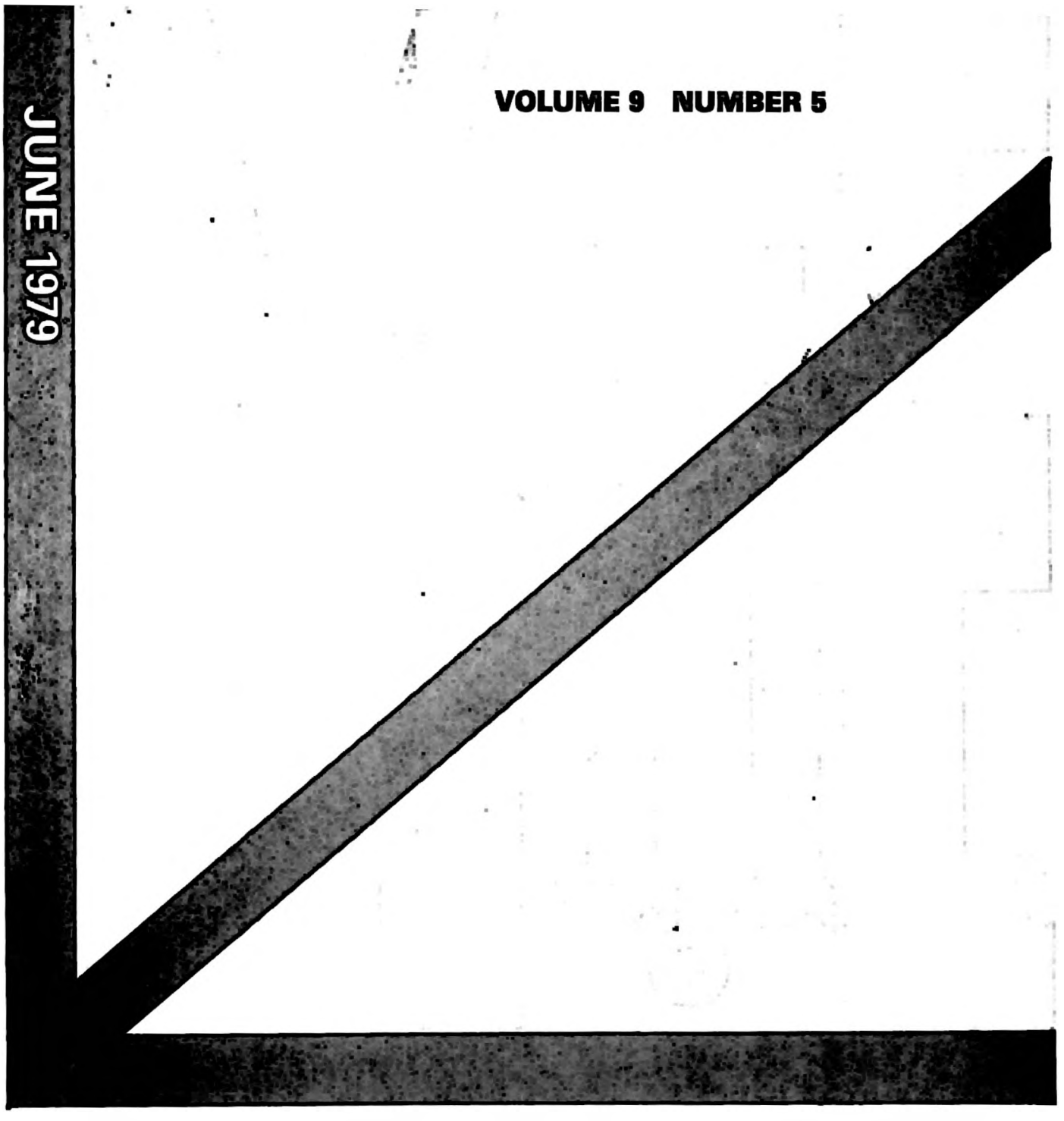
VOLUME 9 NUMBER 5

JUNE 1979

CONVENTION ISSUE



MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group
on Multiple Linear Regression

MULTIPLE LINEAR REGRESSION VIEWPOINTS

Chairman ..... John Williams
University of North Dakota
Grand Forks, ND 58201

Editor ..... Isadore Newman, Research and Design Consultant,
The University of Akron, Akron, OH 44325

Assistant ..... Diane Vukovich
The University of Akron, Akron, OH 44325

Executive Secretary ..... Steve Spaner, Behavioral Studies
University of Missouri, St. Louis, MO 63121

Chairman-elect ..... William Connett
State Department of Education, State Capital, MT 59601

Cover by ..... David G. Barr

EDITORIAL BOARD

TABLE OF CONTENTS

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 9 NO. 5 1979

NONLINEAR TRANSFORMATION OF THE CRITERION

Keith McNeil, NTS Research Corp.
Joe Evans, Evaluation Technology
Judy McNeil, Educational Monitoring Systems

ABSTRACT

The utility of a non-linear transformation of the criterion is established. A well-known law from a field other than education is used as the example to demonstrate the point. The functional relationships may be such (as in the Pythagorean Theorem) that an R^2 of 1.00 cannot be found without making a non-linear transformation of the criterion. The goal of predictability (R^2 = 1.00) thus may not be reached without making a non-linear transformation of the criterion.

There seems to us to be little justification for the present over-dependence in the behavioral sciences on the investigation of linear relationships. Since many of the simplest functional relationships in the physical sciences have been found to be non-linear or interactive, we find it interesting that few non-linear relationships have been established in the behavioral sciences, especially since most behavioral scientists would maintain that human behavior is no less complicated than physical behavior. We believe there are a number of historical reasons for the emphasis on linear relationships and these will be reviewed later.

There are some historical precedents, though, which provide excellent examples of the value of considering non-linear relationships. Newton's development of the law of gravity ($d = \frac{1}{2}gt^2$) within the framework of multiple linear regression has already been discussed elsewhere (McNeil, 1970; and

Presented to Special Interest Group on Multiple Linear Regression: Methodological Advances and Application during the 1979 Annual Meeting of the American Educational Research Association, San Francisco, April, 1979.

reprinted in McNeil, Kelly and McNeil, 1975). Newton's functional relationship relates the criterion (distance) to only one predictor variable, and that one variable is an interaction between a linear term and a non-linear term (gravity and the square of time). No linear interactions, no linear terms, and no "constants" were needed. We have hopes that the Newton discussion has influenced a number of behavioral scientists to investigate non-linear transformations of predictor variables. We turn now to discuss non-linear transformations of criterion variables.

Pythagoras Had a Reason for Considering a Non-Linear Transformation of the Criterion

Pythagoras was attempting to determine the length of the longest side of a right triangle from the lengths of the other two sides. Using regression models with linear terms was leading to only moderate success ($R^2 = .80$):

Model 1: $C_1 = a_0U + a_1A_1 + a_2B_1 + E_1$

Where: C_1 = length of longest side of right triangle;

U = 1 for all triangles;

A_1 = length of one of the shorter sides;

B_1 = length of the other shorter side; and

a_0, a_1, and a_2 are least squares weighting coefficients calculated so as to minimize the error in E_1.

After reading the paper describing Newton's success, Pythagoras considered some non-linear transformations:

Model 2: $C_1 = a_0U + a_1A_1 + a_2B_1 + a_3A_1^2 + a_4B_1^2 + E_2$

Where: A_1^2 = the square of the value in A_1; and

B_1^2 = the square of the value in B_1.

When using the above model, the R^2 increased to only .83, a non-significant increase. Pythagoras got frantic and gave the prediction terms the fifth degree:

$$\text{Model 3: } C_1 = a_0U + a_1A_1 + a_2B_1 + a_3A_1^2 + a_4B_1^2 + a_5A_1^3 + a_6B_1^3 + a_7A_1^4 + a_8B_1^4 + a_9A_1^5 + a_{10}B_1^5 + E_3$$

Alas, the increase in predictability was still only minimal: $R^2 = .84$.

A flash of insight (akin to a falling apple) led Pythagoras to transform the criterion, ending up with the following model:

$$\text{Model 4: } C_1^2 = a_0U + a_1A_1 + a_2B_1 + a_3A_1^2 + a_4B_1^2 + a_5A_1^3 + a_6B_1^3 + a_7A_1^4 + a_8B_1^4 + a_9A_1^5 + a_{10}B1^5 + E_3 + E_4$$

(Actually, the criterion variable was squared accidentally when Pythagoras was hastily transforming the predictor variables. Hence another in a long history of serendipitous findings.) As with most flashes of insight, the results led to an R^2 of .999 ... Pythagoras was delighted in finally reaching the first Goal of Research--Predictability. Pythagoras realized, of course, that the scientific community was not really interested in accounting for variance in C_1^2; their interest was in C_1. But Pythagoras realized that this new model would allow the indirect prediction of C_1 and would do it better than previous models.

Pythagoras also realized that there were other Goals of Research: Parsimony, Replicability, Validity Generalization, Control, and Explanation. (See McNeil, et al, 1975 for discussion of these Goals within the regression framework.) In striving for a more Parsimonious model, Pythagoras considered some restricted cases of Model 4, one of the more parsimonious being:

Model 5: $C_1^2 = a_0U + a_1A_1 + a_2B_1 + a_3A_1^2 + a_4B_1^2 + E_5$

Model 5 continued to yield an R^2 of .999... but contained only <u>five</u> pieces of information. Eliminating the second degree terms from Model 5 reduced the R^2 to .81, a significant and non-acceptable reduction. That is, Pythagoras preferred the less parsimonious model (Model 5) because of its higher degree of predictability.

Fortunately for Pythagoras, a maverick statistician was available for consultation and recommended eliminating either one or both linear terms from Model 5. Most statisticians at that time felt that all lower power terms belonged in a model. The maverick statistician, as well as Pythagoras, realized that the only terms which belonged in a model were the ones which were useful for prediction. Restricting the linear terms from Model 5 yields:

Model 6: $C_1^2 = a_0U + a_3A_1^2 + a_4B_1^2 + E_6$

Model 6 yielded an R^2 of .999... and the numerical values for the weighting coefficients were $a_0 = 0$; $a_1 = 1$; and $a_2 = 1$. Thus the unit vector could also be eliminated from the model, leaving only two pieces of information in the model. Any further restrictions (eliminating either A_1^2 or B_1^2) produced an R^2 too low to be acceptable. Thus, Model 7 was considered to be a better model because it met the Goals of Research of Predictability and Parsimony simultaneously:

Model C: $C_1^2 = a_3A_1^2 + a_4B_1^2 + E_7$

Pythagoras was concerned about Replicability, because Model 7 was obtained through data snooping and may well have been applicable only to the original set of data. Indeed, the high degree of predictbability (including the values of the weighting coefficients) of Model 7 held up under other sets

of data, thus meeting the Goal of Replicability.

The Goal of Validity Generalization did not prove to be of much concern to Pythagoras. Model 7 held up (continued to yield an R^2 of .999...) under a number of diverse conditions--such as color of right triangle, medium on which the right triangle was drawn, and altitude at which the right triangle was measured. (Newton's models did not hold up under varying altitudes and thus had to be modified before the model could validly be generalized across various altitudes.) The fact that Pythagoras' model was applicable to various kinds of right triangles meant that the model was not limited in scope, but was applicable to situations which were of interest to the scientific community at that time.

The Goal of Control was investigated by drawing the two sides which were at 90° to each other and then predicting the length of the hypotenuse. Additionally, varying the length of one of the two 90° sides had predictable effects on the hypotenuse. These studies confirmed that the length of the hypotenuse could be controlled (in a predictable fashion) by manipulation of either or both of the two 90° sides.

Linear and Non-Linear Relationships

The majority of the literature in any field has consisted of linear relationships. Hays (1963) indicates three major reasons why reasearchers have focused of linear relationships, none of which seem to be persuasive reasons to continue the practice.

First, "as a practical first approximation to a complicated relationship, a linear function rule often serves quite well" (Hayes, 1963, p. 565). This may be true in terms of a measure approximating a construct. But in relating one variable to another, many researchers have finally determined

that either a U-shaped or an inverted-U-shaped functional relationship exists between two variables. In such an instance, the bivariate correlation of both the linear component with the criterion and second degree component with the criterion will be close to zero; and if one were using a stepwise regression program neither would end up in the "best model". The "best model" simultaneously requires both the linear and second degree components. Hays (1963) goes on to say that, "the needs of practical prediction may be met quite well by a simple linear regression equation". The point that we would like to make is that settling for a moderate degree of predictability may no longer be an appropriate stance to take. In this day of abundant literature, ease in communication, funding for wide-scale studies, and availability of computer processing, one should attempt to achieve a fairly high degree of predictability--especially when predictability might be achieved by better use of data already collected, not at the expense of extra data collection.

As a second reason for the past focus on linear relationsips, Hays reminds us of the tie between correlation and multivariate normal assumptions. If one's data is normally distributed, then "the absence of a linear relation is the absence of any systematic relation" (Hays, 1963, p. 565). We only need to point out that many interesting criteria and predictors are not normally distributed. In addition, a "normally distributed" criterion could be highly predicted by a set of non-normally-distributed predictors.

Hays' third point is that many measurement and statistical techniques have been developed on the assumption of linear relationships. We would like to point out that the vast majority of computer programs deal only with linear relationships, or at least make it difficult to consider

non-linear relationships. On the other hand, most computer installations have a linear regression program which can investigate non-linear relationships. We find hope in Hays' statement that, "current practice and practical necessity should not be confused with the way nature or man behaves" (1963, p. 565). We have been known to express this same notion a little bit more strongly by pointing out that God did not make the world linear--all that one has to do to verify this point is to look closely at the horizon.

Why Investigate Non-Linear Relationships?

There are two major situations wherein one would investigate non-linear relationships. As discussed in more detail in McNeil, Kelly and McNeil (1975), one situation is when the functional relationship is indeed non-linear. The second situation is when the measure of the construct does not map the construct, and some rescaling of the measure is necessary.

Newton's law of gravity ($d = \frac{1}{2}gt^2$) is an example of a non-linear functional relationship. Any instance wherein raw scores are transformed to standard scores, or any type of scale, would be an example of a rescaling of the measure. Multiple linear regression rescaling applications in the areas of 1) difficulty of items, 2) ceiling effect, 3) guessing effect, and 4) conceptual rescaling are given in McNeil et.al. (1975, pp. 292-300).

The Pythagoras example required a non-linear transformation of the criterion in order to ascertain the functional relationship. Usually, one is interested in accounting for variance in the criterion in order to understand why subjects vary on the criterion. Once Pythagoras made a non-linear transformation on the criterion, the interest was on accounting for variance in the transformed criterion for the sole purpose of under-

standing why subjects vary on the untransformed criterion.

Conclusion

Pythagoras was not interested in accounting for the variance in the non-linear criterion per se. Instead, Pythagoras was interested in determining the functional relationship between the length of the hypotenuse of a right triangle and the length of the other two sides. The use of a transformed criterion led to the solution of the problem. Indeed, Pythagoras could not have obtained an R^2 of 1.00 by considering only the non-transformed criterion. That is, the hypotenuse is only determinable from the two sides of the triangle by the indirect route of the square of the hypotenuse.

Pythagoras was interested in determining the precise relationship between the hypotenuse and the other two sides of a right triangle. He was correct in measuring the two sides and using them as predictors of the hypotenuse. Although the initial results yielded an R^2 or .80, this R^2 was not sufficient for purposes of practical predictability, such as building bridge supports with right triangles. Pythagoras would have been wrong in assuming that the small amount of error existing in the initial results was due to the existance of some unmeasured variable(s) or to error of measurement. The initial results were due to the fact that the original measurements (both predictors and criterion) were not organized into the proper functional relationship.

References

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

McNeil, K.A. Meeting the goals of research with muliple linear regression. Multivariate Behavioral Research, 1970, 5, 375-386.

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

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 9 NO. 5 1979

MULTIPLE REGRESSION AS A TECHNIQUE FOR PREDICTING COLLEGE ENROLLMENT

Ambrose A. Clegg, Jr.
Karen Prichard
Paul Weigand
Kent State University

Abstract: This paper deals with the application of multiple linear regression to the problem of identifying appropriate criterion variables and predicting enrollment in college courses during a period of major rapid decline. Data was gathered on course enrollments for 1972 - 1978 and organized around five criterion variables. Total college enrollment proved to be the best single predictor with correlations of .89 to .99 with each of ten departmental course enrollments. The technique has proved to be 96 to 100 percent accurate in estimating course enrollments in seven of the ten courses. It is also a valuable means for data-based decision making and long range planning when faculty committees must advise on administrative decisions.

A problem faced by most administrators in scheduling staff to meet teaching requirements. In periods of changing enrollments in colleges and universities, it becomes important for administrators to have available a reliable means for predicting enrollment patterns in courses and departmental programs. Such a tool can be used to allocate resources more efficiently (staff and building use), as well as to provide a sound, empirical base for informed decision making--particularly when management decisions are shared with other groups.

This paper deals with the application of multiple linear regression to the problem of identifying appropriate criterion variables and predicting enrollment in college courses. While the initial focus was to develop enrollment patterns and predictions for a single department, the model can be used as a comprehensive tool for colleges or universities.

---

[1]*The authors wish to thank Sonya Blixt and Madeleine Mathis of Kent State University for their help with the statistical analyses.*

Background

In 1972 more than 1,862 students graduated from the College of Education at Kent State University. Of these, 566 were majors in the Department of Elementary Education (ELED). At that time Kent State ranked third in the nation in the number of elementary school teachers graduated. Two factors dramatically reversed this situation. The first was the tragic events of the shooting of four students on the campus on May 4, 1970 which produced a sudden, rapid decline in university enrollments. The second major factor was the growing surplus of teachers in the 1970's. Since 1972 there has been a continual, rapid decline in the number of students enrolled in the college and in elementary education. This decline of about 23 percent per year was evident in the number of new freshmen majors enrolled in ELED, but it was not so readily apparent in ELED class enrollment for two reasons: 1) the impact of loss in freshmen majors was delayed for two years because ELED courses are taught at the junior-senior level; and 2) there was significant increases in the majors in special education and early childhood from 1971 to 1975 in response to national and regional needs. Many of these students took ELED courses as dual majors or as service courses for their ow- major.

In the past years, class scheduling and faculty assignments were made on a short-term basis, usually only one or two quarters in advance. Reasonable estimates could be made using an inventory which provided historical data on class enrollments and on the number of sections offered during the previous six quarters. Estimates could be adjusted on the basis of pre-enrollment figures which were available about six weeks before the start of a new quarter.

In the fall of 1976 freshmen majors in ELED dropped to an all time low of 81. Enrollments in special education and early childhood education also dropped

noticeably. The University's tight financial position and major budget cuts in the College of Education made it evident that no additional faculty appointments could be expected even to replace losses due to retirement or resignation. The purpose of this study, therefore, was to identify those criterion variables that could be used to predict department enrollments in order to provide more orderly planning for scheduling of classes and in the management of faculty assignments.

Source of Data

The data for the study was derived from several official university records for the years 1972 through 1978. These included enrollments by course; the number of students enrolled as majors in elementary education and other departmental areas; total college enrollment; and total university enrollment.

Five independent variables were selected to determine the most likely indicators of future enrollment trends. These were: 1) total departmental majors in elementary education for the years 1972-1978; 2) elementary education majors plus all those enrolled in service courses and allied majors; 3) the number of elementary education freshmen enro-led for each year; 4) total enrollment of the College of Education by year for the period 1972-1978; and 5) total university enrollment.

Data Analysis

Data were analyzed using Multiple Regression Analysis: Subprogram - Regression from the Statistical Package for the Social Sciences (SPSS; Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975). Enrollment totals by year for ten ELED courses were the dependent variables with the five variables listed above as the independent variables. Criterion for inclusion of successive variables in the stepwise regression was $p < .05$.

Results

Results of the multiple regression analysis indicated that the best single predictor was total college enrollment. When other variables were entered in the stepwise regression, their contribution was nonsignificant. As Table 1 indicates correlations between individual courses and total college enrollments were exceptionally high ranging from .89 through .99.

Table 1

Multiple Regression Analysis for Total College Enrollment with ELED Course Enrollment

Course	*Multiple* R	*Multiple* R^2	*Degrees of Freedom*	F
31122	.942	.888	1,4	31.758**
31123	.890	.792	1,4	15.272*
31124	.962	.926	1,4	50.099**
31125	.996	.992	1,4	512.455**
41114	.924	.855	1,4	23.517*
41122	.984	.969	1,4	123.828**
41125	.976	.953	1,4	80.939**
41131	.980	.961	1,4	97.567**
41137	.994	.987	1,4	310.421**
41138	.990	.980	1,4	198.895**

*p .05

**p .01

Examination of the correlation matrix shown in Table 2 indicated that two other varibles (ELED majors, and ELED majors plus those enrolled in service courses and allied majors) correlated highly (.79 to .91) with total college enrollment.

Table 2

Correlation Matrix of Total College Enrollment with Five Variables

Course	*Course Enrollment*	*ELED Major*	*ELED +SVC*	*ELED FRESH*	*Total Univ*
31122	.94	.91	.80	- .70	.23
31123	.89	.91	.87	- .70	.35
31124	.96	.91	.87	- .69	- .41
31125	.99	.91	.87	- .70	.23
41114	.92	.86	.81	- .76	.23
41122	.98	.86	.81	- .76	- .32
41125	.98	.86	.81	- .72	.23
41127	.98	.86	.81	- .76	.23
41131	.98	.86	.81	- .76	.23
41137	.99	.86	.79	- .76	.23
41138	.99	.86	.81	- .76	.23

On the other hand, total university enrollment had only a moderately low correlation (.23 to .35) with college enrollment, suggesting that factors related to college enrollment were probably quite different from those related to total university enrollment. The strong inverse relationships between freshmen in elementary education and total college enrollment probably is accounted for by the fact that all ten courses are offered at junior-senior level and thus no freshmen are included in the enrollment data.

Figure 1 shows the correlations between course enrollment and total college enrollment, and the equations used to predict enrollment.

Figure 1

Correlations and Regression Equations

For Course Enrollments

Course	*Correlation*	*Equation*
31122	.942	PE = .072(TC) + 29.018
31123	.890	PE = .052(TC) + 40.492
31124	.962	PE = .143(TC) - 151.660
31125	.996	PE = .135(TC) + 10.349
41114	.924	PE = .136(TC) + 43.606
41122	.984	PE = .159(TC) - 246.483
41125	.976	PE = .196(TC) - 168.551
41131	.980	PE = .145(TC) - 103.324
41137	.994	PE = .195(TC) - 236.180
41138	.990	PE = .114(TC) - 25.380

Predicting Course Enrollments, 1978-1980

Several steps were involved in determining enrollment estimates for 1978-79 and 1979-80. The first step was to develop an accurate estimate of total college enrollments for the two-year period. This was done by averaging the annual decrease over the years 1974 through 1977. The average decrease was 297. Since it appeared likely that the total college enrollment will continue to drop, a decrease of 297 was estimated for both years 1978 and 1979. While the declines in elementary education and in other areas are more rapid and would indicate a greater decrease, those losses are offset by increases in vocational education where older students are enrolling to seek certification in trades and industries and related programs.

Once total college enrollments were estimated, individual course enrollments were estimated using the general regression equations indicated in Figure 1. In order to allocate the total number of students predicted for the year in any one class to each of the four quarters, a format similar to that shown in Figure 2 was devised.

Figure 2

Estimated Quarter Enrollment Based On

Percentage of Predicted Course Enrollment

	Summer 1978	*Fall 1978*	*Winter 1979*	*Spring 1979*	*Predicted Yearly Enrollment*
% of Yearly Total	07	38	30	25	66
Course Enrollment	4	25	20	17	

This allocation over quarters was baded on an analysis of the preceding five-year enrollment pattern for that particular course for each quarter. As shown in Figure 2 the percentage that each quarter represents of the total year's enrollment is then determined. Using ELED 31122 as an example, Fall quarter 1978 represents 38 percent of the estimated enrollment for that class or 25 students of the predicted total of 66. A similar determination was made for the remaining quarters. Enrollment figures for each quarter were based on similar estimates for each of the departmental courses.

How accurate were these predictions? An analysis of actual enrollment data during Fall quarter 1978 indicated that in seven out of ten of the courses the estimates ranged between 96 and 100 percent accurate. In three courses the margin of error was ten per cent or 90 per cent accuracy. In practical terms this meant that in one course 96 students were predicted and 100 enrolled; in another 93 students were predicted and 96 enrolled. In a third case, 43 were predected and 43 enrolled! In each case actual enrollments were slightly higher than the predicted figure.

Discussion

A number of important implications appear to stem from this study. The accuracy of prediction described above leads to more precise advanced planning of faculty resources. Faculty teaching loads can be worked out for an entire academic year. Where necessary, faculty can be shifted to other responsibilities if enrollment estimates are low. If additional sections of a course are warranted, temporary faculty can be hired well in advance rather than waiting for final registration figures. Rooms and other university resources can also be more efficiently allocated based upon highly accurate predictions. This technique has also proved useful in long-range planning in that enrollment trends can be used to validate decisions whether or not to retain faculty positions when they become vacant through

retirement or resignation, or to reallocate faculty and bedgetary resources to other areas where student enrollment needs are greater.

Another important implication is that use of the technique described here provides an empirical basis for informed decision making when such decisions must be shared with other constituent groups. It also reduces the influence of subjectivity and personality in decision making. For example, during a time of considerable tension and anxiety among faculty about declining enrollment and possible retrenchment, this technique proved an excellent source of data for making difficult decisions about the number of classes to be scheduled, faculty loads, and reassignment to other duties. Data from a preliminary version of this study was used to validate the need for a reduced summer school program and a more limited employment of faculty for summer school teaching. When administrative proposals were accompanied by the data on enrollment predictions as described above, the department executive committee recommended the acceptance of the reduced program and turned its attention to the consideration of other alternatives.

Several other uses might be considered for further application of this model. Whereas this study dealt with the enrollment in a single department, the model could be extended to a number of departments in a college or to an all-university model. Obviously, other relevant variables would need to be examined and selected carefully depending on the circumstances at hand. However, most colleges or universities have easy access to a wealth of statistical information which can be used to determine a long range history for each of the variables. Other uses might include assessing the impact upon enrollment patterns of mandated educational policies such as P.L. 94-142 (mainstreaming) or state required changes in teacher education programs.

In summary, multiple linear regression analysis provides a much better method of predicting college enrollment than the informed "guestimate" method that had been previously employed.

References

Nie, N., Hull, C., Jenkins, J., Steinbrenner, K., and Bent, D. Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill, 1975.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 9 NO. 5 1979

UNMEASURED VARIABLES IN PATH ANALYSIS

Lee M. Wolfe
Virginia Polytechnic Institute and State University

Structural equation models have been useful in estimating parameters of many substantive problems in educational research. Such models have been applied to study the effect of educational attainment in intergenerational occupational mobility (Blau and Duncan, 1967), the social psychological effects of one's best friend's college plans on the respondent's further education (Duncan, Haller, and Portes, 1968), the effect of parents' and teachers' encouragement upon educational attainment (Sewell and Hauser, 1975), and ethnic and social psychological effects upon academic achievement (Anderson and Evans, 1974; Anderson, 1978).

All of the cited analyses (indeed, most studies incorporating regression or structural equation procedures) have rested upon the implicit, but unrealistic, assumption that the independent variables were measured without error (see Blalock, 1964: 49). In practice, measurement errors in independent variables have been ignored, because it was felt that ignoring random measurement errors merely led to conclusions more conservative than would otherwise be the case. For example, it is well known that least-squares estimating procedures yield attenuated estimates of the regression slope and correlation coefficient in the bivariate case (see Appendix). Thus, it has been believed that such results understate the true relationships. In the

---

* This paper was prepared for presentation at the annual meetings of the American Educational Research Association, San Francisco, April 8-12, 1979. Work on this project was in part supported by the National Center for Education Statistics, Department of Health, Education and Welfare (No. 300-78-0516). Susan Rothschild assisted in computations using facilities of Virginia Polytechnic Institute and State University, Northern Virginia Campus. Maureen Moment provided helpful suggestions and criticisms which are gratefully acknowledged.

case of multiple independent variables, however, the existence of measurement errors becomes a serious problem (Blalock, 1965; Bohrnstedt, 1969).

From a formal point of view, assuming measurement error in structural equation models is much the same as assuming variables to be unobserved. In the former case, one assumes that the true variable is observable, but only with error; in the latter, one assumes that the true variable is unobservable, and uses in its place one or more imperfectly measured indicators. Because true variables may never be measured exactly, in a strict sense all true variables are unobserved. In practice, then, observations are collected on manifest variables thought to be related to the latent variable of real theoretical interest.

Measurement errors and unobserved variables constitute a large topic. Indeed, the field of psychometrics addresses itself almost entirely to the problem of errors in variables. In sociology, substantial efforts are under way to estimate error in data collection (for example, Schuman and Presser, 1978), and in estimation procedures (for example, Blalock, Wells and Carter, 1970; Wiley, 1973). In path analysis, models containing unobserved variables have been a part of the literature for years (Hodge and Treiman, 1968; Siegel and Hodge, 1968; Duncan, 1969b; Hauser, 1969; Land, 1970; Wiley and Wiley, 1970; Hauser and Goldberger, 1971; Duncan, Featherman and Duncan, 1972; Hauser, 1973; Otto and Featherman, 1975; Bielby, Hauser and Featherman, 1977). Indeed, one of the earliest substantive applications of path analysis was by Sewell Wright (1925) to the interaction between corn crops and hog prices, and included hog breeding variables which were unobserved.

The application of structural models incorporating unmeasured variables may influence the explanation of educational phenomena.

As Kerlinger (1977) pointed out, models including unmeasured variables make possible the rigorous testing of theories previously difficult to test adequately because of fallible measures of the theoretical variables. And as Cooley (1978) noted, such models now define the "state of the art" in educational research. Unfortunately, both expository articles and reports of substantive applications of structural equation models incorporating unmeasured variables have been rare in educational research literature. This paper first discusses a simple causal model, incorporating a single unmeasured variable for the purpose of exposition. A substantive example will follow, incorporating several unmeasured variables for which multiple indicators were available. This paper thus extends the work of Wolfle (1977) and Williams (1978), who provided introductions to the subject of path analysis from the perspective of regression analysis, Wolfle (1978), who exposited path analysis as a means of substantive interpretation of data, and Anderson (1978), who exposited a nonrecursive equation model.

Let us begin with a simple example. Consider a simple causal chain of the process of intergenerational occupational mobility from father's socioeconomic status (X_3), to respondent's educational attainment (X_2), to respondent's socioeconomic status (X_1). However, let us revise the model such that true educational attainment is not directly observed. Instead, its observed indicator, educational attainment, is contaminated with errors of measurement. We assume that the amount of education actually recorded is caused by the respondent's true educational attainment, in addition to several other factors. For example, the respondent may be more or less ignorant of the number of years of regular school or college he or she completed and got credit for." The respondent may tend to round off

educational attainment to even years, or multiples of four (such as 8, 12, 16, or 20). Some respondents may wish to appear to have acquired more (or less) schooling than was actually the case.

The complete model consists of three equations, the first of which describes the fallible measurement of observed education, while the other two represent the causal model as such. The three equations may be written:

$$x_2 = b_{2\eta}\eta + e,$$

$$\eta = b_{\eta 3}x_3 + u,$$

$$x_1 = b_{1\eta}\eta + v,$$

where x_1 is the respondent's Duncan (1961) socioeconomic index score as revised by Hauser and Featherman (1977), x_2 is the respondent's recorded educational attainment, x_3 is respondent's father's Duncan socioeconomic index score, η is true educational attainment, and e, u, and v are residual errors. All of these are measured as deviations from their means. The b's are, therefore, regression coefficients, and $b_{2\eta} = 1.00$. These structural relationships may be diagrammed, as shown in Figure 1.

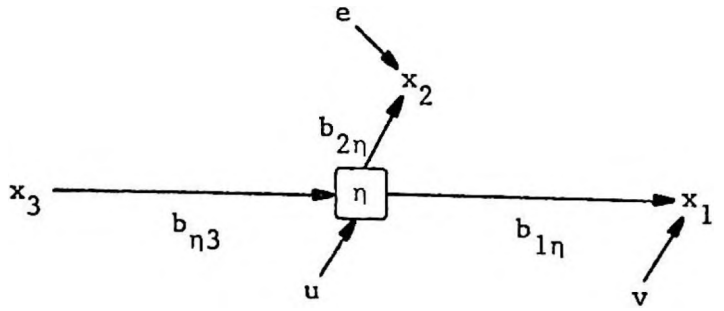


Figure 1. A Causal Chain of Intergenerational Status Attainment

A remarkable property of this model is that unbiased estimates of the coefficients may be obtained for both the causal parameters and the measurement process. This occurs because the model is just identified as a result of some restrictive assumptions imposed on the expected associations of the residual errors with other variables in the model. The usual specifications in such models is that the residual errors are uncorrelated in the population with other, predetermined variables in the equation. Thus,

$$E(x_3u) = E(\eta v) = E(x_3v) = 0.$$

In addition, it is assumed that the residual error of measurement is uncorrelated with the true score, η, and also with both x_1 and x_3:

$$E(\eta e) = E(x_1e) = E(x_3e) = 0.$$

These strong assumptions are roughly equivalent to assuming the error of measurement is random, and not systematic. These oversimplifying assumptions are properties of the model, not necessarily of what the world is really like. In any realistic context, these assumptions are problematic, and must be assessed against the researcher's knowledge of the topic under investigation. For example, it is possible that respondents whose fathers are employed in occupations of low socioeconomic status, or who themselves are employed in such occupations, tend to overstate their educational attainment. Complex models can be constructed which permit the intercorrelation of residuals, but the simple alternatives for this simple example are either to abandon the exercise or to accept the restrictive assumptions.

The three equations in the model may be reduced to three equations with three unknowns. The details of these computations are shown in the Appendix. For purposes of illustration, the model was estimated

with data taken from the 1977 general social survey of the National Opinion Research Center (1977), for whites and blacks. The correlations, means and standard deviations are shown in Table 1, and the results in Table 2.

The upper panel of Table 2 shows the results one would obtain with ordinary least squares (OLS). The structural coefficients (regression coefficients in their original metric) are shown, and below each in parenthesis are the standardized (path) coefficients. When comparing socioeconomic returns across groups, the structural coefficients should be used (see Kim and Mueller, 1976). The coefficients of determination are shown in the right-hand column. For whites, one would conclude that one point of father's socioeconomic index (SEI) returns about .05 years of education, and that one year of education was converted into about 4.5 points of respondent's own SEI. For blacks also, one would conclude that one point of father's SEI yielded about .05 years of education, but that blacks were able to convert one additional year of education into only 3.4 points of their own SEI.

The lower panel of Table 2 shows the results that one would obtain from the model diagramed in Figure 1. Note that measurement errors seem to be larger for blacks than for whites; that is, the standardized coefficient relating true education to observed education is larger for whites than for blacks. Comparing the OLS estimates to the corrected estimates for the regression of education of father's SEI, one should note that the OLS estimate is identical to the corrected estimates. Random measurement error in the dependent variable does not bias the OLS estimate. However, random error of measurement in the independent variable imparts a downward bias to the OLS estimate. And the lower the precision of measurement,

Table 1. Correlations, Means, and Standard Deviations of Status Variables; Whites Above Diagonal (N = 1333), Blacks Below Diagonal (N = 172).

	Father's SEI X_3	Education X_2	SEI X_1	White Mean	White S.D.
Father's SEI (X_3)	1.00	.370	.263	28.05	23.57
Education (X_2)	.255	1.00	.570	11.83	3.14
SEI (X_1)	.279	.560	1.00	37.48	24.53
Black Mean	14.48	10.43	23.39		
Black S.D.	17.90	3.66	21.85		

Table 2. Ordinary Least Square and Corrected Estimates of Parameters in a Causal Chain Model of Intergenerational Status Attainment

Dependent Variable	Predetermined Variables: Father's SEI (X_3)	Observed Education (X_2)	True Education (η)	R^2
Ordinary Least Square Estimates				
Whites				
Observed Education (X_2)	.0493 (.370)	. . .	. . .	.137
SEI (X_1)	. . .	4.451 (.570)	. . .	.325
Blacks				
Observed Education (X_2)	.0521 (.255)	. . .	. . .	.065
SEI (X_1)	. . .	3.346 (.560)	. . .	.314
Corrected Estimates				
Whites				
True Education (η)	.0493 (.413)	. . .	. . .	.171
Observed Education (X_2)	. . .	. . .	1.000 (.895)	.801
SEI (X_1)	. . .	. . .	5.549 (.637)	.406
Blacks				
True Education (η)	.0521 (.356)	. . .	. . .	.127
Observed Education (X_2)	. . .	. . .	1.000 (.715)	.511
SEI (X_1)	. . .	. . .	6.546 (.783)	.613

Note: Standardized (path) coefficients appear in parentheses.

the greater the downward bias. For whites, the corrected regression coefficient leads one to conclude that one year of true educational attainment was converted into 5.5 points of SEI. When measurement errors were ignored for whites, SEI returns to education were underestimated by about 20 percent, assuming the measurement errors were purely random. For blacks, the corrected regression coefficient leads one to conclude that one year of true educational attainment was converted into 6.5 points of SEI. When measurement errors were ignored for blacks, SEI returns to education were underestimated by about 49 percent, again assuming the measurement errors were well behaved.

This analysis was based on the assumption that the only kinds of measurement error for both blacks and whites were random, and not systematic. Our conclusions are, therefore, contingent on the correctness of those assumptions, but suggest that previous studies (for example, Duncan, 1969a) comparing the processes of status attainment for whites and blacks have exaggerated racial differences in returns to education by failing to account for measurement error. Bielby, Hauser and Featherman (1977) estimated status attainment models for whites and blacks incorporating both structural and response components. They found that response errors for whites were random, but were not for blacks. Nonetheless, the substantive consequences were the same as here: ignoring measurement errors exaggerated racial differences.

STRUCTURAL EQUATION MODELS WITH MEASUREMENT ERRORS

Unmeasured variables may also be included in causal models in the more general case in which unmeasured variables appear as underlying causes of several observed variables. This type of model translates into a confirmatory factor-analysis model with an assumed structural order among the factors. A general method has only recently been developed which incorporates recursive path analysis, interdependent econometric models, factor analysis, and analysis of covariance structures. This method, the analysis of linear structural relationships, or LISREL, was introduced by Jöreskog's (1973) technical paper, and has been updated in Jöreskog (1977). Less technical introductions are available in Long (1976) and Jöreskog and Sörbom (1978). A computer program is available (Jöreskog and Sörbom, 1978).

The LISREL model assumes a causal structure among a set of unmeasured, latent variables, some designated as exogenous and others as endogenous. These unmeasured variables are also related to a set of observed variables such that (in the example to follow) the latent variables appear as causes of the observed variables. The LISREL model, therefore, consists of two parts: the measurement model, and the structural equation model (Jöreskog and Sörbom, 1978: 3-4).

By way of example, Lichtman and Wolfle (1978) are studying the processes of educational attainment among several ethnic groups, including whites, blacks, and Hispanics. They proposed to compare structural equation models among ethnic groups in order to determine the extent to which differences exist in the educational returns to socioeconomic background and within-school variables. The population

under study is the high school graduating class of 1972, described in detail by Levinsohn, et al. (1978). One should expect that members of this high school cohort have not all completed their education as of 1976 (the latest followup), so the results pertain to educational returns as attained four years after graduation. The respondents were initially surveyed in 1972, and followed up in 1973, 1974, and 1976. Because some questions were repeated in various questionnaires, or because alternative means exist for constructing manifest variables, in many cases multiple indicators exist for latent variables. This becomes important in light of Bohrnstedt and Carter's (1971: 142) admonition that, "sociologists seem to be blatantly unconcerned with the problems of measurement error." Moreover, Bielby, Hauser, and Featherman (1977) showed that differential measurement errors existed between blacks and whites in the 1973 replication of Blau and Duncan (1967), thus leading to exaggerated racial differences in models ignoring measurement error.

For the expository purposes of this paper, a preliminary model incorporating structural associations among several latent variables, and components of measurement errors, has been constructed. This model, shown in Figure 2, includes two manifest measures for father's educational attainment, mother's educational attainment, and the respondent's high school curriculum. Three manifest measures are included for respondent's educational attainment. Single manifest variables measure father's socioeconomic index, the number of siblings, and high school class ranking. The LISREL model specifies that each manifest variable is generated by a latent factor for that variable, plus a response error which is independent of the latent factor. In LISREL terminology:

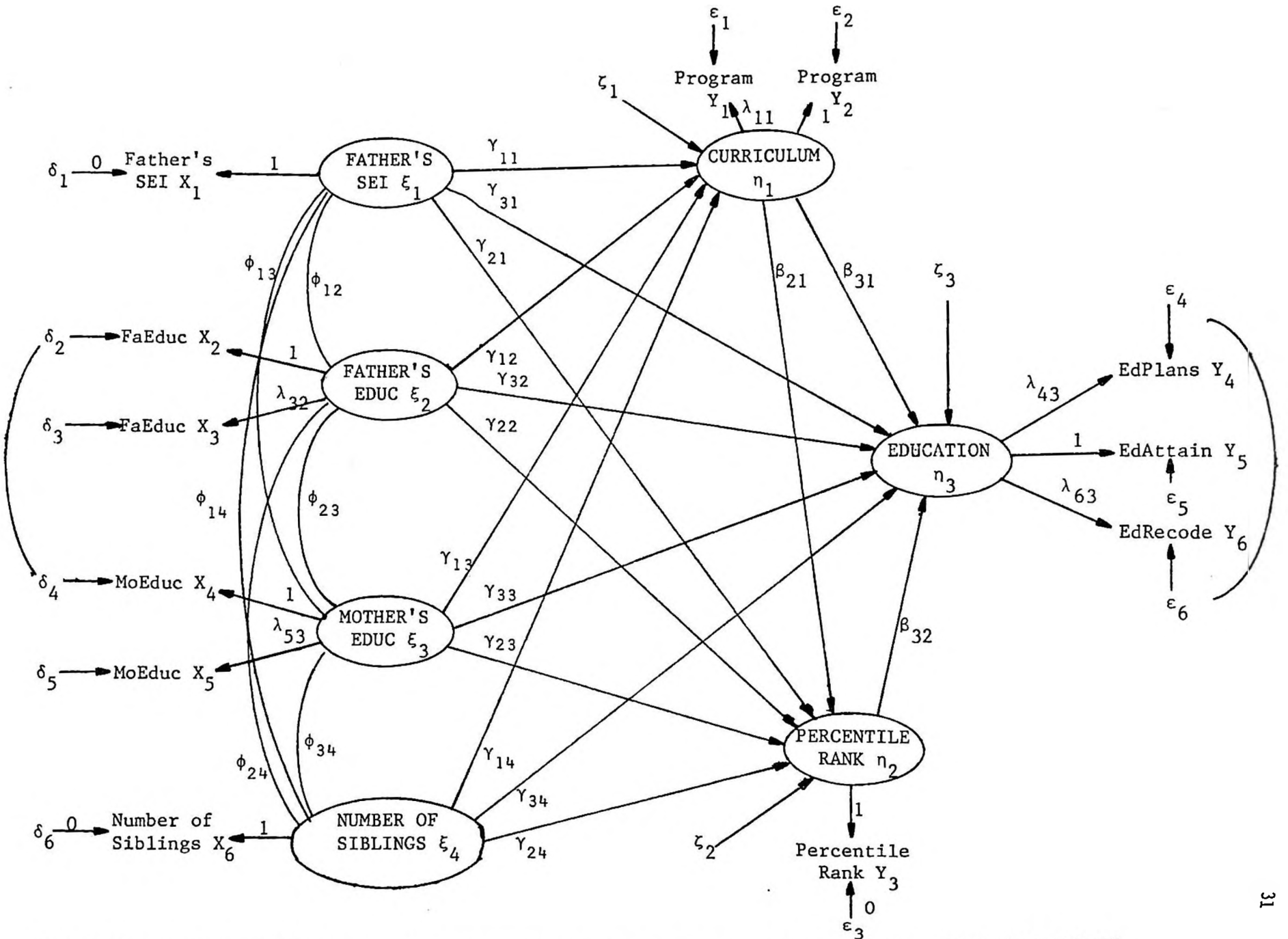


Figure 2. Structural Equation and Measurement Models of Educational Attainment, 1972 High School Graduates

two random vectors $\underset{\sim}{\eta}' = (\eta_1, \eta_2, \eta_3)$, and $\underset{\sim}{\xi}' = (\xi_1, \xi_2, \xi_3, \xi_4)$ represent the latent endogenous and latent exogenous variables, respectively. The model specifies a fully recursive causal structure among the latent variables, such that:

$$\underset{\sim}{\beta}\,\underset{\sim}{\eta} = \underset{\sim}{\Gamma}\,\underset{\sim}{\xi} + \underset{\sim}{\zeta}$$

where $\underset{\sim}{\beta}(3 \times 3)$ and $\underset{\sim}{\Gamma}(3 \times 4)$ are matrices of structural coefficients in which $\underset{\sim}{\Gamma}$ is a full matrix relating the exogenous vector to each of the endogenous latent variables, and $\underset{\sim}{\beta}$ is a matrix relating the endogenous variables to each other. $\underset{\sim}{\zeta}' = (\zeta_1, \zeta_2, \zeta_3)$ is a random vector of residuals uncorrelated with $\underset{\sim}{\xi}$.

The vectors $\underset{\sim}{\eta}$ and $\underset{\sim}{\xi}$ are not observed, but $\underset{\sim}{y}' = (y_1,\ldots,y_6)$ and $\underset{\sim}{x}' = (x_1,\ldots,x_6)$ are observed, such that:

$$\underset{\sim}{y} = \underset{\sim y}{\Lambda}\,\underset{\sim}{\eta} + \underset{\sim}{\varepsilon}$$

and

$$\underset{\sim}{x} = \underset{\sim x}{\Lambda}\,\underset{\sim}{\xi} + \underset{\sim}{\delta}$$

where $\underset{\sim}{\varepsilon}$ and $\underset{\sim}{\delta}$ are vectors of errors of measurement in $\underset{\sim}{y}$ and $\underset{\sim}{x}$, respectively. These errors of measurement represent both specific and random components of variation (see Alwin and Jackson, forthcoming). They are assumed to be uncorrelated with $\underset{\sim}{\eta}$, $\underset{\sim}{\xi}$, and $\underset{\sim}{\zeta}$, but may be correlated among themselves. The matrices $\underset{\sim y}{\Lambda}(6 \times 3)$, and $\underset{\sim x}{\Lambda}(6 \times 4)$ are regression matrices of $\underset{\sim}{y}$ on $\underset{\sim}{\eta}$ and of $\underset{\sim}{x}$ on $\underset{\sim}{\xi}$, respectively.

Let $\underset{\sim}{\Phi}(4 \times 4)$ be the covariance matrix of $\underset{\sim}{\xi}$. Let $\underset{\sim}{\Psi}(3 \times 3) = \text{diag}(\psi_{11}, \psi_{22}, \psi_{33})$ be the variance matrix of $\underset{\sim}{\zeta}$. Let $\underset{\sim\varepsilon}{\Theta}$ and $\underset{\sim\delta}{\Theta}$ be the covariance matrices of $\underset{\sim}{\varepsilon}$ and $\underset{\sim}{\delta}$, respectively. In application, some of the elements of the four regression matrices, and the four covariance matrices, are fixed and equal to assigned values. Other

elements are free parameters to be estimated by the method of maximum likelihood. This defines the LISREL model.

The structural model is presented in the path diagram of Figure 2. The variables enclosed in ellipses are unobserved, latent variables. The manifest variables included in the model are as follows, in which the number in parentheses refers to the variable number as given in Levinsohn, et al. (1978):

X_1 = father's socioeconomic index (V2468),

X_2 = composite measure of father's education (V1627),

X_3 = father's education (V1009),

X_4 = composite measure of mother's education (V1628),

X_5 = mother's education (V1010),

X_6 = sum of older and younger brothers and sisters (V1460 + V1461 + V1462 + V1463),

Y_1 = high school program as reported by respondent (= 1 if academic, = 0 otherwise) (V209),

Y_2 = high school program as reported by school record (= 1 if academic, = 0 otherwise) (V196),

Y_3 = percentile rank in class (V631),

Y_4 = educational plans as of 10/1/76 (V1855),

Y_5 = educational attainment as of 10/1/76 (V1854), and

Y_6 = educational recode (Melone, personal correspondence).

The substantive portion of Figure 2 is a fully recursive model among the latent variables, represented by the following structural equations:

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \gamma_{13}\xi_3 + \gamma_{14}\xi_4 + \zeta_1$$
$$\eta_2 = \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \gamma_{23}\xi_3 + \gamma_{24}\xi_4 + \beta_{21}\eta_1 + \zeta_2$$
$$\eta_3 = \gamma_{31}\xi_1 + \gamma_{32}\xi_2 + \gamma_{33}\xi_3 + \gamma_{34}\xi_4 + \beta_{31}\eta_1 + \beta_{32}\eta_2 + \zeta_3.$$

In algebraic form, the measurement portion of Figure 2 is:

$$x_1 = \xi_1$$
$$x_2 = \lambda_{22}\xi_2 + \delta_2$$
$$x_3 = \lambda_{32}\xi_2 + \delta_3$$
$$x_4 = \lambda_{43}\xi_3 + \delta_4$$
$$x_5 = \lambda_{53}\xi_3 + \delta_5$$
$$x_6 = \xi_4$$
$$y_1 = \lambda_{11}\eta_1 + \varepsilon_1$$
$$y_2 = \lambda_{21}\eta_1 + \varepsilon_2$$
$$y_3 = \eta_2$$
$$y_4 = \lambda_{43}\eta_3 + \varepsilon_4$$
$$y_5 = \lambda_{53}\eta_3 + \varepsilon_5$$
$$y_6 = \lambda_{63}\eta_3 + \varepsilon_6$$

A metric for the latent variables is established by fixing $\lambda_{22} = \lambda_{43} = \lambda_{21} = \lambda_{53} = 1.0$. That is, the metric of the latent variables father's education, mother's education, curriculum, and respondent's education are fixed to be the same as that of the composite measures of education for father and mother, respectively, the school report of curriculum, and educational attainment as of 10/1/76. The metrics of father's socioeconomic status, siblings, and class rank have already been fixed by the algebra of the measurement

model. Normalizations of this kind are necessary because the metric of an unobserved variable is arbitrary. Consequently, the regression slopes of manifest variables on latent variables are identifiable only relative to each other.

The model was estimated with data for white male 1972 high school graduates (N = 2955) with the specification that the response errors were uncorrelated. The resulting $\chi^2 = 150.75$, with df = 38, indicated that the model did not do a very good job of reproducing the observed variance-covariance matrix. Examination of the first-order derivatives indicated the possibility that the specification of uncorrelated response errors may have been untenable. Specifically, the response errors of x_2 and x_4 may be correlated. These variables are the composite measures of father's and mother's education, and apparently systematic errors of construction exist in both variables. Re-estimating the model allowing for correlated response errors between x_2 and x_4 resulted in a $\chi^2 = 95.00$, with df = 37. Because the difference in these chi-squares is itself distributed according to chi-square with one degree of freedom, it is obvious that the correlated response error was statistically significant. Yet once again the model does not do a very good job of reproducing the variance-covariance matrix. Re-examination of the first-order derivatives suggested that y_4 and y_6 had correlated response errors. Re-estimation yielded a $\chi^2 = 73.85$, with df = 36, which became the final model because the addition of the next most likely correlated response error did not significantly reduce the value of χ^2 (see Sörbom, 1975).

Identical models were also estimated for black males (N = 257) with $\chi^2 = 43.45$, with df = 36, and for Hispanic males (N = 125) with $\chi^2 = 51.59$,

with df = 36. Estimates for the measurement model are shown in Table 3 for white males, Table 4 for black males, and Table 5 for Hispanic males. Shown in column 3 of these tables are the standard deviations of manifest variables; column 4 contains the standard deviations of response errors not accounted for by the underlying latent variables; column 5 shows the standard deviations of the latent variables; column 6 contains the relative slopes of the manifest variables regressed on the latent variables; and column 7 shows estimates of the reliability coefficients.

Among white males in the NLS sample, different reports of the same underlying variable were likely to have different slope coefficients. For some variables, such as curriculum track, these different slopes indicate different fits between the manifest and latent variables. For example, among whites the school record measure of curriculum track was a more reliable indicator of the true variable than was the student's own report. For other variables, different slopes reflect different scales of the manifest variables. For example, the composite measures of parental education were scaled from 1 (less than high school) to 5 (MA, or PhD), while the first followup questions of parental education were scaled from 1 (none, or grade school only) to 9 (PhD or equivalent). However, the reliability coefficients for these variables indicate that the composite measures of parental education, which were based on responses to baseyear, first followup, and activity state questionnaires, were more reliable measures of the underlying latent variables than were the first followup questions alone. However, caution should be exercised in generalizing from these preliminary results. The fact that the two measures of parental education differ in their scales of measurement may

Table 3. Measurement Model Parameter Estimates for White Male 1972 High School Graduates (N = 2955).

Variable		Observed SD	SD of Error	SD of True Score	Relative Slope	Reliability Coefficient
True (τ_j)	Observed	σ_i	σ_{e_i}	σ_{τ_j}	λ_{ij}	$(\sigma^2_{\tau_j}/\sigma^2_i)\lambda_{ij}$
FAEDUC ξ_2	X_2(V1627)	1.28	.44 (.11)	1.20	1.00	.88
	X_3(V1009)	2.22	.49 (.19)		1.81 (.02)	.53
MOEDUC ξ_3	X_4(V1628)	1.01	.46 (.11)	.90	1.00	.79
	X_5(V1010)	1.73	.34 (.20)		1.87 (.03)	.51
CURRICULUM η_1	Y_1(V209)	.50	.33 (.06)	.41	.91 (.02)	.61
	Y_2(V196)	.50	.28 (.06)		1.00	.67
EDUCATION η_3	Y_4(V1855)	2.30	1.20 (.25)	1.79	1.10 (.02)	.67
	Y_5(V1854)	1.97	.82 (.20)		1.00	.83
	Y_6(recode)	.70	.29 (.07)		.36 (.01)	* * *

Note. -- Standard errors of parameter estimates appear in parentheses.

Table 4. Measurement Model Parameter Estimates for Black Male 1972 High School Graduates (N = 257)

Variable		Observed SD	SD of Error	SD of True Score	Relative Slope	Reliability Coefficient
True (τ_j)	Observed	σ_i	σ_{e_i}	σ_{τ_j}	λ_{ij}	$(\sigma^2_{\tau_j}/\sigma^2_i)\lambda_{ij}$
FAEDUC ξ_2	X_2(V1627)	.94	.48 (.19)	.82	1.00	.76
	X_3(V1009)	1.67	.13 (.35)		2.04 (.12)	.49
MOEDUC ξ_3	X_4(V1628)	.97	.67 (.23)	.71	1.00	.54
	X_5(V1010)	1.63	.37 (.45)		2.34 (.22)	.44
CURRICULUM η_1	Y_1(V209)	.47	.36 (.12)	.38	.79 (.10)	.52
	Y_2(V196)	.46	.26 (.13)		1.00	.68
EDUCATION η_3	Y_4(V1855)	2.16	1.16 (.51)	1.58	1.15 (.09)	.62
	Y_5(V1854)	1.88	1.01 (.41)		1.00	.71
	Y_6(recode)	.64	.19 (.15)		.38 (.03)	* * *

Note. -- Standard errors of parameter estimates appear in parentheses.

Table 5. Measurement Model Parameter Estimates for Hispanic Male 1972 High School Graduates (N = 125)

Variable		Observed SD	SD of Error	SD of True Score	Relative Slope	Reliability Coefficient
True (τ_j)	Observed	σ_i	σ_{e_i}	σ_{τ_j}	λ_{ij}	$(\sigma^2_{\tau_j}/\sigma^2_i)\lambda_{ij}$
FAEDUC ξ_2	X_2(V1627)	1.08	.53 (.25)	.93	1.00	.74
	X_3(V1009)	2.05	.87 (.47)		1.99 (.17)	.41
MOEDUC ξ_3	X_4(V1628)	.82	.34 (.13)	.75	1.00	.84
	X_5(V1010)	1.51	.33 (.32)		1.98 (.13)	.49
CURRICULUM η_1	Y_1(V209)	.41	.33 (.13)	.32	.76 (.15)	.46
	Y_2(V196)	.46	.33 (.15)		1.00	.48
EDUCATION η_3	Y_4(V1855)	2.19	1.23 (.65)	1.36	1.33 (.18)	.51
	Y_5(V1854)	1.77	1.12 (.49)		1.00	.59
	Y_6(recode)	.54	.32 (.16)		.32 (.04)	* * *

Note. -- Standard errors of parameter estimates appear in parentheses.

indicate that the two observed variables are not manifestations of the same true score.

Caution is particularly recommended in interpreting the results of the measurement of respondent's education. The preliminary model estimated here included two measures of education measured on the same scale: educational plans as of 10/1/76 and educational attainment as of 10/1/76. A third composite measure was included, one suggested by the staff of NCES in which 1 = no higher education, 2 = some college, and 3 = BA and higher. The relative slope of this composite measure varies so much from the other two manifest variable regression slopes, and its error variance is so different, that it should probably not be viewed as a manifest component of the same latent factor that underlies the other two education variables. One result of the mismatch is a meaningless reliability estimate greater than unity.

One further caution of interpretation is worth noting. Classical true score models express an observed score in terms of two orthogonal components composed of a true score and an error score. As a result, errors based on true score models are uncorrelated with true scores and other error scores. However, the vectors of residual factors, $\underset{\sim}{\delta}$ and $\underset{\sim}{\epsilon}$, contain both measurement error and reliable variation specific to each manifest variable (Alwin and Jackson, forthcoming). As a result, it is possible for some or all of the residual errors to be correlated even in the population, much as we have seen that the errors of x_2 and x_4, the two composite measures of parental education, were correlated. Apparently, whatever errors of measurement entered into the construction of one parent's education composite score also entered into the other parent's.

Comparison of the measurement-model results for white males to those of blacks and Hispanics indicates that within each population the most reliable measure of parental education was the constructed composite variable. The most reliable measure of curriculum membership was the school record. The most reliable measure of true education was the respondent's report of his educational attainment as of 10/1/76. Across populations, the reliability coefficients for blacks and Hispanics were lower than those for whites. Both blacks and Hispanics exhibited less variation in the observed measures than did whites. Blacks and Hispanics also exhibited less variation in the latent factor scores; proportionately, there were even greater disparities among the latent variances than among the observed. As a result, the reliability coefficients for blacks and Hispanics were substantially lower than those of whites.

Clearly these findings suggest caution in interpreting models of status attainment among minority groups that do not take account of response error, especially when comparing structural coefficients across groups. Table 6 presents ordinary least square (in parentheses) and corrected LISREL estimates for the structural equation portion of the model represented by Figure 2. Comparison of these estimates provides some indication of the biases encountered when measurement errors are ignored. (Another example has been offered by Bielby, Hauser, and Featherman, 1977.)

First, the ordinary least squares regression of educational attainment on four family background variables and two intervening measures of high school effects accounts for two-fifths of the variance in educational attainment for white males, but only one-fourth of the

Table 6. Corrected (LISREL) and OLS Estimates of Parameters of the Educational Attainment Process for 1972 High School Graduates

Dependent Variable	Predetermined Variables*						R^2
	FaSEI	FaEduc	MoEduc	NumSibs	Curriculum	H.S. Rank	
Whites							
Curriculum	.002	.084	.052	-.025	. . .	. . .	.165
	(.002)	(.068)	(.049)	(-.026)	. . .	. . .	(.109)
H.S. Rank	-.032	.587	-.362	-.160	37.787	. . .	.325
	(.004)	(1.303)	(-.038)	(-.482)	(25.041)	. . .	(.232)
Education	.002	.194	.095	-.063	2.271	.012	.558
	(.003)	(.237)	(.092)	(-.071)	(1.326)	(.020)	(.400)
Blacks							
Curriculum	.003	.063	-.008	-.021	. . .	. . .	.086
	(.002)	(.056)	(.004)	(-.018)	. . .	. . .	(.050)
H.S. Rank	-.074	-2.814	2.952	.723	34.045	. . .	.237
	(-.032)	(-2.306)	(3.115)	(.552)	(24.198)	. . .	(.193)
Education	.012	.010	.344	-.021	1.825	.009	.411
	(.015)	(.013)	(.162)	(-.043)	(1.068)	(.016)	(.249)

Table 6. (continued)

Dependent Variable	Predetermined Variables*						R^2
	FaSEI	FaEduc	MoEduc	NumSibs	Curriculum	H.S. Rank	
Hispanics							
Curriculum	.004	-.000	-.021	.002	. . .	. . .	.062
	(.003)	(-.018)	(.041)	(.006)	. . .	. . .	(.025)
H.S. Rank	.014	-.212	2.996	1.301	50.983	. . .	.407
	(.147)	(.341)	(.172)	(1.227)	(26.166)	. . .	(.255)
Education	-.012	-.053	.168	-.072	2.770	.010	.582
	(-.006)	(-.096)	(.018)	(-.110)	(.558)	(.028)	(.249)

Note. -- The ordinary least squares estimates appear in parentheses below the LISREL estimates. The variables used in the OLS regressions were Father's SEI (X_1: V2468), Father's Education Composite (X_2: V1627), Mother's Education Composite (X_4: V1628), Number of Siblings (X_6: V1460 - V1463), Curriculum {= 1 if Academic; = 0 otherwise} (Y_2: V196), High School Percentile Rank (Y_3: V631), and Educational Attainment as of 10/1/76 (Y_5: V1854).

variance for both blacks and Hispanics. However, these results confound measurement error with true variation and result in coefficients of determination that understate the ratio of explained to total true variance by nearly 30 percent for whites, but by nearly 40 percent for blacks and 60 percent for Hispanics. By taking measurement errors into account, over half the variance in true educational attainment can be explained for Hispanics (R^2 = .582) and for whites (R^2 = .558), but not for blacks (R^2 = .411).

Comparison of the OLS to corrected estimates in the regression of educational attainment again indicates the biases due to ignoring measurement errors. In particular, notice that ignoring measurement errors does not necessarily produce attenuated estimates. Indeed, some of the OLS estimates are substantially larger than the corrected estimates. For all three groups, corrections for measurement error produce nearly identical effects, at least in the direction of the bias. There appear to be downward biases (the OLS estimates underestimate the corrected estimates) for mother's education and curriculum track. There appear to be upward biases for father's education, the number of siblings, and high school class ranking. There is also an upward bias for father's socioeconomic index among whites and blacks, but for Hispanics the OLS estimate for father's SEI understates the corrected estimate.

Although the direction of bias is nearly uniform across ethnic groups, the magnitude is not. One example is provided by the effect of membership in an academic track on educational attainment. Ordinary least squares regressions would indicate that membership in an academic track results in about one and one-third additional years of education

for whites (measured four years after high school graduation), and one additional year for blacks, but only one-half year for Hispanics. However, when the confounding effects of measurement errors are removed, Hispanics are estimated to convert membership in a high school academic track into 2.8 additional years of education, while whites have a comparable estimate of 2.3 years, and blacks 1.8 years.

Another example is provided by the effect of high school rank on educational attainment. Ranking the magnitude of the OLS estimates would lead one to conclude that Hispanics were best able to convert increased high school class ranking into educational attainment, followed by whites, then blacks. However, when corrected for measurement errors, all three groups were apparently equally able to convert class ranking into educational attainment.

Overall, the consequences of ignoring measurement error appear to be greater in the case of Hispanics than of either whites or blacks, and greater for blacks than for whites. Since the biases in structural estimates ignoring measurement error are larger among Hispanics and blacks than among whites, uncorrected ethnic comparisons show unrealistically large differences between ethnic groups in the effects of familial background and high school process effects.

CONCLUSIONS

Educational researchers have long known that ignoring measurement errors will lead to biased estimates of structural effects. However, until recently multivariate analytic procedures which correct for measurement errors were not generally available. Recent developments by Jöreskog

and Sörbom (1978) have made available a general computer program that permits estimation of structural effects corrected for measurement errors. The application of these techniques to a substantive problem in education has indicated the advantages of the LISREL approach, along with several cautionary reservations.

The most important substantive conclusion inherent in this analysis supports the findings of Bielby, Hauser and Featherman (1977): measurement errors differ between blacks and whites; ignoring them leads to biased estimates of structural effects. Moreover, the present analysis shows that Hispanics also report data with inherent measurement errors, and ignoring them will lead to estimates even more biased than among either whites or blacks.

Another set of substantive conclusions could be drawn from the estimated parameters of the Hispanic model of educational attainment. To the best of my knowledge, these are the first estimates, unbiased by measurement error, of the process of status attainment for any ethnic group in America other than whites or blacks. However, I have refrained from discussing Hispanics because the model explicated in this paper was a preliminary construction, and is already outmoded. In particular, the model omits measures of ability, and as Scarr and Weinberg (1978) demonstrated, the omission of ability leads to spurious estimates of causal effects.

Finally, a cautionary note is in order. Kerlinger (1977) correctly pointed out that the LISREL approach toward multivariate analysis contains a great deal of promise for testing theories that have been difficult to test adequately with previously available analytic procedures.

However, application of LISREL (indeed, any analytic procedure) depends upon the collection of appropriate data. Specifically, measures of different variables must be ascertained on different occasions, or by different means, data collection procedures that can be considerably more expensive than the usual survey.

My own view of the utility of LISREL is more skeptical than Kerlinger's (1977). Unless the data to be analyzed have been collected by appropriate procedures, and unless the model is adequately specified, LISREL is unlikely to produce the definitive tests Kerlinger suggests are possible. The past decade has seen recursive path analytic procedures faddishly applied to implausibly constructed models. Except for the inherent difficulties in specifying the model of the computer program, the next decade may see implausible examples of substantive analyses based on LISREL. In the past six months I have twice had manuscripts returned to me with reviewer's naive suggestions that the problems they recognized could be solved by reanalyzing the data with LISREL. They could not. As Cooley (1978: 13) so insightfully pointed out last year, more important than number crunching is the careful measurement of a few "right" variables, variables that permit statistical controls for major alternative explanations. Data analysis may stimulate thinking, but it is not a substitute for it.

APPENDIX

The three equations which define the structural model are:

$$x_2 = b_{2\eta}\eta + e \quad (1),$$

$$\eta = b_{\eta 3}x_3 + u \quad (2),$$

$$x_1 = b_{1\eta}\eta + v \quad (3).$$

The notation may be revised such that x_1, x_2, x_3, and η refer to the standardized values of these variables. Equations (1) through (3) may be rewritten using the using the usual equalities:

$$p_{yx} = b_{yx}(\sigma_x/\sigma_y) \quad (4),$$

$$p_{yu} = \sigma_u/\sigma_y \quad (5).$$

These coefficients were termed path coefficients by Sewell Wright (1921). Rewriting equations (1) through (3) in terms of path coefficients and standardized variables yields:

$$x_2 = p_{2\eta}\eta + p_{2e}e \quad (6),$$

$$\eta = p_{\eta 3}x_3 + p_{\eta u}u \quad (7),$$

$$x_1 = p_{1\eta}\eta + p_{1v}v \quad (8),$$

with the specifications

$$E(x_3u) = E(\eta v) = E(x_3v) = 0 \quad (9),$$

$$E(\eta e) = E(x_1e) = E(x_3e) = 0 \quad (10).$$

To solve the path coefficients in equations (6) through (8), we will multiply through these equations by one or another of the variables, and be taking expectations. Because the covariance of two standardized variables is the coefficient of correlation, taking expectations of a covariance will yield the population correlation coefficient, ρ.

First, multiply equation (6) through by η, and we have:

$$\rho_{2\eta} = p_{2\eta} \quad (11),$$

since $E(\eta\eta) = 1$, and $E(\eta e) = 0$ by assumption. In similar fashion, multiplying equation (7) through by x_3, and equation (8) through by η, yields:

$$\rho_{\eta 3} = p_{\eta 3} \quad (12),$$

$$\rho_{1\eta} = p_{1\eta} \quad (13).$$

Multiplying equation (6) through by x_3, and taking expectations, yields:

$$\rho_{23} = p_{2\eta}p_{\eta 3} \quad (14),$$

because $E(x_3 e) = 0$, by assumption, and $p_{\eta 3} = \rho_{\eta 3}$ by equation (12). Multiplying equation (8) through by x_3, and taking expectations, yields:

$$\rho_{13} = p_{1\eta}p_{\eta 3} \quad (15),$$

and multiplying equation (8) through by x_2 yields:

$$\rho_{12} = p_{1\eta}p_{2\eta} \quad (16),$$

because $E(x_2 v) = 0$, an equality implied by equations (9) and (10).

Equations (14), (15), and (16) form three equations in three unknowns, and:

$$p_{1\eta} = \sqrt{\rho_{12}\rho_{13}/\rho_{23}} \quad (17),$$

$$p_{2\eta} = \sqrt{\rho_{12}\rho_{23}/\rho_{13}} \quad (18),$$

$$p_{\eta 3} = \sqrt{\rho_{13}\rho_{23}/\rho_{12}} \quad (19).$$

Using the sample correlation coefficients given in Table 1 to estimate the population coefficients in equations (17) through (19) give the standardized results presented in Table 2.

The corrected regression coefficient $b_{\eta 3}$ in equation (2), as implied by equation (4), is given by:

$$b_{\eta 3} = p_{\eta 3}(\sigma_{\eta}/\sigma_3) \qquad (20),$$

in which

$$\sigma_{\eta} = p_{2\eta}\sigma_2 \qquad (21),$$

the usual association between the true and observed standard deviations (for example, Gulliksen, 1950: 23). Note therefore that:

$$b_{\eta 3} = p_{\eta 3}(p_{2\eta}\sigma_2/\sigma_3) \qquad (22),$$

but $p_{\eta 3}p_{2\eta} = \rho_{23} = p_{23}$, so that

$$b_{\eta 3} = p_{23}(\sigma_2/\sigma_3) \qquad (23).$$

That is, the corrected regression coefficient of true education regressed on father's SEI is equal to the OLS regression coefficient one would obtain from regressing observed education on father's SEI.

The corrected regression coefficient $b_{2\eta}$ in equation (1) is given by:

$$b_{2\eta} = p_{2\eta}(\sigma_2/\sigma_{\eta}) \qquad (24),$$

but substitution by equation (21) reveals that

$$b_{2\eta} = 1.00 \qquad (25).$$

Finally, let's consider the corrected regression coefficient $b_{1\eta}$ in equation (3). From equation (25), equation (1) may be rewritten $x_2 = \eta + e$, and by substitution, equation (3) becomes:

$$x_1 = b_{1\eta}(x_2 - e) + v \qquad (26),$$

$$x_1 = b_{1\eta}x_2 - b_{1\eta}e + v \qquad (27).$$

Multiplying equation (27) through by x_2 and taking expectations yields:

$$\sigma_{12} = b_{1\eta}\sigma_2^2 - b_{1\eta}\sigma_{2e} + \sigma_{2v} \qquad (28),$$

but $\sigma_{2v} = 0$ as above, and $\sigma_{2e} = \sigma_e^2$, which may be verified by multiplying

equation (1) by e and taking expectations. Thus,

$$\sigma_{12} = b_{1\eta} (\sigma_2^2 - \sigma_e^2) \quad (29);$$

$$b_{1\eta} = \sigma_{12}/(\sigma_2^2 - \sigma_e^2) = \sigma_{12}\sigma_2^2/(\sigma_2^2\{\sigma_2^2 - \sigma_e^2\}) \quad (30),$$

and

$$b_{1\eta} = b_{12} (\sigma_2^2/\{\sigma_2^2 - \sigma_e^2\}) \quad (31).$$

That is, the corrected regression coefficient of respondent's SEI on true education is equal to the OLS estimate only when the error of measurement variance is zero. Otherwise, the greater the variance of errors, the greater will be the downward bias in b_{12}. In this example, because the variance of errors was greater for blacks than for whites, the OLS estimates were more biased for blacks.

REFERENCES

Alwin, Duane F., and David J. Jackson
forthcoming "Measurement error models for survey data: Issues and applications." K. F. Schuessler (ed.), Sociological Methodology 1980. San Francisco: Jossey-Bass.

Anderson, James G.
1978 "Causal models in educational research: Nonrecursive models." American Educational Research Journal 15: 81-97.

Anderson, James G., and Evans, F. B.
1974 "Causal models in educational research: Recursive models." American Educational Research Journal 11: 29-39.

Bielby, William T., Robert M. Hauser, and David L. Featherman
1977 "Response errors of black and nonblack males in models of the intergenerational transmission of socioeconomic status." American Journal of Sociology 82: 1242-1288.

Blalock, Hubert M.
1964 Causal Inferences in Nonexperimental Research. Chapel Hill: University of North Carolina Press.

1965 "Some implications of random measurement error for causal inferences." American Journal of Sociology 71: 37-47.

Blalock, Hubert M., Caryll S. Wells, and Lewis F. Carter
1970 "Statistical estimation with random measurement error." Pp. 75-103 in Edgar F. Borgatta and George W. Bohrnstedt (eds.), Sociological Methodology 1970. San Francisco: Jossey-Bass.

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

Bohrnstedt, George W.
1969 "Observations on the measurement of change." Pp. 113-133 in Edgar F. Borgatta and George W. Bohrnstedt (eds.), Sociological Methodology 1969. San Francisco: Jossey-Bass.

Bohrnstedt, George W., and T. Michael Carter
1971 "Robustness in regression analysis." Pp. 118-146 in Herbert L. Costner (ed.), Sociological Methodology 1971. San Francisco: Jossey-Bass.

Cooley, William W.
1978 "Explanatory observational studies." Educational Researcher 7 (October): 9-15.

Duncan, Otis Dudley
1961 "A socioeconomic index for all occupations." Pp. 109-138 in Albert J. Reiss, Jr., Occupations and Social Status. New York: The Free Press of Glencoe.

1969a "Inheritance of poverty of inheritance of race?" Pp. 85-110 in D. P. Moynihan (ed.), On Understanding Poverty. New York: Basic Books.

1969b "Some linear models for two-wave, two-variable panel analysis." Psychological Bulletin 72: 177-182.

Duncan, Otis Dudley, David L. Featherman, and Beverly Duncan
1972 Socioeconomic Background and Achievement. New York: Seminar Press.

Duncan, Otis Dudley, A. O. Haller, and A. Portes
1968 "Peer influences on aspirations: A reinterpretation." American Journal of Sociology 74: 119-137.

Gulliksen, Harold
1950 Theory of Mental Tests. New York: John Wiley.

Hauser, Robert M.
1969 "Schools and the stratification process." American Journal of Sociology 74: 587-611.

1973 "Disaggregating a social-psychological model of educational attainment." Pp. 255-284 in Arthur S. Goldberger and Otis Dudley Duncan (eds.), Structural Equation Models in the Social Sciences. New York: Seminar Press.

Hauser, Robert M., and David L. Featherman
1977 The Process of Stratification: Trends and Analyses. New York: Academic Press.

Hauser, Robert M., and Arthur S. Goldberger
1971 "The treatment of unobservable variables in path analysis." Pp. 81-117 in Herbert L. Costner (ed.), Sociological Methodology 1971. San Francisco: Jossey-Bass.

Hodge, Robert W., and Donald J. Treiman
1968 "Social participation and social status." American Sociological Review 33: 723-740.

Jöreskog, Karl G.
1973 "A general method for estimating a linear structural equation system." Pp. 85-112 in Arthur S. Goldberger and Otis Dudley Duncan (eds.) Structural Equation Models in the Social Sciences. New York: Seminar Press.

Jöreskog, Karl G.
1977 "Structural equation models in the social sciences: Specification, estimation and testing." Pp. 265-287 in P. R. Krishnaiah (ed.), Application of Statistics. Amsterdam: North-Holland Publishing Co.

Jöreskog, Karl G., and Dag Sörbom
1978 Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Chicago: National Educational Resources, Inc.

Kerlinger, Fred N.
1977 "The influence of research on education practice." Educational Researcher 6 (September): 5-12.

Kim, J., and Charles W. Mueller
1976 "Standardized and unstandardized coefficients in causal analysis: An expository note." Sociological Methods and Research 4: 423-438.

Land, Kenneth C.
1970 "On the estimation of path coefficients for unmeasured variables from correlations among observed variables." Social Forces 48: 506-511.

Levinsohn, Jay R., Louise B. Henderson, John A. Riccobono, and R. Paul Moore
1978 National Longitudinal Study Base Year, First, Second and Third Follow-Up Data File Users Manual, Volumes I and II. Wahsington, D.C.: National Center for Education Statistics.

Lichtman, Marilyn, and Lee M. Wolfle
1978 Analysis of the NLS Data: Processes of Educational and Status Attainment in Distinctive Subgroups: A Proposal in Response to RFP 78-76, Educational and Occupational Development of American Youths: Analyses of the National Longitudinal Study of the High School Class of 1972. Washington, D.C. National Center for Education Statistics.

Long, J. Scott
1976 "Estimation and hypothesis testing in linear models containing measurement error: A review of Joreskog's model for the analysis of covariance structures." Sociological Methods and Research 5: 157-206.

National Opinion Research Center
1977 Codebook for the Spring 1977 General Social Survey. Chicago: National Opinion Research Center.

Otto, Luther B., and David L. Featherman
1975 "Social structural and psychological antecedents of self-estrangement and powerlessness." American Sociological Review 40: 701-719.

Scarr, Sandra, and Richard A. Weinberg
1978 "The influence of 'family background' on intellectual attainment." American Sociological Review 43: 674-692.

Schuman, Howard, and Stanley Presser
1978 "The assessment of 'no opinion' in attitude surveys." Pp. 241-275 in Karl F. Schuessler (ed.), Sociological Methodology 1979. San Francisco: Jossey-Bass.

Sewell, William H., and Robert M. Hauser
1975 Education, Occupation, and Earnings: Achievement in the Early Career. New York: Academic Press.

Siegel, Paul M., and Robert W. Hodge
1968 "A causal approach to the study of measurement error." Pp. 28-59 in H. M. Blalock and A. B. Blalock (eds.), Methodology in Social Research. New York: McGraw-Hill.

Sörbom, Dag
1975 "Detection of correlated errors in longitudinal data." The British Journal of Mathematical and Statistical Psychology 28: 138-151.

Wiley, David E.
1973 "The identification problem for structural equation models with unmeasured variables." Pp. 69-83 in Arthur S. Goldberger and Otis Dudley Duncan (eds.), Structural Equation Models in the Social Sciences. New York: Seminar Press.

Wiley, David E., and James A. Wiley
1970 "The estimation of measurement error in panel data." American Sociological Review 35: 112-117.

Williams, John D.
1978 "Path analysis from a regression perspective." Multiple Linear Regression Viewpoints, Monograph Series No. 3, Volume 9.

Wolfle, Lee M.
1977 "An introduction to path analysis." Multiple Linear Regression Viewpoints 8: 36-61.

1978 "Strategies of path analysis." A paper presented to the American Educational Research Association, Toronto, March, 1978.

Wright, Sewell
1921 "Correlation and causation." Journal of Agricultural Research 20: 557-585.

1925 Corn and Hog Correlations. Washington, D.C.: U.S. Department of Agriculture Bulletin 1300, January 1925.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 9 NO. 5 1979

A MONTE CARLO EVALUATION OF ESTIMATED PARAMETERS OF FIVE SHRINKAGE ESTIMATE FORMULI

Isadore Newman
Gayle A. Seymour
Thomas K. Garver
The University of Akron

ABSTRACT

This study employs a Monte Carlo simulation to determine the accuracy with which the shrinkage in R^2 can be estimated by five shrinkage formuli and cross-validation. The study dealt with the use of shrinkage and cross-validation for different sample sizes, different R^2 values, and different degrees of multi-collinearity.

The present study was designed to determine the relative effects of four different levels of multicollinearity and population R^2's on six different shrinkage estimates. Two of these estimates are used predominately to estimate the population R^2 from the sample (McNemar (M) 1949,1962; and Wherry/McNemar (W/M) 1931). The four others are traditionally used to estimate shrinkage from sample to sample (Lord1 (L_1); Lord2 (L_2), Darlington (D) and Cross-Validation (CV_2). In addition this paper will look at how well the sample to sample shrinkage estimates predict the true population R^2.

This paper was presented at the annual meeting of the American Educational Research Association; April 1979, San Francisco, Ca.

It was assumed that cross-validation would be the least effected by multicollinearity, however, the study did not support this hypothesis. Apparently, multicollinearity had very little effect on each of the shrinkage estimates.

Applied researchers who use multiple regression frequently are interested in:

1) estimating the relationship between a set of predictive variables to a criterion in a sample for the purpose of generalizing to the population

2) estimating the magnitude of relationships from the sample to the population

3) estimating accuracy of prediction from sample to sample and then to the population

In both estimating the magnitude of relationships and the accuracy of prediction, shrinkage estimates are both utilitarian and potentially misleading. Since there are a variety of shrinkage estimating procedures and formuli it is important to understand the differential effects of the shrinkage estimates for different situations and try to identify which is most appropriate under different conditions.

In the classical article by Charles Mosier (1951) he defines six purposes for doing cross validations. Similarly he defines six different types of cross validations.

1) Cross validation using weights determined in one sample applied in a second sample drawn from the same population to determine the differences in R^2 between the two samples (this type of shrinkage is the one that will be used in this study)

2) Validity generalization is similar to the first but the samples come from two distinctly different populations

3) Validity extention in which the samples come from different populations and the two samples criteria are also different

4) Simultaneous validation in which the purpose is to select a set of items (batteries) effectiveness in two different samples from different populations for the purpose of having the single best most useful battery for both populations

5) Replication is a method of determining the best set of weights by drawing a variety of samples from the same population and determining the best combination of weights by using information from these several samples

6) Double-cross validation in which two samples are drawn simultaneously and the weights calculated for each sample are then applied to the other sample

The major problem with cross-validation is that when one estimates shrinkage one tends to estimate the weights upon half of the available data when it is well known that the stability of the weight increases as the ratio of the number of subjects increases in relationship to the number of variables (producing less shrinkage). A suggested path around this dilemma has been to use double-cross-validation. If the weights in each sample are applied to the other sample (weights of Sample 1 to Sample 2 and vice versa) without too much shrinkage (subjective decision) then one may combine the samples and calculate the final weights on the total sample. Another suggestion is to look at which weights (Sample 1 when applied to Sample 2 or Sample 2 when applied to Sample 1) produces the least amount of shrinkage and then use those weights.

The assumption made by Mosier is that the best way to estimate

the accuracy of the weights in one sample to predict the other sample is by using separate samples. He bases this argument on the work done by Cureton (1950) and Kurtz (1948). This assumption is questioned by Dalington (1969), Lord (1950), Schmitt, Coyle and Rauschenberger (1977), and Claudy (1978). If a mathematical correction can be used based upon the entire sample that can predict the shrinkage in R^2 from sample to sample accurately, it would be preferable to use cross-validation procedure which requires two samples. It would also produce more stable weights since it would be done on the entire sample instead of just half the sample. This mathematical correction procedure has been criticized because of some simple misunderstandings.

As stated by others (Newman, 1973; Uhl, 1970, Lord, 1950, Darlington, 1968, etc.), there are a variety of mathematical shrinkage estimate corrections, which we shall call McNemar, Wherry-McNemar, Lord 1, Darlington, and Lord 2.

$$\hat{R}^2 = 1 - (1-R^2)\frac{N-1}{N-K} \quad \text{(McNemar 1949)}$$

$$\hat{R}^2 = 1 - (1-R^2)\frac{N-1}{N-K-1} \quad \text{(Wherry/McNemar (W/M) 1931)}$$

$$\hat{R}^2 = 1 - (1-R^2)\frac{N+K+1}{N-K-1} \quad \text{(Lord 1 } (L_1)\text{)}$$

where:

R^2 = the corrected estimate of the multiple correlation.

R = the actual calculated multiple correlation.

K = the number of independent variables.

N = the number of linearly independent observations.

Also, Schmitt and Rauschenberger (1977) presented several shrinkage estimation formuli which will be included in the analysis. These were:

$$\rho c^2 = 1 - \frac{N-1}{N-p-1}\ \frac{N-2}{N-p-2} \times \frac{N+1}{N}\ (1-R^2) \qquad \text{(Darlington (D))}$$

$$\rho c^2 = 1 - \frac{N-1}{N-p-1}\ \frac{N+p+1}{N}\ (1-R^2) \qquad \text{(Lord 2 } (L_2) \text{)}$$

where:

ρc^2 = the correlated estimate of the multiple correlation.

R = the actual calculated multiple correlation.

N = the number of linearly independent observations.

P = the number of independent variables.

The McNemar and Wherry-McNemar's shrinkage estimate was developed for and is only appropriate to use when one is interested in predicting a population value from a sample. The others, Lord 1, Lord 2, and Darlington, were derived for the purpose of estimating shrinkage from sample to sample, that is the stability and accuracy of the weights from one sample predicting to another.

A Monte Carlo study was run to estimate the efficiency and relationship between these five equations and cross-validation in predicting known population values for differing population R^2's, differing levels of correlations between four predictor variables (degree of multicollinearity) and for four different N sizes (14, 30, 50, and 100). To summarize, the Monte Carlo design will look at six estimates of shrinkage (McNemar, Wherry-McNemar, Lord 1, Lord 2, Darlington, and Cross-Validation) for three different ranges of know population R^2's (Situation 1's R^2's = .06, .07, .06,

.08; Situation 2 R^2's = .31, .32, .33, .34, and Situation 3 R^2's = .45, .47, .46, and .55). For each of the three Situations, there were four differing correlations between the four predictor variables (estimates of multicollinearity ($\bar{r}$)). These were for Situation 2, the correlations ($\bar{r}$) between the four predictor variables were .13, .30, .52, .80; and for Situation 3, the correlations ($\bar{r}$) were .15, .29, .53, .82. For each of the twelve populations (that is a particular population R^2 for a particular level of multicollinearity) there are four different N sizes (14, 30, 50, and 100). Therefore, the Monte Carlo Study is a 3x4x4 design, a range of three sets of population values, a range of four different multicollinearity conditions for four different N sizes. (see Table #1 for results) The data is Table #1 is based upon 100 random samples picked for each situation from each of its 12 specific population. The r for multicollinearity is the average for the 100 samples. In total there were 4,800 samples drawn (3x4x4x100 (samples)).

TABLE OF SHRINKAGE ESTIMATES

Situation I

Population 1		N	R^2_S	R^2_M	$R^2_{W/M}$	$R^2_{L_1}$	R^2_D	$R^2_{L_2}$	R^2_{CV1}	R^2_{CV2}
R^2 = .06	A	14	.33	.13	.04	-.41	-.43	-.31	.69	.15
	B	30	.17	.08	.04	-.16	-.13	-.12	.32	.08
$\bar{r}$ = .03	C	50	.14	.08	.06	-.05	-.05	-.03	.22	.06
	D	100	.09	.06	.05	-.00	.00	.01	.13	.04

Population 2										
R^2 = .07	A	14	.37	.19	.10	-.32	-.50	-.23	.68	.17
	B	30	.19	.10	.06	-.14	-.14	-.10	.32	.09
$\bar{r}$ = .29	C	50	.13	.08	.05	-.06	-.06	-.04	.21	.07
	D	100	.10	.07	.06	.00	.00	.01	.13	.05

Population 3										
R^2 = .06	A	14	.34	.15	.05	-.38	-.48	-.29	.70	.17
	B	30	.19	.10	.06	-.13	-.13	-.09	.31	.08
$\bar{r}$ = .54	C	50	.13	.08	.05	-.06	-.06	-.04	.22	.06
	D	100	.10	.07	.06	.01	.01	.02	.13	.05

Population 4										
R^2 = .08	A	14	.35	.16	.07	-.36	-.51	-.27	.69	.20
	B	30	.20	.11	.07	-.12	-.12	-.08	.36	.08
$\bar{r}$ = .82	C	50	.13	.07	.05	-.06	-.06	-.04	.23	.06
	D	100	.11	.08	.07	.02	.02	.03	.15	.06

Where: R^2_S = The R^2 for the sample

R^2_M = McNemar's shrinkage estimate

$R^2_{W/M}$ = Wherry/McNemar's shrinkage estimate

$R^2_{L_1}$ = Lord 1 shrinkage estimate

continued on next page

Situation II

Population 5		N	R^2_S	R^2_M	$R^2_{W/M}$	$R^2_{L_1}$	R^2_D	$R^2_{L_2}$	R^2_{CV1}	R^2_{CV2}
R^2 = .34	A	14	.50	.35	.28	-.05	-.17	.02	.76	.21
	B	30	.40	.33	.30	.16	.16	.18	.48	.24
$\bar{r}$ = .13	C	50	.39	.35	.34	.26	.26	.27	.43	.29
	D	100	.37	.35	.34	.30	.30	.31	.40	.30

Population 6										
R^2 = .32	A	14	.51	.36	.29	-.04	-.15	.04	.74	.26
	B	30	.41	.34	.32	.18	.18	.20	.51	.21
$\bar{r}$ = .30	C	50	.35	.31	.29	.20	.20	.22	.42	.24
	D	100	.34	.32	.31	.27	.27	.28	.37	.30

Population 7										
R^2 = .31	A	14	.50	.35	.28	-.05	-.12	.02	.76	.23
	B	30	.38	.31	.28	.13	.13	.16	.45	.26
$\bar{r}$ = .32	C	50	.34	.30	.29	.20	.20	.21	.39	.24
	D	100	.33	.31	.30	.26	.26	.26	.36	.27

Population 8										
R^2 = .33	A	14	.50	.34	.27	-.07	-.15	.01	.71	.25
	B	30	.41	.34	.31	.17	.17	.20	.48	.24
$\bar{r}$ = .80	C	50	.36	.32	.30	.22	.22	.23	.42	.25
	D	100	.36	.34	.33	.29	.29	.30	.37	.30

Continued from Table 1:

R^2_D = Darlington's shrinkage estimates

R^2_{CV1} = R^2 calculated on 1/2 of the sample

R^2_{CV2} = The cross validated shrinkage estimate based on the weights from the first sample (CV1) applied to the second sample

Situation III

Population 9		N	R^2_S	R^2_M	$R^2_{W/M}$	$R^2_{L_1}$	R^2_D	$R^2_{L_2}$	R^2_{CV1}	R^2_{CV2}
R^2 = .45	A	14	.59	.47	.41	.13	.05	.20	.80	.33
	B	30	.51	.46	.43	.32	.31	.34	.55	.36
$\bar{r}$ = .15	C	50	.48	.45	.44	.37	.37	.38	.52	.39
	D	100	.46	.45	.44	.41	.41	.41	.46	.41

Population 10										
R^2 = .46	A	14	.61	.50	.44	.18	.08	.24	.80	.29
	B	30	.50	.44	.42	.29	.29	.32	.53	.37
$\bar{r}$ = .29	C	50	.49	.46	.45	.38	.38	.39	.53	.39
	D	100	.47	.46	.46	.42	.42	.43	.49	.45

Population 11										
R^2 = .47	A	14	.62	.50	.45	.19	.10	.25	.81	.34
	B	30	.54	.48	.46	.35	.35	.37	.59	.40
$\bar{r}$ = .53	C	50	.51	.48	.46	.40	.40	.41	.53	.42
	D	100	.49	.48	.47	.44	.44	.45	.50	.45

Population 12										
R^2 = .55	A	14	.66	.56	.52	.29	.20	.34	.84	.43
	B	30	.58	.53	.51	.41	.41	.43	.63	.44
$\bar{r}$ = .82	C	50	.57	.54	.53	.47	.47	.48	.60	.51
	D	100	.55	.54	.53	.51	.51	.51	.56	.53

Table 1 seems to indicate that the McNemar and Wherry/McNemar tend to be relatively stable for the four different N sizes in Situation 1, with the Wherry/McNemar being more consistent across all four N's when the R^2 (population) is close to zero (.06). For Situation 2 the R^2's ranges from .31 to .34 and in Situation 3 the R^2's range from .45 to .55. In both Situations 2 and 3 the relationship of the McNemar and Wherry/McNemar to the population R^2's seem to become even more similar than they were in Situation 1. Lord 1, Lord 2 and Darlington, which are generally used for sample to sample estimates instead of sample to population, tend to have higher R^2's as the N's increase from 14 to 100. However, for all four populations in Situation 1, these are underestimates which one would expect since they are estimates from sample to sample and not sample to population. For Situations 2 and 3 when the N's were 50 and 100, Lord1 and Darlington tended to be fairly decent estimates of the population R^2, but they underestimated the population more than does the McNemar and the Wherry/McNemar.

Surprisingly to us, the cross-validation shrinkage estimate (CV_2) for Situation 1 with an N=14 tends to overestimate the population R^2. For the N=30, N = 50, and N=100, the CV_2 is fairly accurate. For Situations 2 and 3, CV_2 are underestimates of the population R^2's for all N sizes. An interesting effect is that, as the N's become larger, the CV_2 values become smaller, unlike Lord 1, Lord 2, and Darlington, whose values become larger as the N's increase (for Situation 1 only). It seems that the CV_2 values, except for Situation 1, tend to be more consistent than the mathematical values across all N sizes.

However, the shrinkage (amount of shrinkage) between CV_1 and CV_2 tend to decrease as the N's get larger.

The degree of multicollinearity ($\bar{r}$), that is the correlation between predictor variables, seems to have virtually no detectable effect on the accuracy of any of the shrinkage estimates.

Brief Discussion

Assuming that the data have been generated correctly (see the appendix for the computer program) the data from Table 1 would then strongly force the authors to question the preference for using cross-validation as a standard since.

1) The shrunken R does not seem to be more accurate than the mathematical approaches (Darlington (1969), Schmitt and Coyle (1977), and Claudy (1978)).

2) The shrinkage does not decrease as much proportionally as the authors expected when the N's increase

3) Cross-validation forces one to split the sample in half which tends to produce less stability than one would get using the entire sample.

Suggestions for Further Research

Darlington (1978) discusses Stein-type regression which is a shrinkage corrected regression on the weights instead of on the R^2. The partial regression weights are multiplied by 1 - S.

where: S is a correction for shrinkage obtained from the following formula

$$S = \frac{p-2}{n+2-p} \times \frac{(1-R^2)}{R^2}$$

where: p = number of predictor variables

n = size of the sample

R = sample multiple R (not corrected)

This procedure reduces the variability of the weights from sample to sample. As stated previously, in this procedure, all of the weights for the predictor variables are multiplied by the same constant (1-S). The purpose of Stein regression is identical to that of Ridge regression (Newman and Fraas (1979) , Walton, Newman and Fraas (1978); Hoerl and Kennard (1970)) except that in Ridge regression all of the weights are not multiplied by the same constant. Instead they are multiplied by a factor that is determined by the importance of each variable's contribution to the multiple R. It seems that Stein regressions simplicity and ease in determining a constant can be very valuable. The authors are suggesting that one may be interested in investigating the most appropriate shrinkage estimates for the weights, given different conditions, as has been and is being done in investigating the most appropriate shrinkage for the multiple R, given differing constraints.

As with most Monte Carlo studies, one should be careful about not overly generalizing from the artificially generated data to the real world data. It is possible that there is something idiosyncratic in the programs production of the data that one has to at least consider. The data should be replicated to see how well the shrinkage estimates predict from sample to sample.

The major purpose of this study was to check the differential effects of multicollinearity on shrinkage estimates. We found that multicollinearity has virtually no effect on shrinkage estimates. This may be an artifact of the data generated and this should be considered. In addition, a variety of other shrinkage estimates

which were not considered in this paper may be of value for future studies. Such estimates may include:

Burket's (1964) - weight validity estimate:

$$\rho_c^2 = \frac{Nr^2-K}{r(N-K)}$$

Olkin and Pratt (1958):

$$\rho^2 = 1-\frac{N-3}{N-K-1}\ (1-r^2)\ F\ (1,1;\ \frac{N-K+1}{2};\ 1-r^2)$$

Pratt's (1964) approximation of the Olkin and Pratt unbiased estimate of ρ^2:

$$\rho^2 = 1 - \frac{(N-3)(1-r^2)}{N-K-1}\left[1+\frac{2(1-r^2)}{N-K-2.3}\right]$$

Herzberg (1969) which is a further approximation of the Olkin and Pratt:

$$\rho^2 = 1 - \frac{(N-3)(1-r^2)}{N-K-1}\left[1+\frac{2(1-r^2)}{N-K+1}\right]$$

It is hoped that this paper has at least raised some additional questions that may lead to fruitful further investigations.

References

Burket, G.R. A study of reduced rank models for multiple prediction Psychometric Monograph, 1964, 12.

Claudy, J.G. Multiple regression and validity estimation in one sample. Applied Psychological Measurement, 1978, vol 2, 4 595-607.

Darlington, R.B. Reduced-variance regression. Psychological Bulletin, 1978, vol 85, 6.

Herzberg, P.A. The parameters of cross-validation. Psychometric Monograph, 1969, 16.

Hoerland and Kennard Ridge Regression: applications to nonorthogonal problems. Technometrics, 1970, vol 12, 1.

Klein, M. and Newman, I. Estimated parameters of three shrinkage estimate formuli. Multiple Linear Regression Viewpoints, 1974, 4, 6-11.

Lord, F.M. Efficiency of prediction when a regression equation from one sample is used, in a new sample. Research Bulletin, 1950, vol 50, 40, Princeton, N.J.: E.T.S.

McNemar, A. Psychological Statistics (3rd edition) New York: Wiley and Sons 1962.

Mosier, C.I. Problems and designs of cross-validation. Educational and Psychological Measurement, 1951, vol 11, 1-11.

Newman, I. Variations between shrinkage estimation formula and the appropriateness of their interpretation. Multiple Linear Regression Viewpoints, 1973, 4, 45-48.

Newman, I. and Fraas, J.W. Some applied research concerns using multiple linear regression analysis. Paper presented at Psychology Division of The Ohio Academy of Science, Capital University, Columbus, Ohio, April 1977.

Olkin, I. and Pratt, J.W. Unbiased estimation of certain correlation coefficients. Annuals of Mathematical Statistics, 1958, vol 29, 201-211.

Pohlman, J.T., and Moore, J.F. Interval estimation of the population squared multiple correlation. Multiple Linear Regression Viewpoints, 1977, 8, 18-24.

Schmitt, N., Coyle, B.W., and Rauchenberger, J. A Monte Carlo evaluation of three formula estimates of cross-validated multiple correlation. Psychological Bulletin, 1977, 84, 751-758.

Uhl, N. and Eisenber, T. Predicting shrinkage in the multiple correlation coefficient. Educational and Psychological Measurements 1970, vol 30.

Walton, J.M., Newman, I. and Fraas, J. Ridge Regression: A Panacea? M.L.R.V. 1978, vol 8, 2, 118-15.

```
      ∇GAUSS[⎕]∇
      ∇ GAUSS X;M;S;MIN;MAX;POP1;POP;N
[1]   M←X[1]
[2]   S←X[2]
[3]   MIN←X[3]
[4]   MAX←X[4]
[5]   POP←X[5]
[6]   POP1←POP
[7]   V←⍳0
[8]  TRYAGAIN:V←V,⌊0.5+(M+(S×((+/(0.1×((?(POP1,12)⍴11)-1)))-6)))
[9]   N←⍴V←((V<MAX)∧(V>MIN))/V
[10]  POP1←(POP-N)
[11]  →(POP1>0)/TRYAGAIN
[12]  'A POPULATION OF ',(▼POP),' MEMBERS HAS BEEN CREATED AND IS STORED IN THE VARIABLE V'
      ∇

      ∇CREATE[⎕]∇
      ∇ A CREATE B;K
[1]   K←1
[2]   MAT←((1↑⍴A),5)⍴0
[3]  LOOP:K←K+1
[4]   B[1] RAND 1↑⍴A
[5]   MAT[;K]←R+A
[6]   →LOOP×⍳K<5
[7]   B[2] RAND 1↑⍴A
[8]   MAT[;1]←A+R
[9]   10 4 ▼R←CM MAT
[10]  R←(¯4++/+/R[2 3 4 5 ; 2 3 4 5])÷12
[11]  'AVERAGE INTER-CORRELATION (R)= ', 6 4 ▼R
[12]  T← 5 4 3 2 1 REG MAT
[13]  'MULTIPLE R SQ = ', 7 5 ▼T[8;5]
      ∇
```

```
      ∇RAND[⎕]∇
      ∇ NN RAND N;K
[1]   K←0
[2]   R←,⍳0
[3]   →NUL×⍳NN=0
[4]  LOOP:K←K+1
[5]   R←R,?NN
[6]   →LOOP×⍳K<N
[7]   R←R-NN÷2
[8]   →0
[9]  NUL:R←N⍴0
      ∇
```

```
      ∇REG[⎕]∇
      ∇ T←V REG X;N;U;M;R;Q;S;B
[1]   T←((3+⍴V),5)⍴0
[2]   T[(⍳⍴V);1]←V[⍴V],¯1↓V
[3]   M←(+⌿U←X[;V])÷N←(⍴X)[1]
[4]   Q←(⌹R)+.×R←U-(⍴U)⍴M
[5]   T[(3+⍴V);3]←Q[(⍴V);⍴V]
[6]   S←⌹ 0 ¯1 ↓Q← ¯1 0 ↓Q
[7]   B←(-M+.×B,¯1),B←,S+.×Q[;⍴V]
[8]   T[(⍳⍴V);2]←B
[9]   T[(1+⍴V);3]←(1↓B)+.×Q[;⍴V]
[10]  T[((⍴V)+⍳3);2]←((⍴V)-1),(N-⍴V),N-1
[11]  T[(2+⍴V);3]←-⌿T[((3+⍴V),1+⍴V);3]
[12]  T[((⍴V)+⍳2);4]←T[((⍴V)+⍳2);3]+T[((⍴V)+⍳2);2]
[13]  T[(1+⍳(⍴V)-1);3]←(×B)×(|B←T[(2+⍴V);4]×+⌿S×(⍳(⍴V)-1)∘.=⍳(⍴V)-1)*0.5
[14]  T[(1+⍳(⍴V)-1);4]←T[(1+⍳(⍴V)-1);2]+T[(1+⍳(⍴V)-1);3]
[15]  T[(3+⍴V);4]←(×T[(2+⍴V);4])×(|T[(2+⍴V);4])*0.5
[16]  T[(1+⍴V);5]←T[(1+⍴V);4]+T[(2+⍴V);4]
[17]  T[(3+⍴V);5]←T[(1+⍴V);3]+T[(3+⍴V);3]
      ∇
```

```
      ∇SHRINK[□]∇
      ∇ POP SHRINK N
[1]   V←8ρ0
[2]   I←0
[3]   N2←⌈N[1]÷2
[4]  GO:NN←RAND1 N[1]
[5]   SAMP←POP[NN;]
[6]   RS← 5 4 3 2 1 REG SAMP
[7]   V[1]←V[1]+RS[8;5]
[8]   V[2]←V[2]+1-(1-RS[8;5])×(N[1]-1)×÷(N[1]-4)
[9]   V[3]←V[3]+1-(1-RS[8;5])×(N[1]-1)×÷(N[1]-3)
[10]  V[4]←V[4]+1-(1-RS[8;5])×(N[1]+5)×÷(N[1]-3)
[11]  V[5]←V[5]+1-((N[1]-1)×÷(N[1]-3))×((N[1]-2)×÷(N[1]-2))×((N[1]+1)×÷N[1])×(1-RS[8;5])
[12]  V[6]←V[6]+1-((N[1]-1)×÷(N[1]-3))×((N[1]+5)×÷N[1])×(1-RS[8;5])
[13]  CV1← 5 4 3 2 1 REG SAMP[(ι(1+ρSAMP)-N2);]
[14]  V[7]←V[7]+CV1[8;5]
[15]  CV2←SAMP[ιN2;1],[1.5] CV2←RS[1;2]+CV2←SAMP[ιN2; 2 3 4 5]+.×RS[5 4 3 2 ;2]
[16]  CV2← 2 1 REG CV2
[17]  V[8]←V[8]+CV2[5;5]
[18]  →GO×ιN[2]>I←I+1
[19]  V←V÷N[2]
[20]  RS← 5 4 3 2 1 REG POP
[21]  'THE MULTIPLE R SQ FOR THE POPULATION IS ', 6 4 ▼RS[8;5]
[22]  'THE SAMPLE SIZE = ',(5 0 ▼N[1]),'          SAMPLES TAKEN = ', 5 0 ▼N[2]
[23]  'THE AVERAGE SHRINKAGE ESTIMATES FOR THESE SAMPLES ARE:'
[24]  '        RS      RS(W)      RS(M)      RS(L1)     RS(L2)      RS(D)      CV(1)     *CV(2)*'
[25]  10 4 ▼V
      ∇

      ∇CM[□]∇
      ∇ R←CM X;V
[1]   R←R÷(V∘.×V←(1 1)⌽R←(⌽R)+.×R←X-(ρX)ρ(+⌿X)÷(ρX)[1])*0.5
      ∇

      ∇RAND1[□]∇
      ∇ R←RAND1 N;K
[1]   K←0
[2]   R←,ι0
[3]  LOOP:R←R,?500
[4]   →LOOP×ιN>K←K+1
      ∇
```

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 9 NO. 5 1979

MISSING CELLS AND A CURIOUS CASE OF DEGREES OF FREEDOM

John D. Williams
Mohan K. Wali
University of North Dakota

ABSTRACT

An experimental sampling procedure for plant communities on surface mined areas yielded missing cells and caused a further problem of yielding a "total" number of degrees of freedom equal to N rather than the usual N-1. The discrepancy occurred because the degrees of freedom are not necessarily additive for all missing cell designs. A solution which may circumvent this problem is proposed.

While many researchers in applied statistics have recently considered the two (or n) way disproportional fixed effects design, the missing cells case has not been fully reported. Because missing cells do occur either by design, distribution of natural populations, or topographic influences (as in the present case), they deserve due consideration. Perhaps the most extensive treatment of the cases of missing cells has been discussed by Searle (1971).

Missing cells in our study (Wali 1977) occurred in a systematic sampling of plant communities at six sites: five were previously surface mined for coal and one site in an unmined prairie was chosen for comparison. Prior to state and federal regulations on reclamation, these surface mined areas looked like a series of cone-like mounds (commonly referred to as spoil banks). Since topography, slope angle and aspect are important determinants of plant distribution, it was decided to obtain quantitative information on the distribution of plant species from crest, upper slope and mid-slope positions along eight cardinal compass directions (Fig. 1). In the sampling procedure, species composition of each sub-area was determined at every major site. This yielded (Fig. 1) 8 sampling areas in each of the upper and lower slope positions and 1 from the crest for a total of 17. The criterion measures were the types and numbers of plant species in each sample. The two-way classification with the number in each cell is shown in Table 1.

[1] Contribution from "A Systems Approach to Reclamation of Strip-Mined Areas in North Dakota" funded by U.S.D.I., Bureau of Mines, Grant No. 0264001 to M.K.W.

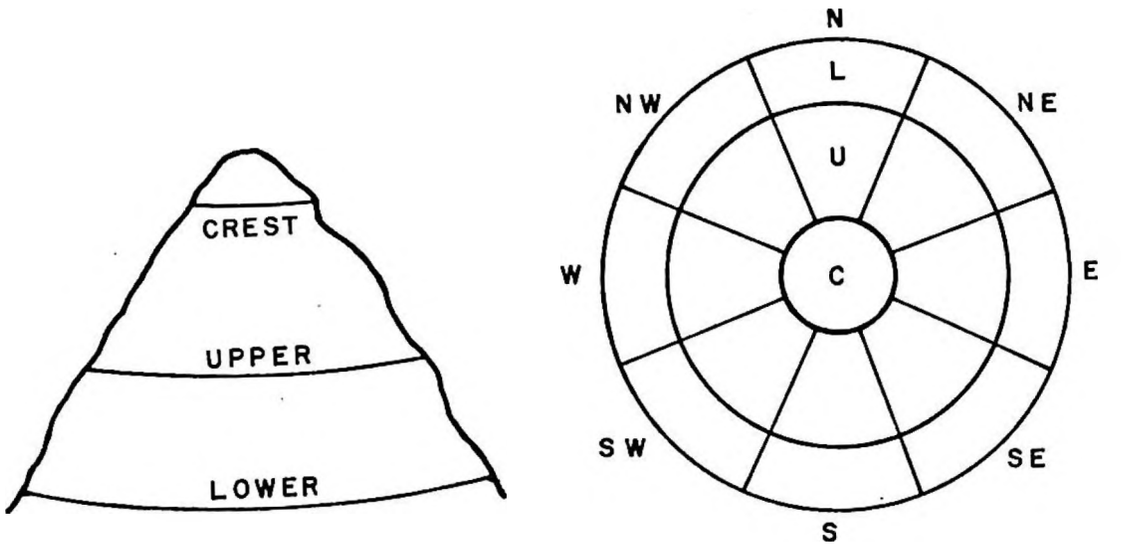


Fig. 1. (a) a diagrammatic representation of a "spoil bank" showing the position of the crest, and the upper and lower slopes; (b) a top view of the same showing sub-areas sampled along compass directions.

Table 1

Number of Samples from Five Strip-Mined and One Unmined Area by Sampling Units

	Direction								
	Crest	N	NE	E	SE	S	SW	W	NW
LEVEL CREST	12	-	-	-	-	-	-	-	-
UPPER	-	12	12	12	12	12	12	12	12
LOWER	-	12	12	12	12	12	12	12	12

NOTE: N = North octant; NE = Northeast octant; E = East octant; SE = Southeast octant; S = South octant; SW = Southwest octant; W = West octant; and NW = Northwest octant.

The unique qualitites of the research design can be seen by examining Table 1. When confronted by data gathered from such a design some might suggest that if the cell in the upper left-hand corner is dropped, there is no problem. However, such a suggestion begs the issue. Also, the actual three-way design is such that the proportionality holds for the other two ways of classification. An additional point is revealed by Table 1; the upper left hand cell does not enter in any way into the interaction. Thus, the interaction sum of squares and degrees of freedom (df = 7) would be the same whether or not the crest data are included.

Models for Two Way Design

First, several variables are defined

Y = the criterion score (number of distinct species in the subsample)
X_1 = 1 if the score is from a crest sample, 0 otherwise;
X_2 = 1 if the score is from an upper sample, 0 otherwise;
X_3 = 1 if the score is from the N octant; 0 otherwise;
X_4 = 1 if the score is from the NE octant; 0 otherwise;
X_5 = 1 if the score is from the E octant; 0 otherwise;
X_6 = 1 if the score is from the SE octant, 0 otherwise;
X_7 = 1 if the score is from the S octant, 0 otherwise;
X_8 = 1 if the score is from the SW octant, 0 otherwise;
X_9 = 1 if the score is from the W octant, 0 otherwise;
$X_{10} = X_2 \cdot X_3$;
$X_{11} = X_2 \cdot X_4$;
$X_{12} = X_2 \cdot X_5$;
$X_{13} = X_2 \cdot X_6$;
$X_{14} = X_2 \cdot X_7$;
$X_{15} = X_2 \cdot X_8$; and
$X_{16} = X_2 \cdot X_9$.

Now several models can be defined:

Row Effect:

$$Y = b_0 + b_1X_1 + b_2X_2 + e_1; \quad (1)$$

Column Effect:

$$Y = b_0 + b_1X_1 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9 + e_2; \quad (2)$$

Row and Column Effects:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9 + e_3; \quad (3)$$

and

Full Model:

$$Y = b_0 + b_1X_1 + \ldots + b_{16}X_{16} + e_4, \quad (4)$$

where the b_i are regression coefficients which can differ from equation to equation (i.e., b_1 will not necessarily be unchanged from equation to equation). Table 2 contains sums of squares generated from equations 1-4.

Table 2

Summary Table for Species Data

	df	SS	MS	F
Rows	2	11.18	5.59	.08
Columns	8	157.24	19.66	.28
Rows & Columns	9	158.93		
Rows (Independent of Columns)	1	1.69	1.69	.02
Columns (Independent of Rows)	7	147.75	21.11	.30
Full Model	16	205.82		
Interaction	7	46.89	6.70	.10
Error	187	13113.76	70.13	
Total	203	13319.58		

While in every case the F values are non-significant, perhaps the traditional hypotheses tested are of less interest than investigating the interrelationships (Table 2). If an unadjusted main effects solution (Williams 1972) is preferred, the sources of variation would include rows, columns, interaction and error; the corresponding degrees of freedom are, respectively, 2, 8, 7, 187, or a total of 204 (apparently 1 df more than customary). If a fitting constants solution (Overall and Spiegel 1969) is preferred, the sources of variation are rows (independent of columns), columns (independent of rows), interaction and error; the degrees of freedom are, 1, 7, 7, 187, or a total of 202 (1 df less than usual) respectively. A hierarchical model (Cohen 1968) would of course be additive; however, there would be attendant complications of the hypotheses tested.

The Use of Full Rank Models

Timm and Carlson (1975) gave full rank models a strong impetus for usage with disproportional data; Dalton (1977) recently affirmed their usage. To use

the full rank models, X_1 - X_{16} are slightly modified:

X_1^* = X_1 except $X_1^* = -1$ for scores in the lower samples;

X_2^* = X_2 except $X_2^* = -1$ for scores in the lower samples;

X_3^* through X_9^* are respectively the same as X_3 through X_9 except X_3^* through $X_9^* = -1$ for NW samples;

X_{10}^* through X_{16}^* are product vectors using X_3^* through X_9^* in a manner analogous to X_{10} through X_{16}.

Because there is some ambiguity as to the nature of the models (due to the crest cell being in both a row and a column by itself) six models are defined:

Full Model:

$$Y = b_0 + b_1X_1^* + b_2X_2^* + \ldots + b_{16}X_{16}^* + e_5, \quad (5)$$

Restricted Model 1 - Rows and Interaction:

$$Y = b_0 + b_1X_1^* + b_2X_2^* + b_{10}X_{10}^* + \ldots + b_{16}X_{16}^* + e_6, \quad (6)$$

Restricted Model 2 - Rows and Interaction, crest excluded:

$$Y = b_0 + b_2X_2^* + b_{10}X_{10}^* + \ldots + b_{16}X_{16}^* + e_7, \quad (7)$$

Restricted Model 3 - Columns and Interaction:

$$Y = b_0 + b_1X_1^* + b_3X_3^* + \ldots + b_{16}X_{16}^* + e_8, \quad (8)$$

Restricted Model 4 - Columns and Interaction, crest excluded:

$$Y = b_0 + b_3X_3^* + \ldots + b_{16}X_{16}^* + e_9, \quad (9)$$

Restricted Model 5 - Rows and Columns:

$$Y = b_0 + b_1X_1^* + \ldots + b_9X_9^* + e_{10}. \quad (10)$$

The results from using the full rank models are shown in Table 3.

Table 3

Results from Full Rank Models

Equation	df	$df_{Full} - df_{Rest}$	SS	$SS_{Full} - SS_{Rest}$
5	16	16	205.82	205.82
6	9	7	58.07	147.75
7	8	8	48.58	157.24
8	15	1	204.13	1.69
9	14	2	194.64	11.18
10	9	7	158.93	46.89

Equations 6 and 8 correspond to the results for columns (independent of rows) and rows (independent columns) for the fitting constants method shown in Table 2; also, equations 7 and 9 correspond to the columns and rows effects for the unadjusted main effects solution in Table 2. The interaction in either case is found from the use of equation 10 together with the full model. It can be seen that, for data with the pecularities of the present design, the full rank models do not yield new sets of results as will occur with either disproportional or proportional (but unequal sized cells) data.

Hypotheses Testing through Restricted Models

Using the methodology of Bottenberg and Ward (1963) and Jennings (1967) 17 cell variables are first defined (Table 4).

Table 4

Representation of Cell Model for Mining Cells Data

	Crest	N	NE	E	SE	S	SW	W	NW
Crest	X_1	-	-	-	-	-	-	-	-
Upper	-	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Lower	-	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}

For example, $X_1 = 1$ if the score is from a crest sample, 0 otherwise. All other variables X_2 - X_{17} are analogously defined (note that variables X_1 - X_{17} presented in Table 4 are different from variables X_1 - X_{17} given earlier).

A full model can be defined by

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_{17}X_{17} + e_{11}, \quad (11)$$

Note that equation 11 does not include a unit vector, implied by the absence of b_0. Likely hypotheses that express the row effect are

$$b_1 = \frac{b_2 + b_3 + b_4 + b_5 + b_6 + b_7 + b_8 + b_9}{8} =$$

$$\frac{b_{10} + b_{11} + b_{12} + b_{13} + b_{14} + b_{15} + b_{16} + b_{17}}{8}.$$

If these restrictions are expressed in terms of two of the coefficients (say b_1 and b_{10}), then

$$b_1 = \frac{b_2 + b_3 + b_4 + b_5 + b_6 + b_7 + b_8 + b_9}{8} \text{ and}$$

$$b_{10} = b_2 + b_3 + b_4 + b_5 + b_6 + b_7 + b_8 + b_9 - b_{11} - b_{12} - b_{13} - b_{14} - b_{15} - b_{16} - b_{17}.$$

If these two restrictions are placed on the full model,

$$Y = \frac{(b_2 + b_3 + b_4 + b_5 + b_6 + b_7 + b_8 + b_9) X_1}{8} + b_2X_2 + \ldots + b_9X_9$$

$$+ (b_2 + b_3 + b_4 + b_5 + b_6 + b_7 + b_8 + b_9 - b_{11} - b_{12} - b_{13} - b_{14} - b_{15} - b_{16} - b_{17})X_{10} + b_{11}X_{11} + \ldots + b_{17}X_{17} + e_{12}. \quad (12)$$

Rearranging terms,

$$Y = b_2 (1/8X_1 + X_2 + X_{10}) + b_3 (1/8X_1 + X_3 + X_{10}) + b_4 (1/8X_1 + X_4 + X_{10}) + b_5 (1/8X_1 + X_5 + X_{10}) + b_6 (1/8X_1 + X_6 + X_{10}) + b_7 (1/8X_1 + X_7 + X_{10}) + b_8 (1/8X_1 + X_8 + X_{10}) + b_9 (1/8X_1 + X_9 + X_{10}) + b_{11} (X_{11} - X_{10}) + b_{12} (X_{12} - X_{10}) + b_{13} (X_{13} - X_{10}) + b_{14} (X_{14} - X_{10}) + b_{15} (X_{15} - X_{10}) + b_{16} (X_{16} - X_{10}) + b_{17} (X_{17} - X_{10}) + e_{12}. \quad (13)$$

Now, 15 variables are defined in relationship to b_2 through b_9 and b_{11} through b_{17}. For example, the variable related to b_2 is $V_2 = 1/8$ if from a crest sample, 1 if from a cell 2 sample, 1 if from a cell 10 sample, 0 otherwise. The restricted model, after creating all 15 variables would be

$$Y = b_2V_2 + b_3V_3 + \ldots + b_9V_9 + b_{11}V_{11} + \ldots + b_{17}V_{17} + e_{12}. \quad (14)$$

Programs such as those of McNeil et al. (1975) LINEAR or Ward and Jennings (1973) DATRAN can use equations like (14) above directly. Most multipurpose multiple linear regression programs automatically include a unit vector, thus including the b_0 term. Any *one* of the variables V_2 - V_9 or V_{11} - V_{17} can be dropped if a multipurpose multiple regression program is used; for the present purpose V_{17} is (arbitrarily) dropped. The restricted model becomes

$$Y = b_0 + b_1V_1 + b_2V_2 + \ldots + b_9V_9 + b_{11}V_{11} + \ldots + b_{16}V_{16} + e_{12}. \quad (15)$$

Using equation 15 as the restricted model, df = 14, SS = 194.64; thus, the df for rows is 2, and SS_{rows} 205.82 - 194.64 = 11.18, identical to those values found both with the use of equation 1 and with the difference between equations 5 and 9.

A similar solution could be set for columns as well. A set of hypotheses likely to be of interest for the column effect is (in terms of the full model, equation 11)

$$b_1 = \frac{b_2 + b_{10}}{2} = \frac{b_3 + b_{11}}{2} = \frac{b_4 + b_{12}}{2} = \frac{b_5 + b_{13}}{2} = \frac{b_7 + b_{15}}{2} = \frac{b_8 + b_{16}}{2} = \frac{b_9 + b_{17}}{2}.$$

In a manner similar to that shown for the row effect a restricted model can be found such that df = 8, SS = 48.58, yielding for the columns effect df = 8, $SS_{columns}$ = 157.24, results similar to those found for equation 2 and for the difference between equations 5 and 7.

Thus, while it may be somewhat troublesome initially to find that the degrees of freedom, if totaled, exceed the total degrees of freedom (by one), this oddity arises from testing what would appear to be a useful set of hypotheses. Except for hierarchical models, those who have used disproportionate cell frequencies are already familar that such data are generally non-additive with respect to both the sums of squares and R^2 values; in some situations with missing cells the degrees of freedom may become non-additive.

The Effect of Dropping a Vector on the Hypotheses Tested

Some readers, particularly those familiar with the paper by Jennings (1978), might wonder about the effect of dropping a vector, such as V_{17} in equation 15.

Strictly speaking, the dropping of the V_{17} from the full model (equation 11) should be written

$$Y = b_0 + c_1X_1 + c_2X_2 + c_3X_3 + \ldots + c_{16}X_{16} + e_{13} \qquad (16)$$

where each $c_i = b_i - b_0$.

Also, $b_0 = \bar{Y}_{17}$; thus each $c_i = \bar{Y}_i - \bar{Y}_{17}$ for equation 16. Letting $b_{17} = 0$ in equation 11 is called setting a side condition; equation 16 is a re-parameterization of equation 11 (meaning that the expected values for all Y_i are the same in each equation; also, $e_{11} = e_{13}$). The null hypothesis corresponding to the row effect would be

$$c_1 = \frac{c_2+\ldots+c_9}{8} = \frac{c_{10}+\ldots+c_{16}+b_0}{8}.$$

It should be noted that this expression of the row effect is testing the same hypothesis as stated earlier; if b_0 is added to each portion of the present expression, the previous expression for the row effect would occur. The restrictions become (in terms of c_1 and c_{10})

$$c_1 = \frac{c_2+\ldots+c_9}{8} \text{ and } c_{10} = c_2 + \ldots + c_9 - (c_{11}+\ldots+c_{16} + b_0).$$

Placing these restrictions on equation 16,

$$Y = \frac{b_0+(c_2+\ldots+c_9)}{8} X_1+c_2X_2+\ldots+c_9X_9 + [(c_2+\ldots+c_9) - (c_{11}+\ldots+c_{16}+b_0)]X_{10}+c_{11}X_{11}+\ldots+c_{16}X_{16}+e_{14}. \qquad (17)$$

Rearranging terms,

$$Y = b_0+c_2(1/8X_1+X_2+X_{10})+c_3(1/8X_1+X_3+X_{10})+c_4(1/8X_1+X_4+X_{10})+c_5(1/8X_1+X_5+X_{10})+c_6(1/8X_1X_6+X_{10})+c_7(1/8X_1+X_7+X_{10})+c_8(1/8X_1+X_6+X_{10})+c_9(1/8X_1X_7+X_{10})+c_{11}(X_{11}-X_{10})+c_{12}(X_{12}-X_{10})+c_{13}(X_{13}-X_{10})+c_{14}(X_{14}-X_{10})+c_{15}(X_{15}-X_{10})+c_{16}(X_{16}-X_{10})+e_{14}. \qquad (18)$$

The similarity between equations 13 and 18 is striking; also $e_{12} = e_{14}$.

An equation similar to equation 14 can be constructed:

$$Y = b_0+c_2V_2+c_3V_3+\ldots+c_9V_9+c_{11}V_{11}+\ldots+c_{16}V_{16}+e_{14}. \qquad (19)$$

Actually, equation 19 is the analog to equation 15; equation 15 would more correctly be said to be the same as equation 19. That is, the values for the b_i in equation 15 are actually the c_i from equation 19. Similar models could be developed for the columns as well.

The hypotheses tested by equation 19 use regression coefficients rather than the means. The hypothesis in terms of the population means corresponding to the restriction on equation 11 (not including the unit vector) is

$$\mu_1 = \frac{\mu_2+\mu_3+\ldots+\mu_9}{8} = \frac{\mu_{10}+\mu_{11}+\ldots+\mu_{17}}{8}. \qquad (20)$$

The restriction on equation 16 (including the unit vector) corresponds to the following hypothesis regarding the population means:

$$(\mu_1-\mu_{17}) = \frac{(\mu_2-\mu_{17})+(\mu_3-\mu_{17})+\ldots+(\mu_9-\mu_{17})}{8} =$$

$$\frac{(\mu_{10}-\mu_{17})+(\mu_{11}-\mu_{17})+\ldots+(\mu_{16}-\mu_{17})}{8} \qquad (21)$$

If μ_{17} is added to each side of the last expression, equation 20 would result.

Other Applications With The Present Design

The present design would be applicable whenever double entry tables are used wherein all sampling units except for a control group are placed into a treatment combination that includes two actual treatments. The present design allows a control group within the context of a two-way analysis of variance; other designs may include a no-treatment for each dimension of the design. If there is an intrinsic interaction, the interaction might be overlooked in designs that include the no-treatment group for both classifications crossed with other treatments in the two-way classification.

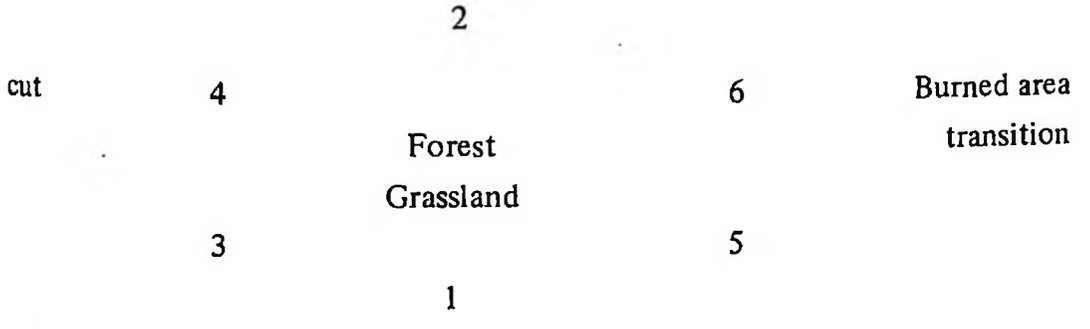
As an example of such an experiment in reclaiming strip-mined land, suppose an amendment program began so that four (non-zero) levels of potassium are used as one way of classification together with the second way of classification consisting of several amendments. If the interest is in looking at possible interactions together with looking for the most effective treatment combination, then perhaps a single control group would serve the experiment as well as trying to incorporate a control group for each way of classification by crossing them with the treatments of the other dimension.

A very different sort of biological application could be made in the area of genetics. Suppose there is interest in cross-breeding an available strain of laboratory mice with three genetically different strains. The intent is to measure hybrid vigor in the offspring when compared to similar offspring in the available colony; further, care is taken to match animals from the four strains such that sex of the parent is also a factor. A table similar to Table 1 could be constructed so that the experiment could be executed:

		1	2	3
Orig.	X_1	-	-	-
Male	-	X_2	X_3	X_4
Female	-	X_5	X_6	X_7

Except for cell 1 (X_1), all other cells would contain hybrids. It would make sense to view cell 1 as a control group to measure the other cells against.

In ecological studies, consider an area such as that shown in the following diagram, in which a natural transition occurs from a grassland to forest. Assume that the burned and mowed/clearcut areas were well defined and were established at the same time. Also assume that a relatively long period of time had passed between establishment and sampling. The objective of the experiment would be to determine the similarities (1) in species composition by some appropriate measurement between the untreated grassland and each of the treatments (burned former grassland, burned former forest, mowed former grassland, clearcut former forest), and (2) the similarities of the untreated forest to each treatment as well as the effects of burning vs. mowing, former forest vs. former grassland.



The table could then be depicted as

	No treatment	Cut	Burned	No treatment
Grassland	X_1	-	-	-
Former grassland	-	X_3	X_5	-
Former forest	-	X_4	X_6	-
Forest	-	-	-	X_2

Educational Applications

While the example given was concerned with sampling plant species in strip-mined land and the other applications had a biological orientation as well, one might ask, "What value is the design in human experimentation?" Within a design sense, an experimenter may have a two-way treatment classification procedure; however, a true control group would be administered neither treatment combination. In such a case, the crest cell would correspond to the control group and the remaining rows and columns would constitute the treatments.

Yet another application might arise from a concern for the cultural bias associated with a given test. Suppose, for sake of argument that a critique of the Iowa Test of Basic Skills (ITBS) asserts that, in regard to district wide means, all the ITBS measures is the distance from Iowa (the further the distance, the lower the mean score). Adjacent states might constitute the "close" states and the next layer of states might constitute the "farther" layer of states (undoubtedly a much more sophisticated sampling process would occur with the actual choosing of subjects). In diagramatic form, the cells would be represented as

	Inner	North	N East	East	South	West	N West
Inner	Iowa	-	-	-	-	-	
Close	-	Minnesota	Wisconsin	Illinois	Missouri	Nebraska	S Dakota
Farther	-	N Dakota	Michigan	Indiana	Arkansas	Wyoming	Montana

Random samples of school districts could be chosen from the listed states and the original hypotheses could be tested.

REFERENCES

Bottenberg, R.A. and Ward, J.H. *Applied multiple linear regression*. Lackland Air Force Base: Personnel Research Laboratory, TDR-63-6, 1963.

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

Dalton, S. Regression approaches and approximate solutions to analysis of variance with disproportionality varied. *Multiple Linear Regression Viewpoints*, 1977, 7, No. 2, 16-32.

Jennings, E. Fixed effects analysis of variance by regression analysis. *Multivariate Behavioral Research*, 1967, 2, 95-108.

Jennings, E. Fixed effects analysis of variance with unequal cell sizes. *Journal of Experimental Education*, 1978, 46, No. 2, 42-51.

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

Overall, J.E. and Spiegel, D.K. Concerning least squares analysis of experimental data. *Psychological Bulletin*, 1969, 72, 311-322.

Searle, S.R. *Linear models*. New York: Wiley, 1971.

Timm, N.J. and Carlson, J.R. Analysis of variance through full rank models. *Multivariate Behavioral Research: Monograph*, 1975, 75-1.

Wali, M.K. *A systems approach to the reclamation of strip-mined areas in North Dakota*. Grand Forks: Project Reclamation, Univ. North Dakota, 1977.

Ward, J.H. and Jennings, E. *Introduction to linear models*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Williams, J.D. Two way fixed effects analysis of variance with disproportionate cell frequencies. *Multivariate Behavioral Research*, 1972, 7, 67-83.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 9 NO. 5 1979

INCREASING POWER AND INTERPRETABILITY IN CERTAIN REPEATED MEASURES DESIGNS

John T. Mouw, Ph.D.
Nu Viet Vu, Ph.D.
Southern Illinois University

Abstract

Repeated measures designs offer a relatively powerful procedure for the analysis of behavioral data. In these designs, research questions involve the change of individuals' patterns of responses across time or across a dimension with intervening treatment effects. The addition of one or more between-subject factors allows for the comparison of treatment effects across the repeated measures between groups of subjects. In most of these researches, the grouping variable has been obtained by arbitrarily dichotomizing a continuous variable. This article presents an alternative analysis of data of certain repeated measures designs where the variable is kept in its natural continuous state instead of being dichotomized. Such an analysis is argued to have two advantages: (a) A more realistic interpretation of the results and (b) A tendency toward an increase in power in the F-tests of the repeated dimension and its interaction.

Dayton (1970) and Lindquist (1953) are at least two authors who advocate the use of repeated measures designs to enhance the work of the researcher. The greater frequency of use of such designs in research tends to indicate their expanding popularity. The usual

This paper originally appeared in the February, 1979, issue of Viewpoints. Due to a misspelling of the author's name, it is being reprinted here.

increase in power of the statistical tests, as well as the advantage of decreasing the number of subjects when more than one observation is taken on a subject, have probably been advantages too great to be passed by in favor of a logically less complex design such as a completely randomized design.

A rather common use of a repeated measures design is found where two or more groups of subjects are defined by median or quartile splits on a continuous variable and observing the performance of the "groups" over a series of treatments or time periods. Spielberger (1966), for example, studied the effects of word-position and stress-non stress conditions on performance in serial-verbal learning for the high (HA) and low (LA) anxiety college males. The subjects (Ss), instead of being classified along the continuum of the anxiety scale, had been separated arbitrarily into two groups according to their raw scores on the Manifest Anxiety Scale. Other common continuous between-factors which have been dichotomized or split are age, IQ, grade and ability. Corrigan (1975), for example, divided his subjects into five groups ranging from 2.5 to 7.5 years old in order to study their use and comprehension of the word "because", and Millar (1971) divided her subjects into two groups of three and four years old in order to determine their use and recognition of visual and haptic stimuli. Vogel (1970) categorized his kindergarteners into groups of High, Medium, and Low intelligence based on their Kuhlmann-Anderson Intelligence Test scores to study the morphology of lower class

children. Youniss et al. (1971) classified their subjects according to both their grade and age to determine the children's inferential size judgment in the figurative or operative aspects or both.

Such a practice of "grouping", which is comparable to the "levels" of a Treatment by Levels design, was probably done in order to make the data layout fit the traditional repeated measures designs described in the major design textbooks (Dayton, 1970; Hays, 1963; Lindquist, 1953; and Winer, 1971). However, with the common availability of computer facilities and the greater flexibility of the multiple regression analysis via the general linear model (Ward and Jenning, 1973; Kerlinger and Pedhazur, 1973), we should be able to take advantage of a more appropriate method of analysis.

This procedure was hinted at in a paper comparing power by Feldt (1958) much earlier. However, whether due to lack of technical facilities or the "zeitgeist" of the field, the technique has not been utilized.

In that same article, Feldt (1958) argues that the "blocking" procedure (on the concomitant variables) yields more power than the use of the concomitant variables as a covariate when the correlation is less than .60. His argument, however, rests partially on two assumptions: (a) The researcher employs a large number of levels depending on the N of the research, and (b) The random variability of the group means on the concomitant variable is a source of error resulting in loss of power. The first assumption appears to be impractical, since researchers seldom use more than three levels. The second argument

does not hold in a repeated measures design, as presented below, where the mean of the concomitant variable is equal across treatment groups.

The following presentation argues for the use of such a concomitant variable in its natural continuous state rather than using the scores to arbitrarily define "groups".

Winer (1971) has presented an analysis of a two-factor experiment with repeated measures on one factor. This usually consists of the observations of J groups of subjects from one factor B, where the groups are designated as b_1, b_2,...b_j, under different treatment conditions of factor A, such as $a_1 \ldots a_i$, which are observations of the same subjects under various treatments. Each level of B consists of K subjects. According to Winer (1971), each observation results from a number of sources of variability which can be represented as follows:

$$Y_{ijk} = \mu + \beta_j + \alpha_i + \pi_{k(j)} + \beta\alpha_{ij} + \alpha\pi_{ik(j)} + e_{ijk} \qquad \text{(Model 1)}$$

Where:

Y_{ijk}: Observation made on subject k on level i under treatment condition j.

μ : Grand mean of all potential observations.

β_j : Effect of factor B under level j.

α_i: Effect of factor A under level i.

$\pi_{k(j)}$: Effect of subject k under level B_j.

$\beta\alpha_{ji}$: Effect of combinations of β under level j and α treatment under level i.

$\alpha_{ik(j)}$: Effect of interaction between subject k with treatment α under level j of β.

e_{ijk}: Experimental error nested within the individual observation.

Given an example where B consists of two levels with three subjects at each level and A of three treatments, the above linear model can be expanded into a general linear model where factor A is broken into linear and quadratic components as follows:

$$y_{ijk} = \mu + \beta + \alpha_{lin} + \alpha_{quad} + \pi + \beta\alpha_{lin} + \beta\alpha_{quad} + \alpha\pi_{lin} + \alpha\pi_{quad} + \varepsilon_2$$

(Model 2)

Models (1) and (2) are similar with the exception that model (2) has been expanded so that the factors have been broken down into linear and quadratic levels according to the number of levels contained across A. If the factors in model (2) are coded orthogonally, an estimate of the model from a sample may be obtained by the model:

$$Y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \ldots + a_7x_7 + a_8x_8 + a_9x_9 + a_{10}x_{10} + \ldots + a_{13}x_{13} + a_{14}x_{14} + \ldots + a_{17}x_{17} + E_3 \quad \text{(Model 3)}$$

Where:

Y: Criterion scores

a_0: Intercept, estimate of μ

a_1-a_{17}: Partial regression weights

X_1: Vector representing the contrast of factor B. ($B_x = +1$, $B_2 = -1$)

X_2, X_3: Vectors representing respectively the linear and quadratic polynomials of factor A.

X_4-X_7: Vectors representing the subjects using effect coding as in Kerlinger and Pedhazur (1973).

X_8: Vector representing the linear component of the AB interaction.

X_9: Vector representing the quadratic component of the AB interaction.

X_{10}-X_{13}: Vectors representing the A linear x Persons interaction (error within).

X_{14}-X_{17}: Vectors representing the A quadratic x Persons interaction (error within).

The effects found in model (2) including A_{lin}, A_{quad}, $A_{lin}B$, and $A_{quad}B$ can be tested with sample data in model (3) through a linear regression procedure. This model is most appropriate with non-continuous factors such as: different conditions (experimental versus control, stress versus non-stress,...,). But as was previously pointed out, many of the between factors that are encountered in behavioral research exist as continuous variables in their natural state. Then model (3) becomes inappropriate because:

1. By dichotomizing or splitting the continuous B factors, we lose the ability to examine the natural rélationship (correlation) between X and Y. This problem becomes especially severe when, as in most designs such as this, the major research interest is in the interaction of A x B, i.e., the difference in performance across A depending on the level of B. Model (3) does not allow the study of such an interaction between the treatment and the natural continuum of factor B.

2. The least-square estimate is best represented by a regression line rather than two or more arbitrarily defined points (Y.j). This results in a smaller error estimate when B is used as a continuous variable, thus decreasing the probability of a type II error and increasing the power of the test.

Designs that have continuous factors are best studied by not dichotomizing or splitting B. This can be done by allowing the B factor to be represented by its raw scores rather than, as in models (2) and (3), by a dichotomy. We can then translate model (2) into model (4) as follows:

$$Y_{i-k} = \alpha + \gamma_{lin} + \gamma_{quad} + \beta(X-\bar{X}) + \pi + \gamma_{lin}\beta(X-\bar{X}) + \gamma_{quad}\beta(X-\bar{X}) + \gamma\pi + \varepsilon_4 \qquad \text{(Model 4)}$$

Where:

Y_{i-k}[1]: Criterion scores.

α: The general intercept of X on Y_{i-k}.

β: The general regression of X on Y_{i-k}.

$(X-\bar{X})$: The deviation of the raw score of the continuum from the mean of factor B.

γ_{lin}: Linear effect of factor A.

γ_{quad}: Quadratic effect of factor A.

By comparing models (2) and (4) it can be seen that μ is represented by α, α by γ, and β by $\beta(X-\bar{X})$.

Model (4) is analogous to an analysis of covariance design where X is the covariate. The overall regression effect of X on Y is contained in $\beta(X-\bar{X})$ of model (4); while the interaction of A x B of model (2) and (3) is contained in $\gamma_{lin}\ \beta(X-\bar{X})$ and $\gamma_{quad}\ \beta(X-\bar{X})$ of model (4). The latter, interaction effects, contain sources of variance which are usually considered heterogeneity of regression in the analysis of

covariance design. Rather than being a nuisance as in ANCOVA, however, interaction effects for this model are often of primary research interest.

The above models may be best understood with the following illustration. An experimenter would like to study the relationship between the performance on a "Reading Achievement" test of six subjects in conjunction with their relative Aptitude score on three days: Day 1, Day 2, and Day 5. The Aptitude scores of the students were obtained before they were given the tests. The following results were obtained:

Table 1
Scores for the Sample Problem

Aptitude Score	Score on "Reading Achievement" Tests		
	Day 1	Day 2	Day 5
12	5	7	9
13	7	10	11
15	8	10	12
24	11	15	16
28	13	18	19
32	13	20	21

If model (2) is used for this illustration, subjects with the first three scores on the Aptitude test could be classified as the "Low Aptitude Group", and the three remaining scores as the "High Aptitude Group". In that model, β (Aptitude) would be orthogonally coded with

B_1 (the low aptitude group) coded as +1 and B_2 (the high aptitude group) coded as -1. Model (4) would utilize the aptitude information by running the variable $(X-\bar{X})$ into the model. In both models (2) and (4), Factor A (day) would be orthogonally coded, with linear and quadratic polynomials, and the criterion scores Y_{ijk} would be the scores on the Reading Achievement test.

Figurative representation of the data from Table 1 is presented in Figures 1 and 2:

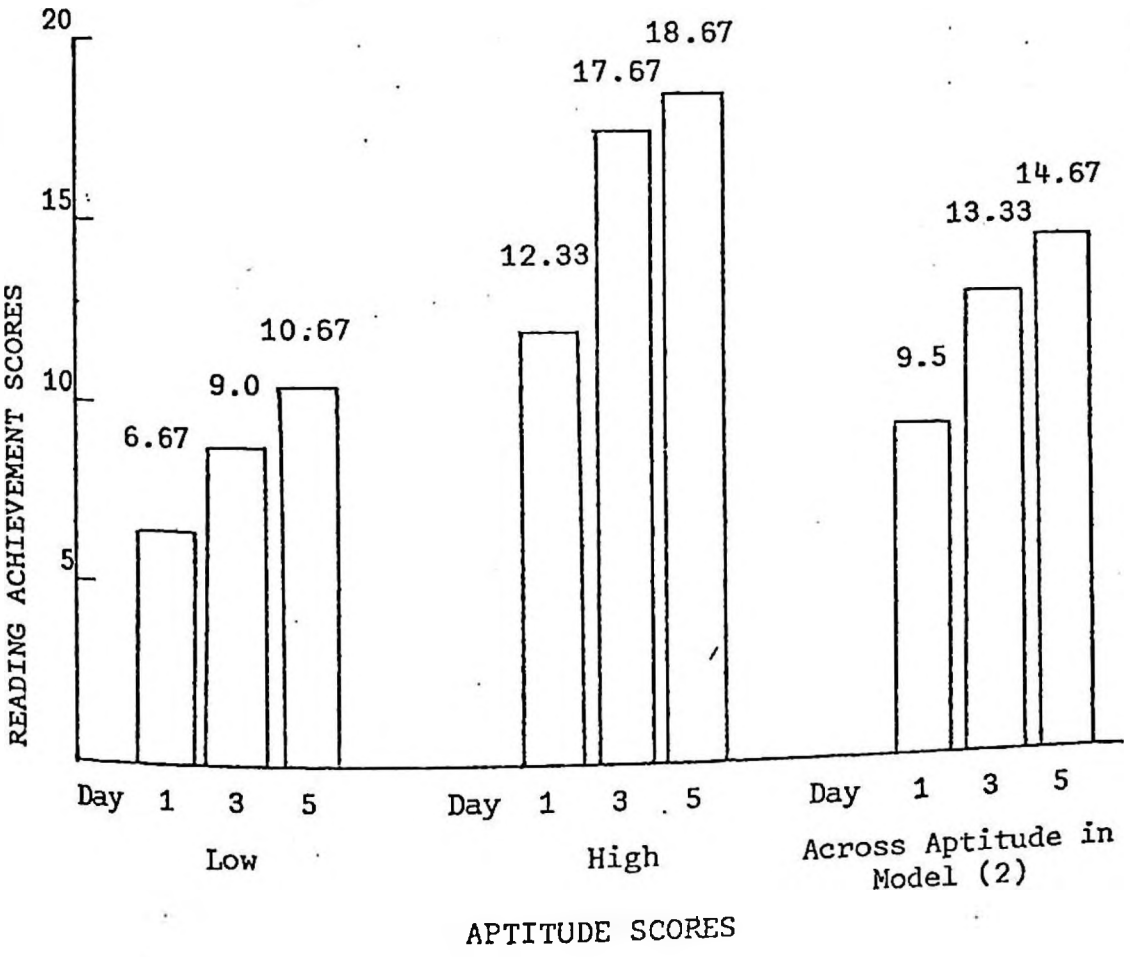


Figure 1. Mean of Reading Achievement Scores in Model (2) for both High and Low Aptitude Groups on Day 1, 3 and 5.

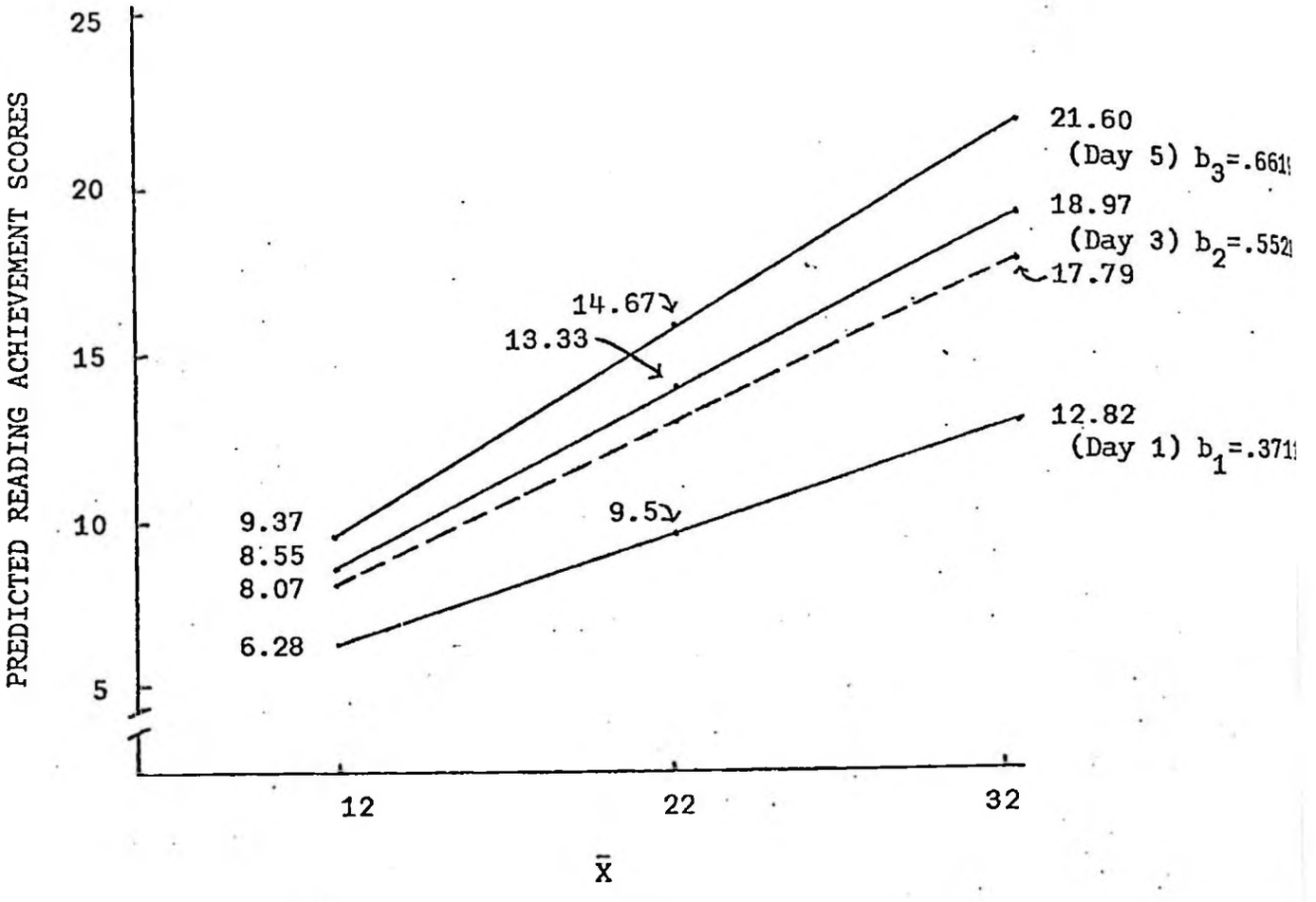


Figure 2. Regression Lines of the Predicted Scores on Reading Achievement and Aptitude Scores from Model (4) on Day 1, 3 and 5

Figure 1 shows the analysis of model (2) by illustrating the means for each of the high and low aptitude groups across days 1, 3 and 5. Figure 2 shows the three regression lines defined by model (4) between Aptitude and Achievement for days 1, 3 and 5.

In Figure 1, each bar represents the means of Reading Achievement for the three scores in the respective high-low groups for day 1, 3 and 5. Whereas, in Figure 2 the natural relationship between Aptitude and

Achievement is defined by a regression line determined by the six pairs of scores for each of days 1, 3 and 5. Although Figure 1 representing model (3) is a more simplistic representation of the data, Figure 2 is a more precise and accurate representation of the relationships, given that the relationship between Y and X is linear.

The day (A) main effect is seen in Figure 1 as the difference between means of the days 1, 3 and 5 across High and Low Aptitude. The same main effect in Figure 2 is seen by the differences among the predicted reading achievement score at the mean of aptitude score ($\bar{X}$) for the different days: 1, 3 and 5.

The Aptitude main effect is seen in Figure 1 as the difference between the means of the Low and High Aptitude groups across the three days; whereas in Figure 2 it is depicted by the common regression line between Aptitude and Achievement across days 1, 3 and 5 (dotted line).

The interaction between Day and Aptitude is depicted in Figure 1 by the differences between the increase in means from day 1, 3 and 5 for Low Aptitude as compared to the increase across Day for High Aptitude. That same interaction is more obviously shown in Figure 2 by the difference among Aptitude Achievement regression slopes among days 1, 3 and 5.

In addition to achieving a more logical and realistic representation of the data, another advantage of using a continuous factor is the increase of power of the statistical tests. This can be seen in the magnitude of sums of squares and the resulting F-tests when comparing the summary Tables 2 and 3.

Table 2

Anova of Data from Model (2)

Sources	SS	df	MS	F-ratios
Between	289.1741	5	57.8348	
B	249.4035	1	249.4035	25.0841
error (b)	39.7706	4	9.9427	
Within	97.3201	12	8.1100	
A	86.3436	2	43.1718	98.2070
A x B	7.4594	2	3.7297	8.4843
error (w)	3.5171	8	.4396	8.4843
Total	386.4978	17	22.7351	

Table 3

Anova of Data from Model (4)

Sources	SS	df	MS	F-ratios
Between	289.1777	5		
$\beta (X_2-\bar{X}_2)$	282.6458	1	282.6458	173.08
error (b)	6.5318	(4)	1.6230	
Within	97.7066	12		
A	86.3436	2	43.1718	182.39
A x $\beta (X_2-\bar{X}_2)$	9.4692	2	4.7346	20.0025
error (w)	1.8938	(8)	.2367	
Total	386.4978	(17)		

By using Aptitude as a continuous vector, rather than dichotomizing, we have increased the sum of squares of B from 249.4035 to 282.6458. Along the same line, the sum of squares of A x B increased from 7.4594 to 9.4692.

The increase of sums of squares by using B as a continuous vector is mainly due to the fact that the regression lines of the Aptitude scores on the criterion (Achievement) is a better fit than a comparison of the means between High and Low Aptitude groups.

Note that the Day factor sum of squares remains the same regardless of which analysis is chosen. This is because the repeated factor (Day) is independent of the continuous factor B (Aptitude Score).

The increase of the power of the F-tests is mainly due to the reduction of the error terms when the continuous B factor is used. In Table 2 and Table 3, the error of the between and within scores decreased in their sums of squares from 39.7706 to 6.5318 and from 3.5171 to 1.8938, respectively.

In consequence of the reduced error terms, the F-tests for this sample data increased drastically from 25.984 to 173.08 for the B effect, from 98.2070 to 182.39 for the A effect and from 8.4843 to 20.0025 for the A x B interaction. In spite of the use of potentially biased sample data, it is apparent that the use of a continuous between factor in its natural state results in an increase of power for the tests of hypotheses.

The variance components for the two models are shown in Table 4. The overall between subject source consists of the variability between the subject means ($\bar{Y}._{jk}$) and the overall mean ($\bar{Y}...$) for either model.

TABLE 4

Variance Components and Degrees of Freedom for the Two Analyses

Source	Dichotomized or Split B factor, Model (2)		Continuous B Factor, Model (4)	
	Variance Component	Degree Freedom	Variance Component	Degree Freedom
Between Subjects	$(\bar{Y}_{.jk}-\bar{Y}...)$	(JK-1)	$(\bar{Y}_{.-k}-\bar{Y}_{.-})$	(K-1)
B effect	$(\bar{Y}_{.j.}-\bar{Y}...)$	(J-1)	$(\bar{Y}'_{.-k}-\bar{Y}_{.-})$	(1)
Subjects (W/B)(residual)	$(\bar{Y}_{.jk}-\bar{Y}_{.j.})$	J(K-1)	$(\bar{Y}_{.-k}-\bar{Y}'_{.-k})$	(K-2)
Within Subjects	$(Y_{ijk}-\bar{Y}_{.jk})$	JK(I-1)	$(\bar{Y}_{i-k}-\bar{Y}_{.-k})$	(K)I-1)
A effect	$(\bar{Y}_{i..}-\bar{Y}...)$	(I-1)	$(\bar{Y}_{i-.}-\bar{Y}_{.-})$	(I-1)
A x B	$(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}...)$	(I-1)(J-1)	$(Y'_{i.k}-\bar{Y}_{.-k}-\bar{Y}_{i..}+\bar{Y}_{.-})$	(I-1)
A x Subjects	$(Y_{ijk}-\bar{Y}_{.jk}-\bar{Y}_{i..}+\bar{Y}...)$	J(K-1)(I-1)	$(Y_{i.k}-Y'_{i-k})$	(K-1)(I-1)

I = number of levels of A

J = number of levels of B

K = number of subjects within each B level

I = number of levels of A

K = number of subjects

No j subscript is needed: a- is used to hold that position.

$\bar{Y}'_{.-k}$ = estimate of the kth subjects mean performance across the I treatments from $b_{between}$

$Y'_{i.k}$ = estimate of the ith scores of the kth subject from b_{within}

The B effect of model (2) consists of the variability between the mean of a given B level ($\bar{Y}._j.$) and the overall mean ($Y...$), whereas for model (4) the variance is found in the deviation of the regression line of X and $\bar{Y}._{-k}$ (the subject mean) from the overall mean ($\bar{Y}...$). The between subject error for model (2) consists of deviation of subject mean ($\bar{Y}._{jk}$) from the B group means ($\bar{Y}._j.$), while the analogous error component for model (4) consists of residuals from the regression of X on $\bar{Y}._{-k}$.

Both the overall within subject effects and the A effect do not change from model (2) to model (4). The A x B effect for model (2) contains the discrepancies of the B effect across A, while in model (4) the A x B effect contains the variability of the simple I regression effect within day from the overall regression effect (X on Y). The within error term for model (2) contains variability between the score (Y_{ijk}) and the A mean ($\bar{Y}_i...$) after differences between subjects have been taken into account. The within error for model (4), on the other hand, consists of residuals from the within A regression effects after subject differences have been accounted for.

The number of degrees of freedom of model (2) does not change in model (4). The estimate of a single parameter is found in the numerator of the F-test for both models. In model (2) we find an estimate of $(\mu._1. - \mu._2.)$, whereas in model (4) the estimate is of $\beta...$ The degrees of freedom for the error between source of variance is obtained from the number of random observations of subject means ($\bar{Y}._{jk}$) minus the number of parameter estimates utilized to obtain such an error estimate. For model (2) the parameter estimates include the two B means $\bar{Y}._j.$, while

for model (4) they are the estimates for α and $\beta..$ from the overall regression effect of X on Y. In either case the example problem contains 1 and 4 degrees of freedom in the F-test for the B effect.

The A main effect for the two types of analyses does not change due to its independence of the B effect. Both the A x B interaction and the error within (A x subjects) are subject to change given the use of B as a continuous factor. The numerator of the F-test for the A x B interaction contains 2 degrees of freedom for both model (2) and model (4). Such estimates in model (2) include the three simple effect estimates for the B effect at a given A ($\bar{Y}_{i1\cdot} - \bar{Y}_{i2\cdot}$) minus the single dependent estimate for the B main effect ($\bar{Y}_{\cdot 1\cdot} - \bar{Y}_{\cdot 2\cdot}$); whereas for model (4), the two estimates include the three regression effects of X on Y at each A level, ($\beta_{j\cdot}$) minus the overall regression estimate for $\beta...$ The error term for the within subject effects, A x Subjects, contains random variability of all observations after the estimates of μ, $\mu_{\cdot j\cdot}$, $\mu_{i\cdot\cdot}$, $\mu_{\cdot\cdot k}$, and A x B interaction are taken into account. As indicated above, the type but not the number of parameter estimate of ($\mu_{\cdot 1\cdot} - \mu_{\cdot 2\cdot}$) is analogous to $\beta...$ If a continuous variable is broken into three or more categories, more degrees of freedom would need to be utilized in the B and A x B effects at the expense of the respective error terms. In such cases the number of parameter estimates and the respective degrees of freedom would be different for model (2) and model (4). General formulae for degrees of freedom are given in Table 4.

The design indicated in model (4) appears to be both a more realistic representation of data and a more powerful test when the B factor is continuous. There is little reason, other than conceptual and communicative

difficulty, why this scheme could not be utilized in more complex designs involving more than one between factor or other extensions. The greatest difficulty may be to overcome the traditional notion of using a continuous variable as a covariate where heterogeneity of regression is a restriction, whereas, in a design like that of model (4), heterogeneity of regression, A x B interaction, becomes a potentially interesting and testable hypothesis.

Bibliography

Corrigan, R. A scalogram analysis of the development of use and comprehension of "Because" in children. Child Development, 1975, 46, 195-201.

Dayton, C.M. The design of educational experiments. New York: McGraw-Hill Book Co., 1970.

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

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

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

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

Millar, S. Visual and haptic cue utilization by preschool children: The recognition of visual and haptic stimuli presented separately and together. Journal of Experimental Psychology, 1971, 12, 88-94.

Spielberger, C. D. & Smith, L. H. Anxiety (drive), stress and serial position effects in serial verbal learning. Journal of Experimental Psychology, 1966, 72, 589-595.

Vogel, P. P. A comparative survey of the morphology of lower class rural kindergarten students in Alachua County, Florida. Unpublished doctoral dissertation, University of Florida, 1970.

Ward, J. H. & Jennings, E. Introduction to linear models. Englewood Cliffs, N.J.: Prentice Hall, 1973.

Winer, B. J. Statistical principles in experimental design. (2nd Ed.) New York: McGraw-Hill, 1971.

Youniss, J. & Dennison, A. Figurative and operative aspects of children's inference. Child Development, 1971, 42, 1837-1847.

Footnotes

[1] : When the B factor, Achievement, is used as a continuous factor, the subscript J is no longer needed. A dash (-) is used to hold its place. Such notation is also used in Table 4.

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

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

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

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

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

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

THE UNIVERSITY OF AKRON
AKRON, OHIO 44325

TABLE OF CONTENTS