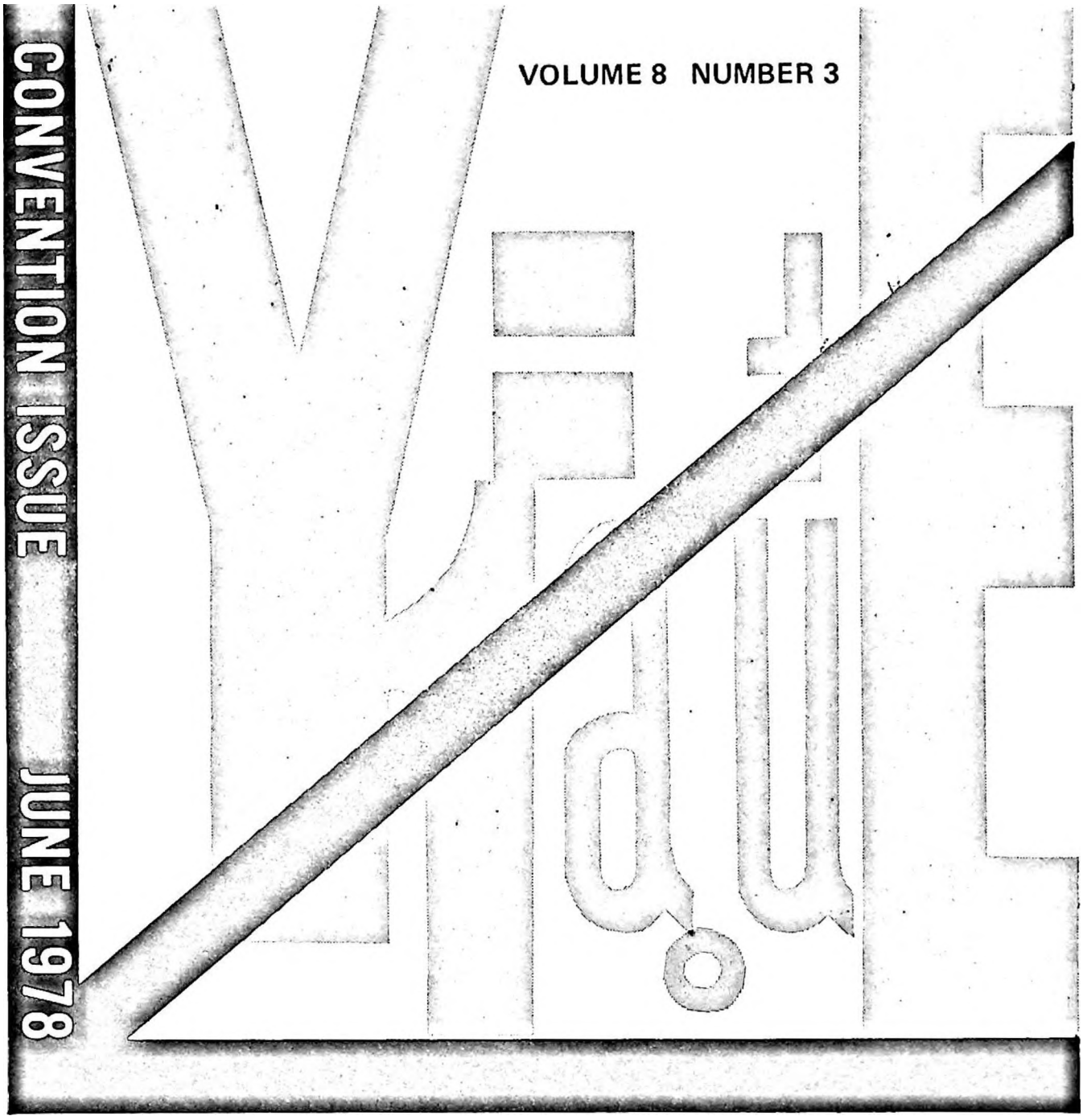
VOLUME 8 NUMBER 3

CONVENTION ISSUE JUNE 1978



MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group
on Multiple Linear Regression

MULTIPLE LINEAR REGRESSION VIEWPOINTS

Chairman ..... Michael McShane, American Association of Medical Colleges
One Dupont Circle, Washington, D.C. 20036

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

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

Secretary and Chairman-elect ..... Steve Spaner, Behavioral Studies,
University of Missouri, St. Louis, MO 63121

Cover by ..... David G. Barr

TABLE OF CONTENTS

TIPLE LINEAR REGRESSION VIEWPOINTS
8 NO.3 1978

RIDGE REGRESSION: A REGRESSION PROCEDURE FOR ANALYZING CORRELATED INDEPENDENT VARIABLES

ERNEST A. RAKOW
Memphis State University

Multiple linear regression is frequently applied in educational research, especially in large scale survey type studies (e.g. Coleman, 1966; Husen, 1967; Peaker, 1975; Madaus, Kellaghan and Rakow, 1975). A difficult problem encountered in such studies has been caused by multicolinearity of independent variables. This multicolinearity creates difficulties in deciding which variables to include in the regression equation, in estimation of the regression coefficients and in the interpretation of relative importance of the variables via these regression coefficients.

The purpose of this paper is to present ridge regression. Ridge regression is an analytic technique to be utilized when predictor variables in the multiple linear regression situation are highly correlated. Ridge regression may predict nearly as accurately as multiple regression as well as yielding regression coefficients which are more easily interpreted.

Multiple linear regression with highly correlated predictor variables has several pitfalls. One is that there may be a suppressor variable for which the regression coefficient and correlation with the criterion may have opposite signs. Interpretation of the importance of a suppressor variable to the equation is very difficult. A closely related pitfall is when two variables are highly correlated and one has a large regression coefficient while the other has a very small coefficient. These differences in magnitude may have little meaning in terms of

the research problem. Still another problem with multicolinearity is that the regression coefficients may be unstable, i.e. slight changes in the data may cause noticeable changes in the regression coefficients. A final difficulty in linear regression with highly correlated predictors is that although the errors of estimation have a small variance in the estimation sample, this variance generally is much larger on a cross-validation sample.

In classical multiple linear regression, a solution to some of these pitfalls has been to follow a step-wise selection procedure. Then, in addition to interpreting the regression coefficients, attention is focused upon the multiple correlation and upon changes in the multiple correlation squared as additional variables are added into the equation. In this type of interpretation, the variabl entered first (either by plan or analytic selection) is interpreted as having exaggerated importance since it is credited with contributing its own unique portion and that shared with all other variables entered later. Variables entered into the regression equation late in the process are given less importance because the portion they share with other variables is already included as part of those variables. To deal with this difficulty, Newton and Spurrell (1967) developed a system for identifying the proportion of unique and shared explained variances. Another major weakness with step-wise regression is that variables not selected for inclusion are assigned regression coefficients of zero while allowing the coefficients for variables selected to be larger than they should be. In such situations the resulting regression coefficients may be very difficult to interprete in that research situation.

Ridge regression analysis (as developed by Draper, 1963; Hoerl, 1962; Hoerl and Kennard, 1970a, 1970b and first presented to educators by Williams, 1975) presents an alternative procedure. It does not select variables, but rather utilizes all variables in the model. This places the burden of variable selection upon the researcher and his development of an adequate theoretical model. For a pair of highly correlated predictor variables, ridge regression would use a portion of each in the predictions. In effect, this procedure equalizes the predictive effects of highly correlated variables. Consequently, this procedure rejects no variables and tends to have no variables with opposite signs for the regression coefficient and correlation.

Ridge regression is performed by adding a constant to the diagonal of the correlation matrix and then calculating the standardized regression coefficients and multiple correlation squared. This may be done iteratively by adding the constant and generating the regression coefficients for a range of constants. The addition of a small positive number to the diagonal functions to reduce the effect of the multicolinearity in the predictor variables. Marquardt and Snee (1975) point out that the lowest constant with stable regression coefficients identifies the best ridge solution. A low constant is desired because as the constant is increased the multiple correlation decreases (i.e. the standard error of estimate increases). Different values from 0.0 to 3.0, in small increments, may be tried as the constant. When zero is used as the constant, the result is identical to multiple linear regression.

Some have pointed out that the regression coefficients in ridge regression are biased estimates but have smaller standard errors than classical regression

(Marquardt and Snee, 1975 and Hoerl, Kennard and Baldwin, (1975). The gain in reduction of error for these coefficients more than compensates for the loss introduced by the bias. The smaller standard errors indicate the regression coefficients are more stable.

The data analyzed to demonstrate an application of ridge regression in this paper are school district means from a statewide assessment program. The 566 school districts with both fourth and seventh grades were systematically divided into two samples by placing every second district into Sample A and the others into Sample B. The same analyses were run on both samples to enable examination of the stability of results from one sample to another and to allow for cross-validation. The first example utilizes the fourth grade district means on four achievement subtests to predict seventh grade Composite Achievement. The second example utilizes five teacher variables to predict seventh grade Composite Achievement. The third example utilizes the nine predictors from the two prior examples (four achievement subtests and five teacher variables) to again predict seventh grade Composite Achievement. The high degree of multicolinearity between the achievement subtests and between the teacher variables would lead to unsatisfactory results with classical multiple linear regression because the regression coefficients would be uninterpretable. These examples demonstrate how this difficulty is met by ridge regression analysis.

In the first example, school district mean level of achievement as measured at grade seven by Composite Achievement is predicted by fourth grade mean levels of achievement as measured by tests of Word Relationships, Reading, English and Mathematics. The correlations among these predictor variables range from a low

of .80 for Mathematics and Reading in Sample A to a high of .90 for Word Relationships and Reading in both Samples A and B. These predictor variables are highly correlated.

Ridge regression analysis analyzes the correlation matrix to obtain estimates of the squared multiple correlation and standardized partial regression coefficients. In the analysis for this paper, the same constant is added to all diagonal elements in the correlation matrix and a regression equation is determined. The values used for this constant ranged from 0.00 to 0.20 in increments of 0.02, from .20 to 1.00 in increments of 0.05 and from 1.00 to 2.50 in increments of 0.10. The many small values are utilized because that is when the estimates show the greatest fluctuation. It should be noted that when the constant is 0.00 a classical multiple regression is obtained.

Figure 1 presents the results of this ridge regression analysis for Sample A. The lines illustrate the changes in the standardized regression coefficients for a variable in relation to changes in the constant added to the diagonal. Since none of the lines cross, the variables remain in the same order of importance. However, the magnitude of some beta weights changes dramatically. The initial beta for Reading is 0.257 when the constant k = 0, but it decreases rapidly to .180 when k = .20 and to .144 when k = .50. Word Relationships also has a large beta initially (.237) which also decreases rapidly to .139 when k = .50. In contrast, the beta for English begins negative (-.069), changing to positive (.003) when k = .08 and reaches a maximum of .077 when k = .90. The line for Mathematics is not similar to any of the others. Its first beta is .079, increases to .090 when k=.50 and then tails of slightly. Examination of this figure leads to the use of k = .50 as perhaps best. From this point the lines are nearly straight.

Since the multiple correlation squared tends to decrease as the constant increases the lowest value of k associated with the straight portions of the lines is best.

Figure 2 presents the corresponding ridge regression analysis for Sample B. For Sample B, Mathematics has a large initial beta (.409). Beta for Mathematics decreases rapidly to .315 for k = .10, .269 for k =.20 and .205 for k = .50. Betas for Word Relationships and English begin at approximately .04 and increase to .11 or .12. Beta for reading drops initially, but stabilizes quickly. Note that in both Figure 1 and Figure 2, when k = 0 some variables have large regression weights while others are small. The effect of ridge regression is to reduce this variability in the weights of correlated variables. Table 1 presents the multiple correlation squared and standardized regression coefficients for three arbitrarily selected values of k.

In the second example, school district mean level of Composite Achievement is predicted by five teacher variables. They are (1) professional instructional staff per 1000 students (INSTRS/K), (2) Teachers per 1000 students (TCHRS/K), (3) years of teaching experience (TCHEXPER), (4) teacher's annual salary (TCHSALRY and (5) percent of teachers with a masters degree (%MA). These variables also are school district means. The correlations among these predictor variables range from -.23 for INSTRS/K with TCHSALRY to a high of .97 for INSTRS/K with TCHRS/K. Both extreme correlations happen to occur in Sample A.

Figure 3 presents the Sample A results of these teacher variables being used as predictors of grade seven achievement. The betas for INSTRS/K and TCHRS/K are of particular interest. For multiple regression their betas are -.675 and .79 respectively. Under ridge regression however, these estimates change rapidly so that when k = .20 these betas have become -.035 and .160. When k = .40 the beta for INSTRS/K becomes positive. The betas for %MA and TCHSALRY also change considerably. When the constant reaches .50 these betas seem to have stabilized.

After this point, the lines all appear relatively straight. Also, the betas for °/oMA and TCHRS/K are nearly equal for all large values of the constant.

Figure 4 presents the corresponding Sample B results. These results are very similar to those for Sample A. Again, INSTRS/K and TCHRS/K change very rapidly for small values of the constant. However, for Sample B the squared multiple correlation is small, initially being .050 and then decreasing.

For the third example the four district mean levels of achievement at grade four are combined with the five district means of teacher variables for prediction of composite achievement at grade 7. These results are in Figure 5, Figure 6 and Table 1.

Figure 5 presents the standardized regression coefficients derived from Sample A. The shape of the line for each variable's beta is very similar to that of each variable in Figure 1 or Figure 3. The major exception is Mathematics which in Figure 5 begins with a beta of -.096. Word Relationships begins with a higher beta of .413 as compared to .237 in Figure 1. Note that after the betas stabilize (about k = .80) some lines are nearly coincident because the betas are so nearly equal, e.g. Reading with Word Relationships, English with Mathematics and with TCHEXPER, and INSTRS/K with °/oMA. These nearly equal standardized regression coefficients would seem to have intuitive meaning. It is counter instructive for district mean levels of fourth grade achievement to have negative weights for prediction of seventh grade achievement. Following ridge regression these variables would have positive betas. Also, the extremely different betas for INSTRS/K and TCHRS/K appear meaningless for two variables so similar in construction. Ridge regression provides more similar, but still unequal, weights for them.

Figure 6 presents the standardized regression coefficients for the same nine variables but from Sample B. Once again the shape of the line for each variable's

betas is very similar to that of the same variable in Figure 2 or Figure 4. The beta for English is initially much lower, but quickly reaches its level of about .11. Word Relationships begins near its desirable level of .12. Again Reading, English and Word Relationships end with nearly coincident lines. The teacher variables INSTRS/K and TCHRS/K appear to have rather different patterns of betas than in Figure 4. Specifically they are much more similar to each other and TCHRS/K starts out below INSTRS/K where previously it began with very large betas.

The beta weights from ridge regression with k = .50 (or even .80) appear to be more easily interpretable on an intuitive level than those from multiple regression. The betas from two different samples are also more similar when k = .50. The price paid for this increase in interpretability appears to be smaller squared multiple correlations. However, according to Hoerl, Kennard and Baldwin (1975) they should more accurately reflect the findings in a cross-validation.

A cross-validation study was done with this data. Every set of standardized regression coefficients calculated via the ridge regression analysis was applied to the standardized scores from both Sample A and Sample B. These results are summarized via the squared multiple correlations presented in Table 2. In the columns headed "Ridge" are squared multiple correlations calculated as part of the matrix manipulations during the ridge regression analysis. The columns headed "Cross" contain squared multiple correlations calculated by predicting standardized scores for each district and dividing the variance of predicted scores by the variance of the actual standardized scores. In all cases the squared multiple correlations decreased as the constant added to the diagonal increased.

When the betas from Sample A were applied to Sample B for prediction, the squared multiple correlations from the cross-validation were less than those from the ridge regression for all of Example 1 and Example 2 and for most of Example 3. The one exception is in Example 3 when k = 0 so that classical multiple regression

betas are utilized. In this case the cross-validation has a higher squared multiple correlation.

When the betas from Sample B were applied to Sample A for prediction, the squared multiple correlations from the cross-validation were less than those from the ridge regression for all of Example 1 and Example 2 and for much of Example 3. For Example 3 the cross-validation yielded higher squared multiple correlations when utilizing the values of the constant from .00 to .35. Note the very sizeable difference from .394 to .532 when cross-validating the betas from classical multiple regression. The reason for this finding when the constant is small and only in this one example is not understood at this time.

When the betas from a sample are applied to the standardized scores in the same sample, it should be possible to verify the squared multiple correlations calculated from the ridge regression. When the constant is zero, the two estimates are equal. As the constant increases so does the difference in the two squared multiple correlations, always favoring the analytical estimate.

The implications of the findings presented here are that researchers using highly correlated independent variables should be cautious when interpreting the regression coefficients resulting from multiple linear regression. In such situations ridge regression could be utilized as an exploratory tool to search for regression coefficients which may yield more meaningful interpretations. However, given the results of the cross-validation, it would seem that further exploratory work concerning the nature of ridge regression needs to be carried out before it is widely applied.

Table 1. Standardized Regression Coefficient From Ridge Regression

SAMPLE A	Constant Added	Multiple Correlation Squared	Word Relation-Ships	Reading	English	Math	Instr. Staff per 1000 Students	Teachers per 1000 Students	Tchrs Years Exp.	Teachers Salary	Percent Teachers w/ Masters
Example 1	.00	.240	.237	.257	-.069	.079					
(Grade 4 Achievement	.50	.204	.139	.144	.071	.090					
	.80	.189	.123	.127	.077	.088					
Example 2	.00	.158					-.675	.797	.197	-.270	.356
(Teacher Variables)	.50	.073					.012	.101	.155	-.096	.115
	.80	.060					.022	.081	.128	-.064	.087
Example 3	.00	.394	.413	.274	-.094	-.096	-.720	.810	.255	-.292	.205
(Grade 4 Achievement	.50	.268	.151	.150	.076	.066	.004	.095	.160	-.125	.048
+ Teacher Variables)	.80	.241	.129	.129	.081	.074	.016	.077	.128	-.090	.035
Correlation with Grade 7 Achievement			.477	.476	.430	.429	.118	.169	.230	-.026	.142
SAMPLE B											
Example 1	.00	.381	.038	.159	.039	.409					
(Grade 4 Achievement)	.50	.325	.110	.125	.121	.205					
	.80	.302	.110	.120	.118	.175					
Example 2	.00	.050					-.212	.237	.160	-.156	.142
(Teacher Variables)	.50	.026					-.009	.047	.110	-.069	.041
	.80	.021					-.001	.037	.091	-.051	.029
Example 3	.00	.417	.136	.175	-.041	.402	.004	-.135	.095	-.231	.130
(Grade 4 Achievement	.50	.342	.124	.127	.115	.203	-.032	-.033	.066	-.085	.031
and Teacher Variables)	.80	.315	.112	.121	.115	.174	-.027	-.022	.057	-.063	.019
Correlation with Grade Achievement			.555	.566	.569	.605	.039	.078	.159	-.052	.043

Table 2. Squared Multiple Correlations from Ridge Equations and from Cross-Validation

		Example 1		Example 2		Example 3	
	Constant	Ridge	Cross	Ridge	Cross	Ridge	Cross
A applied to B	.00	.381	.378	.050	.042	.417	.437
	.25	.348	.320	.033	.021	.371	.343
	.50	.325	.280	.026	.014	.342	.293
	.75	.305	.249	.022	.010	.319	.256
	1.00	.288	.222	.019	.008	.300	.227
B applied to A	.00	.240	.238	.158	.211	.394	.532
	.25	.219	.203	.092	.076	.302	.312
	.50	.204	.179	.073	.051	.268	.255
	.75	.192	.158	.062	.038	.245	.217
	1.00	.181	.142	.054	.030	.226	.189
A applied to A	.00	.240	.240	.158	.158	.394	.394
	.25	.219	.202	.092	.066	.302	.258
	.50	.204	.177	.073	.045	.268	.213
	.75	.192	.157	.062	.033	.245	.184
	1.00	.181	.140	.054	.026	.226	.161
B applied to B	.00	.381	.381	.050	.050	.417	.417
	.25	.348	.322	.033	.023	.371	.337
	.50	.325	.283	.026	.016	.342	.292
	.75	.305	.251	.022	.011	.319	.257
	1.00	.288	.224	.019	.009	.300	.228

Figure 1. Sample A Fourth Grade Tests Predicting Seventh Grade Composite Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.

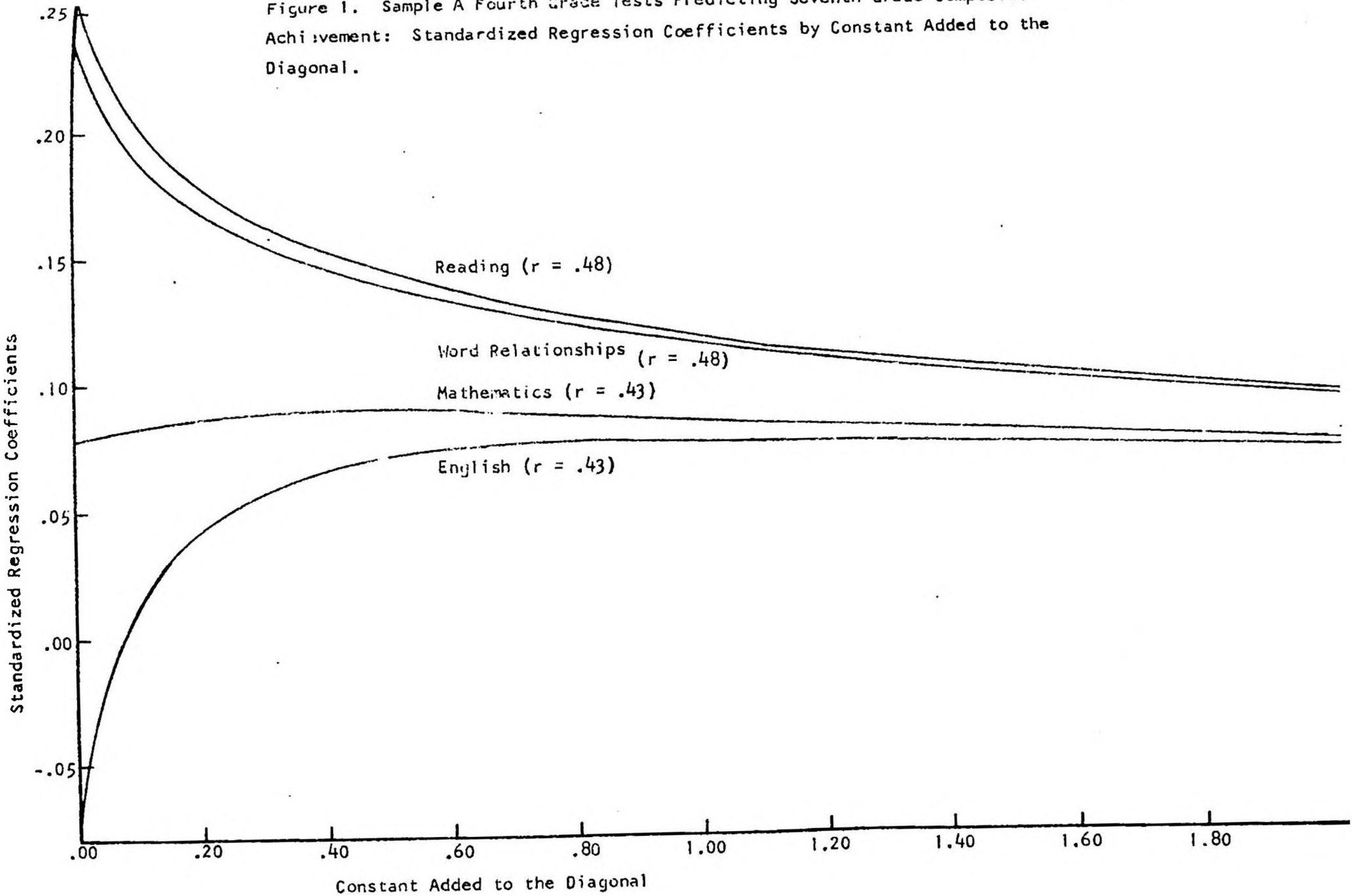


Figure 3. Sample A Teacher Variables Predicting Seventh Grade Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.

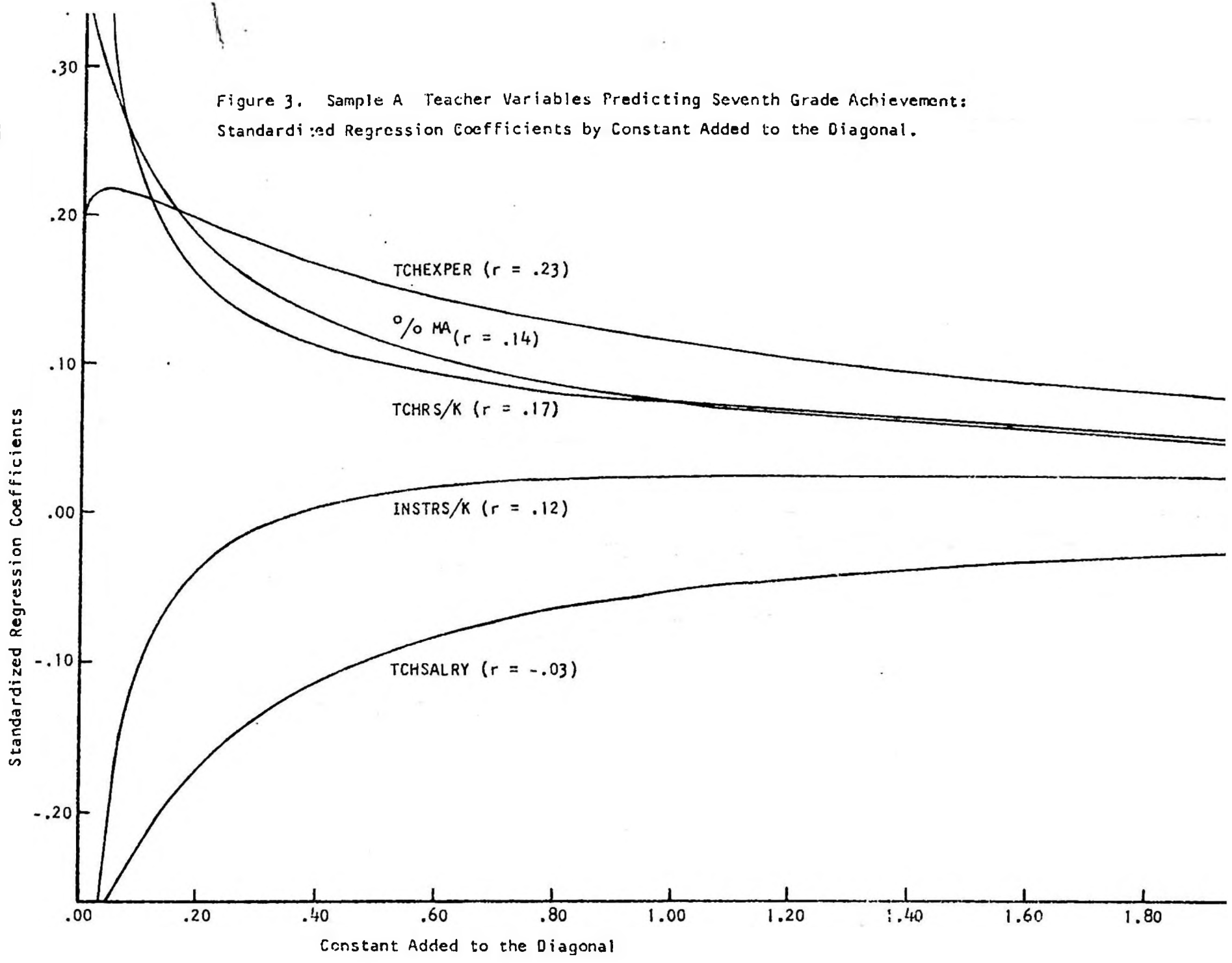


Figure 4. Sample B Teacher Variables Predicting Seventh Grade Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.

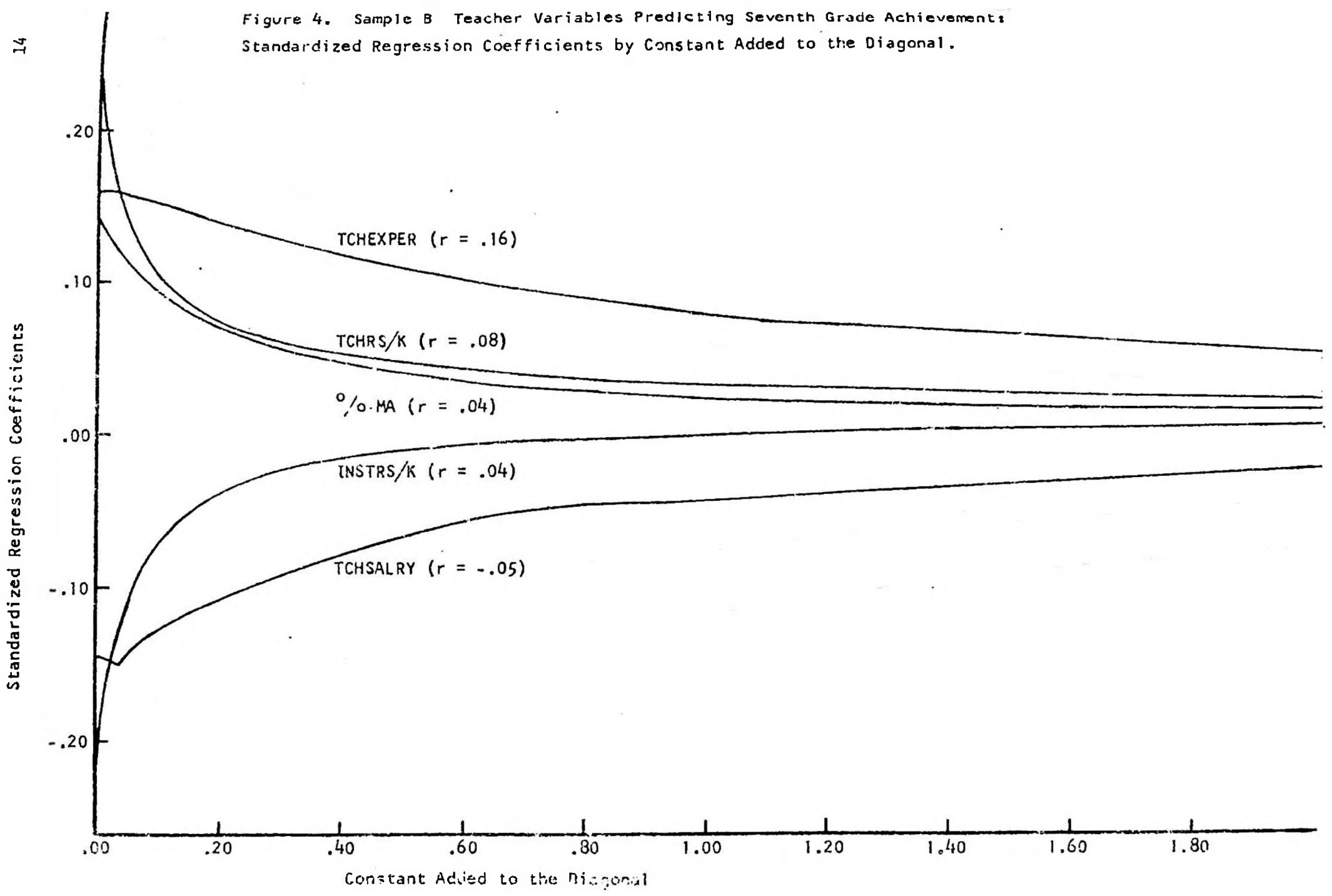


Figure 5. Sample A Teacher Variables and Fourth Grade Tests Predicting Seventh Grade Achievement: Standardized Regression Coefficients by Constand Added to the Diagonal.

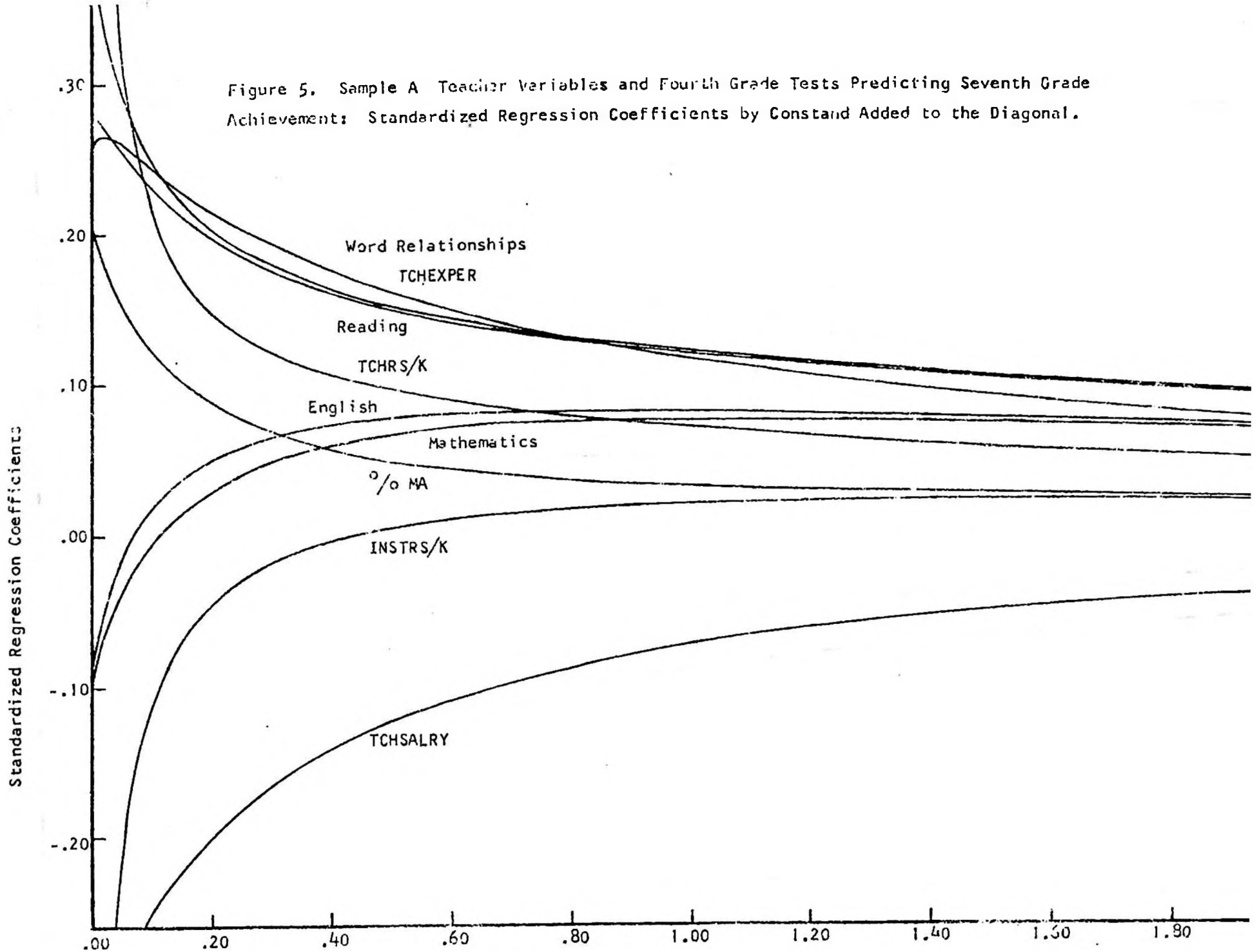
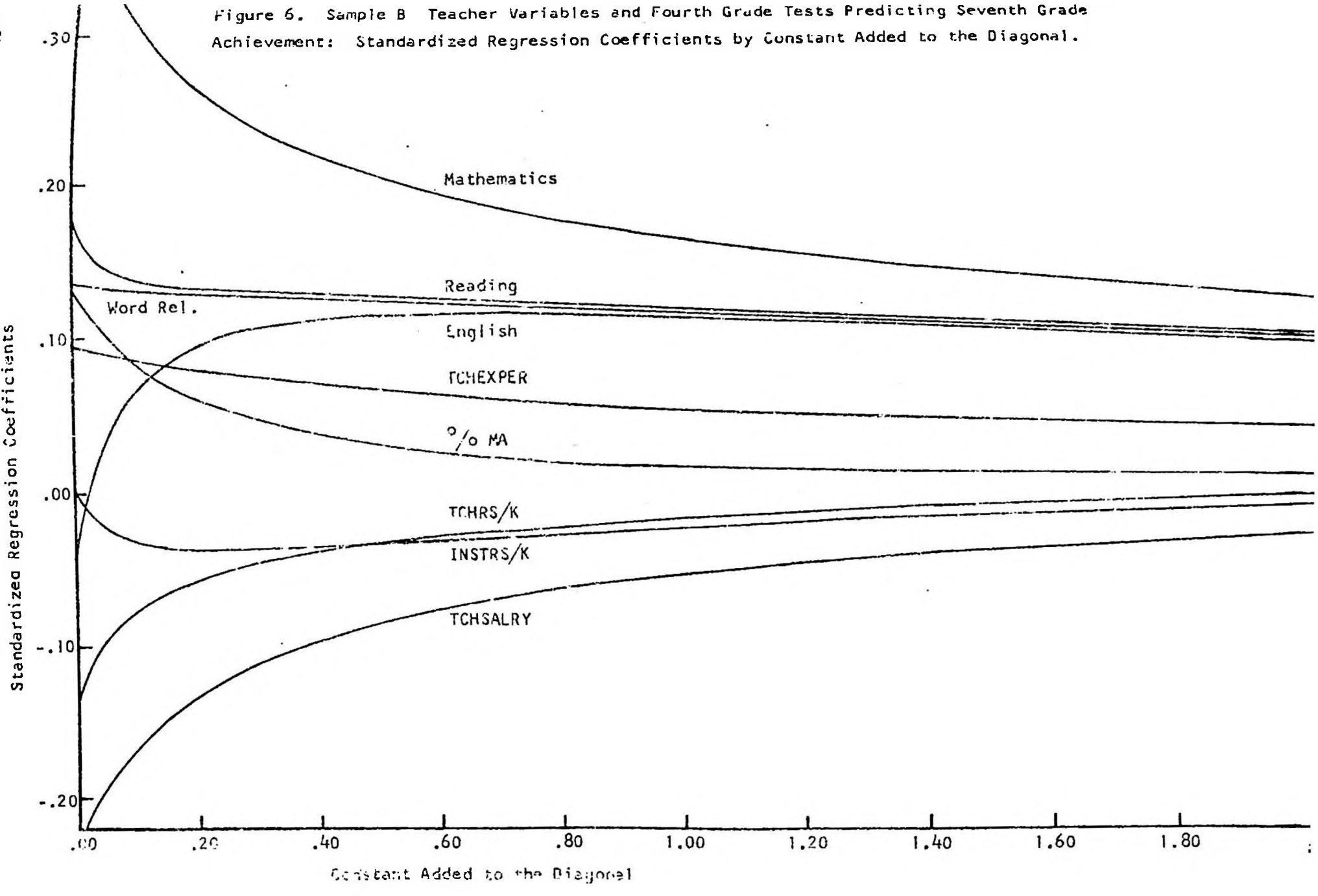


Figure 6. Sample B Teacher Variables and Fourth Grade Tests Predicting Seventh Grade Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.



REFERENCES

Coleman, James S., et al., Equality of Educational Opportunity, Washington D.C.: Office of Education, U. S. Department of Health, Education and Welfare, 1966.

Draper, N. R., "Ridge Analysis"of Response Surfaces. Technometrics, 1963, 5., 469-479.

Hoerl, Arthur E., Application of Ridge Analysis to Regression Problems. Chemical Engineering Progress, 1962, 58, 54-59.

Hoerl, Arthur E. and Kennard, R. W., Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 1970a, 12, 55-67.

Hoerl, Arthur E. and Kennard, R. W., Ridge Regression: Applications to Nonorthogonal Problems. Technometrics, 1970b, 12, 69-82.

Hoerl, Arthur E., Kennard, R. W., and Baldwin, K. F., Ridge Regression: Some Simulations. Communications in Statistics, 1975, 4, 105-123.

Husen, Torsten, International Study of Achievement in Mathematics: A Comparison of Twelve Countries. New York: John Wiley + Sons, 1967.

Madaus, G. F., Kellaghan, T. and Rakow, E. A., A Study of the Sensitivity of Measures of School Effectiveness. Report submitted to the Carnegie Corporation, New York, December, 1975.

Marquardt, D. W. and Snee, R. D., Ridge Regression in Practice. The American Statistician, 1975, 29, 3-19.

Newton, R. G. and Spurrell, D. J., A Development of Multiple Regression for the Analysis of Routine Data, Applied Statistics, 1967, 16, 51-64.

Newton, R. G. and Spurrell, D. J., Examples of the Use of Elements for Clarifying Regression Analysis. Applied Statistics, 1967, 16, 165-172.

Peaker, G. F., An Empirical Study of Education in Twenty-One Countries: A Technical Report. New York: John Wiley + Sons, 1975.

Williams, S. H., An Application of Ridge Regression to the Prediction of Enrollment in Higher Education. Paper presented at Mid-South Educational Research Association Annual Convention, November, 1975.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 8 NO. 3 1978

AN ANALYSIS OF THE HISTORICAL REGRESSION METHOD OF PREDICTING POSTTEST GRADE EQUIVALENTS FOR CATEGORICALLY-AIDED PROGRAMS

THOMAS L. HICK AND DAVID J. IRVINE
New York State Education Department

Abstract

An Analysis of the Historical Regression Method of Predicting Posttest Grade Equivalents for Categorically-Aided Programs

THOMAS L. HICK and DAVID J. IRVINE

New York State Education Department

Historical Regression follows directly from the assumption that, without specific intervention, growth will continue at the rate (grade equivalents per year of schooling) obtained at the time of the pretest. When compared with program-level data (n = 213) it was found that Historical Regression underestimated final achievement for short programs with older children. It overestimated for younger children in long programs. An alternative method was developed which eliminated the bias, removed half of the error, and eliminated much computation since an expected achievement level for each child was not required.

An Analysis of the Historical Regression Method of Predicting Posttest Grade Equivalents for Categorically-Aided Programs

T. L. HICK and D. J. IRVINE
New York State Education Department

T. L. Hick
Prekindergarten Evaluation
State Education Department
Room 562 EBA
Albany, New York 12234

Objectives

To eliminate maturation as a factor in the pretest-posttest design at least one State (New York) recommends a conversion of the pretest to anticipate posttest scores when data are in grade equivalents from standardized tests. This conversion is known as Historical Regression, and follows directly from the assumption that, without specific intervention, growth will continue at the rate (grade equivalents per year of schooling) obtained at the time of pretest.

The District Evaluator's Handbook describes the procedure for obtaining anticipated scores by the following steps:

Step 1. Obtain each pupil's pretest grade equivalent.
Step 2. Subtract 1 (since most standardized tests start at 1.0).
Step 3. Divide the figure obtained in step 2 by the number of months the pupil has been in school to obtain a hypothetical (historical regression) rate of growth per month. (Ignore kindergarten months. One school year = 10 months.)
Step 4. Multiply the number of months of Title I treatment by the historical rate of growth per month.
Step 5. Add the figure obtained in step 4 to the pupil's pretest grade equivalent (step 1).*

This paper examines the Historical Regression method for obtaining predicted posttest grade equivalents to determine its adequacy as a predictive model and to develop an alternative predictive model.

Method

The approach was to (1) express Historical Regression in algebraic terms, (2) from the algebraic formula, produce a linear model for

*Bureau of Urban and Community Programs Evaluation. District Evaluator's Handbook of Selected Evaluation Procedures for Categorically Aided Programs Serving Disadvantaged Learners. Albany: New York State Education Department, Spring 1972.

Historical Regression (this model has assumed weights), (3) produce a Least Squares Historical Regression model that has weights which best fit the data, (4) compare the Historical Regression model with the Least Squares Historical Regression model to determine similarities and differences, (5) develop an alternative model.

Data Source

The data were taken from the reports of Title I Compensatory Education programs filed at the New York State Education Department for the 1972-73 school year. This file contained the following information for 213 programs:

Y - Program mean reading grade equivalent on posttest.
B - Program mean reading grade equivalent on pretest if over grade equivalent of 2.
D - Duration of program in years.
T - Mean previous time spent in school in years.

Characteristics of these variables are reported in Table 1.

Table 1
Means, Standard Deviations and Limits

Variable	M	SD	Minimum	Maximum
Y	4.67	1.34	2.10	9.2
B	3.83	1.23	2.03	7.7
D	.90	.32	.20	1.6
T	6.69	5.72	1.00	11.0

Is the Historical Regression algorithm adequate?

In terms of the defined variables the algebraic expression of the Historical Regression was found to be:

$$Y = B + (B - 1) \times (D/T)$$

In the form of a linear model this Historical Regression Model became:

Model 1. $Y = \text{Zero } U + 1\ (B) + 1\ (B \times D/T) - 1\ (D/T) + E_1$

The weights on the variables were dictated by the Historical Regression procedure. A model with the same variables used in the Historical Regression model but with the weights left for a least squares best fit was expressed as a Least Squares Historical Regression Model:

Model 2. $Y = a_0 U + a_1 (B) + a_2 (B \times D/T) + a_3 (D/T) + E_2$

The solution to the Least Squares Historical Regression model yielded: $a_0 = .68$, $a_1 = .99$, $a_2 = .59$ and $a_3 = -1.02$. The Historical Regression Model assumes that these weights are: $a_0 = 0$, $a_1 = 1$, $a_2 = 1$, $a_3 = -1$. The error sum of squares for the Least Squares model is 43. Upon substitution of the historical regression values, the error sum of squares for the restricted model is 85. With 5 degrees of freedom for the full model and none for the restricted model, the difference between the error sum of squares is tested with F at 5/208 degrees of freedom. The F of 38 is associated with a near zero probability of a chance difference.

Therefore, the Historical Regression Model (1) does a poorer job of predicting posttest grade equivalents than the Least Squares Historical Regression Model (2).

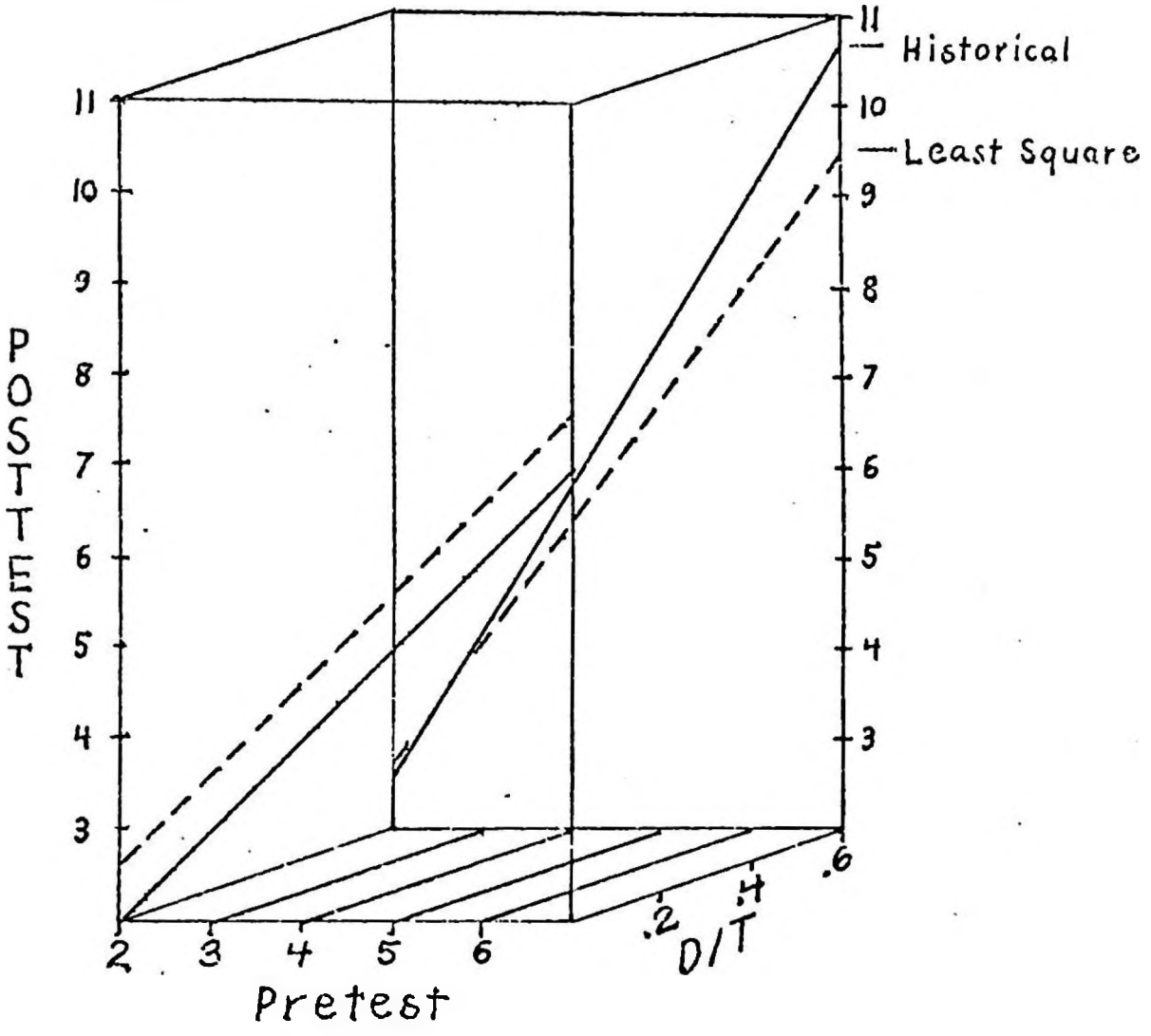
Since the Historical Regression deviates from a Least Squares best fit, Historical Regression was plotted as shown in Figure 1. It may be observed that initial grade equivalents are along the horizontal axis while the ratio of the program duration to the average time the students had previously spent in school is given on the right side of the cube. The historical regression plane twists so that the relationship of pre-test to posttest is 1 to 1 as the ratio D/T approaches zero. This would be the situation of a very short program (e.g., 2 months) for children that had spent a long time in school (e.g., 50 months) yielding a ratio of D/T of .04, a value very close to zero. In this situation, according to historical regression, the children will have the same score at the end of program as at the beginning.

At the other end of the D/T ratio, the situation is shown where the program has lasted as long as the child has previously been in school. This would be the case for young children. At an initial grade equivalent of 2, the final achievement would be approximately 2.4 grade equivalents. At an initial grade equivalent of 3, the final achievement would be approximately 4.0 grade equivalents.

This historical model does fit the data quite well covering 88 percent of the variability in posttest scores, provided it is centered and twisted properly. As previously shown, however, the historical regression model does not match the least squares fit to the data.

Figure 1

TWO TECHNIQUES FOR ESTIMATING POSTTEST SCORES



A Detailed Examination of the Historical Regression Algorithm.

Each of the Historical Regression weights was tested in turn with either the hypothesized weight or the least squares weight retained in accord with the result of testing the historical regression weight. First, the $a_0 = 0$ hypothesis was tested. This hypothesis is that the regression plane passes through zero on the posttest when the pretest is zero and the D/T ration is zero. This turned out to not be the case. The least squares regression plane does not pass through zero at a very high level of probability ($F = 476$ at 1/209 degrees of freedom). The least squares value elevates the plane .68 grade equivalents above zero. The second hypothesis was that $a_2 = 1$. This weight describes the amount of twist in the plane. The weight for twist was not one ($P = 0$). A weight of .59 fits the data much better so that the twist is not as large as assumed by historical regression. The third hypothesis was that the tilt of the plane was 45 degrees when D/T was zero. This was the hypothesis that a_1 was equal to one. Test of this hypothesis failed to reject it, so that a value of one for a_1 was an acceptable fit by least squares ($P = .28$). The last hypothesis that a_3 is -1, also was not rejected ($P = .37$).

The final model is:

$$Y = .68 + B + \frac{((.59B) - 1) \times D}{T}$$

The differences between this equation and the algorithm of historical regression are the addition of .68 to all projected scores and the multiplication of the pretest score by .59 before adjusting it for the beginning at grade 1.

Examination of the difference between the Historical Regression plane in Figure 1 and the least squares best fit plane reveals that the final achievement of older students in short duration programs has been underestimated by as much as 1/2 a year. Younger students in long programs did not fare as well. The effect was to overestimate their achievement by as much as one year, giving the appearance of poor performance.

Development of an Alternative Model

It would seem desirable to simplify the formulation of the expected achievement without losing the predictive power of historical regression. An attempt at this goal was undertaken beginning with this possible model:

$$Y = a_0U + a_iB + a_2D + a_3B^2 + a_4D^2 + a_5B^3 + a_6D^3 +$$

$$a_7BxD + a_8B^2D + a_9BD^2 + a_{10}B^2D^2 + E_3$$

The R^2 for the above model was .896 and is obviously significant. The hypothesis of no curvilinear interaction, was rejected using Model 4:

$$Y = a_0U + a_1B + a_3B^2 + a_4D^2 + a_5B^3 + a_6D^3 + a_7BxD + E_4$$

which yielded an R^2 of .890. The difference between R^2's was significant ($F = 4.12$, df 3/202, $P = .008$); therefore curvilinear interaction was considered as present. The task of locating the interaction remained. The hypothesis that the interaction was very complex was tested with model 5 which was the same as model 3 except for the last term B^2D^2. Complex interaction was considered not present ($F = 4.57$, df = 1/202, $P = .03$) as all relationships are tested at the .01 level of significance. Therefore, model 5 became the full model for a test of the hypothesis that the interaction is curvilinear on program duration. Model 6 expressed this hypothesis:

$$Y = a_0U + a_1B + a_2D + a_3B^2 + a_4D^2 + a_5B^3 + a_6D^3 + a_7BxD + a_8B^2D + E_6$$

This model turned out an R^2 of .889 which was significantly different from model 5 at the .007 level of significance ($F = 7.51$, df = 1/203). Therefore, the term BD^2 was considered significant and included in the next model (7) that was used to test the effect of B^2D. This term was not considered significant ($R^2 = .892$, $F = 2.64$, df 1/203, $P = .10$).

To regain our bearings, the full model is now (#7) which includes third degree polynomial forms:

$$Y = a_0U + a_1B + a_2D + a_3B^2 + a_4D^2 + a_5B^3 + a_6D^3 + a_7BxD + a_9BD^2 + E_7$$

Comparison of this model with a model that lacked the third degree polynomials showed that B^3 and D^3 were not significant ($F = .80$, df 2/204, $P = .45$). The only remaining term not involved in the established interaction was B^2. It was found to be nonsignificant ($F = 1.38$, df = 1/206) and dropped from the final equation.

All the remaining terms were involved in the expression of the interaction and were therefore considered necessary for the expression of the relationship of beginning scores and program duration on posttest scores. The final acceptable model is:

$$Y=-2.14 +1.56B +7.26D -4.04D^2 -1.39BxD +.80BD^2 +E_8$$

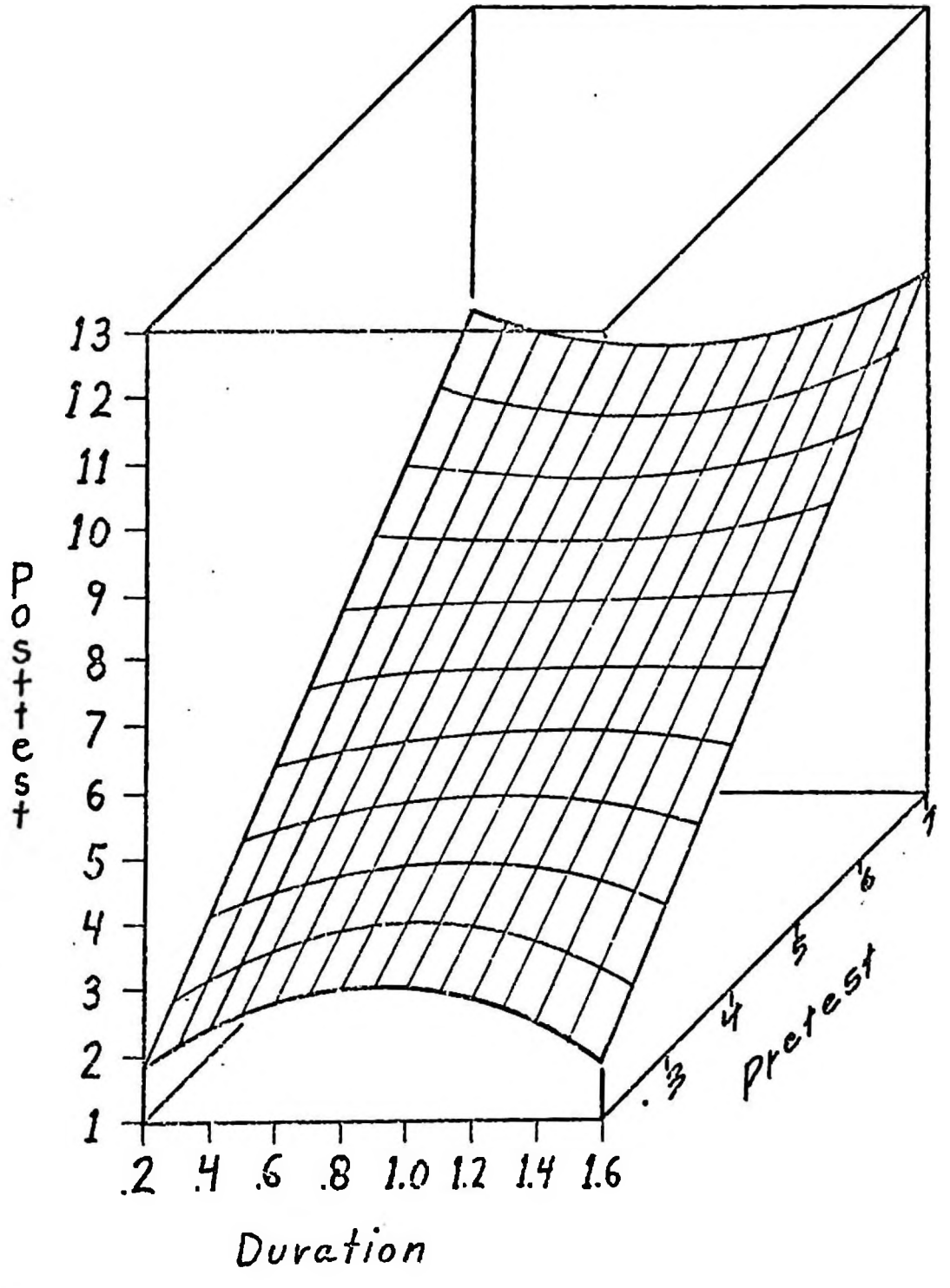
This is a highly complex plane as it describes a different effect for program duration at various levels of the pretest scores. It should be noted that program duration, although highly significant, has an only very subtle relationship with posttest scores after initial skill level is included. This plane, described by the final acceptable model, is shown in Figure 2.

Conclusion

A method for eliminating maturation as a factor in pretest-posttest designs for Title 1 programs was examined in relation to data and an alternative predictive model was developed. The alternative method: (1) makes no assumptions regarding the relationship between pretest level and program duration, (2) does not require the computation of time span in school for each child, (3) does not require an expected achievement level for each child, (4) does not bias against some programs, (5) cuts the error in half, (6) requires only the beginning mean achievement level.

FIGURE 2

LEAST SQUARES MODEL FOR USING PRETEST LEVEL AND PROGRAM DURATION TO ESTIMATE FINAL ACHIEVEMENT FOR THE PROGRAM



MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 8 NO. 3 1978

NEIGHBORHOOD PREDICTORS OF READING ACHIEVEMENT IN SIX BIG CITY SCHOOL DISTRICTS: A PATH ANALYSIS *

CRIS R. KUKUK
DANIEL U. LEVINE
JEANIE KEENY MEYER

The purpose of this study was to examine the direct and indirect effects of neighborhood characteristics on school-level reading achievement and to identify some of the policy implications of the results. Neighborhood characteristics examined in the study included race, family structure, income, and density.

Some previous studies have suggested that socially-disadvantaged racial groups such as black Americans have lower school achievement than whites after controlling for the effects of economic status (e.g. Ogbu, 1978; Pascal, 1977). A number of other studies (e.g. Coleman, 1966; Mayeske, et al., 1973) have indicated that race has little or no independent effect on achievement after controlling for socioeconomic background measured by variables such as income level or percent below the poverty level. Studies of these kinds sometimes have been interpreted as suggesting that educational disadvantages of low-income minority groups can be eliminated by removing the social barriers which impede minority opportunity (e.g. by desegregating schools and/or ensuring equal employment opportunity) and/or by other action to raise their income levels. However, relatively few studies have inquired how race and socioeconomic status relate to other neighborhood characteristics such as family structure and density in affecting academic achievement in the public schools. The present study uses path analysis to explore these relationships.

* Paper presented at the Annual Meeting of AERA, March 27-31, Toronto

SAMPLE AND DATA

Neighborhood data for each city in the study were taken from 4th count tract-level reports from the 1970 Census. Depending on the city, tracts were allocated to schools either according to lists supplied by school district officials or by visual matching of tract and school boundary maps. In order to enhance the reliability of the data, schools serving tracts with less than 1,000 population were eliminated from the study.

Data on grade-level reading achievement were obtained from central office personnel in six big city school districts: Chicago, Illinois; Cincinnati, Ohio; Cleveland, Ohio; Houston, Texas; Kansas City, Missouri; and St. Louis, Missouri. For each city, data were obtained and coded for all elementary schools which included grade six and which reported achievement scores for most of the years included in the data set available for that city. Data analysis reported in the paper used sixth-grade reading achievement means for the following years:

Chicago	- 1970
Cincinnati	- 1972
Cleveland	- 1974
Houston	- 1974
Kansas City	- 1970
St. Louis	- 1973

HYPOTHESES

The first hypothesis explored in the study is that the relationship between race and achievement is eliminated when income, family structure, and density are used as intervening variables, and thus that there is no direct link between race and achievement.

The second hypothesis is that the zero-order relationship between income and achievement is eliminated when race, family structure, and density are controlled, and thus there is no direct link between income and achievement.

Implicit in the preceding hypotheses, family structure and density are viewed as transmitting the effects of income and race on achievement.

ANALYSIS

As expected race and income are highly correlated with achievement among the schools in our six-city sample. As shown in Table 1, these correlations range from -.48 to -.63 for race and achievement, and from .62 to .82 for income and achievement. The major question at issue is whether these relationships will be eliminated when path analysis is used to add controls for the effects of family structure and density.

The strictures of path analysis in its simplest form require that variables be ordered into a set of recursive equations with no feedback loops denoting two-way causation. Obviously this constitutes an oversimplification of reality but is useful for analytic purposes. The variables are so ordered in Figure 1 and the resulting system of equations is given below where:

Race (R) is the percent of blacks in the neighborhood population.

Income (I) is the percent of families three times or more above the poverty level.

Family Structure (F) is the percent of adult females who are separated.

Density (D) is the percent of occupied housing units which have 1.5 or more people per room.

Achievement (A) is a sixth-grade reading comprehension mean score in grade-equivalent or percentile units, depending on the city.

(1) $R = e_1$

(2) $I = \beta_{IR} R + e_2$

(3) $F = \beta_{FI.R} I + \beta_{FR.I} R + e_3$

(4) $D = \beta_{DF.IR} F + \beta_{DI.FR} I + \beta_{DR.FI} R + e_4$

(5) $A = \beta_{AD.FIR} D + \beta_{AF.DIR} F + \beta_{AI.DFR} I + \beta_{AR.DIF} R +$

The B's are standardized partial regression coefficients (Betas) where the first subscript stands for the dependent variable, the second subscript stands for the associated predictor variable and those variables to the right of the dot are controlled for. Thus $\beta_{AR.DIF}$ in the last equation is the standardized partial regression coefficient for achievement predicted by race controlling for density, family status and income. If our theory is correct, that particular B will be effectively zero.

Before reporting the results, however, we should describe a simple equation showing how we tested the first hypothesis. This is the widely-used multiple regression approach for testing the difference between a restricted model and a full model to determine whether variables used in

the full model lead to statistically different results than those in the restricted model:

Full Equation: $A = \beta_{AR} R + e_1$

Restricted Equation: $A = e_2$

$H_0: \beta_{AR} = 0$

$H_1: \beta_{AR} = 0$

We reject the null hypothesis H_0 in favor of the alternative hypothesis H_1 if the associated F statistic is statistically significant. Table 1 shows that the relationships between race and achievement are highly significant.

A three variable model with two predictor variables is tested in a comparable manner. For example, testing for statistically significant effects of race on achievement controlling for income involves the following equations.

Full equation: $A = \beta_{AI.R} I + \beta_{AR.I} R + e_1$

Restricted Equation: $A = \beta_{AI} I + e_2$

$H_0: \beta_{AR.I} = 0$

$H_1: \beta_{AR.I} \neq 0$

Table 3 shows the appropriate Betas for this test. One city (Houston) shows no significant effect of race on achievement when income is controlled five cities (all but Chicago) show no significant effect of race on achievement when family structure is controlled, and all six cities show a direct effect of race on achievement controlling only for density.

As shown in Table 2, all of the relationships between race and other neighborhood characteristics in each of the six cities are statistically significant. This implies that race is highly related to income, family

structure, and density. Generalizing the latter three variables as X, the relationship between X, race, and achievement can be tested as follows:

Full equation: $A = \beta_{AX.R} X + \beta_{AR.X} R + e_1$

Restricted equation: $A = \beta_{AX} X + e_2$

$$H_0: \beta_{AR.X} = 0$$

$$H_1: \beta_{AR.X} \neq 0$$

It would be possible to consider all possible models with three predictor variables, but we choose to skip that step and go to the analysis of the model depicted in Figure 1. This amounts to testing for the statistical significance of each of the B's in equations (3) through (5). The B's for equation (3), (4) and (5) are shown in Tables 4, 5 and 6. The schematic representation of all significant relationships for each of the cities is illustrated in Figure 2. Notice in Table 6 and Figure 2 that in only one of the six cities (Cincinnati) does race have a significant statistical relationship with achievement when income, family status and density are controlled for. This means that by and large hypothesis 1 is supported.

A similar analysis can be conducted to determine whether income has a direct effect on achievement using the path model shown in Figure 1. Results of such an analysis also are given in Table 6 and Figure 2. When controlling for race, family structure, and density, income has a strong independent association with achievement only in Houston and St. Louis. In Chicago it has a barely significant Beta of .243, and in two (Cincinnati and Cleveland) of the remaining three cities where the Beta for income is not significant, this statistic was so small it could not be calculated using the SPSS regression procedure.

Taken together and accepting the causal assumptions built into our path model and our choice of variables and procedures in the analysis, these

results suggest that race and income have few direct effects on achievement but instead work together and through family structure and density to influence achievement levels in big city elementary schools. The importance of density and, particularly, family structure as influences on public school achievement is shown by the fact the former has four significant Betas and family structure has two of the highest three Betas (-.679 and -.743 for Cleveland and St. Louis, respectively) in the full model for which data are reported in Table 6.

Another noteworthy point in the data involves the close connection between race and family structure. As shown in Table 2, the zero-order correlations between these two variables generally are the highest in our data set. As shown in Table 3, controlling for family structure eliminates the association between race and achievement for five of the six cities in the sample, leaving it barely significant only in Chicago where the number of schools is very high. These findings suggest that much of the effect of race on school achievement is transmitted through the forces that result in the characterization of black neighborhoods as being very high in social or family disorganization (as proxied by percent of adult females separated).

DISCUSSION

The major finding of this study is that neighborhood level measures of family structure and density transmit some of the effects of minority racial status and low income level (also measured at the neighborhood level) on achievement in big city schools.

This finding should not be interpreted as meaning that female-headed families (which correlates over .9 with percent of females separated) and high-density housing units necessarily "cause" low achievement. It is

possible that everything else equal, children from female-headed families and or from high-density housing units perform less well academically than children from husband-wife families and/or from low-density housing units. (This might happen because two-parent families are able to give a child more time and attention on the average than can one-parent families, because students in high-density housing units have less space to study than do those in low-density units, or for other reasons.) However, the data reported in this study do not deal directly with such possibilities because they do not describe relationships between family structure or housing density and achievement among individual students.

How, then, should one interpret the relationships and causal hypotheses investigated in the study? All of our variables are highly interrelated: big city neighborhoods which have a high percentage of black residents tend to be low in income and to have high percentages of female-headed families, overcrowded housing units, and low achieving students. Other variables in our data set which correlate very highly with these characteristics include high crime rate and high proportion of deteriorated housing. We believe that such variables taken together denote high levels of social and family disorganization. School achievement is low in neighborhoods which have a large amount of social disorganization.

Elsewhere we have presented data indicating that high density is not associated with low school achievement in high income neighborhoods (Meyer and Levine, 1978) and that achievement in high-income predominantly-black neighborhoods tends to be high (Levine, Kukuk, and Meyer, 1978). Findings such as these indicate that it is the combination of characteristics such as high density, high incidence of female-headed families, low income, and

high percentage of minority residents which denotes a high degree of social disorganization in big city neighborhoods. We also have found that some predominantly low-income white neighborhoods are high on social disorganization defined by variables other than race. Thus it is reasonable to conclude that it is the concentration of poverty residents--particularly those from racial or ethnic minority groups--which results in general social disorganization, and to expect that variables such as family structure and density will transmit much of the association between race or income and achievement.

It is no surprise, of course, to be told that concentrated urban poverty is associated with low school achievement, particularly among segregated minority groups which have been excluded from opportunity in the larger society. Implications for policy, however, are not always viewed in the light of associated measures of social disorganization. Unless social policy deals with the neighborhood-level and institutional-level aspects of educational and other problems in big city poverty neighborhoods, it may have little impact on the long-range situation of the poor and the neighborhoods they inhabit. Providing job training for the poor may have little impact if neighborhood conditions are such that trainees are unable to acquire adequate skills or to put them to good use at work. Providing desegregated schooling for young people in these neighborhoods may not help them much if neighborhood conditions detract from the likelihood that they will succeed in school. Providing more income to the big city poor may have only a temporary and minor impact on the quality of their lives during the next generation if neighborhood characteristics continue to be implicated as a major cause of very low achievement among their children in the public schools.

It should be noted that we have not identified the neighborhood "causes" which work together with race and income to result in low school achievement. We can speculate that such causes may include a wide range of factors such as dysfunctioning public schools, negative peer influences on children and youth, existence of an anti-social "street culture," and political powerlessness among the very poor in big cities. Because we have not specifically investigated such possibilities, our only definite conclusion is that social policies which take account only of race and income are neglecting important determinants of low school achievement and other aspects of social disorganization in big city poverty neighborhoods.

Finally, it should be noted that it would have been much more difficult or impossible to adequately test our hypothesis using methods other than multiple-regression path analysis. It would have been possible to segment our interval variables into categories to be used in the construction of five dimensional tables, and then apply ANOVA or other related techniques to test for main effects and interactions. To do so, however, would have sacrificed the interval-ratio attributes of the variables and thereby made the detection of significant relationships more unlikely. In addition, ferreting out of the directions of relationships is considerably more cumbersome with tables.

FOOTNOTES

[1]For a few schools in our sample, the percentage of black students in the student body is substantially different from the percentage of black residents in the neighborhood a school serves. However, separate analysis for two of the cities on which we have student enrollment data showed the same results as those reported in this paper. In general, student enrollment in the cities in our sample is drawn from the local neighborhood.

[2]One of the methodological decisions we made was to statistically transform predictor variables which consistently had curvilinear relationships with achievement in all the cities. Following this decision, the family structure and density variables used in the analysis were logged. Other important decisions involved selection of the variables to measure income, density, and family structure. Additional analysis we have been conducting shows that these decisions can affect the conclusions obtained in the path analysis. These matters are discussed further in Levine, Kukuk, and Meyer (1977).

REFERENCES

Levine, Daniel U., Cris Kukuk, and Jeanie Keeny Meyer (1977). "Concentrated Poverty Neighborhoods and Academic Achievement in Big City Schools." Kansas City, Missouri: University of Missouri - Kansas City Center for the Study of Metropolitan Problems in Education.

Meyer, Jeanie Keeny and Daniel U. Levine (1978). "The Relationship of Density to School Achievement Within the Context of Neighborhood Type." Paper presented at the annual meeting of the American Educational Research Association, Toronto, March.

Mayeske, George S., et al. (1973). A Study of the Achievement of Our Nation's Students. Washington, D. C.: U. S. Government Printing Office.

Coleman, James S., et al. (1966). Equality of Educational Opportunity. Washington, D. C.: U. S. Government Printing Office.

Ogbu, John U. (1978). Minority Education and Caste: The American System in Cross-Cultural Perspective. New York: Academic Press.

Pascal, Anthony (1977). "What Do We Know About School Desegregation?" California: The Rand Corporation, p. 5777.

TABLE 1

Zero Order Regression Coefficients (Betas) Between Achievement and Race and Between Achievement and Income*

	City					
	Chicago	Cincinnati	Cleveland	Houston	Kansas City	St. Louis
Race	-.522	-.636	-.480	-.627	-.594	-.620
Income	.742	.791	.624	.822	.758	.729

*All relationships significant at the .05 level. Race was measured by percent of black population in elementary school attendance areas. Income was measured by percent of families three time or more above the poverty level.

TABLE 2

Zero Order Regression Coefficients (Betas) Between Race and Other Neighborhood Variables*

Variable	City					
	Chicago	Cincinnati	Cleveland	Houston	Kansas City	St. Louis
Income	-.519	-.483	-.430	-.733	-.588	-.640
Family Structure	.786	.775	.758	.796	.763	.814
Density	.498	.338	.469	.594	.581	.631
N	240	59	108	138	61	67

*All relationships significant at the .05 level.

TABLE 3

Partial Regression Coefficients (Betas) Between Race and Achievement Controlling for One Other Intervening Variable

Intervening Variable	City					
	Chicago	Cincinnati	Cleveland	Houston	Kansas City	St. Louis
Income	-.188*	-.332*	-.264*	-.051	-.283*	-.258*
Family Structure	.194*	.026	.166	-.080	.089	-.006
Density	-.208*	-.405*	-.232*	-.258*	-.302*	-.360*

*Significant at the .05 level.

TABLE 4

Partial Regression Coefficients (Betas) Between Family Structure and Race or Income, Each Controlling for the Other

Variable						
	Chicago	Cincinnati	Cleveland	Houston	Kansas City	St. Louis
Race	.515*	.494*	.519*	.289*	.367*	.449*
Income	-.524*	-.582*	-.554*	-.692*	-.662*	-.570*
R^2	.905	.861	.824	.855	.864	.853

*Significant at the .05 level.

TABLE 5

Partial Regression Coefficients (Betas) Between Density and Each of the Other Three Neighborhood Variables Controlling for the Other Two

Variable	City					
	Chicago	Cincinnati	Cleveland	Houston	Kansas City	St. Louis
Race	-.169*	-.441*	-.069	-.186*	.018	-.172
Income	-.477*	-.539*	-.436*	-.811*	-.689*	-.500*
Family Structure	.533*	.670*	.462*	.232*	.171	.593*
R^2	.717	.838	.648	.796	.730	.863

*Significant at the .05 level.

TABLE 6

Partial Regression Coefficients (Betas) Between Achievement and Each of the Four Neighborhood Variables Controlling for the Other Three

Variable	City					
	Chicago	Cincinnati	Cleveland	Houston	Kansas City	St. Louis
Race	.061	-.440*	.122	-.096	-.096	.027
Income	.243*	.000	.000	.588*	.205	.561*
Family Structure	-.437*	.060	-.679*	.056	-.265	-.743*
Density	-.228*	-.719*	-.193	-.243*	-.308*	.502*
R^2	.647	.820	.563	.690	.644	.635

*Significant at the .05 level.

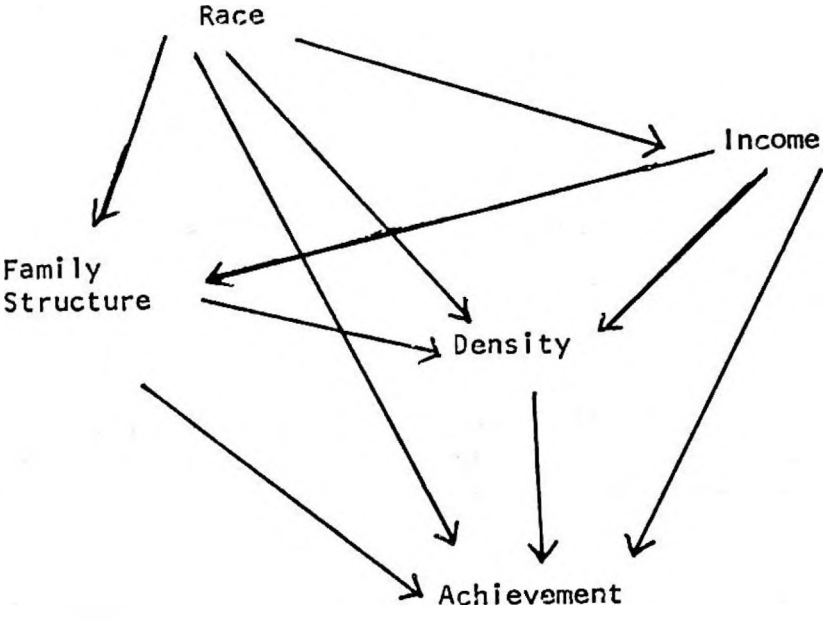


Figure 1. Path analysis model used in the study.

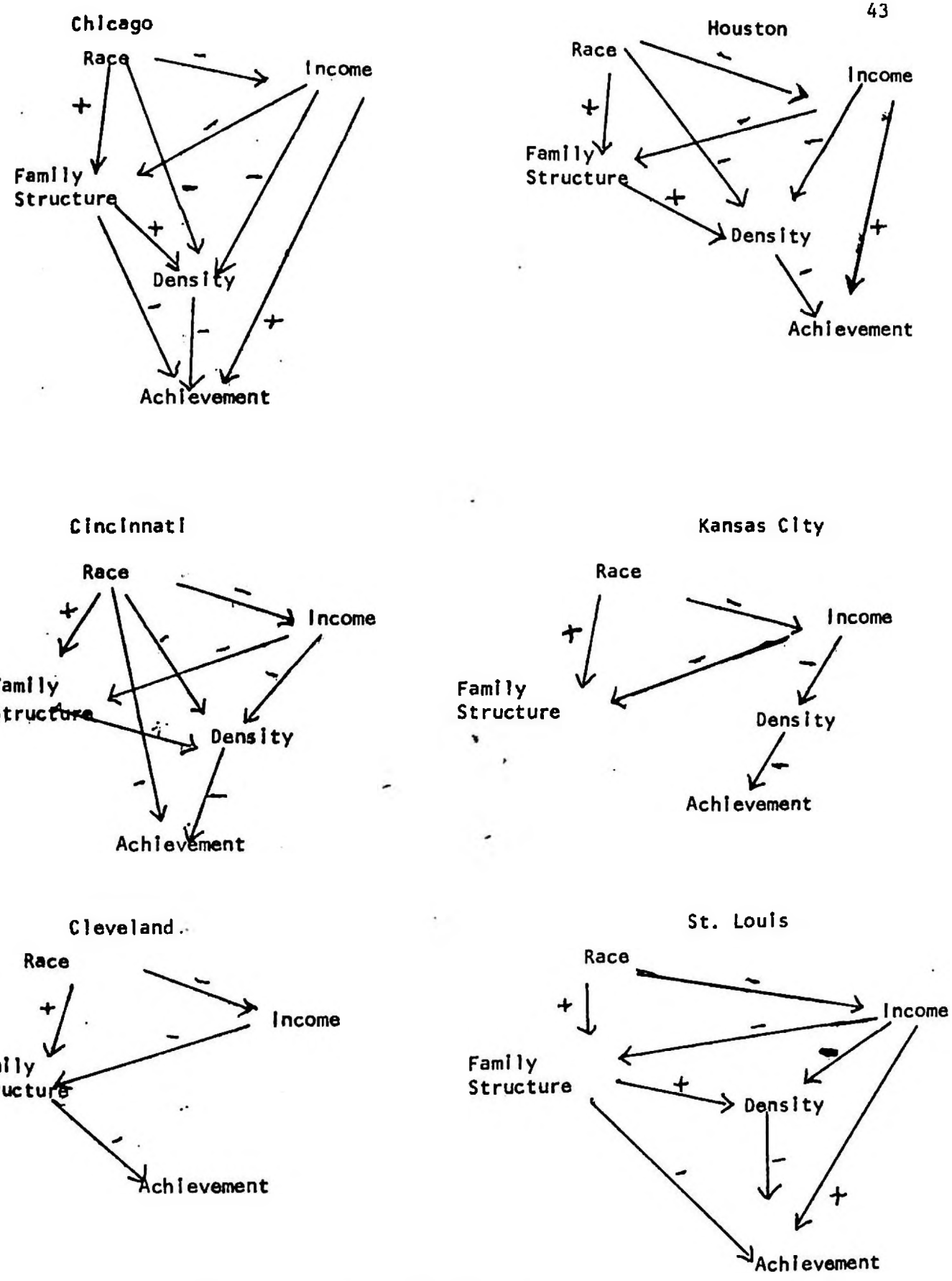


Figure 2. Significant paths to achievement and direction of influence.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 8 NO. 3 1978

EVALUATION OF SEX-RELATED SALARY DISCRIMINATION

P. KENNETH MORSE
Medical College of Georgia

ABSTRACT

Using constructed data, the use of multiple regression is demonstrated for "School A", where salaries are fair but where women have been hired only recently, and for "School B", where there is evidence of sex-related bias in salary. The regression analysis identifies the presence or absence of salary bias, although mean salary by sex presents a different picture.

With virtually every institution of higher education receiving some form of aid from the federal government (and thus subject to Affirmative Action requirements), the ability to identify any sex-related salary discrimination has become increasingly important. However, the very compliance with Affirmative Action guidelines in the employment and promotion of women may tend to increase the apparent relationship of sex to salary level, because faculty salaries are heavily determined by variables (such as rank and years-in-rank) which are often confounded with sex, especially where women are receiving a "catch-up" share of new appointments and promotions.

Probably the most common method of displaying salary data is by computing the mean (or, occasionally, the median) of all salaries within a given category. Often this is supplemented by some measure of variability, if no more than listing the highest and lowest values found in the category. This method is appropriate when the sample is large relative to the number of relevant

category combinations or "cells" and when each "cell" is populated by an adequate sub-sample. However, as the number of relevant variables (and categories within variables) expands, the number of cells expands geometrically. Consider a school with a faculty of 300, which has 10 departments, faculty of 2 sexes, 3 different ranks, and 3 different levels of earned degree (we will say nothing about full-time vs. part-time, time-in-rank, research productivity, or any of the other considerations thought to be related to salary). This school would have to sort its 300 faculty into 10 x 2 x 3 x 3 = 180 discrete combinations of just these 4 variables! Clearly, this is not practical. We typically see figures for mean salary by rank, or by rank by sex. The purpose of the present paper is to demonstrate how such means can be misleading, and how the true state of affairs can be more nearly approximated by the use of multiple linear regression.

METHOD

Faculty staffing patterns were designed for two simulated institutions, "School A" and "School B" as shown in Table 1. The variables in Table 1 were coded as follows for each simulated faculty member:

Variable	Name	Coding
Rank level 1	RANK1	1 if in this rank, otherwise 0
Rank level 2	RANK2	1 if in this rank, otherwise 0
Rank level 3	RANK3	1 if in this rank, otherwise 0
Years in rank 1	TIME1	Number of years if currently in this rank, otherwise 0
Years in rank 2	TIME2	Number of years if currently in this rank, otherwise 0
Years in rank 3	TIME3	Number of years if currently in this rank, otherwise 0
Sex	SEX	1 if male, otherwise 0

and salaries were computed using the formula below:

School A: Salary = 10,000 + 2000(RANK2) + 6000(RANK3) + 300(TIME1)
+ 500(TIME2) + 750(TIME3)

School B: Use same formula but add + 500(SEX).

Since salaries are not usually completely determined by two or three variables, a random error was added to each salary by randomly selecting (from a table of random z-scores) a z-score and multiplying it by 1000. Since z-scores can be either positive or negative, addition of this random error could either increase or decrease a given salary. This random error serves as a proxy for all the other unmeasured variables that help determine salary. For School A the random error had a mean of -4.58 and a standard deviation of 984.93. For School B, the mean and standard deviation were 144.56 and 1065.78 respectively. Neither mean differed significantly from zero (the expected mean of this variable).

"Salaries" generated by the specifications above were recorded and means and standard deviations were computed by school, sex, and rank level. Separately by school, salary was regressed on the variables used in computing salary (except for random error). Separate regressions were computed for each school, both with and without the sex variable. The difference in R^2 produced by adding the sex variable was tested for significance.

RESULTS

Table 2 shows the mean and standard deviation of salaries by school, sex and rank. Within schools, there appears to be a difference due to sex in School A, although we saw that School A was actually free of sex-related bias. We also note that School B appears to be free or nearly free of bias, although there was a systematic bias in the computation of the salaries. Clearly, the data in Table 2 could be misleading.

The regression data, however, as shown in Table 3, successfully detect the true state of affairs. School A is found to be free of bias ($F_{(1,113)} = .404$,

$P > .10$) and School B's bias is duly noted ($F_{(1,113)} = 10.674$, $P < .005$).

DISCUSSION

The data in Table 2 for School A might well be typical of real institutions that have dedicated themselves to Affirmative Action goals by a hiring policy that seeks to increase the proportion of female faculty. Since most newly hired faculty come in at entry level, the very fact of meeting these goals may increase the apparent discrimination in salary. Similarly, the data in Table 2 for School B may be typical of schools which have resisted the promotion of women to higher ranks, and thereby created an apparent absence of bias when measured by mean salary levels.

Since the purpose of this paper was solely to demonstrate the superiority of the multiple regression technique over the simplistic analysis of means, rank was used as one of the variables in the regression analysis. However, rank is often tainted by the same sex-related discrimination as salary; in fact, failure to promote on an equitable basis may a major cause of salary inequity. In the case of School B, the regression analysis (with rank as a predictor) underestimates the total sex-related discrimination. When analyzing real data, rank and other status variables (such as departmental chairmanships) resulting from the institutional decision process should not be utilized in the regression analysis to determine salary equity unless they have first been tested against neutral variables and found to be free from sex-related bias.

SUMMARY AND CONCLUSIONS

Using constructed data, it has been demonstrated that the commonly used method of showing mean salary by sex or mean salary by rank by sex may be misleading; however, a more realistic picture can be obtained using multiple regression analysis.

TABLE 1

Faculty Staffing Design for

Schools A and B: Constructed Data

SCHOOL A

YEARS IN RANK	RANK LEVEL					
	1		2		3	
	M	F	M	F	M	F
0	0	4	1	3	2	2
1	1	3	2	2	3	1
2	2	2	3	1	4	0
3	3	1	4	0	4	0
4	4	0	4	0	4	0

SCHOOL B

YEARS IN RANK	RANK LEVEL					
	1		2		3	
	M	F	M	F	M	F
0	6	0	6	0	6	0
1	6	0	6	0	6	0
2	6	0	6	3	6	3
3	6	3	6	6	6	6
4	6	6	6	3	6	0

TABLE 2

Means and Standard Deviations of Salaries

by School, Rank and Sex

		Rank			
		Level 1	Level 2	Level 3	All Levels
School A:	Male	10,833 (838)	13,353 (1,232)	17,836 (1,275)	14,597 (3,126)
	Female	10,152 (869)	12,136 (1,052)	16,296 (1,199)	11,749 (2,386)
	Both	10,493 (910)	12,988 (1,297)	17,605 (1,368)	13,695 (3,192)
School B:	Male	11,269 (1,121)	13,813 (1,201)	18,195 (1,375)	14,426 (3,135)
	Female	10,965 (1,170)	13,016 (1,167)	18,450 (1,437)	14,031 (3,270)
	Both	11,199 (1,125)	13,585 (1,232)	18,253 (1,375)	14,327 (3,160)

Note. Standard deviations in parentheses

TABLE 3

Significance of the Unique Contribution of the Sex Variable to the Regression of Salary on Rank and Years-in-Rank Variables for Schools A and B

	SCHOOL A	SCHOOL B
Proportion of variance "explained" without sex as variable	R^2 = .91023	R^2 = .87915
Proportion of variance "explained" by adding sex as variable	R^2 = .91055	R^2 = .88958
Unique variance added by sex	.00032	.01043
F for difference between the R^2 values: Degrees of freedom = 1 and 113	.404	10.674
Probability that this difference or a larger difference would occur by chance	P > .10 (Not Significant)	P < .005

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 8 NO. 3 1978

EFFECTS OF STATE-WIDE SALARY EQUITY PROVISIONS ON INSTITUTIONAL SALARY POLICIES: A REGRESSION ANALYSIS

MARY P. MARTIN AND JOHN D. WILLIAMS
Florida State University University of North Dakota

A process of equalization of salaries has taken place in the State of North Dakota for higher education during the 1977-78 school year. An objective observer might easily come to a conclusion that politics played a major role in the decision making process. It is the intent of this paper to document the equalization process as it effected the decision making at a single institution in that the equalization initially was implemented on a statewide scale.

BACKGROUND

The State of North Dakota supports eight institutions of higher education: two universities, four state colleges and two two-year institutions. There has been a long history of disagreement over the average faculty salary figure used in the state formula to allocate salary monies to the three kinds of institutions. The State Board of Higher Education has, in the past, supported the philosophy of retaining a differential, e.g., $2000 difference between the two-year institutions and the universities and $1500 difference between the state colleges and the universities. A 6% salary increase plus implementing a differential would give larger percentage faculty salary appropriations to the state colleges and the two-year institutions.

For the 1977-79 biennium it was the Governor's desire to stop the controversy of the differential and implement "equal pay for equal work". The amount for the biennium that would normally establish the differential was $228,776. An appropriation of $228,776 was made to be used by the State Board of Higher Education for the purpose of creating more equitable salary authorizations to the institutions and addressing primarily the aforementioned objectives.

COMMITTEE PROCEDURE

A statewide faculty equity committee met a total of four times. At the first meeting the State Budget Director and the State Board Budget Director were present to clarify the task of allocating "equal pay for equal work". At this meeting it was decided that the approach of a regression analysis be explored, and the technical advisor was asked to develop a model. This model was reviewed at the second meeting of the committee. It was decided to make a preliminary run with 1976-77 salary data for review and discussion at the third meeting. Some minor modifications were made to redefine the sample and the committee agreed to adopt the regression analysis method for the 1977-78 data at the third meeting. The Committee's last meeting was to finalize the report for the President's Council and State Board of Higher Education. While full agreement was reached on the adoption of the statistical approach, there was dissension on how to distribute the monies as resulted from the regression analysis.

STATEWIDE MODEL

All full time faculty (N=984) at state institutions of higher

education were included. Variables included six categorical variables reflecting years experience at the current institution, four variables reflecting highest degree, four rank variables, highest program level of department (graduate, undergraduate, associate) and 21 Higher Education General Information Survey (HEGIS) Classifications. Note that no traditional outcome variables (research, teaching or service) are included. One might characterize the process as an elaborate salary scale scheme.

RESULTS OF THE COMMITTEE'S ANALYSIS

Initially, inequity was defined to be a negative residual. Thus, if a person's salary exceeded the predicted salary, resulting in a positive residual, no inequity was seen to be present. For each institution the sum of negative residuals was found (but using an overall statewide equation). The total sum of negative residuals, $670,339, obviously exceeded the allocated amount. Each institution was then accorded its "share" of the total amount in relationship to its proportion of the total sum of negative residuals. Then, 25% of the amount to be allocated to the two universities was reallocated to the remaining six institutions; the reasons for the 25% devaluation of university "inequities" reflect more the compromises of the committee than any statistical consideration. Complete details of the preliminary analysis are given in Williams and Martin (1977).

In the final analysis, the University of North Dakota was to receive $51,624 for the biennium for "equity" pay. State Board guidelines included:

1. You will commit 48.5% of your allocation the first year of the biennium.
2. Distribution will only be made to those with negative residuals.
3. No individual will receive more than his or her negative residual.
4. Faculty representation is necessary in the distribution process.

Additional provisions that the University of North Dakota attempted to use included the following:

5. First, all equity monies would be distributed to colleges within the university in proportionto their present salary expenditures without regard to the residuals in the statewide equation.
6. Most available monies were to be given to professors and associate professors. Only if insufficient faculty at higher ranks were available with negative residuals would assistant professors be considered for equity adjustments.
7. In any case, only those who are seen as being especially meritorious should be given equity adjustments.

The guidelines contain sufficient incongruities to insure that they were not always applicable. The vagaries of the regression process insured some interesting adjustments: because those at higher ranks will tend to have higher salaries and hence are less probable to have negative residuals, the most likely recipients are those few higher ranked individuals who have a comparatively lower salary. In at least some cases, the lower comparative salary would

seem to be reflective of lower productivity than their same ranked colleagues. The question then arises; what effect on the overall decision making, particularly as it reflects outcome variables (research, teaching and service) do the equity adjustments cause?

RESEARCH DESIGN

All full time tenured or tenure-track faculty wholly funded on 1977-78 appropriated monies on whom complete data was available were included in the sample. The independent variables were recorded as follows:

	Variable Number		
Degree Level	2	Doctorate	(1 if, 0 if not)
	3	Bachelor's	(1 if, 0 if not)
	4	Professional	(1 if, 0 if not)
Years Experience at Current Institution	5	3-7 years	(1 if, 0 if not)
	6	8-12 years	(1 if, 0 if not)
	7	13-17 years	(1 if, 0 if not)
	8	18-22 years	(1 if, 0 if not)
	9	Over 22 years	(1 if, 0 if not)
Level of Program and Contingent Research Involvement	10	Graduate Program	(1 if, 0 if not)
	11	Sex	(1 if Male, 0 if Female)
Discipline (Hegis Taxonomy)	13	Biological Sciences	(1 if, 0 if not)
	14	Business and Management	(1 if, 0 if not)
	15	Communications	(1 if, 0 if not)
	16	Computer and Information Sciences	(1 if, 0 if not)
	17	Education	(1 if, 0 if not)
	18	Engineering	(1 if, 0 if not)
	19	Fine and Applied Arts	(1 if, 0 if not)
	20	Foreign Languages	(1 if, 0 if not)
	21	Health Professions	(1 if, 0 if not)
	22	Home Economics	(1 if, 0 if not)
	23	Law	(1 if, 0 if not)
	24	Letters	(1 if, 0 if not)
	25	Mathematics	(1 if, 0 if not)
	26	Physical Sciences	(1 if, 0 if not)
	27	Psychology	(1 if, 0 if not)

	Variable Number		
Rank	28	Professor	(1 if, 0 if not)
	29	Associate Professor	(1 if, 0 if not)
	30	Assistant Professor	(1 if, 0 if not)
Research	31	Publications	(1 if, 0 if not)
Teaching	32	Teaching Rating	(1 if, 0 if not)
Service (University Committees)	33	University Senate	
	34	Elected Committees	
	35	Senate Committee	
	36	Presidential Committees	
	37	Appointed Committees	

The dependent variables were the 1977-78 contracted salary and the 1977-78 salary after the equity adjustment. The zero coded variables were: Instructor, Sociology, 0-2 years experience and Undergraduate Degree offered in home department and a holder of a master's degree. The research variable was a measure used at the University of North Dakota; while there are many idiosyncracies involved, 10 points on the research variable would represent publication as a single author of an article published in a refereed journal of national or international status. Other types of publication are covered as well. The teaching variable is the mean rating of the faculty member by students in the most recent administration of the rating scale. Seven faculty did not have a rating, reducing the population from 319 to 312. The service variable was assessed as being number of years serving on five different kinds of university committees: memberbship on the university senate, committees elected on a university wide basis, committees elected by the university senate, committees appointed by the president and other appointed committees; ad hoc committees were not included. Committee memberships for the past five years were included.

RESULTS

Tables 1 and 2 report the results of using the previously described independent variables; these tables respectively report the results for the 1977-78 contracted salary and the 1977-78 salary after the equity adjustments.

Table 1
Multiple Regression for Equity Study
Prior to Adjustments
(N = 312)

VARIABLE NO.	MEAN	STANDARD DEVIATION,	CORRELATION X VS Y	REGRESSION COEFFICIENT	COMPUTED t Value	BETA
2	0.657	0.475	0.360	1050.36	4.299	0.145
3	0.012	0.112	-0.147	545.67	0.604	0.017
4	0.006	0.079	0.068	3203.01	2.577	0.074
5	0.240	0.428	-0.028	121.17	0.387	0.015
6	0.275	0.447	0.069	151.87	0.447	0.019
7	0.134	0.341	0.085	359.17	0.891	0.035
8	0.051	0.220	0.127	915.37	1.736	0.058
9	0.080	0.271	0.173	695.57	1.428	0.054
10	0.868	0.338	0.076	-88.02	-0.277	-0.008
11	0.855	0.351	0.267	390.88	1.307	0.039
13	0.044	0.207	0.116	-154.80	-0.300	-0.009
14	0.105	0.308	0.061	1794.89	4.724	0.160
15	0.009	0.097	-0.047	-1162.32	-1.087	-0.033
16	0.006	0.079	-0.003	2438.64	2.108	0.056
17	0.214	0.411	-0.039	746.16	2.159	0.089
18	0.051	0.220	0.153	316.68	0.637	0.020
19	0.089	0.286	-0.272	-973.55	-2.107	-0.080
20	0.044	0.207	-0.097	-1128.19	-2.154	-0.067
21	0.019	0.137	-0.067	-454.83	-0.603	-0.018
22	0.019	0.137	-0.045	620.30	0.809	0.024
23	0.022	0.148	0.175	7271.58	10.831	0.313
24	0.080	0.271	-0.034	-302.03	-0.717	-0.023
25	0.041	0.200	-0.001	-303.86	0.578	-0.017
26	0.089	0.286	0.150	99.00	0.236	0.008
27	0.041	0.200	0.000	718.88	1.350	0.041
28	0.333	0.472	0.720	8993.55	11.466	1.233
29	0.403	0.491	-0.148	5100.53	6.947	0.728
30	0.237	0.426	-0.521	2682.56	3.932	0.331
31	6.323	13.240	0.113	8.98	1.218	0.034
32	3.077	0.533	-0.126	187.99	1.004	0.029
33	0.503	1.054	0.370	324.20	3.260	0.099
34	0.189	0.670	0.241	-38.82	-0.263	-0.007
35	1.198	1.751	0.216	6.39	0.109	0.003
36	0.637	1.307	0.142	-78.59	-1.037	-0.029
37	0.044	0.346	-0.047	-274.10	-0.995	-0.027
DEPENDENT 1	18536.714	3443.010				
INTERCEPT		10514.50				
MULTIPLE CORRELATION		0.90640				
STD. ERROR OF ESTIMATE		1543.841				
MULTIPLE CORRELATION SQUARED		0.82152				

Table 2
Multiple Regression for Equity Study
After Adjustments
(N = 312)

VARIABLE NO.	MEAN	STANDARD DEVIATION,	CORRELATION X VS Y	REGRESSION COEFFICIENT	COMPUTED t Value	BETA
2	0.657	0.475	0.365	1079.99	4.589	0.149
3	0.012	0.112	-0.148	557.57	0.641	0.018
4	0.006	0.079	0.067	3196.81	2.670	0.074
5	0.240	0.428	-0.026	158.40	0.526	0.019
6	0.275	0.447	0.074	221.49	0.676	0.028
7	0.134	0.341	0.090	442.96	1.141	0.044
8	0.051	0.220	0.125	906.15	1.784	0.058
9	0.080	0.271	0.169	668.25	1.424	0.052
10	0.868	0.338	0.078	-91.01	-0.297	-0.008
11	0.855	0.351	0.268	361.03	1.253	0.037
13	0.044	0.207	0.118	-100.49	-0.202	-0.006
14	0.105	0.308	0.061	1808.37	4.941	0.162
15	0.009	0.097	-0.049	-1139.43	-1.106	-0.032
16	0.006	0.079	-0.005	2361.01	2.118	0.054
17	0.214	0.411	-0.042	724.52	2.176	0.086
18	0.051	0.220	0.156	401.19	0.838	0.025
19	0.089	0.286	-0.273	-964.05	-2.165	-0.080
20	0.044	0.207	-0.098	-1162.17	-2.303	-0.070
21	0.019	0.137	-0.063	-337.29	-0.464	-0.013
22	0.019	0.137	-0.048	543.25	0.735	0.021
23	0.022	0.148	0.176	7342.87	11.354	0.317
24	0.080	0.271	-0.036	-317.21	-0.782	-0.025
25	0.041	0.200	0.001	-235.44	-0.465	-0.013
26	0.089	0.286	0.154	141.75	0.351	0.011
27	0.041	0.200	-0.004	691.00	1.347	0.040
28	0.333	0.472	0.723	9050.52	11.979	1.245
29	0.403	0.491	-0.142	5174.23	7.316	0.740
30	0.237	0.426	-0.530	2698.64	4.106	0.335
31	6.323	13.240	0.111	7.89	1.110	0.030
32	3.077	0.533	-0.124	219.33	1.216	0.034
33	0.503	1.054	0.367	298.29	3.114	0.091
34	0.189	0.670	0.240	-41.61	-0.292	-0.008
35	1.198	1.751	0.217	2.12	0.037	0.001
36	0.637	1.307	0.149	-61.61	-0.844	-0.023
37	0.044	0.346	-0.049	-286.46	-1.079	-0.028
DEPENDENT 12	18612.140	3431.767				
INTERCEPT		10418.76				
MULTIPLE CORRELATION		0.91287				
STD. ERROR ESTIMATE		1487.210				
MULTIPLE CORRELATION SQUARED			0.83333			
ONE MINUS CORRELATION SQD			0.16667	F = 39.427		

Table 1 reports decision making that took place at the University of North Dakota prior to any involvement (or knowledge) or the equity process; Table 1 could be seen as reflecting the priorities (even as they are changing) in the decision making process prior to the intervention. The question that arises regarding Table 1 is, "Has the intervention (equity) process changed to any significant degree the decision making process?"

A direct answer to the quesiton can be made by imposing the regression coefficients established in Table 1, plus the mean increase $75.19 to the data set that formed Table 2, using the following (Bottenberg and Ward, 1963) equation: $F = \frac{(q_2 - q_1)/(df_1)}{(q_1)/(df_2)}$, where

q_1 is the sum of squared deviations from the regression line for Table 2, and q_2 is the sum of squared deviations for the imposed equation; the F value is within rounding error of zero, showing a close fit.

While considerable difficulty is encountered in trying to interpret each coefficient in the tables, the following directions of change can be noted from Table 1 to Table 2: the research and teaching variables (Variables 28 and 31) seem to be not being attended to in the decision making process (looking at the correlation coefficient, the regression coefficient and the beta weight). On the other hand, being on the University Senate made a significant contribution. A variable included in the analyses that logically does not belong is the sex variable. It is included as a control variable; because discrimination is being attended to on a national level, the variable may have some predictive value due to efforts to eliminate discrimination. The drop in the size of the coefficient for

sex from Table 1 to Table 2 seems to be indicative of a slight attempt to address this issue. It appears that most attention is focused upon rank, degree level, and on some of the HEGIS categories, notably law (Variable 13). Using the degree variables, rank variables and six HEGIS categories (14, Management; 16, Computer and Information Sciences; 17, Education; 19, Fine and Applied Arts; 20, Foreign Languages and 23, Law) a total of twelve variables, results in only a minimal drop in the R^2, from .83333 to .81421, using the salary after the equity adjustment as criterion. Using rank and the HEGIS category for Law (a total of four variables), R^2 = .74206, amounting to slightly less than a nine per cent drop in accounted variance despite the dropping of 31 variables.

It could be argued that the reason the outcome variables have so little impact in complex equations such as are demonstrated in Tables 1 and 2 is that rank is in fact due to the outcome variables and, therefore, reduces their apparent impact. Accordingly, the outcome variables reported in Tables 1 and 2, and the equity salary variables were analyzed by rank in a one-way analysis of variance (See Table 3).

TABLE 3

Academic Rank and Outcome Variables and Salary Variables

(N = 312)

VARIABLE	Professor N=104	Associate Professor N=126	Assistant Professor N=74	Instructor N=8	F
Research	8.99	6.44	3.06	.13	3.58
Teaching	2.99	3.09	3.15	3.34	2.01
University Senate	.87	.48	.09	0	9.08
Elected Committees	.48	.06	.01	0	10.88
Senate Committees	1.56	1.43	.42	.13	8.67
Presidential Committees	.82	.77	.22	.13	4.18
Appointed Committees	.04	.05	.03	.25	1.01
1977-78 Contracted Salary	22037	17918	15321	12521	186.43
1977-78 Salary (After Equity Adjustment)	22119	18020	15350	12521	196.90
1977-78 Salary Residual (Positive Only)	407	682	708	2465	15.62
Salary Equity Adjustment	82	102	29	.00	4.11

The professors tend to have spent considerably more time in committee activity than is true of other ranks, but one might rationally question whether service on committees justifies either promotion in rank or considerably higher salaries. While professors have a higher mean research output, this difference is offset by two considerations; publication activity clearly is non-normally distributed (the standard deviation is approximately twice the mean) and the mean output is less than a single article a year. Also, the publication activity of faculty at all ranks tended, for the typical faculty member, to be almost non-existent. Previously, Martin (1977) was unable to discern any plausible explanation for promotion, other than having served in some administrative capacity. Thus, the "policy" regarding salaries is both more complex and simpler than is shown in Tables 1 and 2. It is simpler, in that knowing a person's degree, rank and whether or not they teach in a few selected departments can give almost as good an indication toward salary as is knowing the complete set of information used in Tables 1 and 2. It is more complex in that the variables that determine rank are not sufficiently known to be of much predictive value.

What then could be said about the adjustment process, in an overall sense? Apparently, no major effects on the decision making machinery has taken place. Even when a deliberate attempt is made to implement a policy at the local level, some non-compliance occurs. Note that associate professors received higher equity increases than professors even though the intention was the opposite. Also, implicit in the State Board action was that the negative residuals be closely looked at in the decision

making process; other than a limiter of the possible adjustment, the residuals were not highly related to actual increases. Using only those who had a negative residual (N=172), r = .06 between negative residuals and equity increase. Interestingly, the correlations between equity increase and the research and teaching variables were respectively -.05 and .06. However, 24 of the 36 professors with negative residuals received equity increases from $52 to $870 with a mean of $355 for those receiving raises; 46 of the 74 eligible associate professors received equity increases from $80 to $572 with a mean of $279; 9 of the 54 eligible associate professors received equity increases from $70 to $443 with a mean of $241. All eight of the instructors had negative residuals; none had an equity increase. Looking at the data yet another way, 21 faculty had 20 or more points on the research scale (two referred articles by a single author per year); 10 of these faculty were eligible for equity increases. Three such faculty actually received equity increases ($80 to $120 with a mean of $100). Clearly, research played no important part in the equity adjustment process. Teaching success, as measured by the rating scales, fared scarcely better.

Referring back to the original guidelines, the four provisions of the State Board were closely followed (with the possible exception of faculty representation); the first two of the three provisions imposed by the University itself were adhered to. Only the last provision ("... only those who are seen as being especially meritorious should be given equity adjustments.") was seemingly violated; unfortunately, the one provision that was violated might be seen by many to be the most important.

REFERENCES

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

Martin, M. Promotion policies examined by rank and sex at the University of North Dakota. Journal of Teaching and Learning, 1977, 3, No. 1, 41-46.

Williams, J. D. and Martin, M. Equalization of salaries for higher education in North Dakota; or, equal pay for equal work. Presented at the Association of Institutional Researcher - Upper Midwest meeting, Cedar Falls, Iowa, November 2, 1977.

THE USE OF PREDICTION INTERVALS IN MULTIPLE REGRESSION ANALYSIS*

ELLEN STOREY VASU
University of North Carolina at Chapel Hill

OBJECTIVES

With the development of such easily accessible statistical programs such as SPSS (Nie, et. al, 1975) and SAS (Barr et. al., 1976), the techniques of multivariate analysis are available to an increasingly larger number of individuals engaged in research. The proper application and interpretation of regression techniques are more complicated than the novice researcher initially may believe. Before the statistics resulting from a regression analysis computer run can be validly interpreted, a proper model must be selected which adequately defines the regression relation under examination. Once the aptness of this model has been determined, confidence intervals can be constructed around the weights of the equation and the importance of the independent variables in the prediction of the dependent variable can be assessed. In addition to defining the approximate form of the relation existing between the set of predictors and the dependent variable, individual predictions can be made about the value of the dependent variable for new cases not originally included in the determination of the regression equation.

The major objective of this paper is the explanation and application of 'prediction intervals' in regression analysis (Neter & Wasserman, 1974) which may be constructed for a new case or set of cases for which we have measurements on our independent variables. In order to demonstrate their calculation and interpretation most clearly, a small simulated dataset with a high R^2 value will be used. All calculations were performed using the program SAS (Statistical Analysis System, Barr, et. al., 1976) using PROC MATRIX and PROC STEPWISE.

The mathematics for the following three cases will be presented:

(1) The construction of a prediction interval estimating a range of values on the dependent variable (Y) for a new observation with specific fixed

*Revised version (April, 1978).

values on each of the independent variables [(1 - α) confident coefficient].

(2) The construction of a prediction interval for m new observations where the mean Y value for these new observations is to be estimated. (A range of values will be determined for the mean Y value). The m new observations all have the same specific fixed values on each of the independent variables [(1 - α) confidence coefficient].

(3) The construction of either Bonferroni or Sheffé simultaneous prediction intervals estimating a range of values on the dependent variable for s new observations with s different sets of specified fixed values (in other words s different profile types) on the independent variables [(1 - α) family confidence coefficient].

THEORETICAL FRAMEWORK

Population regression equation

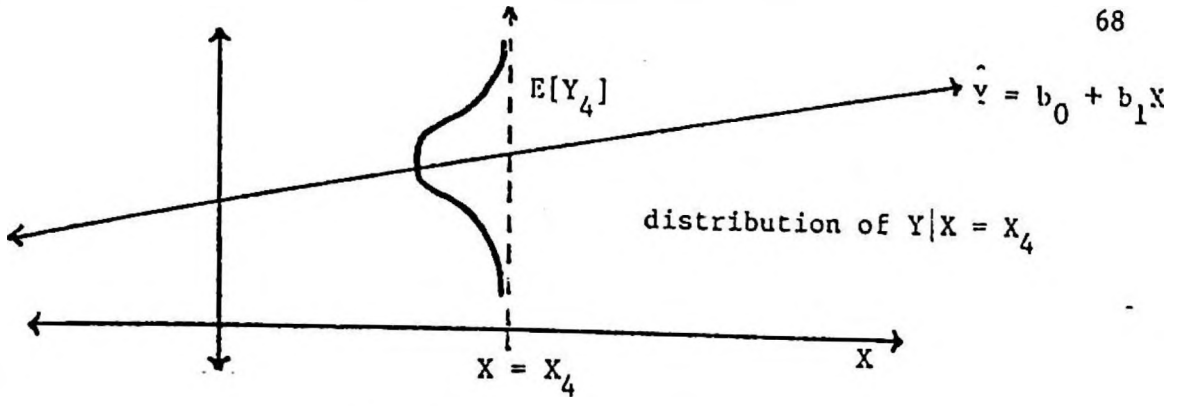
The theoretical framework within which this paper is presented is classical regression analysis in which we have one dependent variable (Y) and one set of fixed values on our independent variables $(X_1,\ldots,X_{p-1})$, such that:

$$\underset{nx1}{Y} = \underset{nx1}{X\beta} + \underset{nx1}{\varepsilon}$$

where the ε_1 are independent $N(0,\sigma^2)$

This equation defines the best fitting regression model for our population of values, since the regression weights presented are population parameters. The equation does not define a functional relationship between the independent vari and the dependent variable, since we have allowed for error. Therefore, even i we had every single case in our population and fitted this equation to our population of values, there still would be some errors of measurement since we are essentially dealing with a relation rather than a function. If our model involved only one independent variable, we could illustrate the fit

of the model to our population of values as indicated below:



Notice that the distribution of Y_4 is normal with a mean and variance:

$$E[Y_4]$$
$$\sigma^2_{Y|X = X_4}$$

The errors of prediction are essentially errors in the fit of the population regression equation to the population data matrix.

Sample regression equation

Assume we have calculated a regression equation utilizing sample data and have arrived at the following:

$$\underline{Y} = \underline{Xb} + \underline{e}$$

Let us also assume that this final model we have derived is the best fitting model for our sample data.

This equation which we defined is an estimate of the population regression equation, therefore, if we took another random sample from the population and calculated a new regression equation the new regression equation would probably have a different $\underline{b}$ vector. Since our $\underline{b}$ vector consists of statistics which are estimates of the corresponding population vector $\underline{\beta}$, any regression equation which we develop is subject to sampling error, and the estimates may fluctuate from sample to sample.

For the equation involving one independent variable, if we are interested in predicting the Y value for the case whose X value equals X_4, not only do we

have the error in the fit of the population regression equation to the population of data, but also the error introduced by estimating that population regression equation. In other words we have introduced variability within the distribution of Y, and also error in locating the E[Y] which lies directly on our population regression line. (If we have more than one independent variable our equation will not represent a line but a plane or hyperplane.) Let us assume that the final model we derive by statistical testing is the best fitting model for our data on hand. This model fitting process includes an examination of the residual in addition to statistical comparisons among potential equations. Once our final model is selected we can then focus on the prediction of Y for new observations whose X values lie within the range of the X values included in the original sample which produced our sample regression equation. If we use the developed regression relation and solve for $\hat{Y}$ we would arrive at a vector consisting of n elements. Each of these particular predicted Y values is based upon a linear combination of <u>specific</u> X values. For example, examine the equation below:

$$\underline{\hat{Y}} = \underline{Xb}$$

Where:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} \\ 1 & X_{21} & X_{22} & X_{23} \\ 1 & X_{31} & X_{32} & X_{33} \\ 1 & X_{41} & X_{42} & X_{43} \end{bmatrix} \qquad b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

If we look at the predicted Y value for the third case, we see that it is based upon a linear combination of specific X values, namely those X values which correspond to the third case.

$$\hat{Y}_3 = b_0 + b_1X_{31} + b_2X_{32} + b_3X_{33}$$

But what if we wish to make a prediction on the variable Y for a new (independent) individual whose profile on the X variables does not correspond to either case(1), case(2), case(3), or case(4)? This individual can be classified as a new observation. Assuming that the model which we have defined is the best fitting model for our data, we can make predictions for new observations whose scores on the X variables lie within the region jointly defined by the values of our original independent variables. If we make predictions for observations outside this region, we cannot be confident that our model is appropriate (Neter & Wasserman, 1974) since we never tested its fit. Because of the variability both within the distribution of Y and the variability of the regression line itself, estimates of any specific Y value are more accurate if we calculate a range of possible values within which Y will fall with a given probability. This range of values for Y given a specific set of values on the independent variables is called a prediction interval. It differs from a confidence interval in that we are not estimating a population parameter, but rather an individual outcome for a specific observation or set of observations.

METHODS

Assume the following model is appropriate:

$$\underline{Y} = \underline{X\beta} + \underline{\varepsilon} \quad \text{where } \varepsilon_i \sim N(o, \sigma^2) \text{ and independently}$$

We will examine the calculation of prediction intervals for three cases:

Case 1: The prediction of $Y_i | \underline{X}_i$ where $\underline{X}_i$ corresponds to a vector of fixed values on the independent variables and:

$$\hat{Y}_i - t_* s_{Y_i} \leq Y_i \leq \hat{Y}_i + t_* s_{y_i}$$

where $$s_{y_i} = [MSE\ (1 + \underline{X}_i' (\underline{X}'\underline{X})^{-1} \underline{X}_i)]\ 1/2$$

$$t_* = t(1 - \alpha/2, n - p)$$

Case 2: The prediction of $\bar{Y}_i$, the mean Y_i value for m cases where each case has the profile $\underline{X}_i$ on the independent variables, and:

$$\hat{Y}_i - t_* s_{\bar{Y}} \leq \bar{Y}_i \leq \hat{Y}_i + t_* s_{\bar{Y}}$$

where $$s_{\bar{Y}_i} = [MSE\ (1/m + \underline{X}_i'\ (\underline{X}'\underline{X})^{-1} \underline{X}_i]^{\frac{1}{2}}$$

$$t_* = t(1 - \alpha/2,\ n - p)$$

Case 3: The creation of s simultaneous prediction intervals for s new observations at s different levels of $\underline{X}_i$, where each interval is:

$$\hat{Y}_i - Cs_{y_i} \leq Y_i \leq \hat{Y}_i + Cs_{y_i}$$

where s_{y_i} is as defined above

and C = Min $|S,B|$ where:

$S^2 = sF\ (1 - \alpha,\ s,\ n - p)$ (Sheffé)

$B = t\ (1 - \alpha/2s,\ n - p)$ (Bonferroni)

Assume we are interested in predicting ability level in introductory statistics based upon measures of both spatial ability and mathematical ability.

RESULTS (full matrix manipulations presented in Appendix)

Table 1 contains hypothetical data on statistical performance, mathematical ability and spatial ability.

In order to calculate our sample regression equation:

$$\underline{Y} = \underline{Xb} + \underline{e}$$

We must solve the following equation:

$$\underline{b} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$$

Which yields:

$$\underline{b} = \begin{bmatrix} 18.2457 \\ .4960 \\ .0184 \end{bmatrix}$$

In order to determine the aptness of the model we can examine the source table presented in table 1. Both X1 and X2 contribute significantly to the

prediction of Y, and were added stepwise to the equation. ($F = MSR/MSE = 5679.47$; $p < .0001$). The final R^2 value (SSR/SSTO) equals .9989 which indicates a close fitting equation. Examination of the distribution of the residuals did not seem to indicate any major deviation from normality or homoscedasticity.

Case 1

Suppose we are interested in predicting the statistical performance level of a new entering student who has a math ability score of 390, and a spatial ability score of 460:

$$\underline{X}_1 = \begin{bmatrix} 1 \\ 390 \\ 460 \end{bmatrix}$$

Setting $\alpha = .05$ we can then calculate our prediction interval:

$$\hat{\underline{Y}}_1 = \underline{X}_1'b = 220.151$$

$$s^2_{Y_1} = [MSE\ (1 + \underline{X}_1'\ (\underline{X}'\underline{X})^{-1}\underline{X}_1)]^{1/2} = [18.961189(1.172238)]^{1/2} = [22.227]^{1/2} = 4.715$$

$$t(.975,\ 12) = 2.179$$

Therefore:

$$\hat{Y}_1 - t(1 - \alpha/2;\ n - p)S_{Y_1} \leq Y_1 \leq \hat{Y}_1 + t(1 - \alpha/2;\ n - p)S_{Y_1}$$

$$220.151 - 2.179\ (4.715) \leq Y_1 \leq 220.151 + 2.179\ (4.715)$$

$$209.877 \leq Y_1 \leq 230.425$$

With confidence coefficient .95, we predict that the final math exam score for the next student with a math ability score of 390 and a spatial ability score of 460 will be somewhere between 209.9 and 230.4.

Case 2

Assume our measure of statistical performance is actually the final score total for each student in the introductory statistics course in which a score total below 230 is considered failure. Let us calculate how students with the following profile can be expected to perform in the course on the average.

Assume we are considering 3 students with very low math ability scores, and low spatial ability scores. Let $\alpha = .05$, $m = 3$ (3 students):

$$\underline{X}_2 = \begin{bmatrix} 1 \\ 125 \\ 770 \end{bmatrix}$$

$$\hat{Y}_2 = \underline{X}_2'\underline{b} = 94.4137$$

$$S_{\bar{Y}_2} = [MSE\ (1/m + \underline{X}_2'\ (\underline{X}'\underline{X})^{-1}\underline{X}_2]^{\frac{1}{2}} = [18.961189(.4886353)]^{\frac{1}{2}} = [9.26511]^{\frac{1}{2}} =$$

$$t\ (.975,\ 12) = 2.179$$

$$\hat{Y}_2 - t(.975,\ 12)\ S_{\bar{Y}_2} \leq \bar{Y}_2 \leq \hat{Y}_2 + t(.975,\ 12)\ S_{\bar{Y}_2}$$

$$94.414 - 2.179\ (3.044) \leq \bar{Y}_2 \leq 94.414 + 2.179\ (3.044)$$

$$87.781 \leq \bar{Y}_2 \leq 101.047$$

With confidence coefficient .95, we predict that the average final math exam scores for the three new pupils with math ability scores of 125 and spat ability scores of 770 will be between 87.8 and 101.0 which is failing.

Case 3

Suppose we are interested in predicting performance level for a number of difference profile types. Specifically, we would like to assign a family confidence coefficient to the simultaneous prediction intervals created for the following two ($s = 2$) profile types:

$$\underline{X}_3 = \begin{bmatrix} 1 \\ 700 \\ 1460 \end{bmatrix} \qquad \underline{X}_4 = \begin{bmatrix} 1 \\ 400 \\ 950 \end{bmatrix}$$

$$S_{Y_3} = [MSE\ (1 + \underline{X}_3'\ (X'X)^{-1}X_3)]^{\frac{1}{2}} = [18.961189\ (1.417022)]^{\frac{1}{2}} = [26.8684]^{\frac{1}{2}} = 5.$$

$$\hat{Y}_3 = \underline{X}_3'\underline{b} = 392.311$$

$$S_{Y_4} = [MSE\ (1 + \underline{X}_4'\ (X'X)^{-1}\underline{X}_4]^{\frac{1}{2}} = [18.961189\ (1.14631)]^{\frac{1}{2}} = [21.7354]^{\frac{1}{2}} = 4.662$$

$$\hat{Y}_4 = \underline{X}_4'\underline{b} = 234.126$$

We will calculate both the Scheffé (S) and Bonferroni (B) multipliers and use the smallest of the two in the construction of the intervals.

$$S = 2F(.95, 2, 12) = 2(3.88530) = 7.7706$$

$$B = t(1 - .05/2(2); 12) = t(.9875; 12) = 2.56003$$

$$\text{Min}|\ S, B\ | = B = 2.56003$$

$$392.311 - 2.560\ (5.183) \leq Y_3 \leq 392.311 + 2.560\ (5.183)$$

$$379.043 \leq Y_3 \leq 405.579$$

$$234.126 - 2.560\ (4.662) \leq Y_4 \leq 234.126 + 2.560\ (4.662)$$

$$222.191 \leq Y_4 \leq 246.061$$

With family confidence coefficient .95 we predict that the final math scores for these two profile types will be within the above specified limits. In other words, if repeated samples are taken and the regression coefficients are calculated for each of these, 95 per cent of the samples would yield situations in which both of the specified prediction intervals would be correct for the specific $\underline{X}_3$ and $\underline{X}_4$ vectors stipulated above.

DISCUSSION

Whenever we make predictions concerning the value of our dependent variable for observations independent of our original sample, it is more useful and less misleading to attach confidence coefficients to our estimates by creating intervals on our dependent variable within our specific case or cases can be expected to fall. The preceding example was based upon a dataset which had a high R^2 value between the two variables under examination. In addition, note that the MSE term for the dataset was relatively small (18.96). The variability of each of the independent variables involved was quite large (s_{x1}=233.66; s_{x2}=730.64). This results in a more stable estimate

for our regression line (Neter & Wasserman, 1976), since the levels of each X_i were spread along their respective axes.

This fact can be made clearer by examining the regression equation involving only one independent variable. The variance of b_0 and b_1 are:

$$\sigma^2(b_0) = \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X - \bar{X})^2}\right]$$

$$\sigma^2(b_1) = \frac{\sigma^2}{\Sigma (X - \bar{X})^2}$$

The farther the levels of the X variable are apart, the larger the deviations about the mean ($\bar{X}$) become. This will result in larger terms in the denominator of the formulae above which in tern results in a smaller variance. Therefore, placing the levels of the X variable far apart will result in minimizing the variance of the regression equation. Reducing this variability which is a component of the prediction interval formulae, will result in narrower prediction intervals.

Therefore, we can see that both the MSE term, and the spacing of the level of the independent variable have an affect upon the utility of our resulting regression equation. Many times these levels of our independent variables are not under the control of the researcher, if the design is not one of a true experiment. Nevertheless, rather than making predictions for individual cases in the usual manner which results in a point estimate, the usefulness and accuracy of our prediction equation can be better demonstrated by employing prediction intervals rather than point estimates. A high R^2 value does not necessarily imply that the developed regression equation is useful (Neter & Wasserman, 1974, p. 229). When the R^2 value of a sample regression equation is extremely high, point estimation alone may leave the impression that the estimate itself is an extremely good one, and, therefore, close to the real Y value of the case involved. Researchers should begin to examine not only R^2 values, but also the variability of the regression coefficients and

the size of the prediction intervals in order to make more valid interpretations.

REFERENCES

Barr, A. J. et. al. A User's Guide to SAS 76. Raleigh, N.C.: SAS Institute Inc., 1976.

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

Kelley, F. J., Beggs, D. L., et al. Multiple Regression Approach. Carbondale, Ill.: Southern Illinois University Press, 1969.

Neter, J., and Wasserman, W. Applied Linear Statistical Models. Homewood, Ill.: Richard D. Irwin, Inc., 1974.

Table 1*

Data Matrix, Statistics, and Source Table for Dataset 1.

Statistical Performance		Math Ability	Spatial Ability		
Y =	X =			e =	
274	1	500	410	.2086
190	1	312	1214	-5.3346
396	1	702	1762	-2.8588
212	1	362	798	- .4812
84	1	124	307	-1.3986
288	1	482	1742	-1.3697
112	1	148	968	2.5361
334	1	612	410	4.6600
182	1	342	97	-7.6640
60	1	58	520	3.4190
454	1	812	1980	-3.4301
414	1	696	2387	6.6184
238	1	424	620	-1.9587
156	1	266	48	4.9339
374	1	692	565	2.1239

$\bar{Y} = 251.2$ $\quad \bar{X}_1 = 435.5$ $\quad \bar{X}_2 = 921.87$

$S_Y = 124.1$ $\quad S_{X_1} = 233.66$ $\quad S_{X_2} = 730.64$

RANGE

Y: 60 - 454
X1: 58 - 812
X2: 48 - 2387

CORRELATION MATRIX

Y	1.000		
X1	.995	1.000	
X2	.639	.569	1.000

SOURCE	DF	SS	SS	MS	F	P
Regression	2	SSTO - SSE	215378.87	107689.43	5679.47	.0001
Error	12	Y'Y - b'X'Y	227.53	18.96		
Total	14	$Y'Y - n\bar{Y}^2$	215606.40			

$R^2 = .9989$

	b	S(b)
b_0	18.2457	
b_1	.4960	.0061
b_2	.0184	.0019

*Matrix manipulations and regression output were accomplished via the program SAS (Statistical Analysis System, Barr et. al., 1976).

APPENDIX

DATASETS MANIPULATIONS

The program listed below was written for SAS and was used to calculate the following:

$$\hat{\underline{Y}} = \underline{X}\underline{b}$$

$$\underline{b} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$$

$$\underline{e} = \underline{Y} - \hat{\underline{Y}}$$

$$SSE = \underline{Y}'\underline{Y} - \underline{B}'\underline{X}'\underline{Y}.$$

and also to compare this $\underline{b}$ with the $\underline{b}$ output from the SAS procedure STEPWISE to check for accuracy. $(\underline{X}'\underline{X})^{-1}$, $\underline{b}$, and the MSE error were needed to form the prediction intervals, and were used as input for the programs listed on the following pages.

```
**********************************************************************************
1          DATA; INPUT ID V1 VX1 VX2 @ @;
5          CARDS;

NOTE: SAS WENT TO A NEW LINE WHEN INPUT STATEMENT
      REACHED PAST THE END OF A LINE.
NOTE: DATA SET WORK.DATA1 HAS 15 OBSERVATIONS AND 4 VARIABLES.
NOTE: THE DATA STATEMENT USED 0.14 SECONDS AND 96K.
23         PROC MATRIX PRINT;
24         FETCH FULL DATA=DATA1;
25         Y=FULL(*,2);
26         IND=FULL(*,3 4);
27         U=1/1/1/1/1/1/1/1/1/1/1/1/1/1/1;
28         X = U || IND;
29         XTY=(X')*(Y);
30         XTX=(X')*(X);
31         XTXI=INV(XTX);
32         B=(XTXI)*(XTY);
33         YP=(X)*(B);
34         E=Y-YP;
35         YT=(Y');
36         YTY=(YT)*(Y);
37         UNIT=1 1 1 1 1 1 1 1 1 1 1 1 1 1 1;
38         MMEANY=(UNIT)*(Y);
39         SSE=(YTY)-(B')*(XTY);
40         OUTPUT FULL OUT=REG (RENAME=(COL2 =Y1 COL3=XX1
41         COL4 =XX2));

NOTE: DATA SET WORK.REG HAS 15 OBSERVATIONS AND 5 VARIABLES.
NOTE: THE PROCEDURE MATRIX USED 0.66 SECONDS AND 124K AND PRINTED PAGES 3 TO

42         PROC STEPWISE DATA=REG;
43         MODEL Y1 = XX1 XX2/MAXR;

NOTE: THE PROCEDURE STEPWISE USED 0.26 SECONDS AND 112K AND PRINTED PAGE 8.
NOTE: SAS USED 124K MEMORY.
```

Dataset Manipulations (continued)

Our regression coefficients for the dataset are:

$$\underline{b} = \begin{bmatrix} 18.2457 \\ .4960 \\ .0184 \end{bmatrix}$$

In calculating prediction intervals for cases (1), (2), and (3) we used the following vectors:

$$\underline{X}_1 = \begin{bmatrix} 1 \\ 390 \\ 460 \end{bmatrix} \quad \underline{X}_2 = \begin{bmatrix} 1 \\ 125 \\ 770 \end{bmatrix} \quad \underline{X}_3 = \begin{bmatrix} 1 \\ 700 \\ 1460 \end{bmatrix} \quad \underline{X}_4 = \begin{bmatrix} 1 \\ 400 \\ 950 \end{bmatrix}$$

We also needed to calculate:

$$\underline{X}_1'(\underline{X}'\underline{X})^{-1}\underline{X}_1 \text{ and } \hat{Y}_1 = \underline{X}_1\underline{b}$$

$$\underline{X}_2'(\underline{X}'\underline{X})^{-1}\underline{X}_2 \text{ and } \hat{Y}_2 = \underline{X}_2\underline{b}$$

$$\underline{X}_3'(\underline{X}'\underline{X})^{-1}\underline{X}_3 \text{ and } \hat{Y}_3 = \underline{X}_3\underline{b}$$

$$\underline{X}_4'(\underline{X}'\underline{X})^{-1}\underline{X}_4 \text{ and } \hat{Y}_4 = \underline{X}_4\underline{b}$$

The following SAS programs run under release 76.5 of the interactive version of SAS were used. Some of the output appears on page A3 of this Appendix.

****

```
PROC MATRIX PRINT;
X2=1/125/770;
X2T=(X2');
B=18.2457/.496/.0184;
XTX2I=.319071 -.000517808 -.000029198/
-.000517808 .0000019332 -.0000004 /
-.000029198 -.0000004 .0000000002;
Y2P= (X2T)*(B);
VYBAR2= (X2T)*(XTX2I)*(X2);
X3= 1/390/460;
X3T=(X3');
X4=1/400/950;
X4T=(X4');
Y3P=(X3T)*(B);
YP4=(X4T)*(B);
VYBAR3=(X3T)*(XTX2I)*(X3);
VYBAR4=(X4T)*(XTX2I)*(X4);
```

```
PROC MATRIX PRINT;
X1=1/390/460;
X1T=(X1');
XTXINV=.319071 -.000517808 -.000029198/
-.000517808 .0000019332 -.0000004 /
-.000029198 -.0000004 .0000000002;
V1=(X1T)*(XTXINV)*(X1);
X2=1/700/1460;
X2T=(X2');
V2=(X2T)*(XTXINV)*(X2);
X3=1/400/950;
X3T=(X3');
V3=(X3T)*(XTXINV)*(X3);
B=18.2457/.496005/.0183982;
YP1=(X1T)*(B);
YP2=(X2T)*(B);
YP3=(X3T)*(B);
```

MATRICES FROM

DATASET

$(\underline{X}'\underline{Y})$ = XTY	COL1
ROW1	3768
ROW2	2044964
ROW3	4285118

(X'X) = XTX	COL1	COL2	COL3
ROW1	15	6532	13828
ROW2	6532	3608824	7380538
ROW3	13828	7380538	20221188

$(\underline{X}'\underline{X})^{-1}$ = XTXI	COL1	COL2	COL3
ROW1	0.319071	-.000517808	-.000029198
ROW2	-.000517808	.0000019332	-3.5151E-07
ROW3	-.000029198	-3.5151E-07	1.9772E-07

B	COL1
ROW1	18.2457
ROW2	0.496005
ROW3	0.0183982

$\underline{Y}'\underline{Y}$ = YTY	COL1
ROW1	1162128

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 8 NO. 3 1978

A STUDY OF THREE TREATMENTS FOR MENSTRUAL DIFFICULTIES: AN ANALYSIS USING MULTIPLE LINEAR REGRESSION

WILLIAM ROSENTHAL
St. Louis Community College at Meramec

STEVEN D. SPANER
University of Missouri at St. Louis

INTRODUCTION

The purpose of this investigation was to develop a regression approach to a treatment design which would traditionally have been analyzed through analysis of variance procedures. Two approaches are possible: developing direct regression equivalents of ANOVA designs and developing regression approaches without regard to traditional approaches. We have attempted to deal with both in the paper.

Perhaps a disclaimer is appropriate at this point. We have attempted neither to develop new regression formulations nor to demonstrate the mathematical equivalence of various models or designs. Our concerns have been premised at the purely pragmatic level of the practitioner rather than that of the advanced statistician. Therefore, we have taken the regression models directly (or with very slight modification) from Testing Research Hypotheses Using Multiple Linear Regression (McNeil, Kelly, McNeil), and our test of equivalence has been our ability to produce similar results using commonly available statistical packages (SPSS, SAS, VELDMAN, DPLINEAR).

METHOD

The data is from a dissertation by Ruth Rosenthal (in progress) in clinical psychology. (We should point out that we are not attempting to present her results but to use the material of her study as the basis for discussing the statistical approaches).

Subjects

There were thirty-one subjects, essentially self selected, who fit the following criteria: ages eighteen to forty, no children, and not taking any hormones. Rosenthal's intentions were to be able to generalize to women in the general child bearing range; other studies had concentrated on college age females.

Instruments

This investigation uses data from only one of Rosenthal's instruments, the Menstrual Distress Questionnaire (Moos, 1968) because we are interested in the models rather than in reporting her results. The MDQ (her primary criterion measure) is a retrospective questionnaire with questions for three phases: most recent menstrual period (MENSTRUAL), week before menstrual period (PREMENSTRUAL), and remainder of the cycle (INTERMENSTRUAL). The MDQ has eight factors for each part of the cycle,all scored so that lower numbers indicate lower levels of distress save for the arousal scale, which is scored in reverse of the other seven. For this reason and because of the highly intercorrelated information, Rosenthal collapsed the results into three scores, one for each of the phases, each containing the total of the seven consistently scored scales. They are referred to as MENSEVN, PREMEVN, and INTSEVN (Menstrual scale comprising total of seven subscales and so on) in this presentation.

Method

The subjects were divided into three treatment groups receiving deep muscle relaxation, systematic desensitization, (behavioristic therapies) and a non-directive (talking) therapy (10,11, and 10 members respectively).

For purposes of this investigation, we removed one case from the first treatment group in order to create unequal cells across the board, feeling that this situation would provide the most difficult situation for finding equivalent results between programs and models.

The MDQ was administered seven weeks prior to treatment, immediately before treatment (to test for stability), and seven weeks after treatment.

Analysis

This was a problem which was originally posed for Analysis of Variance solutions, and we will present the result of those solutions and compare them with the solutions as they would have been formulated and run as Regression solutions. We will also present some Regression solutions which are not direct translations of the ANOVA solutions. A by product of this investigation is the opportunity to compare results and output between the several computer packages which we used.

The direct comparisons are with a one way analysis of variance with three groups; analysis of covariance (covarying the premeasure against the postmeasure); and a repeated measures (three groups by three trials) design.

The regression model representations are adaptations of those described by McNeil, Kelly, and McNeil (1975).

An initial concern would be that there be no difference between the groups on the MDQ prior to treatment. A one way analysis of variance is traditional. The one way equivalent is taken from Generalized Research Hypothesis 6.2 (p. 195).

Non-directional research hypothesis: the three treatments are equally effective on MENSEVN.

Statistical hypothesis: The three treatments are not equally effective on MENSEVN.

Full model: $Y_1 = a_0U + a_1G_1 + a_2G_2 + a_3G_3 + E_1$

Restricted model: $Y_1 = a_0U + E_2$

where:

Y_1 = mensevn score

U = 1 for all subjects

G_1 = 1 if criterion from subject in group 1, zero otherwise

G_2 = 1 if criterion from subject in group 2, zero otherwise

G_3 = 1 if criterion from subject in group 3, zero otherwise

a_0,1,2,3 are the least squares weighting coefficients

Degrees of freedom numerator: $(m_1 - m_2) = 3 - 1 = 2$

Degrees of freedom denominator: $(N - m_1) = 30 - 3 = 27$

The analysis of covariance models are taken from Applied Research Hypothesis 8.3 (p. 328). We are assuming here for demonstration purposes that the two premeasures are stable and are using the first.

Non-directional research hypothesis: Treatments differ on the post test over and above any differences observed on the prescores.

Statistical hypothesis: Treatments do not differ on the post test over and above any differences observed on the pre test.

Full model: $Y_2 = a_0U + b_1M_1 + a_1G_1 + a_2G_2 + a_3G_3 + E_1$

Restricted model: $Y_2 = a_0U + b_1P_1 + E_2$

where:

Y_1 = cmensevn (trial 3) score

U = 1 for all subjects

M_1 = mensevn (trial 1) score

G_1 = 1 if criterion from subject in group 1, zero otherwise

G_2 = 1 if criterion from subject in group 2, zero otherwise

G_3 = 1 if criterion from subject in group 3, zero otherwise

a_0, 1,2,3, and b_1 are least squares weighting coefficients

Degrees of freedom numerator: $(m_1 - m_2) = 4 - 2 = 2$

Degrees of freedom denominator: $(N - m_1) = 30 - 4 = 26$

The following model is a directional treatment of the previous hypothesis and, as such, is more in line with MLR thinking:

Research hypothesis: subjects are lower on the MDQ as a result of the experimental treatments over and above the pre test levels.

Statistical hypothesis: treatment group subjects are the same as control group subjects on the post test over and above pre test differences.

Full model: $Y_1 = a_0U + m_1M_1 + g_1G_1 + g_2G_2 + E_1$

Restricted model: $Y = a_0U + m_1M_1 + E_2$

where:

Y_1 = CMENSEVN score (post test)

M_1 = MENSEVN score (pre test)

G_1 = 1 if group one or group two (treatment groups)

G_2 = 1 if group three (control group)

$df_n = 3 - 2 = 1$

$df_d = 30 - 3 = 27$

A second directional hypothesis to measure change would be as follows:

Directional research hypothesis: Subjects in treatment group one score lower on the MDQ than subjects in treatment group two over and above pre test levels.

Statistical hypothesis: Subjects in treatment groups one and two are equal on the MDQ over and above pre test levels.

Full model: $Y_1 = a_0U + g_1G_1 + g_2G_2 + m_1M_1 + E_1$

Restricted model: $Y_1 = a_0U + E_1$

where:

Y_1 = CMENSEVN score (post test)

G_1 = 1 if subject from treatment group one, zero otherwise

G_2 = 1 if subject from treatment group two, zero otherwise

M_1 = MENSEVN score (pre test)

$df_n = 2$

$df_d = 27$

Also required in this sequence would be a test of the stability of the two pretreatment measures (McNeil, Kelly, and McNeil, p. 324).

Research hypothesis: Subjects are equal on the MDQ at the two pre test assessment periods.

Statistical hypothesis: Subjects are not equal on the MDQ at the two pre test assessment periods.

Full model: $Y_1 = a_0U + t_1T_1 + t_2T_2 + p_1P_1 + p_2P_2 + p_nP_n + E_1$

Restricted model: $Y_1 = a_0U + p_1P_1 + p_nP_n + E_2$

where:

Y_1 = score on MDQ (MENSEVN and BMENSEVN read in consecutively)

T_1 = 1 if score taken at time one, zero otherwise

T_2 = 1 if score taken at time two, zero otherwise

P_1 = 1 if score taken from subject one, zero otherwise

P_n = 1 if score taken from subject n, zero otherwise

$df_n = (n + 1) - n = 1$

$df_d = (n + n) - (n + 1) = 60 - 31 = 29$

A traditional approach to this problem would include a repeated measures, groups by trials, design. McNeil, Kelly, and McNeil present a model including person vectors, which should approximate the results of a standard repeated measures program such as the one in Veldman's ANOVAR. (332)

Non-directional research hypothesis: The three treatments are different over the three measurement periods over and above individual differences.

Statistical hypothesis: The three treatments are the same, over and above individual differences.

Full model: $Y_1 = a_0U + a_1I_1 = a_2I_2 + \ldots. a_9I_9 + p_1P_1 + \ldots.p_{30}P_{30} + E_1$

Restricted model: $Y_1 = a_0U + g_1G_1 + g_2G_2 + g_3G_3 + t_1T_1 + t_2T_2 + t_3T_3 + p_1P_1 \ldots + p_{30}P_{30} + E_2$

where: $I_1 = G_1 * T_1$

$I_2 = G_1 * T_2$

and so on

G_1, G_2, G_3 = treatment groups one to three

T_1, T_2, T_3 = times one to three

$df_n = 36 - 32 = 4$

$df_d = 90 - 36 = 54$

RESULTS

Tables 1 through 4 show the results of the oneway analysis of variance on the menstrual scale (MENSEVN) as solved by SPSS program ONEWAY, Veldman program ANOVAR, Veldman program REGRAN, and DP LINEAR. The first two are analysis of variance routines; the last two are hypothesis testing regression routines. With very minor exceptions, the four solutions calculate the same F test and probability results, although the kinds and amount of information are different, particularly between the ANOVA and regression outputs. ANOVA programs routinely report sums of squares, mean squares, degrees of freedom, F ratios, probabilities; regression routines routinely report degrees of freedom, multiple correlation squared, F ratios, probabilities and an intercorrelation matrix. ANOVA routines usually report cell means; regression routines usually do not.

Table 5 and 6 show the results of the analysis of covariance models as run on SPSS and the over and above hypothesis run on DP LINEAR. Neither does SPSS include a hypothesis testing regression program nor Veldman a covariance routine in the ANOVAR program. As a result, both models could not be tested on the same program library; however, F test results from the two different programs are virtually the same.

Tables 7, 8 and 9 show the results of the t-test on the means of a repeated administration of the MDQ as run on SPSS T-Test program and the regression formulation with person vectors as run on DPLINEAR. Results are the same. The correlation reported by the SPSS program was also duplicated by DPLINEAR using a model in which the measures at time two were used to predict the measures at time one.

Tables 10 through 12 show the results of the repeated measures (groups by trials) questions as run on SAS (Macro written by George Gantner, see Appendix A), ANOVAR, and on DPLINEAR. The F test is the same for all three.

In addition, for those who may not be familiar with some recent additions to SPSS, Appendix B shows results produced by program RELIABILITY, which allows for a repeated measures design but does not allow for a second factor. By imposing a subfile structure, it is possible to produce a repeated measures analysis for each of several groups; a result from that approach is presented in Appendix B.

DISCUSSION

The results of this project will provide no great surprises for the experienced student of regression. That there are valid regression substitutions has been discussed in many sources including the one with which we have worked. For some researchers who have been trained in analysis of variance techniques and for whom regression is a new approach, we would hope that the comparison of actual results is helpful. Our own experience has been that it is possible to find regression approaches to use in lieu of more traditional ones and that it is possible, as we have tried to demonstrate, to find directly comparable solutions. It is not, however, always easy to do the latter. We spent some trial and error hours finding the appropriate repeated measures model, having been seduced by the fact that three groups times three trials looks like the interaction model on page 210 of McNeil, Kelly, and McNeil when, in fact, the more complex model we have presented (taken from page 332) is the appropriate one. We also found that the many data transformations necessary to create vectors for each person as well as for each interaction introduce

considerable room for error in keypunching. Granted that one ought not to make such errors, nonetheless, the careful researcher should note that a mistake in that part of the program will produce incorrect results and that careful proofreading and crosschecking are essential.

References

Barr, A. J., Goodnight, J. H., Sall, J. P., & Helwig, J. T. A User's Guide to SAS 76. Raleigh, N. C.: SAS Institute, Inc., 1976.

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

Moos, R. H. The development of a menstrual distress questionnaire. Psychosomatic Medicine, 1968,30,853-867.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical Package for the Social Sciences, 2nd ed. New York: McGraw Hill, Inc., 1975.

Veldman, D. J. FORTRAN Programming for the Behavioral Sciences. New York: Holt, Rinehart and Winston, 1967.

ANOVA VERSION

FILE MENSTRUA (CREATION DATE = 01/23/78)

- O N E W A Y - - - - - - - - -

VARIABLE MENSEVN

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES
BETWEEN GROUPS	2	817.8027	408.9011
WITHIN GROUPS	27	23277.3867	862.1252
TOTAL	29	24095.1875	

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM
GRP00	9	113.3333	32.9659	10.9886	53.0000
GRP01	11	117.9091	29.2761	8.8271	76.0000
GRP02	10	105.5000	25.8468	8.1755	74.0000
TOTAL	30	112.4000	28.8248	5.2627	53.0000

RATIO	F PROB.
0.474	0.6274

Table 1

*** CLIPPER PLOT PROGRAM ANOVA ***

TEST OF ANOVA HALF OF ANOVA-REGRESSION COMPARISON

PARAMETERS
COL 1- 5 = 5
COL 6-10 = 5
COL 11-15 = 1
COL 16-20 = 0
COL 21-25 = 0
COL 26-30 = 5

DATA FORMAT = (A9,3F3.0)

GROUP 1 9 SUBJECTS.

GROUP 2 11 SUBJECTS.

GROUP 3 10 SUBJECTS.

Table 2

VARIABLE 1 ANALYSIS.

SOURCE	MEAN SQUARE	D.F.	F-RATIO	P
TOTAL	830.8706	29.		
GROUPS	409.7963	2.	0.474	0.6328
ERROR (G)	862.1272	27.		

G MEAN	1	2	3
	113.3333	117.9091	105.5000

Table 3

*** OUTPUT FROM PROGRAM REGRAN ***

REGRESSION HALF OF REGRESSION ANOVA COMPARISON

PARAMETERS
COL 1- 5 = 6
COL 6-10 = 30
COL 11-15 = 3
COL 16-20 = 3
COL 21-25 = 1
COL 26-30 = 5

DATA FORMAT = (A9,3F3.0,3F1.0//)

INTERCORRELATION ANALYSIS.

MODEL 1 M1 CRITERION = 1

PREDICTORS = 0- 6
P = 6 RSQ = 0.0296
P = 5 RSQ = 0.0339

R = 0.3682 RSQ = 0.0339

V	BETA	R
4	0.0	0.0
5	0.0770	4.5756
6	-0.1303	-7.8333
REG. CONST. =		113.3333

MEANS	1	2	3	4	5	6
	112.0000	93.0333	61.3333	0.3000	0.3667	0.3333

SIGMAS	1	2	3	4	5	6
	29.3905	33.1100	18.7995	0.4583	0.4819	0.4714

R MATRIX	1	2	3	4	5	6
1	1.0000	0.3822	0.5180	0.0216	0.1479	-0.1722
2	0.3822	1.0000	0.3697	0.3636	-0.0296	-0.3232
3	0.5180	0.3697	1.0000	-0.0193	0.1190	-0.1028
4	0.0216	0.3636	-0.0193	1.0000	-0.4981	-0.4629
5	0.1479	-0.0296	0.1190	-0.4981	1.0000	-0.5380
6	-0.1722	-0.3232	-0.1028	-0.4629	-0.5380	1.0000

F-TEST 1 REDUCED BY GROUPS
RSQ FULL = 0.0339 MODEL 1
RSQ REDUCED = 0.0 MODEL 0
DIFFERENCE = 0.0339
DEN = 2. DEN = 27. F-RATIO = 0.474 P = 0.6328

CORRELATIONS

	1	2	3	4	5	
1	1.0000	0.3822	0.5180	0.0216	0.1479	-0
2	0.3822	1.0000	0.3697	0.3636	-0.0296	-0
3	0.5180	0.3697	1.0000	-0.0193	0.1190	-0
4	0.0216	0.3636	-0.0193	1.0000	-0.4981	-0
5	0.1479	-0.0296	0.1190	-0.4981	1.0000	-0
6	-0.1722	-0.3232	-0.1028	-0.4629	-0.5380	1

....GROUPSON M 0.99999997E-05
CRITERION 1
PREDICTORS 4 - 6

2 ITERATIONS

RSQ = 0.03393997

VAR. NUMBER	STD. WT.	ERROR
4	0.0	0.00000003
5	0.07780540	-0.00000012
6	-0.13029736	0.00000006

VAR. NUMBER	WEIGHT
4	0.0
5	4.57574654
6	-7.83332348
CONSTANT=	113.33331299

****

****

GROUPSONM7 F = 0.4703 D.F. NUM.= 2 D.F. DEN.= 27

******

1 0.07390 99 0.0 PROB = 0.6270

Table 4

ANCOVA VERSION

FILE MENSTRUA (CREATION DATE = 01/26/78)

* * * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * *

CMENSEVN
BY CGROUP
WITH MENSEVN

* *

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGN OF
COVARIATES	6227.250	1	6227.250	7.110	0.0
MENSEVN	6227.250	1	6227.250	7.110	0.0
MAIN EFFECTS	5074.059	2	2537.029	2.897	0.0
CGROUP	5074.059	2	2537.029	2.897	0.0
EXPLAINED	11301.309	3	3767.103	4.301	0.0
RESIDUAL	22771.609	26	875.831		
TOTAL	34072.918	29	1174.928		

COVARIATE	RAW REGRESSION COEFFICIENT
MENSEVN	0.508

30 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

ANCOVA VERSION

FILE MENSTRUA (CREATION DATE = 01/26/78)

* * * M U L T I P L E   C L A S S I F I C A T I O N   A N A L Y S I S * *

CMENSEVN
BY CGROUP
WITH MENSEVN

* *

GRAND MEAN = 98.97

VARIABLE + CATEGORY	N	UNADJUSTED DEV'N	ETA	ADJUSTED FOR INDEPENDENTS DEV'N	BETA	ADJUSTED F INDEPENDEN + COVARIAT DEV'N	BE
CGROUP							
0	9	9.26				8.71	
1	11	-14.06				-17.26	
2	10	7.13				11.14	
			0.32				0.
MULTIPLE R SQUARED							0.3
MULTIPLE R							0.5

Table 5

CORRELATION:

	1	2	3	4	5
1	1.0000	0.4275	0.0216	0.1479	-0.1722
2	0.4275	1.0000	0.1798	-0.3174	0.1497
3	0.0216	0.1798	1.0000	-0.4981	-0.4629
4	0.1479	-0.3174	-0.4981	1.0000	-0.5360
5	-0.1722	0.1497	-0.4629	-0.5360	1.0000

....COVAREQUIV 0.99999997E-05
CRITERION 2
PREDICTORS 1 - 1
4 - 6

RSQ = 0.33167762 5 ITERATIONS

VAR. NUMBER	STD. WT.	ERROR
1	0.00643082	-0.00190293
4	-0.37402387	-0.00257266
5	0.03174704	-0.00000018
6	0.0	0.0

VAR. NUMBER	WEIGHT
1	0.57844377
4	-26.21316525
5	2.26962376
6	0.0
CONSTANT=	42.80458069

****

....RESTRICTED 0.99999997E-05
CRITERION 2
PREDICTORS 1 - 1

RSQ = 0.10276334 1 ITERATIONS

VAR. NUMBER	STD. WT.	ERROR
1	0.32750829	0.0

VAR. NUMBER	WEIGHT
1	0.50837547
CONSTANT=	41.82527161

****

COVAREQUIV F = 2.8966 D.F. NUM.= 2 D.F. DEN.= 26

1 0.33168 2 0.10276 PROB = 0.07316

Table 6

Table 7

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

FILE MENSTRUA (CREATION DATE = 02/08/78) STUDY

- T - T E S T - -

VARIABLE	NUMBER OF CASES	MEAN	STANDARD DEVIATION	STANDARD ERROR	(DIFFERENCE) MEAN	STANDARD DEVIATION
MENSEVEN	31	112.7742	29.437	5.287	8.7419	20.370
BMENSEVN		104.0323	29.501	5.299		

STANDARD ERROR	CORR.	2-TAIL PROB.	T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.
3.659	0.761	0.000	2.39	30	0.023

Table 8

CORRELATIONS

	1	2
1	1.0000	0.7574
2	0.7574	1.0000

[illegible]

[illegible] ITERATION

[illegible]

[illegible]

[illegible]

CORRELATION F = [illegible] D.F. NUM.= 1 D.F. DEN.= 28

*****

1 0.97368 99 0.0 PROB = 0.00000

PREDICTORS 3 - 34

RSQ = 0.88109595

VAR. NUMBER	STD. WT.	ERROR
3	0.14648181	-0.00056219
4	0.0	0.00056219
5	-0.02457492	-0.00000066
6	0.29297614	0.00019693
7	0.36671370	-0.00002044
8	0.27018416	-0.00106913
9	-0.16444850	-0.00126433
10	0.22444123	-0.00057840
11	0.25375968	0.00106454
12	0.02183281	0.00019616
13	0.35869384	0.00124556
14	-0.09184122	-0.00265115
15	-0.02242427	0.00222409
16	0.26567400	0.00064021
17	0.14266354	0.00088441
18	0.0	-0.00007766
19	0.34285557	0.00079834
20	-0.06783980	-0.00013399
21	0.18033159	-0.00158498
22	0.13463247	-0.00104860
23	0.03023295	0.00251114
24	0.42181337	-0.00039327
25	-0.05362921	0.00181705
26	0.17140424	-0.00125755
27	0.38953400	-0.00191164
28	0.15088463	-0.00017291
29	-0.01997955	-0.00162178
30	-0.02059705	0.00092685
31	-0.02188374	-0.00040424
32	0.10446620	0.00280653
33	0.06951225	-0.00147867
34	0.38556516	0.00035787

****

...TTESTRESTM 0.99999997E-05
CRITERION 2
PREDICTORS 5 - 34

RSQ = 0.85948616

VAR. NUMBER	STD. WT.	ERROR
5	-0.01739192	0.00292158
6	0.29722655	0.00008541
7	0.37089032	-0.00020826
8	0.27417290	-0.00145131
9	-0.16098297	-0.00218773
10	0.22867155	-0.00071061
11	0.25789106	0.00082994
12	0.02848206	0.00256604
13	0.36299896	0.00119066
14	-0.08562195	-0.00072581
15	-0.01852488	0.00174952
16	0.26986152	0.00046366
17	0.14644986	0.00029263
18	0.00451702	0.00008672
19	0.34713221	0.00071400
20	-0.06335276	-0.00000072
21	0.18617505	-0.00004852
22	0.14012367	0.00012337
23	0.03304263	0.00090927
24	0.42701262	0.00047684
25	-0.05109492	-0.00006974
26	0.17523217	-0.00180612
27	0.39433539	-0.00145310
28	0.15412974	-0.00132437
29	-0.01539097	-0.00138336
30	-0.01647887	0.00067860
31	-0.01753479	-0.00041372
32	0.10583609	-0.00028479
33	0.07406390	-0.00127852

****

TTESTEQUIV F = 5.2705 D.F. NUM.= 1 D.F. DEN.= 29 1 0.88110 2 0.85949

PROB = 0.02911

Table 9

RATIONS

```
      *** OUTPUT FROM PROGRAM ANOVAR ***
THREE BY THREE GROUPS ON VELOMAN ANOVA VERSION

PARAMETERS
COL  1- 5 =     1
COL  6-10 =     3
COL 11-15 =     3
COL 16-20 =     0
COL 21-25 =     0
COL 26-30 =     5

DATA FORMAT = (A3,F3.0/9X,F3.0/9X,F3.0)

GROUP 1      9 SUBJECTS.

GROUP 2     11 SUBJECTS.

GROUP 3     10 SUBJECTS.


VARIABLE  1 ANALYSIS.

SOURCE      MEAN SQUARE      D.F.      F-RATIO        P

TOTAL          973.7036       89.
BETWEEN       2056.3318       29.
 GROUPS       1429.2188        2.        0.680      0.5195
 ERROR (G)    2102.7847       27.
WITHIN         450.4331       60.
 TRIALS       1387.4688        2.        3.701      0.0303
 G BY T       1001.1563        4.        2.670      0.0411
 ERROR (T)     374.9338       54.

G MEAN              1              2              3
                112.6296       98.7576       105.2000

T MEAN              1              2              3
                112.4000       103.8333      98.9667

G BY T              1              2              3
      1        113.3333       116.3333      108.2222
      2        117.9091        93.4545       84.9091
      3        105.5000       104.0000      106.1000
```

Table 10

SAS

GROUPS BY TRIALS
TWO FACTOR ANOVA - GROUP BY TRIAL
< TRIAL IS REPEATED MEASURE>

GENERAL LINEAR MODELS PROCEDURE

ENDENT VARIABLE: MEASURE

URCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALU
DEL	59	86659.60000000	373.70337079	99999.9
OR	0	0.00000000	0.00000000	
RECTED TOTAL	89	86659.60000000		

RCE	DF	TYPE I SS	F VALUE	PR > F	D
P	2	2858.44309764	.	.	
JECT(GROUP)	27	56775.42356902	.	.	2
L	2	2775.26666667	.	.	
P*TRIAL	4	4004.38922559	.	.	
L*SUBJECT(GROUP)	54	28245.67744108	.	.	5

S OF HYPOTHESES USING THE TYPE IV MS FOR SUBJECT(GROUP) AS AN ERROR TERM

CE	DF	TYPE IV SS	F VALUE	PR > F
P	2	2858.44309764	0.68	0.5152

S OF HYPOTHESES USING THE TYPE IV MS FOR TRIAL*SUBJECT(GROUP) AS AN ERROR TERM

CE	DF	TYPE IV SS	F VALUE	PR > F
L	2	2775.26666667	3.70	0.0312
P*TRIAL	4	4004.38922559	2.67	0.0418

Table 11

****....GPTIPERSON 0.99999997E-05
CRITERION 2
PREDICTORS 11 - 49

RSQ = 0.76633173

VAR. NUMBER	STD. WT.	ERROR
11	-0.00819026	-0.00057101
12	0.02205475	-0.00093412
13	-0.05564047	-0.00013083
14	0.08593541	-0.00077212
15	-0.17676663	0.00050527
16	-0.26632231	0.00124145
17	0.0	-0.00100359
18	-0.01820346	0.00175096
19	0.0	-0.00169718
20	-0.13450277	-0.00172287
21	0.19648767	0.00156850
22	0.30133814	0.00231695
23	0.24706143	-0.00196731
24	-0.17355102	-0.00022757
25	0.15234715	-0.00020903
26	0.03685201	-0.00000002
27	-0.03286016	-0.00030422
28	0.11825687	0.00043106
29	-0.10030907	0.00101173
30	-0.07442808	-0.00014162
31	0.16699147	0.00225025
32	0.04705375	-0.00014177
33	-0.04518731	0.00018591
34	0.31647390	-0.00069934
35	-0.10099673	0.00030029
36	0.03272816	0.00299174
37	0.04064761	-0.00277911
38	-0.01372488	-0.00117791
39	0.36340785	-0.00002122
40	-0.13782346	0.00018030
41	0.06386989	-0.00062245
42	0.17426652	-0.00012118
43	0.06233719	-0.00021323
44	-0.06841391	0.00017142
45	-0.09606159	-0.00050277
46	-0.09438932	0.00122714
47	0.04336784	0.00011102
48	0.11226928	-0.00241816
49	0.26744831	0.00052506

GPTIPERSON F = 2.6691 D.F. NUM.= 4 D.F. DEN.= 54

****....RESTGTPERS 0.99999997E-05
CRITERION 2
PREDICTORS 5 - 10
20 - 49

RSQ = 0.72013313

VAR. NUMBER	STD. WT.	ERROR
5	0.0	0.00055534
6	0.0	-0.00000435
7	0.0	-0.00053471
8	0.19673830	-0.00185752
9	0.06935561	0.00228198
10	-0.00638164	-0.00042439
20	-0.04889162	0.00138348
21	0.27812380	0.00056285
22	0.38095069	-0.00078207
23	0.33155382	-0.00001818
24	-0.09023714	0.00050247
25	0.23457319	-0.00060445
26	0.11989707	0.00045177
27	0.05041094	0.00038141
28	0.20000452	-0.00045925
29	-0.07682508	0.00041342
30	-0.05214722	-0.00198460
31	0.18959761	0.00074375
32	0.06980479	-0.00149845
33	-0.02415153	-0.00294518
34	0.34183466	0.00064355
35	-0.07618934	0.00107088
36	0.05389842	-0.00000030
37	0.06875110	0.00140120
38	0.01136464	-0.00011551
39	0.38967586	0.00226033
40	-0.05197487	0.00018835
41	0.15118605	0.00090366
42	0.26022792	0.00000364
43	0.14696985	-0.00146304
44	0.01629545	-0.00099891
45	-0.00922751	0.00052476
46	-0.00976519	-0.00003141
47	0.13100088	0.00196516
48	0.20007032	-0.00039017
49	0.35074693	-0.00210458

Table 12

4 0.76633 5 0.72013

PROB = 0.04182

```
MACRO ANOVU2R1
PROC GLM DATA=DATASET;
CLASSES FAC1 FAC2 SUBJECT;
MODEL VARLIST=
FAC1 SUBJECT(FAC1)
FAC2 FAC1*FAC2 FAC2*SUBJECT(FAC1);
TEST H=FAC1 E=SUBJECT(FAC1);
TEST H=FAC2 FAC1*FAC2 E=FAC2*SUBJECT(FAC1);
TITLE2 TWO FACTOR ANOVA  -  FAC1 BY FAC2;
TITLE3 (FAC2 IS REPEATED MEASURE);
%
MACRO MEANS2
PROC MEANS DATA=DATASET;VAR VARLIST;
TITLE2 OVERALL MEANS;
PROC SORT DATA=DATASET;BY FAC1;
PROC MEANS DATA=DATASET;BY FAC1;VAR VARLIST;
TITLE2 CELL MEANS - FAC1 BY FAC2;
PROC SORT DATA=DATASET;BY FAC2;
PROC MEANS DATA=DATASET;BY FAC2;VAR VARLIST;
PROC SORT DATA=DATASET;BY FAC1 FAC2;
PROC MEANS DATA=DATASET;BY FAC1 FAC2;VAR VARLIST;%
DATA SUBSTITU;
INPUT SUBJECT 1-3 TRIAL 4 MENSEVN 10-12 PRESEVN 13-15 INTSEVN 16-18 GROUP 20;
CARDS;
```

APPENDIX A

REPEATED MEASURES ON MENSEVN USING PROGRAM RELIABILITY

FILE RUTHOISS (CREATION DATE = 01/29/78)
SUBFILE CROUPONE

* * * * * * * * R E L I A B I L I T Y A N A L Y S I S F O R S C

MEANS

1.	TRIAL1	113.33333
2.	TRIAL2	116.33333
3.	TRIAL3	108.22222

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SS	DF	MEAN SQUARE
BETWEEN PEOPLE	19764.29630	8	2470.53704
WITHIN PEOPLE	9616.00000	18	534.22222
BETWEEN MEASURES	302.74074	2	151.37037
RESIDUAL	9313.25926	16	582.07870
TOTAL	29380.29630	26	1130.01140

GRAND MEAN = 112.62963

RELIABILITY COEFFICIENTS

N OF CASES = 9.0 N OF ITEMS = 3

ALPHA = 0.76439

A L E (R E A C T I O N) * * * * * * * * *

STD DEV	CASES
32.96589	9.0
30.01666	9.0
40.58256	9.0

F	PROBABILITY
0.26005	0.77421

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOL. 8 NO. 3 1978

SELECTING AN APPROPRIATE MODEL FOR DATA ANALYSIS

PATRICIA COHEN
New York State Department of Mental Hygine

Let me preface my remarks with what is, I suspect, the most important point one can make about multiple regression or any other complex means of handling data. That is, these procedures are tools which may enable you to obtain the best possible, most valid and informative answer to questions which you as a substantive expert have thoughtfully asked. Thus, the whole point of the analysis is, as Jacob Cohen entitled an address, "Multiple regression in the service of the ego". The justification for using multiple regression is that it enables one to answer the questions of interest, however its appropriate application presupposes a relevant selection of data for analysis.

Unfortunately, complex analytical methods sometimes seduce unwary researchers by their fascinating inner workings to the extent that these researchers neglect to pay sufficient attention to the overall logic of the research. It is I fear, this too frequent tendency to "forget the forest while counting rings on trees" which has given MRC, factor analysis and in general procedures which require coefficients beyond means and frequency differences a bad name among some thoughtful researchers. Perhaps it is the relative novelty to many users of these analytical procedures which produces these tendencies. Let me urge all of you to read in the micro-economic literature where you will find an easy familiarity with regression which allows the method to simply

serve as a means for illuminating substance. In this context, let me comment on this afternoon's papers, roughly in the sequence in which they were presented.

First, on the paper on ridge regression, let me say how very useful it is to the field to accumulate empirical examinations of the procedures which have been proposed on theoretical grounds as ways of coping with certain analytical problems. Ridge regression is just such a fascinating and promising development and the study's failure to find the expected improvement on cross validation is quite interesting. However, two difficulties make the implications of this finding unclear. First, we note that the two subsamples (each of 280 cases) gave us very different R^2 on the 4th grade achievement variables (.24 and .38 respectively) and on the Teacher variables (.16 and .05 respectively). These large differences make one concerned that something may have gone awry in the assignment of alternate districts to the two samples. One would be happier if other, truly random, ways of selecting subsamples were to yield the same findings.

Second, perhaps more serious, are problems which result from a failure to provide justification for the variable choice. Just why are we interested in the unique predictive value of each of the 4th grade scores, especially when only about 1% is added to the predicted variance by the three remaining variables after the best predictor? Similarly, staff per 1000 is very highly correlated with teachers per 1000 students. The issues involved would be better assessed by

separate measures of teachers per 1000 students and non-teacher instructional staff per 1000 students which, no doubt, would be much less correlated. As Marquardt and Snee (1975) state one should turn to ridge regression only after all unnecessary multicollinearity has been eliminated, and only because your theoretical model demands all of the remaining variables.

If I read the paper on the Historical Regression Method correctly I think the analysis obscures a critical problem. The prediction that the best estimate of future progress in the educational system is the rate of past progress is explicitly under the "no special program" condition. However, the data in Table 1 suggest that most children entered the "compensatory programs" during their first year in school (mean years in school = .56). If this is so, clearly no base rate of educational progress can have been established. An appropriate comparison would have been the achievement of comparable but "no treatment" classes.

Moving on to the neighborhood predictors of reading achievement, I would like to say that I liked this analysis and personally would like to see more like it. There are several reservations about the conclusions, however, that I would like to voice. First, one cannot be sure whether between-city apparent differences in findings were attributable to low power because of small sample size, on the one hand, or to low β because of lesser variability on the

"independent" variable on the other. We may note, for example, that the most significant paths were found in the city (Chicago) with the largest sample size.

The second problem, the possible non-comparability of β because of differences in variability, would have been solved by the use of unstandardized (raw) regression coefficients. B weights are generally much to be preferred whenever they can be meaningfully interpreted and especially when comparing populations. In this case B coefficients would have also been substantively interesting (e.g. the direct and indirect effects of a $ 1000 increase in median family income on reading achievement scores). They also could have been tested in a more elegant and powerful analysis by pooling the cities data and testing for interactions.

Another problem in this study is that no rationale is provided for the selection of the variables over other available alternatives. Which income measure did they use - % poverty, % children in poverty, median family income, mean earned income - and why? It is of critical importance in any causal analysis to consider these alternatives from a theory standpoint, and to provide evidence whenever possible which bears on the validity of the choice. These choices may be even more important when compounded by the issue of aggregated relationships, as in this study.

Finally, the authors refer to other findings indicating that the high density - low achievement relationship disappears

in high income neighborhoods. Since most of the analyses reported here find that these two variables each have an effect on achievement when the other is controlled the authors must mean that there is also an interaction between income and density in the prediction of achievement. Interactions can be handled in path analysis only by drawing separate path diagrams for each variable conditional on the other. Unfortunately the tests for the significance of interactions seem to have been omitted in the analysis.

Moving on to the evaluation of sex-related salary discrimination, although this example uses fictitious data we again encounter the problems created by an inadequate consideration of the consequences of variable selection. The moral of the story is that if you ignore the effect of a variable which is relevant and either redundant with or suppressing another variable which you have included, you will have errors of identification due to errors of specification. For example, if time in rank was itself the result of sex discrimination we will have made the correct conclusion in institution A. Indeed we will have underestimated the sex discrimination effect by including rank as a variable in the analysis, or overestimated it by excluding considerations such as time in the field or productivity.

The empirical analysis of higher education salaries presented by Martin and Williams I found fascinating, particularly as it raises some of the same knotty problems

we have been trying to cope with in attempts to assess equity and inequity in the provision of health and mental health services. By using the empirical regression as a standard one succeeds in identifying inequities apparently suffered by individuals relative to the entire system (assuming that appropriate unmeasured variables do not account for this). One does not, however, correct inequities connected with an entire class. For example, we are implicitly treating the inequities between programs as proscriptive as well as descriptive. One might alternatively use the empirical values as a starting point for the proscriptive values. For example, it hardly seems sensible that most kinds of service should go unrewarded. Why shouldn't an overall service index be constructed from the 5 variables and an appropriate weight be assigned proscriptively. Similarly, one may question the significance criterion as relevant to certain issues - for example, an appropriate weight for sex is zero. Finally, it may be noted that age would have been quite significant had it been represented linearly (or perhaps quadratically) rather than as five categorical variables. Only if one employs an empirical model such as the one here to create a proscriptive model may one avoid the problem of making current inequities "institutionalized" and even more pervasive.

Let me now skip ahead for a moment, to the paper on prediction intervals. Unfortunately there seem to be some serious errors which have led the author to conclude that confidence limits in cases with nearly perfect prediction are much larger

than my calculations indicated they were. Contrary to assertion, for a constant R^2 mean square error is entirely a function of raw unit scale. It may be quite useful for those who use multiple regression equations to estimate individual scores to point out that this prediction will be less accurate for scores at the extremes of the distributions than for those near the means (see Cohen and Cohen, 1975, p. 114).

Finally, I would like to discuss the comparison of the log linear and regression models lucidly presented by Jay Magidson.

Whenever we have developed alternative means of analyzing certain kinds of data sets, it is of extreme value to have a variety of empirical comparisons of the relative advantages and disadvantages of each. These empirical studies give us some means of assessing the practical importance of various factors known to be theoretically important. In comparing a log-linear analysis with a regression analysis there are several reasonable criteria available for assessing their relative advantages.

1. First, of course, we are critically concerned with the nature of the variables, and most particularly with the nature of the dependent variable. If the dependent variable is a true dichotomy, and especially if it is one on which we find a distinctly uneven split, the log-linear model would have a distinct initial edge in our preference rating. The reason for caring about the split is the expectation that

when splits are no more extreme than 3/1 the two methods will give similar results and other considerations enter in. If, however, the dependent variable is only a "dichotomy of convenience" on a variable which may take a range of values, the a priori preference for a log linear model is not so clearcut. Sometimes it is not so easy to be sure whether a variable approaches a true dichotomy or not (as in the current case of camp location preference for which we may clearly at least have a variety of strengths of preference).

The nature of the independent variables would also influence one's a priori preference. In the current study we have three fairly straightforward dichotomies (Region of origin, present location, race). In many studies the independent variables are not at all confined to truly categorical variables, and in fact frequently require division of ordered variables into categories - a procedure which either capitalizes on chance or risks making inappropriate a priori cuts. These problems are compounded when we have more than three or four independent variables because unless our sample is exceedingly large and the variables are not substantially correlated we will run into the problem of empty cells using the log linear model.

2. The second major criterion for preferring one analysis mode rather than the other has to do with the replicability of the findings. If it could be shown that the odds ratios obtained from a variety of studies were more homogeneous than

were the regression coefficients we would prefer the former, and vice versa. One circumstance in which we may expect odds ratios to be unstable is when the dependent dichotomy is an estimate of a normally distributed variable. Mosteller (1968) has shown the odds ratio in such cases to be sensitive to the choice of breaking point, thus we may expect that to the extent that investigators may vary on this choice the odds ratio will vary. Of course, in this circumstance, although we would expect stability of the regression coefficient for the true dependent variable we may not have a stable regression of the dichotomy on the independent variables either.

An issue related to replicability is the validity of the significance tests. Here the question is usually whether the known failure to meet the assumptions of the regression model result in estimated P values which will seriously mislead. It would appear thus far that the answer is "not usually", but we need more information on the limits here.

3. Finally we have the question of parsimony of the required model, the criterion applied in this paper. Parsimony is, by its very nature, a ceteris paribus conditon. That is, we prefer the simpler model providing we have no good reason not to. In this study I see no particular advantage to complexity, therefore I would agree that on this criterion we would prefer the log linear findings.

REFERENCES

Cohen, J. and Cohen, P. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Lawrence Erlbaum Assoc., Hillsdale, N.J. 1975.

Marquardt, D.W. and Snee, R.D. Ridge regression in practice. The Amer. Stat., 1975, 29: 3-20.

Mosteller, F. Association and estimation in contingency tables. J. A. S. A., 1968, 63: 1-28.

Minutes of the 1978 Multiple Regression Special Interest Group
B u s i n e s s - M e e t i n g

Approximately 25 persons were in attendance at the annual meeting.

The Chair, Steve Spaner, brought the meeting to order.

Announcements regarding the MLRSIG sponsored social and path analysis training session were made.

The report from the Viewpoints editor, Izadore Newman, indicated that the journal has received and accepted enough articles to fill this next year's four issues. Discussion was raised concerning the appropriate group of members to receive the MLRSIG convention proceedings: last year's paid members or this comming year's paid members. Since one of the irticements to join MLRSIG is the receipt of the proceedings it was suggested and approved to send the proceedings to all the past year members and those new members who have paid by June 1 of the year.

The report from the Executive Secretary, Steve Spaner, followed next. The members were appraised of a letter of criticism received by Steve Spaner presenting a view that the MLRSIG Executive Officers and the Viewpoint's Articles were representative of only a specific faction and school of thought within the field of multiple linear regression techniques. The letter writer in essence called for a broader representation of the MLR practioners in both the leaderships and the published articles of the SIG. Steve Spaner indicated his response to the writer and his agreement in principle with the writer. Steve pointed out the efforts which he had made to broaden the attraction of and input to the MLRSIG; namely the convention programs sponsored by the SIG. (Parenthetically, both the symposium and the Path Analysis Training Session were very successful, attracting 80 and 65 AERA conventioners respectively.)

Steve gave the treasurers report next. The treasury currently is balanced at $478.06 but this is before the fourth quarter interest earnings, which was an asset of $7.05, and the publication costs of the convention proceedings, which is an expected liability of about $400.00. The hoped for carry over would then be in the $85.00 range.

The members were next informed of a letter received by the Chair, Steve Spaner, from Dick Schutz, AERJ Editor, requesting names of persons that our SIG would consider qualified to be referees for AERJ. The letter spoke of a desire on the part of AERJ to broadly reflect the interests and inquiry of the AERA Special Interest Groups. The letter encouraged SIG members to send their articles to AERJ. Steve asked if any members of the audience wished to have their name submitted; receiving no volunteers, Steve indicated his willingness to convey the names and credentials of any MLRSIG member who feels qualified and interested in becomming an AERJ referee.

The next item of business was the report of an Executive Board recommendation. The Chair reported that he had posed the problem of inequity in the SIG's allocation for the social event at each annual meeting and the actual costs that were being incurred. The practice has become that the Chair rents a suite in one of the hotels for one night during the convention. The suite serves as a meeting place but also as the residence for the chair for one night; thus, theoretically, giving our chair a free night of lodging in consideration for their work on the convention program and in recognition of their obligation to be at the convention. The allocation for this event was set at $65.00 at the 1973 meeting. However, in no year since then has the allocation covered the cost of the suite let alone provided funds for refreshments. Previous chairs and the present MLRSIG chair have paid the difference on the suite out of their own pockets and conducted BYOB socials.

After polling the executive board on the issue, the board recommendation was to increase the allocation to $85.00. The motion to increase the allocation for the MLRSIG social to $85.00 was seconded and passed. The discussion brought out the possibility of the SIG approaching the convention hotel, independent of the AERA channels, to explore the posibility of a complimentary suite for the SIG chair. This often is the practice if the organization can guarantee or promise a certain number of their group as paid guests.

The next business item was concerned with the possible selling of our membership labels for a mailing on a workshop being conducted. The issue was recognized as being twofold: whether to release the mailing addresses and to whom and whether to charge for the addresses and how much. The group voted to release the mailing addresses to MLRSIG members only at no cost. If requests become numerous this policy will be reviewed and possibly revised by the executive board.

Elections were the last item of business. Two replacements were to be elected for the Executive/Editorial Board (Drs. Pohlman and Jennings terms were expiring). The nominations committee provided a slate of eight nominees and two nominations were made from the floor. Note was made that not all the nominees were paid-up MLRSIG members. No deletions were made on this account but a policy decision was called for from the executive board. Winners of the election for the Executive/Editorial Board were Dennis Leitner of Southern Illinois University at Carbondale and Leigh Burstein of the University of California - Los Angeles.

A chair-elect was the next position to be filled. The nominations committee provided a slate of three nominees and one nomination was made from the floor. The new chair-elect is Bill Connett of the Montana State Department of Education.

The meeting was adjourned to be reconvened by the new chair, John Williams, next year in SAN FRANCISCO!

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08540

May 2, 1978

Dr. Isadore Newman
College of Education
University of Akron
Akron, Ohio 44325

Dear Isadore:

Multiple Linear Regression Viewpoints will be indexed cover-to-cover for the Current Index to Journals in Education (CIJE). The August 1977 issue, Volume 8, Number 1, will be the first to be indexed in CIJE.

Thank you for bringing viewpoints to our attention.

Best wishes,

Eleanor V. Horne
Associate Director

EVH/bg

TELEPHONE: 609-921-9000

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

Title
Author and affiliation
Single-space indented abstract (the rest of the manuscript should be double-spaced)
Introduction (purpose—short review of literature, etc.)
Method
Results
Discussion (conclusion)
References

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

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

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

"Membership in the *Multiple Linear Regression Special Interest Group* is renewed yearly at the time of the American Educational Research Association Convention. Membership dues pay for a subscription to the *Viewpoints* and are divided into two categories: individual = $2.00; and institutional (libraries and other agencies) = $12.00. Membership dues and subscription requests should be sent to the Secretary of the MLRSIG."

THE UNIVERSITY OF AKRON
AKRON, OHIO 44325

TABLE OF CONTENTS