Multiple Linear Regression Viewpoints
Volume 3, Number 1
May, 1972

A publication of the Special Interest Group on Multiple Linear Regression

of the American Educational Research Association

Table of Contents

ON PAYING DUES

The SIG on multiple linear regression may be the only organization in the U. S. to lower it's dues. Previously, the dues have been $2.00. For the year 1972, the dues have been reduced to $1.00. If you have not already paid your 1972 dues, send your dollar to either Bill Connett, The University of Northern Colorado (Greeley) or Judy Lyon, CIRCE (St. Louis). This issue is being sent to several people who have not yet paid their 1972 dues and other likely members, in addition to being sent to paid up members.

WRITING FOR VIEWPOINTS

Every member of the SIG is encouraged (pleaded?) to make a written contribution to Viewpoints. Preference is given to the short papers (one, two, three or four pages, typewritten).

Also, members are encouraged to send in lists of publications in multiple linear regressions. There will be some attempt by the SIG to put together one or more books of readings on multiple linear regression. Your thoughts and contributions are sincerely encouraged. Send $1.00 a page directly to the editor when you send in your contributions to Viewpoints. Get those papers rolling!

A COMPARISON OF RAW GAIN SCORES, RESIDUAL GAIN SCORES, AND THE ANALYSIS OF COVARIANCE WITH TWO MODES OF TEACHING READING

John D. Williams
The University of North Dakota

Roger T. Maresh
Rockwood School District

James D. Peebles
University City, Missouri Public School System

The measurement of change has been seen to be one of the most difficult issues in psycho-educational research (Harris, 1963). Several different solutions have been proposed, and almost simultaneously, have been criticized. When pre and post-testing have taken place, an intuitively pleasing approach has been the use of raw gains (that is, the post-test score minus the pre-test score for each subject). The use of this measure has been severely criticized. Ruch (1970) has indicated his displeasure with gain scores because of their disregard for the psychology of learning. Because learning, in its latter phases, is often characterized by a negatively accelerated curve, those students who enter an experiment with more practice in the skill or concept being tested will be handicapped by the gain score approach. The student who has a smaller amount of prior practice enters the experiment during the initial phase of learning, which will allow him to be in a period of rapid acceleration in regard to measured learning.

A common approach to the problem of measuring change when a pre and post-test have been used is the analysis of covariance. The analysis of covariance is often used when the assignment of subjects to an experiment has been made on some basis other than strict randomization. The analysis of covariance takes into account the correlation between the pre-test and the post-test. More specifically, it is helpful to look at the process of the analysis of covariance as it can be generated through the use of linear models. Because the present application is concerned with two modes of instruction, one mode being the vertically grouped method of teaching reading, and the second method being the more typical graded method of teaching reading, and because a pre- and post-test are being used, the linear models developed here will represent that specific situation.

First, a full model can be defined. A full model is essentially a model that contains all the information relevant to the data analysis. For this specific situation, the full model is:

$$Y = b_0 + b_1X_1 + b_2X_2 + e_1, \qquad (1)$$

where

Y = the post-test score,

X_1 = the pre-test score,

X_2 = 1 if the score is from a member of the vertical group; 0 otherwise,

b_0 = the Y-intercept,

b_1 = the regression coefficient for X_1,

b_2 = the regression coefficient for X_2, and

e_1 = the error in prediction with the full model.

If this model is solved using a multiple linear regression computer programming routine, part of the output includes the multiple correlation coefficient (R). For the present usage, since a full model is being used, the R value found from the use of equation 1 will be labeled R_{FM}.

Similarly, a restricted model can be developed, using the pre-test as the predictor variable:

$$Y = b_0 + b_1X_1 + e_2, \qquad (2)$$

where

Y = the post-test score,

X_1 = the pre-test score,

b_0 = the Y-intercept (the b_0 value for equation 2 will, in general, be different from the b_0 value in equation 1),

b_1 = the regression coefficient for X (again, the b_1 value for equation 2 will, in general, be different from the b_1 value found in equation 1), and

e_2 = the error in prediction with the restricted model.

The restricted model will also yield an R value, and it will be labeled R_{RM}.

The F test for the analysis of covariance is given by:

$$F = \frac{(R^2_{FM} - R^2_{RM}) / 1}{(1 - R^2_{FM}) / N - 3} \quad . \qquad (3)$$

This F test is specific for this situation. A more general F test would be given by:

$$F = \frac{(R^2_{FM} - R^2_{RM})/(k - 1)}{(1 - R^2_{FM})/(N - C - k)} \quad , \qquad (4)$$

where

k is the number of groups,

N is the number of subjects, and

C is the number of covariates.

It is also possible to find adjusted means for the analysis of covariance.

DuBois (1957, 1970) has worked extensively with the residual gain analysis. Essentially, the residual gain analysis can be conceptualized as a part correlation between the group membership variable(s) and the residual in the post-test data when using the pre-test as the predictor. As a model, this can be accomplished easily in two stages with an ordinary multiple regression program. The first model is:

$$Y = b_0 + b_1X_1 + e_3, \qquad (5)$$

where

Y = the post-test score,

X_1 = the pre-test score,

b_0 = the Y-intercept (the value for b_0 in equation 5 will, in general, be different than previously defined b_0 values),

b_1 = the regression coefficient for X_1 (the value for b_1 in equation 5 will, in general, be different than previously defined b_1 values), and

e_3 = the error in prediction for this model.

The focus in the residual gains analysis is on the residual errors (e_3) for each subject. These residual errors become the criterion scores, and the group membership variable(s) are used to complete the residual gain analysis. The model is as follows:

$$Y' = b_0 + b_2X_2 + e_4, \qquad (6)$$

where

Y' = the residual errors found from the use of equation 5,

X_2 = 1 if the score is from a member of the vertical group; 0 otherwise,

b_0 = the Y-intercept (the b_0 value in equation 6 will, in general, be different than the previous b_0 values),

b_2 = the regression coefficient for X_2 (the b_2 value in equation 6 will, in general, be different from the b_2 value in equation 2), and

e_4 = the error in prediction for this model.

The use of the residual gain analysis has been based upon the following considerations: the residual gain scores will be uncorrelated with initial status, whereas it can be expected that the raw gain scores will show a negative correlation with initial states; whenever all subjects do not start at a common point (so that the methods of common points of mastery could not be used), the residual score nevertheless:

1. can be defined precisely and accurately,
2. the residual does not require the use of a ratio scale to measure initial and final states, and
3. higher ordered residual gains can be found.

Carver (1970) has compared the residual gain analysis to the method of common points of mastery, initially proposed by Ruch (1936). Conceptually, both of these measures were employed to overcome the difficulties involved with the raw gain scores. Employing both methodologies on empirical data, Carver was able to find only moderate correlations between the measures.

Subjects

The subjects for this study included 165 students in 8 rural North Dakota schools. All the students were enrolled in learning situations where the instructor was an intern (or in some cases, graduates) from the New School of the University of North Dakota, an experimental program funded by the United States Office of Education. The vertically grouped subjects were those students who were enrolled in a classroom setting that allowed a non-graded approach to instruction in several areas. Thus, the reading instruction took place in a homogeneous group rather than an age (or graded) group. The second group of students received their reading instruction in a graded group (i.e., Grade Four, Grade Five, etc). The grade levels involved were Grades Two through Grade Six.

Method

Two instruments were administered on a pre and post-test basis. Pre-tests were administered in October, 1970, and post-tests were administered in May,

1971. The vocabulary and comprehension sections of the California Reading Test (Tiegs and Clark, 1957) was used at all five grade levels. The Attitudes Toward Reading Inventory (Hunt, 1961) was used with grades four, five, and six. The Attitudes Toward Reading Inventory has two subtests, Attitudes Toward Reading, and Attitudes Toward Reading Class.

Results

Tables 1-6 show the analysis of the data. Each table includes means for the pre-test and post-test, adjusted means, raw gain, and residual gain for the two modes of instruction in reading. Included also are the F values, R, R^2, and SS_T (Sum of Squares Total) for each analysis. This method of presentation is used for economy of space and to allow for ease in comparing the different results. Actually, a summary table could be generated for all five different sets of data analyses. In the following tables the R value is the correlation between the dichotomous predictor (group membership) and the criterion scores, with the exception of the analysis of covariance (illustrated here under the name adjusted means), which is completed as it was described earlier. While there are different approaches to measuring the strength of relationships with dichotomous information, using Walberg's (1971) approach, the R^2 value is interpreted as being the amount of criterion variance accounted for by group membership. Also included in each table is some indication of significance. There is a slight discrepancy with the analysis of covariance (adjusted means) and the residual gains analysis. The degrees of freedom for the analysis of covariance and the residual gains analysis in this situation will actually be one less than the degrees of freedom listed under each table. In that no interpretations are changed in the present situation in regard to the differences in degrees of freedom, that slight difference in degrees of freedom is not indicated in the tables.

TABLE 1

SUMMARY DATA RELATING TO SECOND GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 2 (N = 35)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	2.359	3.194	3.229	.835	-.022
Graded Group	2.611	3.306	3.273	.694	.021
$F = t^2$	1.840	.695	.119	.581	.112
R	.230	.144	Full .392 Rest .388	.132	.059
R^2	.053	.021	Full .154 Rest .151	.014	.003
SS_T	10.535	5.267	4.475	10.022	4.474

Critical value for significance at .05 level with df = 1, 33 is 4.14.
Critical value for significance at .01 level with df = 1, 33 is 7.47.

TABLE 2

SUMMARY DATA RELATING TO THIRD GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 3 (N = 48)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	3.667	4.270	4.300	.603	-.049
Graded Group	3.789	4.483	4.433	.694	.082
$F = t^2$	.701	2.246	1.537	.603	1.512
R	.123	.216	Full .688 Rest .675	.114	.180
R^2	.015	.047	Full .473 Rest .456	.013	.032
SS_T	11.192	11.000	5.989	7.212	5.988

Critical value for significance at .05 level with df = 1, 46 is 4.05.
Critical value for significance at .01 level with df = 1, 46 is 7.21.

TABLE 3

SUMMARY DATA RELATING TO FOURTH GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 4 (N = 37)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	5.244	5.900	5.748	.656	-.134
Graded Group	4.986	5.876	5.992	.890	.102
$F = t^2$	1.161	.005	1.328	1.290	1.283
R	.179	.012	Full .771 Rest .761	.189	.191
R^2	.032	.001	Full .594 Rest .579	.036	.036
SS_T	18.829	33.243	14.015	14.016	14.014

Critical value for significance at .05 level with df = 1, 35 is 4.12.
Critical value for significance at .01 level with df = 1, 35 is 7.42.

TABLE 4

SUMMARY DATA RELATING TO FIFTH GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 5 (N = 27)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	5.880	6.580	6.536	.700	.120
Graded Group	5.800	6.318	6.344	.518	-.071
$F = t^2$	.064	.599	.954	.869	.952
R	.050	.153	Full .829 Rest .821	.183	.195
R^2	.003	.023	Full .687 Rest .674	.033	.038
SS_T	15.816	18.534	6.038	6.234	6.038

Critical value for significance at .05 level with df = 1, 25 is 4.24.
Critical value for significance at .01 level with df = 1, 25 is 7.77.

TABLE 5

SUMMARY DATA RELATING TO SIXTH GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 6 (N = 28)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	6.356	7.089	7.162	.774	.033
Graded Group	6.479	7.147	7.113	.669	-.016
$F = t^2$	.153	.027	.044	.079	.045
R	.077	.032	Full .771 Rest .770	.055	.042
R^2	.006	.001	Full .594 Rest .593	.003	.002
SS_T	15.847	20.297	8.248	8.507	8.248

Critical value for significance at .05 level with df = 1, 26 is 4.22.
Critical value for significance at .01 level with df = 1, 26 is 7.72.

TABLE 6

SUMMARY DATA RELATING TO SECOND GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 2 (N = 35)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	2.047	3.094	3.216	1.047	.057
Graded Group	2.550	3.194	2.079	.644	-.054
$F = t^2$	7.93**	.318	.594	4.838*	.478
R	.440	.098	Full .481 Rest .466	.358	.121
R^2	.194	.010	Full .231 Rest .217	.128	.015
SS_T	11.419	9.228	7.226	11.084	7.225

*Significant at .05 level. Critical value for significance at .05 level with df = 1, 33 is 4.14.
**Significant at .01 level. Critical value for significance at .01 level with df = 1, 33 is 7.47.

TABLE 7

SUMMARY DATA RELATING TO THIRD GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 3 (N = 48)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	3.678	4.307	4.321	.627	-.035
Graded Group	3.744	4.439	4.415	.694	.058
$F = t^2$	.203	.890	.696	.308	.693
R	.066	.138	Full .624 Rest .617	.082	.123
R^2	.004	.019	Full .389 Rest .381	.007	.015
SS_T	9.548	10.358	6.419	7.780	6.419

Critical value for significance at .05 level with df = 1, 46 is 4.05.
Critical value for significance at .01 level with df = 1, 46 is 7.21.

TABLE 8

SUMMARY DATA RELATING TO FOURTH GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 4 (N = 37)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	5.237	6.244	6.205	1.006	.189
Graded Group	5.176	5.843	5.872	.667	-.144
$F = t^2$	.046	1.160	2.719	2.847	2.714
R	.036	.179	Full .850 Rest .837	.274	.272
R^2	.001	.032	Full .723 Rest .701	.075	.074
SS_T	25.910	45.490	13.616	13.923	13.614

Critical value for significance at .05 level with df = 1, 35 is 4.12.
Critical value for significance at .01 level with df = 1, 35 is 7.42.

TABLE 9

SUMMARY DATA RELATING TO FIFTH GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 5 (N = 27)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	6.330	7.070	6.626	.740	.149
Graded Group	5.394	6.053	6.314	.659	-.087
$F = t^2$	8.124**	10.778**	2.031	.159	1.501
R	.495	.549	Full .864 Rest .851	.079	.243
R^2	.245	.301	Full .747 Rest .742	.006	.059
SS_T	22.485	21.617	5.954	6.567	5.952

Critical value for significance at .05 level with df = 1, 25 is 4.22.
**Significant at .01 level. Critical value for significance at .01 level with df = 1, 25 is 7.77.

TABLE 10

SUMMARY DATA RELATING TO SIXTH GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 6 (N = 28)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	6.367	7.378	7.512	1.011	.201
Graded Group	6.616	7.274	7.210	.658	-.095
$F = t^2$	.598	.101	2.094	2.746	2.043
R	.150	.062	.786 .765	.309	.275
R^2	.023	.004	.618 .585	.095	.076
SS_T	16.864	17.138	7.096	7.978	7.097

Critical value for significance at .05 level with df = 1, 26 is 4.22.
Critical value for significance at .01 level with df = 1, 26 is 7.72.

TABLE 11

SUMMARY DATA RELATING TO FOURTH GRADE
ATTITUDES TOWARD READING SCORES

	Attitudes Toward Reading Scores - Grade 4 (N = 37)				
	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	24.50	26.000	25.816	1.500	.490
Graded Group	24.048	24.810	24.950	.762	-.374
$F = t^2$	.114	.755	.756	.492	.754
R	.057	.145	Full .706 Rest .698	.118	.147
R^2	.003	.021	Full .498 Rest .487	.014	.022
SS_T	570.808	610.105	312.619	356.755	312.615

Critical value for significance at .05 level with df = 1, 35 is 4.12.
Critical value for significance at .01 level with df = 1, 35 is 7.42.

TABLE 12

SUMMARY DATA RELATING TO FIFTH GRADE
ATTITUDES TOWARD READING SCORES

	Attitudes Toward Reading Scores - Grade 5 (N = 27)				
	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	26.100	26.700	24.339	.600	.588
Graded Group	21.412	21.765	23.154	.354	.346
$F = t^2$	6.718*	6.232*	.586	.031	.460
R	.460	.447	Full .793 Rest .787	.035	.137
R^2	.212	.200	Full .629 Rest .619	.001	.019
SS_T	653.407	768.516	292.629	306.665	292.626

*Significant at .05 level. Critical value for significance at .05 level with df = 1, 25 is 4.24.
Critical value for significance at .01 level with df = 1, 25 is 7.77.

TABLE 13

SUMMARY DATA RELATING TO SIXTH GRADE
ATTITUDES TOWARD READING SCORES

Attitudes Toward Reading Scores - Grade 6 (N = 28)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	24.444	25.000	24.175	.555	.197
Graded Group	22.053	23.474	23.864	.526	-.093
$F = t^2$	1.896	1.092	.064	.455	.060
R	.261	.201	Full .625 Rest .624	.131	.049
R^2	.068	.041	Full .391 Rest .389	.017	.002
SS_T	514.105	342.962	215.518	334.712	215.518

Critical value for significance at .05 level with df = 1, 26 is 4.22.
Critical value for significance at .01 level with df = 1, 26 is 7.72.

TABLE 14

SUMMARY DATA RELATING TO FOURTH GRADE
ATTITUDES TOWARD READING CLASS SCORES

Attitudes Toward Reading Class Scores - Grade 4 (N = 37)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	38.375	40.188	39.450	1.813	.103
Graded Group	36.619	37.000	37.562	.381	-.78
$F = t^2$	1.312	2.862	1.474	.847	1.418
R	.190	.275	Full .641 Rest .621	.154	.200
R^2	.036	.076	Full .411 Rest .386	.024	.040
SS_T	774.702	20.670	750.256	787.997	750.254

Critical value for significance at .05 level with df = 1, 35 is 4.12.
Critical value for significance at .01 level with df = 1, 35 is 7.42.

TABLE 15

SUMMARY DATA RELATING TO FIFTH GRADE
ATTITUDES TOWARD READING CLASS SCORES

Attitudes Toward Reading Class Scores - Grade 5 (N = 27)	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	37.600	39.100	37.434	1.500	.819
Graded Group	33.882	35.000	35.980	1.118	-.482
$F = t^2$	2.947	4.289*	1.176	.076	1.046
R	.325	.383	Full .815 Rest .805	.055	.204
R^2	.106	.147	Full .664 Rest .648	.003	.042
SS_T	825.183	722.754	254.847	305.184	254.844

Critical value for significance at .05 level with df = 1, 25 is 4.24.
Critical value for significance at .01 level with df = 1, 25 is 7.77.

TABLE 16

SUMMARY DATA RELATING TO SIXTH GRADE
ATTITUDES TOWARD READING CLASS SCORES

Attitudes Toward Reading Class Scores - Grade 6 (N = 28)	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	35.667	36.000	36.726	.333	-.164
Graded Group	37.053	37.316	36.972	.263	.077
$F = t^2$	.556	.404	.025	.002	.025
R	.145	.124	Full .699 Rest .699	.009	.032
R^2	.021	.015	Full .489 Rest .489	.0001	.001
SS_T	560.677	690.678	353.465	381.713	353.462

Critical value for significance at .05 level with df = 1, 26 is 4.22.
Critical value for significance at .01 level with df = 1, 26 is 7.72.

Discussion

It should be abundantly clear from the 16 tables that the three approaches to psycho-educational change are different. While this set of data does not exhibit strong relationships between the dichotomous predictor and the various criteria, the use of the statistical significance approach would occasionally yield different interpretations. Perhaps the most objective comparison between the three measures would be the R^2 term (for the analysis of covariance, or adjusted means approach, $R^2_{FM} - R^2_{RM}$). Only one significant difference is found in the three measures. In Table 6, the raw gain is significant ($p < .05$), but, under exactly the conditions that would tend to make this occur, the vertical group was significantly smaller than the graded group on the pre-test, but this difference was almost erased on the post-test. In terms of the raw gains score, this produced a significant difference in favor of the vertical group.

In general, the interpretations of the tests would be in the same direction, although the reverse is true in Table 1. In Table 1, the raw gain scores favor the vertical group, while the analysis of covariance (adjusted means) and the residual gain scores favor the graded group.

LINEAR MODELS UNDERLYING THE ANALYSIS OF COVARIANCE, RESIDUAL GAIN SCORES AND RAW GAIN SCORES

Earl Jennings
The University of Texas at Austin

The problem of investigating "change" or "gain" that can be attributed to "treatments" has been discussed extensively over a number of years without a noticeable concensus emerging. Cronbach (1970) has even suggested that many questions that appear to involve "change" can be effectively resolved without reference to the concept of change.

This paper has the modest purpose of identifying the linear models that can be viewed as the basis for some of the commonly used procedures.

Assumption: The expected value at Time 2 for a member of Group j with a Time 1 of _q_ is

$$\alpha j + \beta j q$$

where αj and βj are unknown parameters and j ranges from 1 to the number of groups. Denote this value.

$$E(j, q)$$

Problem:

1. What is a good number to characterize or describe amount of change for the various combinations of group membership and levels of initial or Time 1 performance.
2. What is a good way to evaluate the hypothesis that the change is the same for all groups.

Argument:

1. The quantity

$$E(j, q) - q$$

is a reasonable number to characterize the change for a member of group $\underline{j}$ with a Time 1 score of $\underline{q}$.

2. Test the hypothesis

$$\left[E(i, q) - q \right] - \left[E(k, q) - q \right] = 0$$

where i and k range from 1 to the number of groups and $i \neq k$.

Consider the case of two groups. If the assumptions are true then a least squares solution to the following model should produce good estimates of the α's and β's and of the expected values.

Model 1. $$T^{(2)} = a_1 G^{(1)} + a_2 G^{(2)} + b_1 (G^{(1)} * T^{(1)}) + b_2 (G^{(2)} * T^{(1)}) + E^{(1)}$$

where

$T^{(2)}$ is a column vector of dimension $\underline{n}$ containing measures obtained at Time 2.

$G^{(1)}$ is a column vector of dimension $\underline{n}$ containing a one if the corresponding element in $T^{(2)}$ was observed on a member of Group 1; zero otherwise.

$G^{(2)}$ is a column vector of dimension $\underline{n}$ containing a one if the corresponding element in $T^{(2)}$ was observed on a member of Group 2; zero otherwise.

$T^{(1)}$ is a column vector of dimension $\underline{n}$ containing Time one measures arranged in the same order as $T^{(2)}$.

$E^{(1)}$ is the residual vector.

The a's are estimates of the α's and the b's are estimates of the β's.

Thus:

(1, q) is estimated by $a_1 + b_1 q$

and

(2, q) is estimated by $a_2 + b_2 q$

and the hypothesis

$$(a_1 + b_1 q - q) - (a_2 + b_2 q - q) = 0 \qquad (1)$$

implies that

$$a_1 = a_2 = a_3 \text{ a common value}$$

$$b_1 = b_2 = b_3 \text{ a common value}$$

Imposing these restrictions on Model 1 yields

Model 2. $$T^{(2)} = a_3(G^{(1)} + G^{(2)}) + b_3 T^{(1)} + E^{(2)}$$

and $$F_{12} = \frac{(ESS_2 - ESS_1) / (4 - 2)}{ESS_1 / (N - 4)}$$

An undesirable characteristic of this test in evaluating group differences in change is that a statistically significant F_{12} can be the result of $a_1 \neq a_2$ or $b_1 \neq b_2$ or both. Note that in (1), not only does the amount of change within a group depend on q but also the difference in change between the two groups depends on q. In other words, the amount that an individual can be expected to change depends on where he started (q) and also whether or not he can be expected to change by the same amount as an individual in the other group who started with the same value (q) depends on what that value is.

Suppose we assume only that

$$b_1 = b_2 = b_3 \text{ a common value.} \qquad (2)$$

Then (1) can be written

$$(a_1 + b_3 q - q) - (a_2 + b_3 q - q) = 0$$

or $$a_1 - a_2 = 0 \qquad (3)$$

Thus if (2) is accepted the amount of change still depends on q but the difference in change between the two groups does not.

The hypothesis implied by (2) can be evaluated by imposing the restriction on Model 1 yielding

Model 3 $$T^{(2)} = a_1 G^{(1)} + a_2 G^{(2)} + b_3 T^{(1)} + E^{(3)}$$

$$F_{13} = \frac{(ESS_3 - ESS_1) / (4 - 3)}{ESS_1 / (N - 4)}$$

A rejection of (2) would lead to the inference that an unqualified conclusion about group differences in change is not justified (i.e., there may be some value of q at which the expected change is the same for both groups and other values of q where the expected change is not the same.)

If (2) is acceptable then (3) can be evaluated by imposing the restriction on Model 3 yielding Model 2 and

$$F_{32} = \frac{(ESS_2 - ESS_3) / (3 - 2)}{ESS_3 / (N - 3)}$$

A rejection of (3) would lead to the inference that two individuals from different groups but with the same initial performance do not change by the same amount and the difference between a_1 and a_2 in Model 3 is an estimate of the expected difference in change.

Notice that Models 1, 2 and 3 are identical to those ordinarily associated with the analysis of covariance. F_{13} is used to evaluate the question of homogeneous slopes and F_{32} is used to evaluate the question of homogeneous intercepts (or homogeneous adjusted means).

Notice that in Model 3 the predicted changes

$$a_1 + b_3 q - q$$

and

$$a_2 + b_3 q - q$$

still depend on initial performance, q. These expressions can be written

$$a_1 + q(b_3 - 1)$$

and

$$a_2 + q(b_3 - 1)$$

Thus, if b_3 can be assumed equal to one then the expected change does not depend on the value of q. This assumption can be tested by imposing the restriction

$$b_3 = 1 \qquad (4)$$

on Model 3 yielding

Model 4

$$T^{(2)} - T^{(1)} = a_1G^{(1)} + a_2G^{(2)} + E^{(4)}$$

and

$$F_{34} = \frac{(ESS_4 - ESS_3) / (3 - 2)}{ESS_3 / (N - 3)}$$

In terms of Model 4 the expected change for Group 1 is estimated by a_1 and the expected change for Group 2 is a_2 and the hypothesis of equal change can be tested by imposing the restriction

$$a_1 = a_2 = a_3 \text{ a common value} \qquad (5)$$

on Model 4 yielding

Model 5

$$T^{(2)} - T^{(1)} = a_3(G^{(1)} + G^{(2)}) + E^{(5)}$$

and

$$F_{45} = \frac{(ESS_5 - ESS_4) / (2 - 1)}{ESS_4 / (n - 2)}$$

Notice that in Model 4 and 5 the criterion vector has as each element the difference between Time 2 and Time 1 scores. Thus a comparison of Model 4 and 5 is identical to performing a one way analysis of variance on difference scores.

Consider the following model:

Model 6

$$Y = a_{11}X^{(11)} + a_{12}X^{(12)} + a_{21}X^{(21)} + a_{22}X^{(22)}$$
$$+ c_1P^{(1)} + \cdots + c_nP^{(n)} + E^{(6)}$$

where Y is a column vector of dimension 2n containing both Time 1 and Time 2 scores on n individuals.

$X^{(IJ)}$ is a column vector of dimension 2n containing a one if the corresponding observation in Y was observed on an individual in group I at Time J. (I = 1, 2; J = 1, 2).

$P^{(K)}$ is a column vector of dimension 2n containing a one if the corresponding observation in Y was observed on individual K.

(K = 1, 2, ..., n). Note that each P vector contains two ones and 2n - n zeroes.

Note that Model 6 has only n + 2 linearly independent predictors.

Model 6 is the full model that is used in a type of analysis that goes by different names including Lindquist Type I Design, two groups two times repeated measurements analysis of variance, and a Groups by Trials analysis of variance. A particular comparison that can be referred to as the test for a Groups by Time interaction is evaluated by imposing the restriction

$$a_{11} - a_{12} = a_{21} - a_{22} \qquad (6)$$

on Model 6 which yields a restricted model (Model 7) with one less parameter.

$$F_{45} = F_{67} = \frac{(ESS_7 - ESS_6) \;/\; (n + 2 - n - 1)}{ESS_6 \;/\; (2n - n - 2)}$$

Thus the F test resulting from a one way analysis of variance of difference scores is identical to the Groups by Time test in a Groups by Trials analysis when there are only two times.

In view of the fact that many writers caution against using Models 4 and 5 (e.g., Edwards, 1960), it is somewhat surprising that similar cautions are infrequently urged with respect to Models 6 and 7 even though they lead to the same result.

The models used in analysis of residual gains can be written as:

Model 8 $$T^{(2)} = a_0 U + b_1 T^{(1)} + E^{(8)}$$

Model 9 $$E^{(8)} = a_1 G^{(1)} + a_2 G^{(2)} + E^{(9)}$$

The restriction to test for equality of mean residual gain is

$$a_1 = a_2 = a_3 \text{ a common value}$$

yielding

Model 10 $$E^{(8)} = a_3 U + E^{(10)}$$

and the test statistic is calculated by

$$F_{9,10} = \frac{(ESS_{10} - ESS_9) / 1}{ESS_9 / (n - 2)}$$

I have looked in vain for an argument that persuades me of the desirability of conducting such an analysis. It is true that $E^{(8)}$ is orthogonal to and uncorrelated with $T^{(1)}$. It may be desirable to have a "measure of change" that is uncorrelated with initial performance but it certainly cannot be argued that any set of numbers that can be shown to be uncorrelated with initial performance are a "measure of change."

It is easy to construct a set of data in which the slopes for the two groups are different that lead to conclusions using models 8, 9 and 10 that are flatly contradicted by the data. Moreover, even when there is good reason to believe that the slopes are equal (see F_{13}) the value of b_3 in Model 3 may be quite different from the value of b_1 in Model 8 if the groups are not matched on initial performance.

REFERENCES

Cronbach, L. J., and Furby, Lita. "How Should We Measure 'Change' --- Or Should We," Psychological Bulletin, Vol. 74, No. 1, July, 1970. (68-80)

Edwards, A. L. Experimental Design in Psychological Research, (Third Edition), Holt, Rinehart and Winston, New York, 1968. (Pp. 343-344)

AUTOMATIC INTERACTION DETECTOR AID-4

Janos B. Koplyay

I. INTRODUCTION

The primary value of AID-4 to the task scientist is its ability to identify the maximum amount of variance in the criterion which can be accounted for by the predictors available; it relieves the task scientist of the trial-and-error task of attempting to identify the various relevant combinations of linear and non-linear interaction terms presently required by the multiple linear regression technique. The splitting process of AID-4, being based upon maximizing the between sums-of-squares and minimizing the within-sums-of-squares, automatically takes all present interactions into account, indicating the maximum variance predictable in the criterion from the predictors. The interactions and patterns or trends are identifiable from the AID-4 output, however, no interactions sums of squares are provided the user.

The major advantage accruing to the task scientist is the evaluation of the maximum R^2 without the impossible task of generating all patterns of simple and complex interaction for entry into the linear regression system.

II. APPROACH

On the following pages there is an example for a simple two predictor, one criterion multiple linear regression model in the form of

<u>Model 1:</u> $$Y = a_0 U + a_1 X^{(1)} + a_2 X^{(2)} + e_1$$

where $X^{(1)}$ and $X^{(2)}$ are predictors (independent variables)

Y = dependent variable

which after the conventional multiple regression solution yielded an R^2 of .750847. Using the same data, AID-4 produced an R^2 of .900306. It will be demonstrated that the following polynomial regression model was used to obtain this R^2 = .900306:

Model 2: $Y = a_0U + a_1X^{(1)} + a_2X^{(2)} + a_3\left[X^{(1)}\right]^2 + a_4\left[X^{(1)}\right]^2 + a_5\left[X^{(1)}\right]^2 \cdot \left[X^{(2)}\right] + e_2$

It is obvious that without definite a priori knowledge of these additional and complex interaction terms, the researcher would have to try all kinds of combinations of all kinds of interaction terms to arrive at an R^2 = .900306. He would not know what the magnitude of the maximum R^2 might be, thus he could be satisfied with the original R^2 = .750847, reaching the conclusion that the predictor variables cover 75% of the variance of the criterion variable. He might abandon further investigation from the conclusion that the variables are not strong predictors. Little would be known about the maximum R^2 of .900306 identified by AID-4.

Since we know that there are only 6 groups in this example problem (3 educational levels and 2 rating status levels), we could solve this problem simply by solving a regression model of

$$Y = a_1X^{(1)} + a_2X^{(2)} + a_3X^{(3)} + a_4X^{(4)} + a_5X^{(5)} + a_6X^{(6)} + e$$

where X_1 = 1 if Group 1, zero otherwise

X_2 = 1 if Group 2, zero otherwise

.

.

.

X_6 = 6 if Group 6, zero otherwise

However, if the number of variables is large (say 80) and each variable has many levels or categories (say 10 categories each), in order to exhaust the system and arrive at the maximum R^2, one would have to generate 10^{80} categorical variables. Obviously most of these categories would have no cases in them (empty cells) but without a priori knowledge of the number and type of non-empty cells all 10^{80} groups would have to be considered. A distribution of cell frequencies could solve this problem to identify non-empty cells, nevertheless it would constitute much more labor and groundwork than AID-4 which requires no such identification of empty cells or generation of categorical variables.

Many additional and useful bits of information are provided by the output of AID-4; some of which are (1) at each split, the increased present total explained variance (R^2) is printed, together with a statistical test of significance for this R^2, (2) the splits are in a descending order of importance, that is, the first split identifies that variable which contributes the most to the explained variance; the second split identifies the second variable or a subset of the first split as the next important contributor to the explained variance; and so on. This hierarchy is very useful especially if after a few splits a reasonably high R^2 is obtained, thus giving the researcher an option of using only a few of the predictors in the actual prediction system if data collection on the rest of the variables is costly; (3) the branching pattern of splits reflects trends of characteristics specific to the groups split, that is, it can serve as an "eyeball" pattern analysis. Following the path of each branch of the split-tree, one can identify major characteristics of the final groups on which they differ the most in light of the criterion measure, (4) cross-validation and double-crossvalidation option. This feature splits the original sample into 2 random samples, treats each random sample separately deciding the best split pattern for each and the associated maximum R^2. Then it forces the split pattern of Sample 1 upon Sample 2 and vice-versa computing the R^2 for these forced splits. The difference between the maximum R^2 for each sample and the corresponding R obtained by forced splitting is a good indicator of the stability of the system, (5) selective or "partial" effects of the predictors are identifiable meanings that even if the so called main effect in a complex analysis of variance results in a non-significant F-ratio, AID-4 selectively indicates the level on the other variable at which this non-significant effect becomes significant. In the example which follows it will be seen that setting the level of significance at .01 there was no significant overall row effect in a two-way analysis of variance, however, it was significant at two of the three levels of the columns.

Let us take a classroom example taken from Hays (1963), page 403 with the following as given: (1) 60 observations, (2) one criterion assumed to be Air Force Qualifying Test Score (AFQT), (3) two predictor variables; education (3 levels) and pilot status (2 levels). Table 1 shows the AFQT scores at different levels of the two predictor variables.

TABLE 1

PILOT STATUS $X^{(2)}$	Education $X^{(1)}$: 1(E1)	2(E2)	3(E3)
1 (P)	52	28	15
	48	35	14
	43	34	13
	50	32	21
	43	34	14
	44	27	20
	46	31	21
	46	27	16
	43	29	20
	49	25	14
2 (NP)	38	43	23
	42	34	25
	42	33	18
	35	42	26
	33	41	18
	38	37	26
	39	37	20
	34	40	19
	33	36	20
	34	35	17

AID-4 SOLUTION

Note that no more than 5 splits can be expected as only 6 possible groups are initially defined.

The AID-4 output "Split Summary" (Figure 1) summarizes the resultant splits and "Split Diagram 1" (Figure 2) can then be drawn.

The "Split Summary" represents the most compact format of AID-4 output, however, the user has the option of requesting a detailed printed sequence of the whole splitting process if desired.

Note that with a final R value of .900306, approximately 90% of the variance is accounted for by the final groups numbered 6 through 11.

CONVENTIONAL REGRESSION SOLUTION

The first regression model was formulated without interaction terms.

<u>Model 1:</u> $y = a_0 u + a_1 x^{(1)} + a_2 x^{(2)} + e_1$

where the y's are AFQT scores, u is the unit vector, $x^{(1)}$ is the education variable, e_1 is the error vector and a_0, a_1, a_2 are the unknown parameters to be computed in the least-square sense. The resulting R^2 was .750847.

Next, the interaction term for $x^{(1)}$ and $x^{(2)}$ were generated as a cross product of the two:

<u>Model 2:</u> $y = b_0 u + b_1 x^{(1)} + b_2 x^{(2)} + b_3 z^{(1)} + e_2$

where $Z^{(1)} = \left[X^{(1)}\right] \cdot \left[X^{(2)}\right]$

The resulting R^2 increased to .818364.

At this point, assuming that the maximum R^2 of .900306 was unknown, one would have probably stopped pursuing the issue and conclude that considering

FIGURE 1

AID-4 SPLIT SUMMARY

SPLIT SUMMARY

Split group	On predictor	Predictor name	Into group	With codes	And group	With codes
1	1	EDUCATION LEVEL	2	0 1	3	2
2	1	EDUCATION LEVEL	4	0	5	1
4	2	PILOT STATUS	6	0	7	1
5	2	PILOT STATUS	8	0	9	1
3	2	PILOT STATUS	10	0	11	1

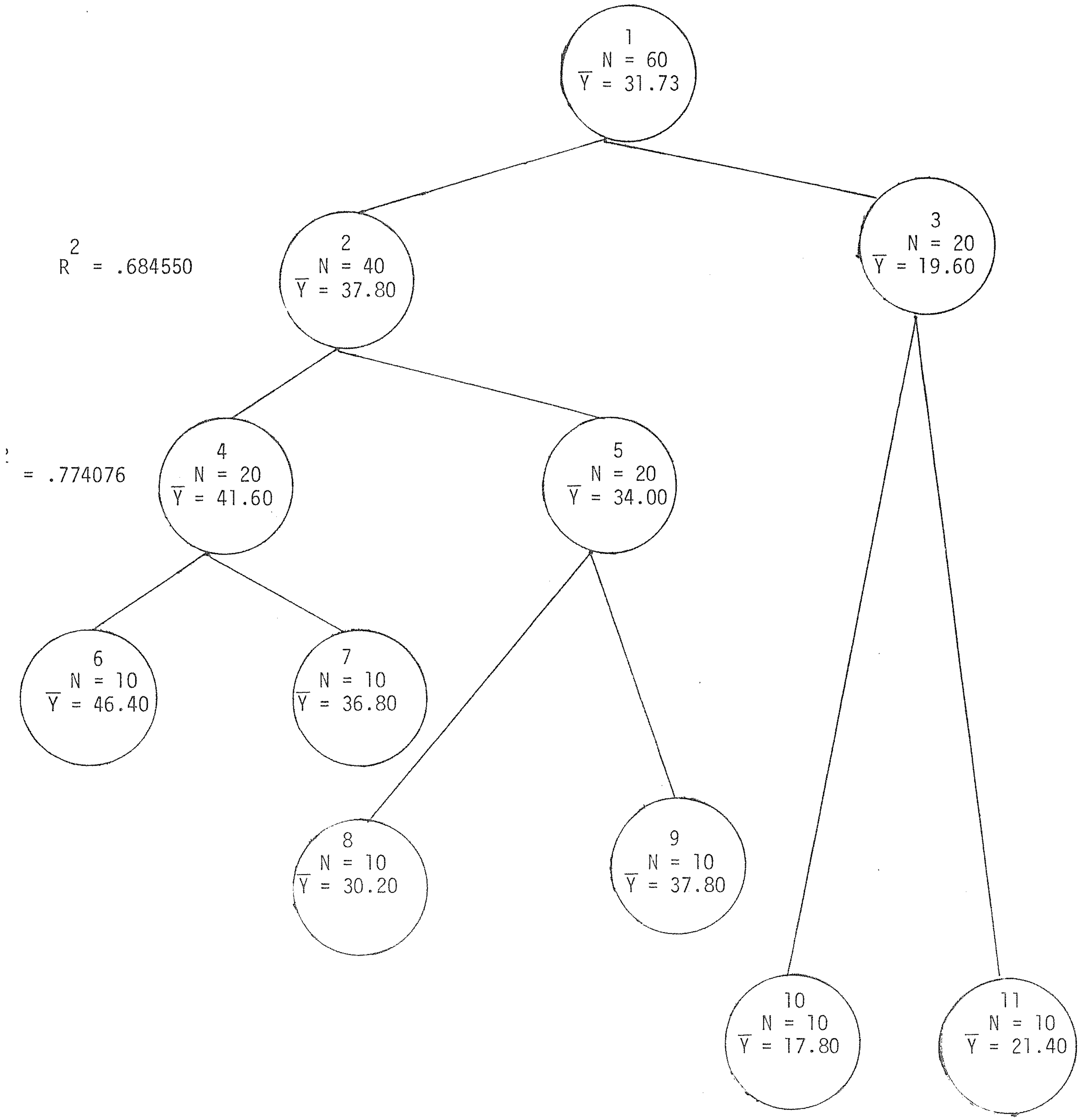
GROUP	N=	MEAN=	GROUP	N=	MEAN=	T=	R=	F-RSQ=
2	40	37.80	3	20	19.60	11.22	0.827	125.9
4	20	41.60	5	20	34.00	4.326	0.880	22.59
6	10	46.40	7	10	36.80	6.376	0.920	25.89
8	10	30.20	9	10	37.80	4.870	0.944	22.43
10	10	17.80	11	10	21.40	2.303	0.949	5.440

FINAL SUMMARY

MCF	TOTAL TSS	TOTAL BSS	TOTAL WSS	R-SQUARED	R	F-ANOVA	DF1	DF2
6	6451.7339	5808.5331	643.20081	0.90030574	0.94884	97.5312	5	54

FIGURE 2

SPLIT DIAGRAM 1



interactions between the two predictors, approximately 81% of the criterion variance is explainable with Model 2.

In order to make the contrast between conventional regression technique and AID-4 stronger, Model 3 was formulated. This model had no logical bases other than duplicate the $R^2 = .900306$ of AID-4.

Model 3:

$$Y = c_0U + c_1X^{(1)} + c_2X^{(2)} + c_3\left\{\left[X^{(1)}\right]\cdot\left[X^{(2)}\right]\right\} + c_4\left\{\left[X^{(1)}\right]^2\cdot\left[X^{(2)}\right]\right\} + c_5\left[X^{(1)}\right]^2 + e_3$$

or

$$Y = c_0U + c_1X^{(1)} + c_2X^{(2)} + c_3Z^{(1)} + c_4Z^{(2)} + c_5Z^{(3)} + e_3$$

where

$$Z^{(1)} = \left[X^{(1)}\right]\cdot\left[X^{(2)}\right]$$
$$Z^{(2)} = \left[X^{(1)}\right]^2\cdot\left[X^{(2)}\right]$$
$$Z^{(3)} = \left[X^{(1)}\right]^2$$

Solving for the unknown parameters c_1, c_2, c_3, c_4, and c_5 in the least square sense by the conventional iterative technique bringing in one variable at a time in descending order of importance of contribution to the explained variance, one would obtain an R^2 of .900288 in 494 iterations. The maximum $R^2 = .900306$ achieved by AID-4 will not be reached because some of the predictors are highly correlated and the iterative algorithm terminates or "hangs-up" by cycling back and forth between predictors and thus the stop criterion; i.e., the increase in the amount of explained variance becomes lower that that specified for the algorithm. This condition, however, can be remedied by a modified algorithm which takes three variables at the time into consideration. With this latter algorithm the resulting R^2 reaches the optimum of .900306 in ten iterations.

It is obvious that the liklihood is very small that a researcher would identify interaction terms as included in Model 3 above.

III. ANALYSIS AND RESULTS

Using Split Diagram 1, one can start asking meaningful questions in terms of linear regression models and arrive at the necessary weights and prediction equations.

Proceedings from bottom to the top of the diagram, the full model consists of all the final groups; i.e., groups 6, 7, 8, 9, 10, and 11. The original model then becomes

Model 4: $y = a_1 x^{(1)} + a_2 x^{(2)} + a_3 x^{(3)} + a_4 x^{(4)} + a_5 x^{(5)} + a_6 x^{(6)} + e_4$

where:

$x^{(1)}, x^{(2)}, \ldots, x^{(6)}$ are categorical vectors with the value of 1 if belongs to groups 6, 8, 10, 7, 9, 11 respectively; zero otherwise. y is the criterion vector (AFQT score); $a_1, a_2, \ldots, a_6$ are the unknown parameters.

Solving Model 4 for the unknown parameters in the least-square sense resulted in $R^2 = .900306$. Split Diagram 1 suggests that one would first test the hypothesis about pilot-status in educational level 3 (college graduates) by assuming that the respective parameters of a_3 (Group 10) and a_6 (Group 11) are equal in Model 4.

The resulting model is:

Model 5:

$$y = b_1 x^{(1)} + b_2 x^{(2)} + b_3 z^{(1)} + b_4 x^{(4)} + b_5 x^{(5)} + e_5$$

where

$$z^{(1)} = x^{(3)} + x^{(6)}$$

Solving this model for the unknown parameters, $R^2 = .890262$. One can test for significance between the R^2's of Model 4 and Model 5 by

$$F = \frac{(R^2_{\text{Model 4}} - R^2_{\text{Model 5}}) / (6-5)}{(1 - R^2_{\text{Model 5}}) / (60-6)} = 5.44$$

with a probability of p = .02 which is not significant at .01 level of confidence.

In a similar manner one can proceed and assume that the unknown coefficients of Groups 8 and 9 are equal (a_1 and a_4 respectively). This restricted model becomes

Model 6:

$$y = c_1 x^{(1)} + c_2 z^{(2)} + c_3 z^{(1)} + c_4 x^{(4)} + e_6$$

where

$$z^{2} = x + x^{(2)} + x^{(5)}$$

$$R^2 = .845499$$

$$F = \frac{(R^2_{\text{Model 5}} - R^2_{\text{Model 6}}) / (5-1)}{(1 - R^2_{\text{Model 5}}) / (60-5)}$$

$$p = 0.0000$$

significant beyond the .01 level of confidence

COMPARISON OF AID-4 WITH THE CONVENTIONAL REGRESSION TECHNIQUE

Table 2 summarizes the results of the two different approaches.

TABLE 2

Iterative Regression			AID-4		
Variable	Iteration	R^2	Split	Variable	R^2
$\left[X^{(1)}\right]^2$	1	.773221	1	E1, E2 vs E3	.684550
$\left[X^{(1)}\right]^2 \cdot \left[X^{(2)}\right]$	2	.795602	2	E1 vs E2	.774076
$X^{(2)}$	3	.805800	3	E1P vs E1NP	.845499
$\left[X^{(1)}\right]^2$	4	.822057	4	E2P vs E2NP	.890262
$X^{(2)}$	5	.823394	5	E3P vs E3NP	.900306
$X^{(1)}$	6	.823973			
⋮					
$\left[X^{(1)}\right]^2 \cdot \left[X^{(2)}\right]$	493	.900285			
$X^{(2)}$	494	.900287			

Symbols: $X^{(1)}$: EDUCATION predictor

$X^{(2)}$: PILOT-STATUS predictor

E1 : Educational level 1 (non-high school graduate)

E2 : Educational level 2 (high school but not college graduate)

E3 : College Graduate

P : Pilot

NP : Non-pilot

Example: E3NP : College graduate and non-pilot

IV. CONCLUSIONS

Looking at Table 2, it is obvious that in case of the regression model it would be very difficult to put any practical meanings to the entering variables such as

$$\left[X^{(1)}\right]^2$$

or

$$\left[X^{(1)}\right]^2 \cdot \left[X^{(2)}\right]$$

On the other hand AID-4 is easily interpretable. The first split simply implies that separating college graduates (E3) from non-college graduates (E1 and E2) explains 68% of the variance of the criterion variable. Further separation of high school graduates from non-high school graduates increases the explained variance to 77%. The effect of pilot status is most important is separating non-high school graduates (Split 3) and increases the explained variance to 84%.

In short, one can easily interpret the meanings and relative importance of the variables in the AID-4 splits while in the iterative regression scheme it is an almost impossible task. In addition, AID-4 needed only 5 splits to contrast to 494 iterations with a simple iterative procedure or 10 iterations with a modified version considering triplets of variables at a time (this latter procedure is not included in Table 2).

This procedure can be continued until there are only two groups remaining (Groups 1 and 2) in which case the F-test is simply the result of a one-way analysis of variance.

Split Diagram 1 suggests all kinds of interesting hypotheses to be tested. It identifies trends, gives cumulative explained variances and could conceivably be a very valuable tool for a researcher.

All the properties of the AID-4 approach discussed above and exemplified in this section can be generalized to problems of a more complex nature where attempts to include all possible combinations of interaction terms represent a practical impossibility. Appendix I contains an example of a real-life research project using AID-4 as a basic research tool.

REFERENCE

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