-TR-73-17). Lackland Air Force Base, Texas: Personnel Research Division, 1973.

Lewis, D. Quantitative methods in psychology. New York City: McGraw Hill, 1960.

Lindquist, E.F. Design and analysis of experiments in psychology and education. Boston, Massachusetts: Houghton Mifflin Co., 1953.

McNeil, K.A. Position statement on the roles and relationships between stepwise regression and hypothesis testing regression. Multiple Linear Regression Viewpoints, 1976, 6, 46-49.

McNeil, K.A. & Kelly, F.J. Express functional relationships among data rather than assume "intervalness." The Journal of Experimental Education, 1970, 39, 43-48.

McNeil, K.A., Kelly, F.J. & McNeil, J.T. Testing research hypotheses using multiple linear regression. Carbondale, Illinois: Southern Illinois University Press, 1973.

McNeil, K.A. & Spaner, S.D. Brief report: highly correlated predictor variables in multiple regression models. Multivariate Behavioral Research, 1971, 6, 117-125.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R. & Ervin, L. Type VI error: inconsistency between the statistical procedure and the research question. Multiple Linear Regression Viewpoints, 1976, 6, 1-19.

Norton, D.W. An empirical investigation of some effects of non-normality and heterogeneity on the F-distribution. An unpublished dissertation, 1952. Summarized in Lindquist, E.F. Design and analysis of experiments in psychology and education (pp. 78-90). Boston, Massachusetts: Houghton Mifflin Co., 1953.

Pohlman, J.T. & Newman, I. Regression effects when the assumption of rectilinearity is not tenable. Multiple Linear Regression Viewpoints, 1973, 4, 51-59.

Roscoe, J.T. Fundamental research statistics for the behavioral sciences (2nd ed). New York City: Holt, Rinehart & Winston, 1975.

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

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

Wainer, H. Estimating coefficients in linear models: it don't make ho nevermind. Psychological Bulletin, 1976, 83, 213-217.

Ward, J.H. & Jennings, E. Introduction to linear models. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.

Williams, J.D. Should a first course in ANOVA be through MLR? Multiple Linear Regression Viewpoints, 1976, 6, 42-45.

Young, R.K. & Veldman, D.J. Heterogeneity and skewness in analysis of variance. Perceptual and Motor Skills, 1963, 16, 588.

Young, R.K. & Veldman, D.J. Introductory statistics for the behavioral sciences (2nd ed). New York City: Holt, Rinehart, and Winston, 1972.

Multiple Linear Regression Viewpoints
Vol. 7 No. 2 1977

EMPIRICAL EXAMPLE OF THE USE OF INTERACTION TERMS IN THE MULTIPLE REGRESSION MODEL*

Thomas E. Dinero
Kent State University
Kent, Ohio

* Presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, April, 1976. This has not been reviewed.

The objective of the present study was to compare empirically the results from an analysis of variance and a multiple regression solution when appropriate interaction terms are included in the regression model. To the present author the study was revealing in that while the two models are logically and mathematically identical, the translation to raw data involves so many decisions that the identicalness may be lost. The following presentation thus may be incomplete in some respects since not all options could be taken. It will include the background to the present data collection,* the decisions made for the several analyses included here, and finally the various forms of the data.

While the theoretical rationale and methodology of the present study have been documented elsewhere (Schindler, 1975), a brief description of the variables may prove informative. The original study focused on the effects of teacher pressure on the socially desirable behavior of fourth grade children. A lie scale was devised which was made of the names of 30 television shows, 15 of which were fictitious: the directions for the scale involved requiring the subjects to "check those shows which [they] watched." The number of fictitious titles checked became that child's lie score. Classrooms in the experiment were randomly allocated to one of two treatments: the control group (n = 100) received the standard directions while the experimental group (n = 113) was told that a group

*Data used with the kind permission of Dr. Jean Schindler.

previously given the scale had not watched as many shows as the experimenter thought they might and that it was hoped that "they (the experimental group) would do better."

In addition to these two variables (the dependent variable "Lie" and the independent variable "Group"), two other variables were used; these were intelligence (measured by the Lorge-Thorndike) and locus of control (measured by the Children's Locus of Control Scale). When intelligence and locus of control were used to predict the lie score for the experimental and control groups separately, apparent differences were found in the two regression equations. It was these differences that led to the present consideration of the interaction of the predictors with experimental condition (Group) as a way of exploring the differences statistically.

Methodology and Rationale

Since the independent variables are measured variables, the regression analysis is appropriate. The ANOVA requires categorical variables, and thus the data had to be categorized in order to allow comparison.

There were thus three analyses done with the data collected. The first was a step-in regression using the raw data as collected. Because this model was most appropriate for the data, the step-in regression using the raw data became the analysis to which the others would be compared. The second was a 2x5x4 analysis of variance using the same variables as independent variables but after they had been broken into categories. The third analysis was another regression analysis using the categorized data instead of the raw data. This analysis was run

to study the information loss due to this categorization.

Regression Analysis on Original Data

Any statistical analysis is at heart a variance analysis procedure. In a regression setting, this becomes painfully clear with R^2 present wherever one turns. As with its twin brother, the linear model, the regression model brings with its flexibility a set of decisions many researchers in the past have either ignored or been unaware of.

The use of interaction terms in the regression model is one decision. This area appears to involve many points, the most salient of which seems to be (to this author) whether to include such terms and which ones to include.

As was said above, the difference in regression lines seemed to indicate that an interaction term was needed. However, the exact nature of the interaction term was another question. When one has truly nominal predictors, such as in this case, experimental and control, then contrast coding, dummy coding, or orthogonal coding yield the same summative results (through the overall F) but not the same (explanative) regression equation (Kerlinger and Pedhazur, 1973; McNeil, Kelley, and McNeil, 1975). With nominal independent variables, one predictor is created for each of the J levels of the first factor (J-1 in some analytic systems) and K (or K-1) for each of the K levels of the second factor. The number of interaction terms is JxK [or J-1) x (K-1) which could be used as predictors.

Now that one has decided to use interaction terms in his prediction model, he has to decide which ones to include. The predictors raised to the first power, these variables squared or cubed or any of their cross-products may be used. These terms all describe response surfaces such as those discussed by Ezekiel and Fox (1959) and Finn (1974). Any or all of these terms may be necessary to describe maximal variance. McNeil, Kelly, and McNeil (1975) have presented an excellent introduction to the correspondence between the mathematical formula and the physical surface itself.

For the present analysis, the most typical interaction term was used. This is the cross-product of the two independent variables. The most serious objection to this term is that it is dependent on the means and variances of the original scores (Glass, 1968; Kerlinger and Pedhazur, 1973). Thus, for these analyses the raw variables were standardized first and the z scores used to form the cross-product terms (Finn, 1974).

This rationale was further used in the creation (and potential interpretation) of the three-way interaction. In general, significant three-way interaction is seen to reflect different two-way interactions: If the ABC interaction is significantly different from zero, then either AB varies across C, AC varies across B, or BC varies across A. In any case, these differences would be manifest by significant cross-products of the standardized predictors.

Another major decision that must be made occurs when the predictors are correlated. Kerlinger and Pedhazur (1973) discussed this particular decision under the topic of unequal n. When one has, for whatever reason, a factorial design with different subsample sizes, he has a non-orthogonal design, one in which the "independent" variables are correlated. They thus share variance, and the order in which they are stepped into a regression equation becomes important. Kerlinger and Pedhazur suggested two alternatives to deal with this problem. Both were used in the present study. The first they call the experimental design approach. Here, a series of step-in analyses is performed such that every independent variable (not necessarily covariates) is entered last. In this way the unique (non-redundant) contribution of each variable is ascertained. In the second approach, the a priori ordering approach, one performs (generally) a single regression analysis, stepping in the variables as predetermined by one's research hypotheses. In the present study there were no research hypotheses as a guide. There was, however, a logic to the order in which the terms were stepped into the model. In standard ANOVA computation (Scheffe,1959) all independent variables are treated as if they were independent, and the interaction terms are adjusted for confounding with the main effects. In addition, three-way interaction is computed with adjustments for the main effects and the two-way interactions. This suggested that the main effects be entered first, the two-way interaction next, and the three-way last. Obviously, if the terms were independent as calculated in the regression analysis, ordering would be a moot point.

The results of these decisions led to a series of regression solutions for each of the predictors alone (seven models) for each two-way interaction after all unitary variables had been stepped in, and finally for the three-way interaction after all the other terms had been stepped in (the total model).

Analysis of Variance

In a factorial ANOVA, a cross-break is analyzed which has been created by two or more nominal partitions. The immediate problem for the present analysis involved translating the two continua to discrete variables. There was no a priori rationale evident, so it was decided to retain as many levels for each variable as would allow fewest (or no) empty cells. As can be seen from Table 2 , Intelligence and Locus of Control had a Pearson correlation of .077 for this sample. Recalling that these were two treatment conditions, it can be seen that the problem involved placing 100 and 113 people into IxJ categories. After some juggling, Intelligence was broken into five levels and locus of control into 4; this left one empty cell. Its value was computed from the marginal means (Winer, 1971), and, since this estimation method assumes no interaction, it was felt to be conservative.

A 2x5x4 fixed effects ANOVA was run on the Lie Scale data using the approximate method of unweighted means. The tabled degrees of freedom reflect the adjustment for the one missing cell: Note that the degrees of freedom for the three-way interaction effect are 1x4x3-1 = 11, instead of 1x4x3 = 12.

Regression Analysis Using Coded Data

Since there was no a priori research hypothesis, there was no basis on which to select a coding schema. Cohen (1968) has indicated that as long as one wants an over-all F or estimate of the R for a particular independent variable, any set of numbers may be used. Hence, since for the ANOVA the levels had been indicated by ordinal numbers, these were used in the present analysis.

Thus, for this last analysis, the first regression methodology was followed using the ordinal data instead of the raw data. Note that the input information for this analysis was exactly the same as that for the ANOVA.

Results

The results of the analyses are presented in the twelve tables appended to the report. They will be discussed in two parts, an initial description of salient aspects of the data and a final behavioral discussion.

Table 1 presents the means and standard deviations of the seven predictors used in the raw data regression analysis. It should be noted that the mean lying score (1.6009) and the high standard deviation suggest that the data are highly positively skewed. Table 2 presents intercorrelations among these predictors.

Eleven runs were made using variously selected predictors. These were done to demonstrate the contribution of each predictor in isolation (redundant with the first-order correlations) and then for each interaction term stepped in after the main effects; this method was an effort to duplicate the computation of the ANOVA. The standardized regression coefficients are presented in Table 3.

Here the terms for any "model" are presented as they had been stepped into the model (left to right). For example, there are four coefficients for Model 8: the first three, reflecting the main effects, are followed by the unique contribution of Group x Intelligence (ignoring the remaining terms). Table 4 gives more information about the last model which used all seven terms. Note that the three main effects explained 19.92 percent of the total variance, with the interaction terms adding only 3.45 percent in addition.

The results of the experimental design approach are presented in Table 5. The unique contribution of each predictor is shown by both the percent of variance and a significance test. To be noted here is the lack of significance of Locus of Control (compared with a p value of .0254 in the step-in analysis).

Table 6 is the ANOVA table for the analysis of variance. The percentages of variance accounted for are merely the sums of squares divided by the total sum of squares.

Tables 7 through 11 present the results of the regression analysis on the ordinally categorized data.

With three sets of results there are three pairs of comparisons, the two regressions with the ANOVA and the two regressions with each other. Any differences in the results would have to be due to analytic artifacts since the dependent variable and the sample are common to all.

For both the ANOVA and the raw score regression, Group and Intelligence were significant although Locus of Control did not appear to have unique variance. Neither the ANOVA nor the step-in regression picked this up.

The (non-significant) interactions showed inconsistencies across all three analyses. No meaningful interpretation seemed possible.

The categorized regression (Tables 10 and 11) did not parallel the ANOVA either, although the percent unique variances were closer to the ANOVA than were those for the raw data. Here, somewhat surprisingly, Locus of Control did have significant unique variance (see Table 11): any that was not unique was redundant with the other two main effects and not with the interactions (compare Locus of Control in Table 10 with that in Table 11).

Comparing the entries of Table 5 with those of Table 11 indicates that, on the whole, surprisingly little information is lost by using few levels of a variable instead of the raw data if this information is the variance accounted for. Note in contrast that the F values for Locus of Control, GxI, GxLC, and the three-way interaction did change.

Behaviorally, decisions stemming from the three present analyses and based on both variance accounted for and significance tests might have been as follows. The ANOVA was the most liberal, rejecting all three main effect hypotheses which accounted for the most variance (17.54 percent). The regression using raw data was the most conservative: it rejected only two hypotheses (for a combined 10.79 percent of the variance--see Table 5). The categorized regression analysis (Table 11) rejected three hypotheses explaining a combined total of 13.67 percent of the criterion variance.

Discussion

The following discussion is only a set of conjectures aimed at the results and should be taken only as that. Lector caveat.

Subjects were randomly selected by classrooms which were then randomly assigned to experimental treatment. Thus, the only "assigned" characteristic of the situation was "Group Membership." All others were naturally occurring and thus, for better or worse, have a degree of external validity. Given this natural situation, with its mixture of measured and assigned variables, unequal n, and unequal within variance estimates, regression is the correct analytic model. This model will thus be the focus in the present discussion.

In the ANOVA, it is felt that there were two main reasons for discrepancies from the regression. The first involved the coding of the Locus of Control. At first the continuum from 8 to 21 (the sample range) was broken down into unit intervals. With the presence of far too many empty cells, it was further collapsed to five levels. The continuum was further collapsed to four levels when it was observed that there were four empty cells at the highest level of the Control group. Thus a strong imbalance was created among the levels of Locus of Control. This may have had an influence in the hypotheses tests of this factor.

This same problem was noted across all cells in the design. The lack of orthogonality in the design can be noted statistically by the correlations among the factors (Table 8) and descriptively by the cell frequencies of the three two-way crossbreaks (Table 12). McNemar suggests that non-orthogonality may result from the number of levels, the forms of the distributions, and the overall n. These were certainly active in the present study. While the correlation between Intelligence and Locus of Control was low ($r = .077$), the disproportionality of level 4 of Locus of Control may have been an influence on the analyses.

The correlations among the predictors (Table 2) suggests that the interaction terms are confounded with the main effects. Thus, when each interaction was used alone, Group x LC was significant. It even contributed to the prediction above and beyond the three main effects (Table 9) but did not really have unique variance able to be considered non-chance (Table 11).

The ANOVA solution used here was only"approximate" (unweighted means analysis) and may have contributed to observed discrepancies. For the present analyses, the ANOVA almost always over estimated the variance accounted for (excepting Group x Intelligence where the difference of .07 may be due to calibration or rounding error). It may be that the covariances among the factors were not accurately partialled out. If this were a characteristic of such approximate solutions as used here, then this over estimate may lead to more Type I errors than stated alpha levels would suggest.

To some, the discrepancies between the three analyses would shed great doubt on the credibility of statistical analyses. With the same 113 (or 114) Lying scores analyzed in all three analyses, one would expect the same results for all analyses. However, the partitioning of the variance can be done in several ways, and it would seem that the "robustness" of any data analytic system is based on each characteristic of that system taken singly and in combination.

Thus, it may have been premature to assume that since the two regression models <u>appeared</u> different, when the two were combined, a significant and meaningful interaction term would emerge.

A host of conditions might have obviated this, including those conditions which restrict the correlation coefficient (restriction of range, non-symmetry, etc.) and incorrect interaction term. This last point deserves more comment. Given the problem of shrinkage, any regression analysis should be run in two phases, the first to estimate and the second to corroborate. This being the case, it may be just as wise to explore with the data of the first phase, to the extent of plotting the scatter diagrams, and use this information to select the interaction term to be used in the second phase. This type of exploration would seem to be almost a necessity in educational and psychological studies where there is little such comparative data available, where interaction has been something more to be avoided than awaited, and where complex aptitude-treatment interactions could bring exciting new interpretations to old data.

The suggestion resulting from all the above is that made by many authors (McNeil, Kelly, and McNeil, 1975): use one of the variants of the regression model. Where the design is purely experimental, watch the cell sizes: with equal n, the two models would yield exactly the same solutions. When non-orthogonality has entered the design for any reason, the regression analysis is the only appropriate solution. Rather than committing a Type VI error (Newman, Deitchman, Burkholder, and Sanders, 1976) by allowing analytic artifacts decide the results, et the research question (conveniently omitted in the present study) be the determiner.

TABLE 1

Means and Standard Deviations for All Variables (Total Group) Computed Under Regression Model

Predictor	Mean	Standard Deviation
Lying	1.6009	2.5412
Group Membership	1.4695	.5002
Intelligence	102.7559	17.1320
Locus of Control	13.9296	2.5734
Group x Intelligence	0.2161	.9579
Group x Locus of Control	-0.4229	.8943
Intelligence x Locus of Control	.0769	9950
GM x I x LC	-.0063	.9876

TABLE 2

Matrix of Intercorrelations Among the Seven Predictors and the Criterion

	Criterion	Predictor						
	Lying	Group Membership	Intelligence	Locus of Control	GxI	GxLC	IxLC	GxIxLC
Lying	1.000							
Group Membership	-.349**	1.000						
Intelligence	-.313**	.217**	1.000					
Locus of Control	-.014	-.425**	.077	1.000				
GxI	-.164*	.028	-.264**	-.007	1.000			
GxLC	-.138*	-.058	-.007	-.145*	.196**	1.000		
IxLC	.042	-.006	.097	-.098	-.472**	.148*	1.000	
GxIxLC	.042	-.078	-.442**	.101	.048	-.100	-.283	1.000

* Significant at $\alpha = .05$
** Significant at $\alpha = .01$

TABLE 3

Standardized Full and Semi-Partial Regression Coefficients For Each Step of the Step-Wise Analysis

Model Number	Predictor: Group Membership	Intelligence	Locus of Control	Group x Intelligence	Group x L.C.	Locus x Intelligence	Three-Way Interaction
1.	-0.349**						
2.		-0.313**					
3.			-0.014				
4.				- .164*			
5.					- .138*		
6.						.042	
7.							.045
8.	-0.383**	-0.184**	-0.161**	0.125			
9.	-0.351**	-0.226**	-0.132*		-0.097		
10.	-0.361**	-0.228**	-0.146*			0.048	
11.	-0.339**	-0.240**	-0.114*	0.146	0.050	0.088	0.091

* Significant at $\alpha = .05$
** Significant at $\alpha = .01$

TABLE 4

Step-In Regression Using All Seven Terms
In their Raw Score Form

Variable	Percent Additional Variance	F ratio	d.f.	p
Group	12.19	29.29	1,211	.0001
Intelligence	5.89	15.09	1,210	.0002
Locus of Control	1.84	4.81	1,209	.0294
Group x Intelligence	1.44	3.80	1,208	.0526
Group x Locus of Control	0.49	1.30	1,207	.2557
Intelligence x Locus of Control	1.01	2.70	1,206	.1017
Three-Way Interaction	0.51	1.36	1,205	.2457
Total Model	23.37	8.93	7,205	.0001

(R = .4834)

TABLE 5

Percent of Unique Variance Attributable to Each Predictor of the Total Model (Raw Data)

Predictor	Percent Unique Variance	F ratio	d.f.	p
Group	7.46	19.96	1,205	.0001
Intelligence	3.33	8.91	1,205	.0032
Locus of Control	.90	2.41	1,205	.1219
Group x Intelligence	1.27	3.40	1,205	.0668
Group x Locus of Control	0.21	0.56	1,205	.4568
Intelligence x Locus of Control	0.49	1.30	1,205	.2552
Three-Way Interaction	0.51	1.36	1,205	.2457

TABLE 6

Summary Table of Three Factor Analysis of Variance

Source	SS	d.f.	MS	Percent of Variance	F	P
Group	116.42	1	116.42	9.27	23.07	.00004
Intelligence	59.41	4	14.85	4.73	2.94	.02158
Locus of Control	44.47	3	14.82	3.54	2.94	.03402
Group x Intelligence	15.07	4	3.77	1.20	0.75	.56383
Group x Locus of Control	33.18	3	11.06	2.64	2.19	.08933
Intelligence x Locus of Control	46.48	12	3.87	3.70	0.77	.68405
GM x I x LC	63.43	12	5.29	5.05	1.05*	.40769
Error	877.82	174	5.04	69.87		
TOTAL	1256.28	213				

* With the d.f. corrected to 11 for the single estimated mean, the F becomes 1.14.

TABLE 7

Means and Standard Deviations for All Variables
Computed on Categorized Data Under Regression Model

(N = 214)

Variable	Mean	Standard Deviation
Lying	1.6028	2.5354
Group Membership	1.4720	0.5004
Intelligence	3.0514	1.4347
Locus of Control	2.7664	1.1914
Group x Intelligence	0.1548	0.9824
Group x Locus of Control	-0.2886	0.7896
Intelligence x Locus of Control	0.0507	0.8375
GM x I x LC	-0.0339	0.8391

TABLE 8.

Matrix of Intercorrelations Among the Seven Predictors and the Criterion (Categorized Data)

	Criterion	Predictor					
	Lying	Group Membership	Intelligence	Locus of Control	GxI	GxLC	IxLC	GxIxLC
Lying	1.000							
Group Membership	-.344**	1.000						
Intelligence	-.255**	.156	1.000					
Locus of Control	-.067	-.356**	.062	1.000				
GxI	.128	.015	-.111	-.010	1.000			
GxLC	.226**	-.241**	-.043	-.074	.120	1.000		
IxLC	.091	-.039	.124	-.006	-.404**	.197	1.000	
GxIxLC	-.100	.057	-.386**	.160	-.145*	-.022	.103	1.000

* Significant at $\alpha = .05$
** Significant at $\alpha = .01$

TABLE 9

Standardized Full and Semi-Partial Regression Coefficients For Each Step of the Step-Wise Analysis Using Categorized Independent Variables

	Predictor						
	Group Membership	Intelligence	Locus of Control	Group x Intelligence	Group x L. C.	Locus x Intelligence	Three-Way Interaction
1.	-.347						
2.		.255**					
3.			-.067				
4.				.128			
5.					.226**		
6.						.091	
7.							-.100
8.	-.393**	-.169**	-.195**	.113			
9.	-.352**	-.182**	-.192**		.147*		
10.	-.387**	-.176**	-.194**			.052	
11.	-.334**	-.216**	-.159*	.109	.115	.084	-.130

* Significant at $\alpha = .05$
** Significant at $\alpha = .01$

TABLE 10

Step-In Regression Using All Seven Terms in Categorized Form

Variable	Percent Additional Variance	F ratio	d.f.	p
Group	12.07	29.10	1,212	.0001
Intelligence	4.14	10.42	1,211	.0015
Locus of Control	3.24	8.43	1,210	.0041
Group x Intelligence	1.27	3.34	1,209	.0692
Group x Locus of Control	1.68	4.49	1,208	.0353
Intelligence x Locus of Control	0.59	1.59	1,207	.2090
Three-Way Interaction	1.24	3.38	1,206	.0673
Total Model	24.22			
	(R = .4922)	9.41	7,206	.0001

TABLE 11

Percent of Unique Variance Attributable to Each Element of the Total Model (Categorized Data)

Predictor	Percent Unique Variance	F ratio	d.f.	p
Group	8.33	22.65	1,206	.0001
Intelligence	3.35	9.12	1,206	.0029
Locus of Control	1.99	5.42	1,206	.0209
Group x Intelligence	.87	2.36	1,206	.1263
Group x Locus of Control	1.13	3.08	1,206	.0810
Intelligence x Locus of Control	0.53	1.44	1,206	.2318
Three-Way Interaction	1.24	3.38	1,206	.0673

TABLE 12

Frequencies With n the Two-Way Cross-breaks of the Main Effects

Intelligence

	1	2	3	4	5	
Group E	30	22	18	22	21	113
Group C	12	19	24	20	26	101
	42	41	42	42	37	214

$\chi^2 = 21.4$ $\phi = .316$ $\phi^2 = .1$
Significant at $\alpha = .01$

Locus of Control

	1	2	3	4	
Group E	13	22	11	67	113
Group C	32	25	24	20	101
	45	47	35	87	214

$\chi^2 = 28.676$ $\phi = .366$ $\phi = .134$
Significant at $\alpha = .01$

Intelligence

Locus of Control	1	2	3	4	5	
1	9	10	8	8	10	45
2	8	10	14	10	5	47
3	6	9	5	6	9	35
4	19	12	15	18	23	87
	42	41	42	42	47	214

$\chi^2 = 9.844$ $\phi = 214$ $\phi^2 = .046$
Not significant

REFERENCES

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 70(6), 426-443.

Ezekiel, M. and Fox, K. A. Methods of correlation and regression analysis. (3rd ed.), New York: Wiley, 1959.

Finn, J. A general model for multivariate analysis. New York: Holt, Rinehart and Winston, 1974.

Glass, G. V. Correlations with products of variables: Statistical formulation and implications for methodology. American Educational Research Journal, 1968, 5, 721-728.

Kerlinger, F. N., and Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart, and Winston, 1973.

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

McNemar, Q. Psychological statistics. New York: Wiley, 1969.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R., and Ervin, L. Type VI Error: Inconsistency between the statistical procedure and the research question. Multiple Linear Regression Viewpoints. Vol. 6, 4, 1-19, April, 1976.

Scheffe, H. The analysis of variance. New York, John Wiley & Sons, 1959.

Schindler, W. J. Locus of control and intelligence as predictors of overt lying behavior in nine through thirteen-year old children. Unpublished doctoral dissertation, Kent State University, 1975

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

INDEX OF ARTICLES IN VIEWPOINTS

1973 - 1976

Listed by Auther Alphabetically

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

Single-space indented abstract (the rest of the manuscript should be double-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, double-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 = $2.00; and institutional (libraries and other agencies) = $12.00. Membership dues and subscription requests should be sent to the Secretary of the MLRSIG."

THE UNIVERSITY OF AKRON
AKRON, OHIO 44325

TABLE OF CONTENTS