VOLUME 9 NUMBER 3

FEBRUARY 1979

MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group on Multiple Linear Regression

MLRV Abstracts appear in CIJE, the ERIC System, and microform copies are available from University Microfilms International

MULTIPLE LINEAR REGRESSION VIEWPOINTS

TABLE OF CONTENTS

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Volume 9 Number 3 Feb. 1979

"CONTRASTS WITH UNEQUAL N BY MULTIPLE LINEAR REGRESSION"

Dr. John D. Williams

It is shown that some of the more simplified methods for contrasts with equal N result in erroneous calculations when applied to data sets with unequal N. Instead, the methodology given earlier by Bottenberg and Ward (1963) is effective for finding values for contrasts (where $t = \sqrt{F}$). Also, the unweighted means solution for maximized Scheffé contrasts is shown to fail in finding the maximized contrast with unequal N.

In Chapter IV of a monograph on multiple comparisons, Williams (1976) shows a simplified way to accomplish complex multiple comparisons (contrasts) using multiple linear regression. A major limitation of the simplified method was not given; the method works for equal N cases only. Similarly the methodology of Chapter V in the same monograph for finding maximized Scheffé contrasts is inappropriate for unequal N; the use of the methodology for maximized Scheffé contrasts continues to yield a value equal to R^2_{Full} for the model corresponding to the coefficients found; however, the coefficients do not yield a contrast for unequal N, since Σc_j in general is not equal to zero.

An Example

Table 1 is taken from Williams (1976, p. 7). In the original data set, four groups of five subjects were given; they are repeated in Table 1; six of these subjects are noted by asterisks and are excluded from any further analysis. Both the 20 subject data set and the 14 subject data set have the following means: $\bar{X}_1 = 6$, $\bar{X}_2 = 7$, $\bar{X}_3 = 12$ and $\bar{X}_4 = 13$. The overall mean for the 20 subject data set is 9.5 and the mean for the 14 subject data set is 10.42857.

TABLE 1

DATA FOR CONTRASTS

Group 1	Group 2	Group 3	Group 4
9*	8	13	15
8	7	10	12
6*	8*	12*	10
3*	6*	11	17
4	6	14	11

*Score excluded for unequal N

First, an analysis of variance (achieved by using multiple linear regression) can be performed. The following variables can be defined:

Y = the criterion score;

X_1 = 1 if the score is from a member of group 1, 0 otherwise;

X_2 = 1 if the score is from a member of group 2, 0 otherwise;

X_3 = 1 if the score is from a member of group 3, 0 otherwise; and

X_4 = 1 if the score is from a member of group 4, 0 otherwise.

Also the b_i are the regression coefficients.

A full model can be defined as

$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_1$. (1)

Most regression programs include a unit vector for finding the constant (or Y- intercept) term. In that case, the full model can be defined as

$Y = b_o + b_1X_1 + b_2X_2 + b_3X_3 + e_1$. (2)

The summary table for the analysis of variance is given in Table 2. It should also be mentioned that the degrees of freedom associated with equation 2 (df_2) will be the same as would be true in the usual analysis of variance. Since there are four groups, $df_2 = 3$. It can also be noticed that df_2 are equal to the number of non-redundant predictors.

TABLE 2

ANALYSIS OF VARIANCE FOR THE DATA IN TABLE 1

SOURCE OF VARIATION	df	MS	MS	F
Among	3	117.428	39.143	7.249
Within	10	54.000	5.400	
Total	13	171.428		

Also, $R^2 = .68500$.

When equation 2 is used, the regression coefficients take on a most useful form:

$$Y = \overline{Y}_4 + (\overline{Y}_1 - \overline{Y}_4)\ X_1 + (\overline{Y}_2 - \overline{Y}_4)\ X_2 + (\overline{Y}_3 - \overline{Y}_4)\ X_3 + e_1. \quad (3)$$

The tests of significance of the partial regression weights for b_1, b_2 and b_3 are directly interpretable as tests of significance for the multiple comparisons, provided that appropriate tables are used for judging significance. The use of the test of the partial regression weights will remain computationally valid even with unequal N, although the use of the tabled values may in some cases cause the test to become an approximate test rather than an exact test, such as with Tukey's (1959), Dunnett's (1955) or Duncan's (1955) tests.

CONTRASTS

If more complex comparisons (called contrasts) are used, the situation is not quite so simple. Suppose the following contrast is of interest:

$$\Psi = \tfrac{1}{2}\overline{Y}_1 + \tfrac{1}{2}\overline{Y}_2 - \tfrac{1}{2}\overline{Y}_3 - \tfrac{1}{2}\overline{Y}_4.$$

Using the methodology of Bottenberg and Ward (1963, see also Ward and Jennings, 1973), the following restriction can be placed on equation 1:

$\frac{1}{2}b_1 + \frac{1}{2}b_2 - \frac{1}{2}b_3 - \frac{1}{2}b_4 = 0.$ (4)

Then $b_1 = b_3 + b_4 - b_2$; substituting back into equation 1,

$Y = (b_3 + b_4 - b_2) X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_2,$

$Y = b_2 (X_2 - X_1) + b_3 (X_3 + X_1) + b_4 (X_4 + X_1) + e_2.$ (5)

Then three additional variables can be defined:

X_5 = 1 if a member of group 2, -1 if a member of group 1, 0 otherwise;

X_6 = 1 if a member of either group 3 or group 1, 0 otherwise; and

X_7 = 1 if a member of either group 4 or group 1, 0 otherwise.

An equation incorporating X_5, X_6 and X_7 is

$Y = b_5X_5 + b_6X_6 + b_7X_7 + e_2.$ (6)

Equation 6 is appropriate if the regression program does not use a unit vector. Otherwise, any two of X_5, X_6 or X_7 can be used:

$Y = b_0 + b_5X_5 + b_6X_6 + e_2.$ (7)

The use of equation 7 results in $R_7^2 = .03045$.

The significance of the restriction given by equation 4 can be tested by

$$t = \sqrt{F} = \sqrt{\frac{(R_2^2 - R_7^2)/(df_2 - df_7)}{(1 - R_2^2)/(N - df_2 - 1)}}$$

$t = \sqrt{\frac{(.68500 - .03045)/1}{(1 - .68500)/10}} = 4.558$, which can be compared to an appropriate table, to maintain the apparent error rate.

Had the methodology for finding contrasts (Williams, 1976) been used directly, a new variable can be defined as $X_8 = \frac{1}{2}$ if the score is from a member of either group 1 or group 2; $-\frac{1}{2}$ if the score is from a member of either group 1 or group 2; $-\frac{1}{2}$ if the score is from a member of either group 3 or group 4.

Then $Y = b_0 + b_8X_8 + e_3.$ (8)

Using equation 8 yields $R_8^2 = .66504$;

$t = \sqrt{\dfrac{.66504/1}{(1 - .68500/10}}$ = 4.595, which is close to, but not precisely equal to the correct value found through the restriction given in equation 4. Thus, if contrasts beyond simple comparisons of means are needed when unequal sample sizes are present, the use of full and restricted models is recommended rather than shortcut procedures which are appropriate for equal sample sizes, but inappropriate for unequal sample sizes.

MAXIMIZED SCHEFFÉ CONTRASTS

As indicated earlier, the methodology in Williams (1976, Chapter V) fails to yield contrasts with unequal N. A method for finding a maximized Scheffé (1959) contrast can be found in Winer (1971, p. 177). The maximized contrast coefficients are $c_j = \dfrac{n_j\ (\overline{Y}_j - \overline{Y}_T)}{\sqrt{SS_a}}$, where $\overline{Y}_T = 10.42857$.

For the data in Table 1, the maximized contrast is

$\Psi_2 = -.81735\overline{Y}_1 - .94918\overline{Y}_2 + .58005\overline{Y}_3 + 1.18647\overline{Y}_4$.

To test Ψ_2, regarding its being a maximized contrast, it is necessary for the restricted model to have $R^2 = 0$. The restriction on equation 1 for Ψ_2 is $-.81735b_1 - .94918b_2 + .58005b_3 + 1.18647b_4 = 0$.

or

$b_1 = -1.16129b_2 + .70967b_3 + 1.45161b_4$. (9)

Substituting into equation 1

$Y = (-1.16129b_2 + .70967b_3 + 1.45161b_4)\ X_1 + b_2X_2 + b_3X_3 + e_4$;

$Y = b_2\ (X_2 - 1.16129X_1) + b_3\ (X_3 + .70967X_1) + b_4\ (X_4 + 1.45161X_1) + e_4$. (10)

Then three additional variables can be defined:

X_9 = 1 if the score is from a member of group 2,

-1.16129 if the score is from a member of group 1, 0 otherwise;

X_{10} = 1 if the score is from a member of group 3, .70967 if the score is from a member of group 1, 0 otherwise; and X_{11} = 1 if the score is from a member of group 4, 1.45161 if the score is from a member of group 1, 0 otherwise. Any two of X_9, X_{10} and X_{11} can be used as predictors (when using a program that includes the unit vector); so

$Y = b_0 + b_9X_9 + b_{10}X_{10} + e_4.$ (11)

As expected, $R^2_{11} = 0$. Therefore,

$$t = \sqrt{\frac{(R^2_2 - R^2_{11})/df_2 - df_{11})}{(1 - R^2_2)/(N - df_2 - 1)}} = \sqrt{\frac{(.68500-0)/1.}{(.31500/10}}$$

Then t = 4.663. Whenever the overall F test yields significance, the maximized Scheffé contrast is guaranteed to be significant; whether or not the maximized Scheffé test would ever have any practical value is an entirely different concern.

Any multiple of Ψ_2 will also prove to be a maximized Scheffe contrast. Hollingsworth (1978) claims that the use of an unweighted means solution also yields a maximized contrast; her contrast coefficients are given by

$$c_j = \sqrt{\frac{\tilde{n}\ (\bar{Y}_j - \bar{T})}{SS_a}}$$

Where $\tilde{n}$ is the harmonic mean of the sample sizes: $\tilde{n} = \dfrac{4}{1/2 + 1/3 + 1/4 + 1/5}$ = 3.11688; and $\bar{T}$ is the unweighted mean of the four group means: $\bar{T} = 9.5$.

The resulting contrast is

$\Psi_3 = -.57022\overline{X}_1 - .40730\overline{X}_2 + .40730\overline{X}_3 + .57022\overline{X}_4.$

A restriction on the full model can be made to correspond to Ψ_3:

$-.57022b_1 - .40730b_2 + .40730b_3 + .57022b_4 = 0$

$b_1 = -.71429b_2 - .71429b_3 + b_4.$ (12)

Following in the same manner as before, the restricted model yields $R^2 = .03154$. In that R^2 is not zero, Ψ_3 is not a maximized contrast. The resulting t value, $t = 4.555$, though it is close to the maximized t, clearly is less than the maximum value.

REFERENCES

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

Duncan, D. B. Multiple range and multiple F tests. Biometrics, 1955, 11, 1-42.

Dunnett, C. W. A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 1955, 50; 1096-1121.

Hollingsworth, H. The coefficients of a maximum contrast as interpretable statistics. Paper presented at the American Educational Research Association Annual Meeting, Toronto, March, 1978.

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

Tukey, J. W. The problem of multiple comparisons. Dittoed, Princeton Univ., 1953.

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

Williams, J. D. Multiple comparisons by multiple linear regression. Multiple Linear Regression Viewpoints, Monograph Series #2, 1976, 7.

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

John D. Williams

The University of North Dakota

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Volume 9 Number 3 Feb. 1979

"IS THE PhD RESEARCH TOOL USED IN THE DISSERTATION?"

Dr. Ernest Lewis and Dr. Dennis Leitner

Abstract: Students taking Multiple Regression as the PhD research tool from 1970 through 1975 tended to use Multiple Regression at the data analytic tool in the dissertation.

Most teachers are concerned with what becomes of their students after the students leave the classroom and with whether the students use what they learned in the classroom. The Guidance and Educational Psychology Department at SIU-C has been teaching educational statistics and research courses, which in many cases serve as the PhD student's research tool. Since SIU-C was one of the first schools to implement courses in multiple linear regression as a data analytic tool, we had a unique opportunity to determine the extent to which our teaching affected students and the extent to which the research tool requirement is a valuable aspect of the PhD program.

The majority of faculty advisors at this campus has been unfamiliar with multiple linear regression. As a result, application of the technique in a dissertation could be attributed to the student's training in the regression course. The lack of familiarity with the technique on the part of the faculty and the limited use in the research literature until recently have served as limitations of the impact of the regression course. Many advisors are reluctant to permit use of unfamiliar and unconventional analysis procedures, even though those procedures may be just as appropriate, and perhaps as easily understood by students and researchers.

In order to investigate the impact of the regression course, we examined the grade lists of all instructors teaching a course in multiple linear regression during the period 1970 to 1975, inclusive. During Spring Semester, 1976, we examined the library's dissertation volumes at SIU-C to determine how many of the doctoral students who had taken the regression course had a dissertation on file. (All students must deposit one copy of the dissertation in the library before they are cleared for graduation.) Each dissertation was examined for the type of statistical analysis used and the type of statistical references cited. The results of this analysis are presented below.

During the period 1970-1975, 763 students completed the multiple linear regression course, roughly evenly divided between masters degree students (51.6%) and PhD students. Fifty dissertations were on file in the library. (The authors realize that many of the students who took the course in late 1974 and 1975 have not yet graduated, since there is about a two year interval between taking the course and filing the dissertation.) The table below presents a breakdown of the major departments of the graduating students:

Department	Frequency	Percentage
Education	36	72%
Journalism	4	8
Psychology	3	6
Speech	2	4
Geology	1	2
Government	1	2
Speech Pathology	1	2
Zoology	1	2
Unclassified	1	2
	N = 50	100%

Of the 50 dissertations, 41 reported descriptive statistics, 38 reported inferential statistics, and 3 were primarily library research. Of the 38 dissertations employing inferential statistics, 20 (52.6%) actually used multiple linear regression analysis. Of these 20, 18 used the technique for prediction purposes, while 2 used it to perform a traditional analysis of variance.

Further indication of the impact of the course in providing students a useful tool would be indicated by the types of statistical references in the dissertation. The 20 dissertations which employed regression analysis cited at least one regression text as a reference (e.g., Kelly, Beggs, McNeil, Eichelberger and Lyon, 1969). A total of 22 dissertations used a statistical reference which would be considered traditional (e.g., Lindquist, 1953).

The results indicate that the statistical courses at SIU-C are having a significant impact upon the techniques used for data analysis by doctoral students at SIU-C. The research tool was used directly in more than half of the dissertations found which employed any inferential statistics. One particularly disturbing finding, however, was that by Spring, 1976, only 50 of 369 doctoral students (13.6%) who took multiple linear regression analysis between 1970 and 1975 had graduated.

References

Kelly, F. J., Beggs, D. L., McNeil, K. A., Eichelberger, R., and Lyon, J. Research Design in the Behavioral Sciences: Multiple Regression Approach. Carbondale: Southern Illinois University Press, 1969.

Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton-Mifflin, 1953.

Ernest Lewis
Dennis Leitner

Southern Illinois University

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Volume 9 Number 3 Feb. 1979

"INCREASING POWER AND INTERPRETABILITY IN CERTAIN REPEATED MEASURES DESIGNS"

Dr. John T. Mouw and Dr. View Nu

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

Dayton (1970) and Lindquist (1953) are at least two authors who advocate the use of repeated measures designs to enhance the work of the researcher. The greater frequency of use of such designs in research tends to indicate their expanding popularity. The usual increase in power of the statistical tests, as well as the advantage of decreasing the number of subjects when more than one observation is taken on a subject, have probably been advantages too great to be passed by in favor of a logically less complex design such as a completely randomized design.

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

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

Jenning, 1973; Kerlinger and Pedhazur, 1973), we should be able to take advantage of a more appropriate method of analysis.

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

In that same article, Feldt (1958) argues that the "blocking" procedure (on the concomitant variables) yields more power than the use of the concomitant variables as a covariate when the correlation is less than .60. His argument, however, rests partially on two assumptions: (a) The researcher employs a large number of levels depending on the N of the research, and (b) The random variability of the group means on the concomitant variable is a source of error resulting in loss of power. The first assumption appears to be impractical, since researchers seldom use more than three levels. The second argument does not hold in a repeated measures design, as presented below, where the mean of the concomitant variable is equal across treatment groups.

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

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

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

Where:

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

μ : Grand mean of all potential observations.

β_j : Effect of factor B under level j.

α_i: Effect of factor A under level i.

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

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

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

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

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

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

(Model 2)

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

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

Where:

Y: Criterion scores

a_0: Intercept, estimate of μ

a_1-a_{17}: Partial regression weights

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

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

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

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

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

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

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

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

1. By dichotomizing or splitting the continuous B factors, we lose the ability to examine the natural rélationship (correlation)

between X and Y. This problem becomes especially severe when, as in most designs such as this, the major research interest is in the interaction of A x B, i.e., the difference in performance across A depending on the level of B. Model (3) does not allow the study of such an interaction between the treatment and the natural continuum of factor B.

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

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

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

Where:

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

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

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

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

γ_{lin}: Linear effect of factor A.

γ_{quad}: Quadratic effect of factor A.

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

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

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

Table 1
Scores for the Sample Problem

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

If model (2) is used for this illustration, subjects with the first three scores on the Aptitude test could be classified as the "Low Aptitude Group", and the three remaining scores as the "High Aptitude Group". In that model, β (Aptitude) would be orthogonally coded with B_1 (the low aptitude group) coded as +1 and B_2 (the high aptitude group) coded as -1. Model (4) would utilize the aptitude information by running the variable $(X-\bar{X})$ into the model. In both models (2) and (4), Factor A (day) would be orthogonally coded, with linear and quadratic polynomials, and the criterion scores Y_{ijk} would be the scores on the Reading Achievement test.

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

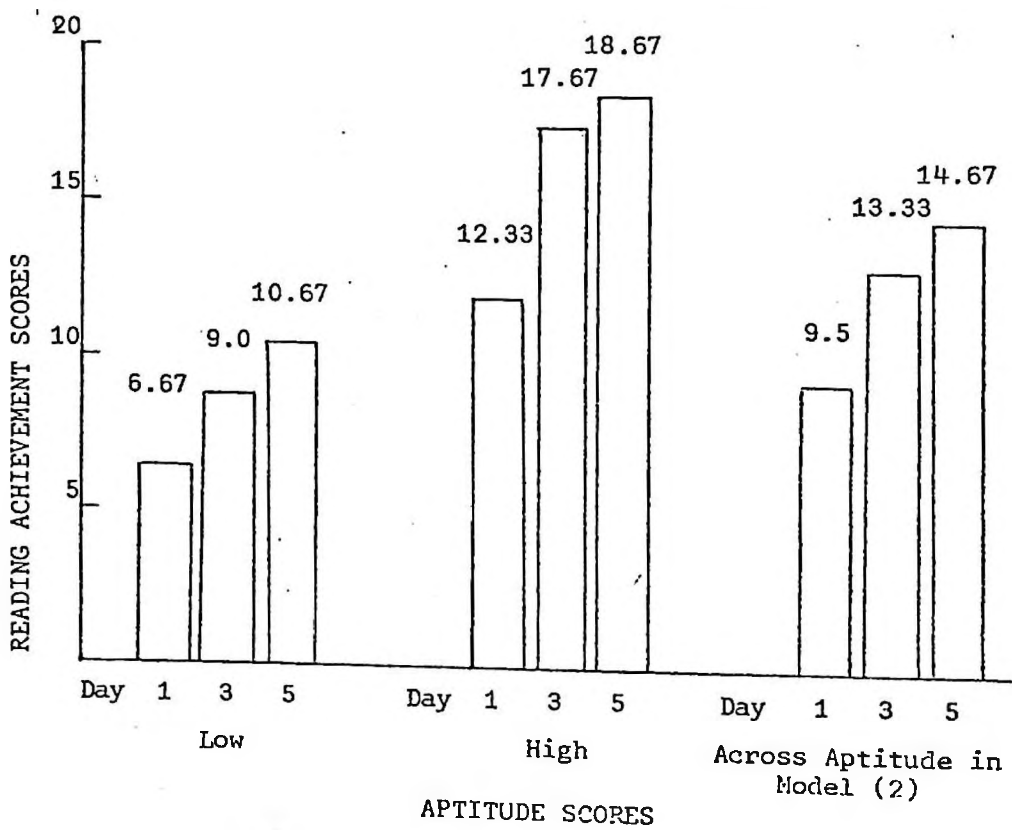

Figure 1. Mean of Reading Achievement Scores in Model (2) for both High and Low Aptitude Groups on Day 1, 3 and 5

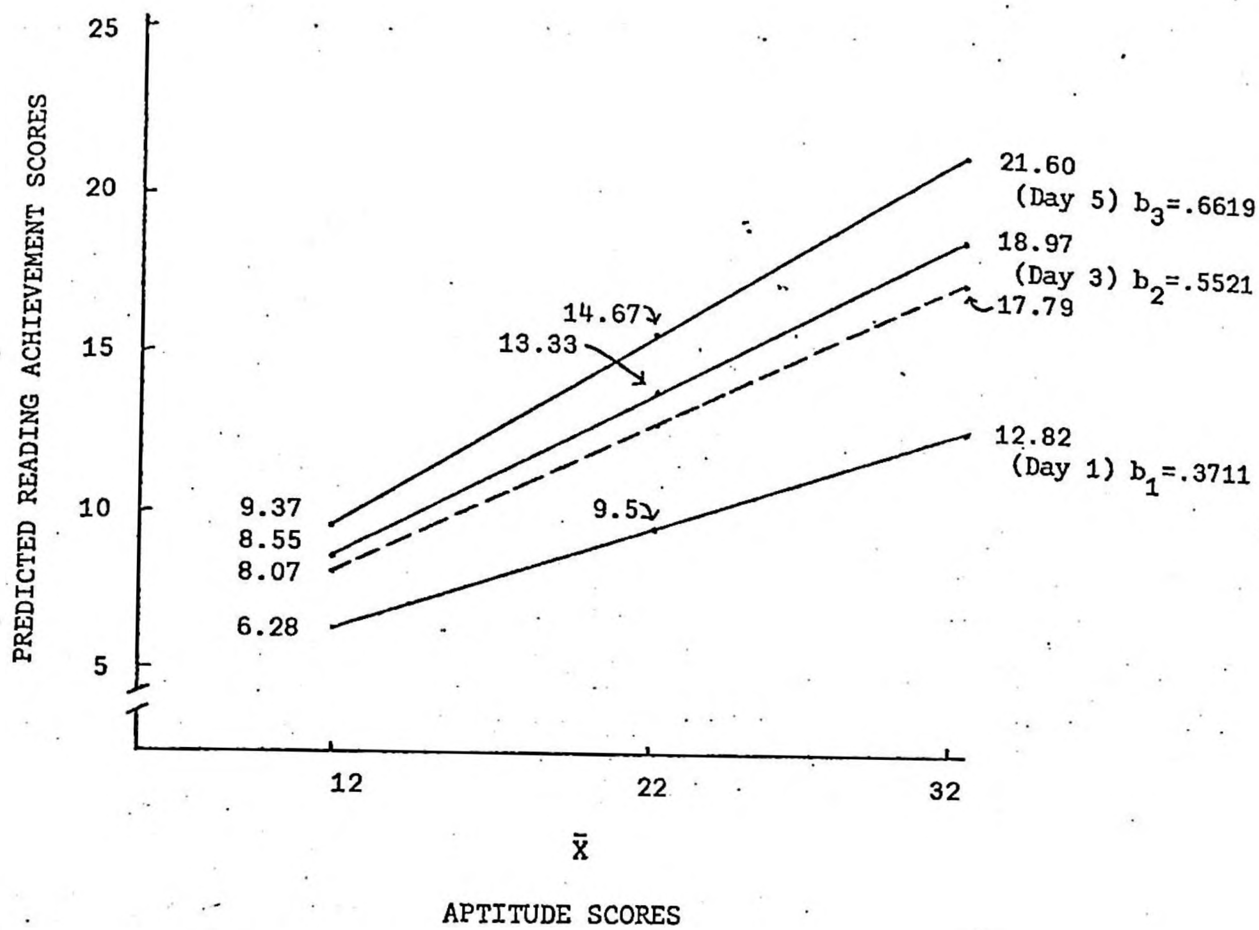

Figure 2. Regression Lines of the Predicted Scores on Reading Achievement and Aptitude Scores from Model (4) on Day 1, 3 and 5

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

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

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

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

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

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

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

Table 2

Anova of Data from Model (2)

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

Table 3

Anova of Data from Model (4)

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

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

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

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

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

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

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

TABLE 4

Variance Components and Degrees of Freedom for the Two Analyses

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

I = number of levels of A

J = number of levels of B

K = number of subjects within each B level

I = number of levels of A

K = number of subjects

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

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

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

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

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

The number of degrees of freedom of model (2) does not change in model (4). The estimate of a single parameter is found in the numerator of the F-test for both models. In model (2) we find an estimate of ($\mu_{\cdot 1 \cdot} - \mu_{\cdot 2 \cdot}$), whereas in model (4) the estimate is of $\beta\ldots$ The degrees of freedom for the error between source of variance is obtained from the number of random observations of subject means ($\bar{Y}_{\cdot jk}$) minus the number of parameter estimates utilized to obtain such an error estimate. For model (2) the parameter estimates include the two B means $\bar{Y}_{\cdot j \cdot}$, while for model (4) they are the estimates for α and $\beta_{..}$ from the overall regression effect of X on Y. In either case the example problem contains 1 and 4 degrees of freedom in the F-test for the B effect.

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

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

Bibliography

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

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

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

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

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

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

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

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

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

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

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

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

Footnotes

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

John T. Mouw
View Nu

Southern Illinois University

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Volume 9 Number 3 Feb. 1979

"EFFECTS OF DIFFERENT TYPES OF SCORES ON MAGNITUDES OF COMPUTED R^2"

Gary D. House

This study compared the magnitudes of R^2 values computed through multiple linear regression models using grade equivalent scores versus raw scores, standard scores, and percentiles as both criterion and predictor variables. It was found that grade equivalent and standard score modes produced similar and higher R^2 values than did raw scores or percentiles.

Introduction

After having completed a paper (House, 1978) which utilized multiple linear regression analyses to study the impact of a remedial arithmetic program on eligible pupils across a three year time span, House was thrown into a temporary state of dissonance when his advisor showed him an abstract of another paper (Hogan and Beck, 1974) which concluded in part that grade equivalent scores were likely to lead to less precise results than were raw scores. As might be guessed, House had used grade equivalent scores (Iowa Tests of Basic Skills Arithmetic Total Subscales, Levels 10, 11 and 12, 1971, Houghton Mifflin Company) as both criterion and covariate measures in his paper.

House did not enter lightly into the decisions a) to use grade equivalent scores, or b) to use multiple linear regression analyses. He looked over the literature rather throughly and became aware of large scale controversies over whether norm referenced tests were better than criterion referenced tests, whether grade equivalent scores were of any value at all, whether gain scores, adjusted gain scores, or simply regressed-upon criterion scores should be analyzed, or finally, whether any examination of achievement growth was worth the effort. Side issues such as regression-to-mean effects, true score *estimation, testing bias and other more obscure matters added fog to an already obtuse and confusing measurement scene (not to mention Rasch).*

Four papers were most influential as House made his decisions: 1) Linn and Slinde (1977) pointed out that difference (gain) scores are commonly unreliable, frequently correlate negatively with pretest scores, and are often derrived from two measures which lack common measurement traits and scales. They also discussed at some length alternatives to the use of difference scores and concluded that regression techniques treating a post test as the criterion, and the pretest as one of several predictors, minimiz "many of the difficulties that are introduced by gain scores".

2) Cronback and Furby (1970) discussed in some detail technical issues relative to the measurement of change, vis a vis sampling effects, reliabili procedures for correcting for unreliability. They concluded that although it would be desirable to obtain true scores for individuals, there are techniques (i.e. multiple regression [covariance] analysis) which admit the futility of obtaining those theoretical values and which reduce the bias resulting from measurement and sampling error.

3) Fennessey (1973) in his monograph on measuring achievement growth in practical educational settings, sets two necessary criteria for any scale which may be utilized: a) It should measure the variable in question; b) It should have equal interval properties. Although he recognized the common misuses and misinterpretations of norm referenced grade equivalent scores, Fennessey recommended their use in tracing individual student growth across time and in comparing programs designed for similar pupils.

4) Finally, Echnernacht and Plas (1977) provided a straightforward discussion of different types of scores and their potential usefulness. The described grade equivalent scores as somewhat different from T scores and other standard scores in their equal interval properties, but nevertheless *indicated that grade equivalent scores take into account item difficulties across grade levels since they are based upon expanded scale scores of published survey tests.*

These papers led House to his decisions, and he felt reasonably comfortable in proceding. The Hogan and Beck paper made him worry that the grade equivalent scores decision might have been a bad one, even though his multiple linear regression decision was fairly sound. After a careful reading of the Hogan and Beck paper, however, House had serious questions about their conclusions.

Problem Statement

Theirs was a Monte Carlo type study in which they used Total Reading and Total Math Subtests from the Metropolitan Achievement Tests-Elementary Battery. Specifically, they generated multiple random sets of Reading and Math raw score distributions each for 3 theoretical groups of pupils numbering 25, 100 and 1000 in order to produce predetermined Pearson Product-Moment Correlation Coefficients of .50, .75 and .90, and predetermined t values of 1.00, 2.00 and 5.00. They then converted those raw score distributions to corresponding grade-equivalent scores, standard scores, percentiles and stanines. Subsequently, they used these converted scores to recompute r's and t's finding a) comparable results for obtained levels of t and r for percentiles and raw scores; b) lower obtained r's for stanines than for raw scores, but comparable t results; c) comparable results for t's obtained using standard versus raw scores, but lower r's; d) lower t's and r's from the use of grade equivalent scores than from raw scores. Hogan and Beck recommended that grade equivalent scores should be used with caution, in hypothesis testing, but suggested that further investigation was needed.

From what House knew about scaling, it seemed to him odd that raw scores should be used to compute r's and t's from a survey test whose scaling properties and norming populations are known. Obviously, those raw scores could not be an interval scale due to the range of item difficulties of the test. Percentiles by definition are not scaled in equal intervals. Althou stanines are gauged in equal intervals, they would lack the precision of most standard scores or grade equivalents because of their limited range. Grade equivalent scores are scaled from expanded standard scores of one kin or another, and although they might not be interval level values, they shou be close.

It was therefore House's hypothesis that if pre and post raw scores, percentiles, standard scores and grade equivalent scores obtained from a non referenced measure on a population of pupils were used in statistical testin the standard scores and grade equivalent scores would yield similar results and the raw scores and percentiles would yield similar results. It was also his hunch that the grade equivalent scores and standard scores would yield higher values of test statistics than would either raw scores or percentiles, since they would be sensitive to item difficulities and therefore more precis

Methods

House therefore recalculated R^2 values for all full and restricted linear models used in hypothesis testing sections of his paper. He then compared each R^2 value obtained using grade equivalent scores to those obtain using raw scores, percentiles and standard scores.

Results

In the process of obtaining the scores other than grade equivalents, House lost 35 of 400 cases due to record keeping errors. Nevertheless, the obtained R^2 values for each type of score for both full and restricted models are shown in Table I according to the years for which the data were analyzed and the 365 cases which remained from the original population.

TABLE I

R^2 VALUES OBTAINED BY FULL AND RESTRICTED MULTIPLE LINEAR REGRESSION MODELS USING DIFFERENT SCORE MODES FOR ANALYZING THE IMPACT OF A COMPENSATORY ARITHMETIC PROGRAM

Year/ Grade Analyzed	G.E. Scores		Raw Scores		Percentile Scores		Standard Scores*	
	Full	Rest.	Full	Rest.	Full	Rest.	Full	Rest.
1975/ 4	.292	.291	.279	.279	.274	.273	.292	.291
1975/ 5	.206	.201	.194	.189	.193	.187	.208	.203
1976/ 6	.185	.168	.188	.176	.187	.171	.188	.170

Table I shows very similar results for both full and restricted R^2 values between grade equivalent and standard scores on the one hand, and between raw scores and percentiles on the other. In two out of three analysis years the R^2 values obtained from grade equivalent/standard scores are higher than those obtained using raw scores or percentiles.

The average magnitude of the effects of chosing raw scores, percentiles or standard scores as opposed to grade equivalent scores for analysis are summarized in Table II.

---

*Normal curve equivalent

TABLE II

DIFFERENCES IN OBTAINED R^2 VALUES USING GRADE EQUIVALENTS VERSUS OTHER SCORE MODES IN A STUDY OF COMPENSATORY ARITHMETIC ACHIEVEMENT

(R^2 G.E.) Minus...		(R^2 R.S.)...	(R^2 %ile)...	(R^2 S.S.)
1975	Full	.013	.018	.000
	Restricted	.012	.018	.000
1976	Full	.012	.013	-.002
	Restricted	.012	.014	-.002
1977	Full	-.003	-.002	-.003
	Restricted	-.008	-.003	-.002
	ΣX	.038	.058	.009
	$\bar{X}$	.006	.010	.002

Table II demonstrates that the average loss in explained variance using standard scores versus grade equivalents was virtually zero, while the average loss using either raw scores or percentiles was approximately 1 percent.

Conclusions

It is House's conclusion that in norm referenced achievement measures scaled scores of either type are likely to be better predictors of subsequently obtained scaled scores of the same type, than are non scaled scores. The better predictive precission of scaled scores justifies their use in the analysis of achievement where the criterion measure has been scaled, and has appropriate content validity. While House acknowledges the need for more valid and reliable measures of achievement and for better methods of analyzing achievement growth, he cautions against any impulsive bandwagon jumping-on.

Currently there is much debate over the grade-equivalent score, and it has obviously been misused. This, however, does not make it totally useless, and certainly does not justify its replacement by such imprecise, power losing scales as raw scores and percentiles. It is imperative that the user of grade equivalent scores become familiar with their characteristics and adjust to their limitations and/or benefits.

References

Cronback, Lee J. and Furby, Lita, How We Should Measure "Change" - Or Should We?, Psychologial Bulletin, 1970, Vol. 74, No. 1, pp. 68-80

Echternacht, Gary, Grade Equivalent Scores and Plas, Jeanne M., If Not Grade Equivalent Scores - Then What?, NCME Special Publication, 1977, National Council on Measurement In Education

Fennessey, James, Using Achievement Growth To Analyze Educational Programs, John Hopkins University, Baltimore, Md., Center for the Study of Social Organization of Schools

Hogan, Thomas P. and Beck Michael, Comparability of Different Score Modes for Analyses of Group Data; Paper presented at AERA, Chicago, Illinois, April, 1974

House, Gary D., A Three-Year Ex Post Facto Study of Arithmetic Achievement For Elementary Pupils Eligible For A Remedial Arithmetic Program, St. Louis Public Schools, Division of Evaluation, 1978

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

Gary D. House

St. Louis Public Schools

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Volume 9 Number 3 Feb. 1979

"A MULTIPLE REGRESSION MODEL FOR RESEARCH ON TEACHER EFFECTS"

Dr. Barry J. Fraser

A description is given of a model for research on teacher effects in which the variance in student outcome posttest performance is attributed to pretest performance, to separate construct domains of student, instructional, and teacher variables, and to interactions between variables in these three construct domains. When the model was applied with a sample of 780 Australian seventh grade pupils, it was found that pretest, an instructional variable, a block of teacher variables, a block of instruction-student interactions, and a block of instruction-teacher interactions were each significant independent predictors of a student attitudinal posttest.

A number of extensive literature reviews have indicated that research into the relationship between teacher variables and student learning outcomes has generally been disappointingly unproductive in isolating specific teacher variables which are consistently linked to student outcomes across a broad range of students and contexts (see Dunkin, 1976; Dunkin & Biddle, 1974; Medley & Mitzel, 1963; Rosenshine & Furst, 1971). According to Good and Power (1976), the major reason for the unproductiveness of this research is that researchers have unrealistically sought general teacher effects which would be applicable to all students in all situations. Instead of this naively simple conception, Jackson contends that the complexity of naturalistic classroom settings demands that research questions be phrased in terms of whether

> certain _types_ of teachers, using certain _types_ of methods, work best with certain _types_ of students, given certain _types_ of educational goals. (Jackson, 1962, p. 86)

The aims of the present paper are to describe a multiple regression model for research on teacher effects which incorporates each of the major classes of variables advocated by Jackson (1962), and to illustrate the application of this model with data from a study of teacher effects on the attitudes of seventh grade students.

The Model

The variance in student learning outcome posttest performance can be conceptualised as being a linear combination of the variance (R^2) associated with eight distinct construct domains:

$$R^2(Po_h) = R^2(Pre_h) + R^2(S_i) + R^2(I_j) + R^2(T_k) + R^2(I_jS_i) + R^2(S_iT_k) + R^2(I_jT_k) + R^2(S_iI_jT_k)$$

where Po_h stands for a multidimensional set of learning outcome <u>posttests</u>, Pre_h represents the corresponding <u>pretests</u>, S_i represents <u>student</u> variables, I_k represents <u>instructional</u> variables, and T_k represents <u>teacher</u> variables. Furthermore, as suggested by the symbols h, i, j, and k, the symbols Po_h, Pre_h, S_i, I_j, and T_k are all meant to represent numerous operational representations of variables. In the above equation, the variance in posttest performance (Po_h) is conceptualised as being attributable to pretest (Pre_h), to the three classes of variables suggested by Jackson, namely student variables (S_i), instructional method variables (I_j), and teacher variables (T_k), and to the four sets of interactions between variables in the different construct domains, namely instruction-student interactions (I_jS_i), student-teacher interactions (S_iT_k), instruction-teacher interactions (I_jT_k), and student-instruction-teacher interactions ($S_iI_jT_k$).

The model can be further clarified by considering some specific examples of variables embraced by each construct domain. Po_h and Pre_h, respectively, could include pretest and posttest measures of student achievement of any cognitive, affective, or psychomotor aims. S_i would include student classroom behaviors and individual differences (or aptitudinal) variables such as age, sex, general ability, or personality. I_j would include instructional variables like curriculum materials used or choice of instructional approach (e.g. discovery versus expository approaches). T_k would embrace teacher characteristics such as experience, sex, attitudes, or personality, and specific teacher behaviours (e.g. use of praise or questions) commonly involved in classroom interaction research (see Rosenshine & Furst, 1971). It should be noted also that the symbols S_i, I_j, and T_k are intended to embrace interactions among main effects within the same construct domain as well as the main effects themselves. For example, S_i might include an age-sex interaction as well as age and sex main effects. Similarly, the individual terms S_i, I_j, and T_k need not necessarily be linear and could include variables in quadratic, cubic, or other curvilinear form (see Nuthall, 1974). For example, S_i might involve the square of age.

Each of the first five clusters of independent variables in the model, when considered in conjunction with the dependent variables, corresponds to a well-known type of study in educational research. The term Pre_h in conjunction with Po_h corresponds to studies of changes in student outcomes over time. S_i in conjunction with Po_h corresponds to correlational studies of prediction or selection. I_j in conjunction with Po_h corresponds to classical experimental studies comparing alternative instructional treatments. T_k in conjunction with Po_h

corresponds to studies of teacher effects. I_jS_i in conjunction with Po_h corresponds to aptitude-treatment interaction (ATI) research. Whereas the first five clusters of independent variables correspond to common research approaches, the last three terms in the model - namely student-teacher interactions (S_iT_k), instruction-teacher interactions (I_jT_k), and student-instruction-teacher interactions ($S_iI_jT_k$) - seem to have received considerably less attention from researchers. In fact, although research interest in the individuality of students is reflected in the attention given to ATI research, Jackson (1962) has noted that it is surprising to find the lack of corresponding concern for the individuality of teachers. Certainly it is intuitively plausible that teachers would vary in their preferences for different instructional materials or approaches and that different students would benefit differentially from different teachers. For this reason, the last three domains of interactions have been included specifically to permit the investigation of whether optimal combinations of teacher, student, and instructional variables lead to more favourable student outcomes than do other combinations.

An important merit of the proposed model is that it accommodates a very comprehensive range and variety of independent variables, including interaction and non-linear terms, which influence student outcomes. Nevertheless, by employing clusters of variables, the model offers some degree of simplicity and clarity in conceptualizing the quite complex relationships which exist between student outcomes and student, instructional, and teacher variables. It is also important to emphasise that, although the model can accommodate very large numbers of independent variables, research-strategic and practical considerations would normally place limitations on the number of variables actually included in a particular study. For example, the use of certain sampling procedures or statistical controls could obviate the need to include measures of pretest achievement or student characteristics. Similarly, in studies involving a reasonably large number of main effects, it would normally be practically feasible to include in the study only a small proportion of the large number of possible interaction terms. Nevertheless, despite the fact that specific studies might often exclude some of the possible terms, a major advantage of the proposed model is that it reflects the complexity of research on teacher effects and accommodates a comprehensive range of variables which should be taken into consideration by the researcher.

A Priori Ordering of Construct Domains

It should be emphasised that it is only in the special case where all independent variables are mutually uncorrelated (e.g. balanced factorial designs) that the magnitude of R^2 associated with a particular construct domain is invariant to the ordering of construct domains in the model (see Kerlinger & Pedhazur, 1973). For example, in the nonorthogonal case, estimates of the variance associated with teacher effects would usually vary considerably depending whether or not the variance attributable to pretest or student characteristics was first removed.

Although theoretical perspectives should dictate what is the best a priori ordering in any particular study, the ordering suggested in the proposed model is likely to be suitable for many studies of teacher effects. In this ordering, the variance attributable to the pretest is first removed since one is interested in the influence of teacher effects in bringing about changes in student outcomes during the time between pretesting and posttesting. Next the variance attributable to student aptitudinal variables is removed since one normally requires an estimate of the variance explained by instructional or teacher variables over and above that explained by differences in pretest and student variables. The variance attributable to instructional variables is removed next so that the variance due to teacher effects can be estimated as a quantity over and above that accounted for by pretest, student, and instructional variables. Moreover, by entering teacher effects into the model only after all other main effects, a conservative test of relationships between teacher effects and student outcomes is provided. Finally, the variances due to clusters of interactions are estimated (in the order corresponding to the order of entry of main effects) after the removal of variance associated with all main effects. That is, for reasons of simplicity, main effects are first explored and then interaction effects are estimated in terms of the variance they account for over and above that due to main effects (see Cohen & Cohen, 1975).

Application of the Model

The proposed model was employed in conceptualising a study into the relationship between teacher effects and a student attitudinal outcome among some Australian students. The sample consisted of 31 teachers and their 31 seventh grade science classes (780 students), each in a different coeducational high school in the Melbourne metropolitan area. Although schools were not randomly chosen, they were still spread representatively over the range of socioeconomic and geographical areas around Melbourne.

The outcome investigated was student enjoyment of their science lessons. The instrument used to measure this attitude was a scale described by Fraser (1977) which has an α reliability of 0.85 and consists of seven items with a five-point Likert response format. This scale was administered as a pretest at the beginning of a school year and then again as a posttest at the end of the same year.

In addition to teacher effects, both student and instructional variables were included in the research. The student variables (S_i) included were socioeconomic status (SES), general ability, and sex. These three variables were selected because Lavin's (1965) comprehensive literature review indicated that these three characteristics have consistently been found to be related to a variety of student outcomes. While SES was measured using Congalton's (1969) classification of Australian occupations, general ability was measured with a version of the Otis test. The instructional variable (I_j) was a dichotomous variable designating whether students followed traditional science materials or Australian Science Education Project (ASEP, 1974) materials in science lessons during the time of the study.

A set of four teacher variables (T_k) was included in the study: sex, experience, attitude to pupil-centredness, and attitude to structure. The two teacher attitude variables were measured, respectively, by the 10-item and five-item factor-analytic scales described in Fraser and Northfield (1976). While high scores on the attitude to pupil-centredness scale involved agreement with statements that pupils should work in small groups and be permitted to have a choice of classroom activities, high scores on the attitude to structure scale involved agreement with statements that students should be taught ideas which are drawn from each major scientific discipline and which provide a sound preparation for future science study.

Because the number of interaction terms, many of high order, possible with eight main effects (instruction, three student variables, and four teacher variables) is excessively large, severe restrictions had to be placed on the number of interaction terms included in the analysis. First, all interactions of order greater than two-way were excluded from consideration. Second, interactions among variables in the same construct domain were deemed of lesser importance than interactions between terms in different construct domains. Third, it was decided to omit student-teacher interactions because the number of possible two-way interactions, namely 12, was considered excessively large. After excluding these interactions, the following seven interaction terms remained: three instruction-student interactions (instruction-SES, instruction-general ability, and instruction-student sex) and four instruction-teacher interactions (instruction-teacher sex, instruction-experience, instruction-attitude to pupil-centredness, and instruction-attitude to structure). The inclusion of these interaction terms permitted investigation of the differential effectiveness of ASEP and traditional materials for different students and different teachers. In terms of the proposed model, then, I_jS_i and I_jT_k interactions were included in the research while S_iT_k and $S_iI_jT_k$ interactions were excluded.

Data Analysis

Multiple regression analysis was a particularly appropriate data analytic technique for several of the reasons outlined by Cohen and

Cohen (1975). First, multiple regression techniques provide a number of solutions when predictor variables are intercorrelated. Second, estimates of R^2 are automatically provided. Third, multiple regression techniques permit the inclusion of interactions yet readily allow the exclusion of some of the interaction terms normally present in analysis of variance solutions. Fourth, the statistical power of the analysis could be maximised by maintaining all quantitative predictor variables as such (rather than reducing them to categories) both when estimating main effects and when forming interactions by taking products of predictor variables.

Good and Power (1976) have claimed that subgroups of students within classes are more appropriate units of analysis for research on teacher effects than either individual students or intact classes. For this reason, the unit of analysis used in the present investigation was the subgroup within the class formed by grouping students according to similarities in SES, general ability, and sex. Each student in the sample was classified as either high or low SES and as either high or low general ability according to whether his or her scores were above or below the whole sample's median for SES and general ability. Each student within a given class was assigned to one of eight possible subgroups according to his or her dichotomous scores on the three variables of SES, general ability, and sex. When this procedure was used, it was found that, either because of absences during testing or because of the distribution of SES, general ability, and sex in some classes, the total sample consisted of 231 subgroups which is only 17 less than the maximum number possible if each of the 31 schools provided all eight subgroups.

The 16 independent variables were entered into the regression equation in six stages corresponding to the ordering of terms in the proposed model. Pretest scores (Pre_h) were entered first, the cluster of student variables (S_i) was entered next, then the instructional variable (I_j), then the group of four teacher variables (T_k), then the block of three instruction-student interactions (I_jS_i) and, lastly, the block of four instruction-teacher interactions (I_jT_k). Table 1 shows the cumulative value of R^2 at each of the six stages, the increment in R^2 produced by the addition of each of the six clusters of variables, and the F value associated with each increment in R^2 calculated using Overall and Spiegel's (1969) Method 3. This table shows that pretest scores accounted for 25.0 per cent of the variance in posttest performance, student variables accounted for a further 1.2 per cent, the instructional variable accounted for another 3.7 per cent, the set of teacher variables accounted for another 5.0 per cent, the set of instruction-student interactions accounted for a further 2.4 per cent, and the set of instruction-teacher interactions accounted for a further 3.3 per cent of the variance in posttest scores. Together, the whole set of 16 predictor variables accounted for a total of 40.6 per cent of posttest variance. Furthermore, Table 1 indicates that the successive addition of the six clusters of variables was associated with a significant increment in variance at the 0.05 level of confidence for all clusters except the cluster of student variables.

Table 1. Percentage of variance in enjoyment scores and F values associated with six sets of predictors

Block of Variables Added	Cumulative R^2 (%)	Increment in R^2: ΔR^2 (%)	F	df	Unique Contribution of Individual Variables in Blocks: Variable	R^2 (%)	Significant F value
Pretest (Pre_h)	25.0	25.0	87.1**	1,214			
Student variables (S_i)	26.2	1.2	1.4	3,214	SES	0.4	
					General ability	0.5	
					Pupil sex	0.0	
Instruction (I_j)	29.9	3.7	12.9**	1,214			
Teacher variables (T_k)	34.9	5.0	4.3**	4,214	Teacher sex	2.4	8.4**
					Experience	0.4	
					Att. to pupil-centredness	1.7	5.9*
					Att. to structure	0.3	
Instruction-student interactions (I_jS_i)	37.3	2.4	2.7*	3,214	Instr x SES	1.3	4.5*
					Instr x gen ability	0.1	
					Instr x pupil sex	0.6	
Instruction-teacher interactions (I_jT_k)	40.6	3.3	11.5**	4,214	Instr x teacher sex	0.3	
					Instr x experience	0.5	
					Instr x att. to pupil centredness	0.5	
					Instr x att. to structure	1.4	4.9*

At 0.05 level of confidence, critical $F(1,214) = 3.9$, $F(3,214) = 2.6$, $F(4,214) = 2.4$

To facilitate interpretation of findings, the increment in variance associated with each cluster of predictors was further partitioned into unique amounts accounted for by each individual predictor within a cluster. However, although an a priori ordering was used for entering the six different clusters of variables into the regression equation, it was considered that no a priori ordering could be justified for the individual predictors within a particular cluster. Instead, an estimate was made of the amount of variance uniquely accounted for by a particular individual predictor over and above that accounted for by all other variables in the same block (and all variables in preceding blocks). Because some correlation existed between variables within a given block, however, it can be seen from Table 1 that the sum of the unique contributions to variance made by individual predictors in a given block is a little smaller than the increment associated with the block as a whole. Table 1 shows the unique contribution to variance in posttest performance attributable to each individual predictor together with F values for those individual variables which were significant independent predictors of posttest scores. Results indicate that a significant relationship emerged for an individual variable within a cluster in four cases: two teacher variables (sex and experience), an instruction-student interaction (instruction-SES), and an instruction-teacher interaction (instruction-attitude to structure).

The interpretation of the significant result for the instructional variable is that, when pretest enjoyment was held constant, pupils following ASEP materials expressed greater posttest enjoyment than pupils following traditional materials. With pretest enjoyment held constant, the interpretations of the two significant findings for teacher effects are as follows: female teachers' classes expressed greater posttest enjoyment than did male teachers' classes, and posttest enjoyment was greater in classes of more pupil-centred teachers than in classes of less pupil-centred teachers. The interpretation of the instruction-SES interaction is that, with pretest constant, posttest enjoyment was almost independent of SES among pupils using traditional materials, while ASEP pupils of higher SES expressed greater posttest enjoyment than ASEP pupils of lower SES. The interpretation of the instruction-attitude to structure interaction is that, with pretest constant, student posttest enjoyment was almost independent of teacher attitude to structure in the ASEP group, while posttest enjoyment in the control group was greater for teachers with more favorable attitudes to structure than for teachers with less favorable attitudes to structure.

Conclusion

A description was given of a multiple regression model for research on teacher effects in which the variance in student outcome posttest performance is conceptualized as a linear combination of the variance associated with the following eight construct domains: pretest performance, student variables, instructional variables, teacher variables, instruction-student interactions, student-teacher interactions, instruction-teacher interactions, and student-instruction-teacher interactions. This model was employed in conceptualizing a study of the relationship between teacher effects and student enjoyment of their science lessons in some Australian seventh grade classrooms. It was found that pretest, an instructional variable, a block of four teacher variables, a block of three instruction-student interactions, and a block of four instruction-teacher interactions were all significant independent predictors of student attitudinal posttest scores. When pretest enjoyment scores were held constant, student posttest enjoyment scores were found to be related to the instructional variable (involving choice of curriculum materials), teacher sex, teacher attitude to pupil-centredness, an instruction-student socioeconomic status interaction, and an instruction-teacher attitude to structure interaction.

The author gratefully acknowledges the comments of Michael Dunkin on a draft of this manuscript.

References

ASEP (Australian Science Education Project). A guide to ASEP. Melbourne: Government Printer, 1974.

Congalton, A. A. Status and prestige in Australia. Melbourne: Cheshire, 1969.

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. New York: Wiley, 1975.

Dunkin, M. J. Problems in the accumulation of process-product evidence in classroom research. British Journal of Teacher Education, 1976, 2, 175-187.

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt, Rinehart & Winston, 1974.

Fraser, B. J. Selection and validation of attitude scales for curriculum evaluation. Science Education, 1977, 61, 317-329.

Fraser, B. J., & Northfield, J. R. A study of some aspects of ASEP during its first year of availability. Report to the Australian Curriculum Development Centre, Canberra, 1976.

Good, T. L., & Power, C. N. Designing successful classroom environments for different types of students. Journal of Curriculum Studies, 1976, 8, 45-60.

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

Lavin, D. E. The prediction of academic performance. New York: Sage, 1965.

Jackson, P. W. The teacher and individual differences. Sixty-first yearbook of National Society for Study of Education, Part I. Chicago: University of Chicago Press, 1962.

Medley, D. M., & Mitzel, H. E. Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Nuthall, G. A. Is classroom interaction research worth the effort involved? New Zealand Journal of Educational Studies, 1974, 9, 1-17.

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

Rosenshine, B., & Furst, N. Research on teacher performance criteria. In B. O. Smith (Ed.), Research on teacher education. New Jersey: Prentice Hall, 1971.

Barry J. Fraser

Macquarie University

Sidney, Australia

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Volume 9 Number 3 Feb. 1979

"A NOTE ON THE CALCULATION OF DEGREES OF FREEDOM FOR POWER ANALYSIS USING MULTIPLE LINEAR REGRESSION MODELS"

Dr. Isadore Newman and
Jay Thomas

This note presents fifteen examples worked by Cohen in which he uses different formulas to calculate degrees of freedom, depending on the power analysis situation. It is then demonstrated that the same results can be obtained by using a more general formula for calculating degrees of freedom. It was felt that this information may be of pedagogical value.

Cohen (1972) has made the use of power analysis more common and more frequently referenced to in the behavioral science literature. There is obviously nothing new in the power analysis concepts Cohen presents. His work is based on the earlier work by Neyman and Egan Pearson. However, we believe Jacob Cohen has made a very significant contribution in popularizing its use for the applied researcher.

Before the work of Jacob Cohen many more applied researchers felt the power curves were only useful if they knew the population parameters. Cohen has developed a variety of tables and many examples which make it very easy to understand and use.

In Jacob Cohen's book Statistical Power Analysis for the Behavioral Sciences, revised edition 1977, he includes a chapter (chapter 9) on doing power analysis using multiple regression for fixed models. In this chapter he presents 15 examples of different situations in which regression was used. Depending upon the example, different formulas for calculating degrees of freedom are incorporated.

We feel it may be easier or at least more effective from a pedagogical point of view to use one formula that is consistent for all cases. For this purpose we have recalculated the degrees of freedom (df) for the examples 9.1 - 9.15 presented in Cohen (1977, ch. 9) using the formula for df_1 and df_2 given in McNeil, Kelly, McNeil (1975).

$$df_1 \text{ in Cohen} = u$$

$$df_2 \text{ in Cohen} = v$$

The following formula was used:

$$df_1 = M_1 - M_2$$

$$df_2 = N - M_1$$

where M_1 = number of linearly independent vectors in the full model

M_2 = number of linear independent vectors in the restricted model

N = number of independent replicates (subjects)

The following are the examples from Cohen along with their page numbers. Please note only in example 9.9 is there a difference. This is found in df_2, where our calculations for df_2 differs by 1. We feel that either Cohen or we made a mistake in the calculations of df_2. In either case the examples support the contention that Cohen's method for calculating power analysis can be made even simpler to use by employing the method of calculating degrees of freedom that the applied researcher is most familiar with. For example, if one is familiar with the McNeil *et al* text approach that procedure can be substituted for Cohen's U and V.

Examples

page 419

9.1 N=95 M_1=6 M_2=1

$u = 6 - 1 = 5$
$v = 95 - 6 = 89$ check with Cohen

9.2 N=90 M_1=4 M_2=1

$u = 4 - 1 = 3$
$v = 90 - 4 = 86$ check with Cohen

page 420

9.3 N=90 M_1=4 M_2=1

$u = 4 - 1 = 3$
$v = 90 - 4 = 86$ check with Cohen

9.4 N=326 M_1=12 M_2=1

$u = 12 - 1 = 11$
$v = 326 - 12 = 314$ check with Cohen

page 421

9.5 N=80 M_1=4 M_2=1

$u = 4 - 1 = 3$
$v = 80 - 4 = 76$ check with Cohen

page 422

9.6 N=50 M_1=2 M_2=1

$u = 2 - 1 = 1$
$v = 50 - 2 = 48$ check with Cohen

9.7 $N = 150$ $M_1 = 2$ $M_2 = 1$

$u = 2 - 1 = 1$
$v = 150 - 2 = 148$ check with Cohen

Model 1 $R^2y.A, B$ -- $Ry^2 \cdot A$

$u = k_B - 1$
$v = N - u - w - 1.$

9.8 $N = 90$ $M_1 = 6$ $M_2 = 4$

$A = 3,\ B = 2$

$u = 6 - 4 = 2$
$v = 90 - 6 = 84$ check with Cohen

9.9 $N = 80$ $M_1 = 5$ $M_2 = 2$

$A = 1,\ B = 3,\ E = 1$

$u = 5 - 2 = 3$
$v = 80 - 5 = 75$

OK
Note: Cohen uses $A = 2$, so he gets $u = 74$. I think his A is wrong.

9.10 $N = 90$ $M_1 = 10$ $M_2 = 6$

$A = 5$ $B = 4$ $E = 1$ (Note: B=4 because B is coding for 5 ethnic groups)

$u = 10 - 6 = 4$
$v = 90 - 10 = 80$ check with Cohen

9.11 $N = 148$ $M_1 = 11$ $M_2 = 9$

$A = 8$ $B = 2$ $E = 1$

$u = 11 - 9 = 2$
$v = 148 - 11 = 137$ check with Cohen

9.12 $N = 95$ $M_1 = 6$ $M_2 = 4$

$A = 3$ $B = 2$ $E = 1$

$u = 6 - 4 = 2$
$v = 95 - 6 = 89$ Check with Cohen

9.15 $N = 90$ $M_1 = 4$ $M_2 = 3$

$A = 2$ $B = 1$ $E = 1$

References

Cohen, Jacob. "F Test of Variance Proportions in Multiple Regression/ Correlation Analysis." in _Statistical Power Analysis for the Behavioral Sciences_, rev. ed. New York: Academic Press, 1977.

McNeil, Kelly, and McNeil. _Testing Research Hypotheses Using Multiple Linear Regression_. Carbondale: Southern Illinois University Press, 1969.

Isadore Newman
Jay Thomas

The University of Akron

HILTON (DOLORES, BALLROOM FLOOR)

Tuesday, April 10, 1979

12:25 to 3:25 p.m.

SESSION TITLE - MULTIPLE REGRESSION -

Methodological Advances and Applications

SPONSOR SIG on Multiple Linear Regression

Chair John D. Williams, University of North Dakota

Unmeasured Variables in Path Analysis. Lee M. Wolfle, Virginia Polytechnic and State University.

On the Use of Two Common Measures of Importance in Linear Causal Models. Lloyd Bond, University of Pittsburg, Learning Research and Development Center and James Algina, Florida State University.

An Analysis with Illustrations of Suppressor Variables in Social Science Research. Daniel U. Leving, School of Education, University of Missouri-Kansas City, Cris Kukuk, Social Impact Research, Inc. and Jeanie Keeny Meyer, Kansas City (MO) Police Dept.

Non-linear Transformation of the Criterion. Keith McNeil, NTS Research Corp., Joe Evans, Evaluation Technology and Judy McNeil, Educational Monitoring Systems.

A Monte Carlo Evaluation of Estimated Parameters of Five Shrinkage Estimate Formuli. Isadore Newman, Gayle A. Seymour, and Thomas K. Garver, the University of Akron.

Missing Cells and a Curious Case of Degrees of Freedom. John D. Williams, and Mohan K. Wali, the University of North Dakota.

Using Multiple Regression to Interpret Chi-Square Contingency Table Analysis. Dennis Leitner, Southern Illinois University of Carbondale.

Multiple Regression as a Technique for Predicting College Enrollment. Ambrose A. Clegg, Jr., and Paul Weigand, Kent State University.

Discussants Leigh Burstein, UCLA and Paul Games, Penn State University

General SIG Meeting
Chair: John D. Williams, the University of North Dakota
Chair-Elect: William Connett, State Dept. of Education
Helena, Montana

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

Title

Author and affiliation

Indented abstract (entire manuscript should be single spaced)

Introduction (purpose—short review of literature, etc.)

Method

Results

Discussion (conclusion)

References

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

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

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

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

THE UNIVERSITY OF AKRON
AKRON, OHIO 44325

AKRON OHIO
U.S. POSTAGE 40
P.B. METER 672269

157SPAN9
SPANER, STEVEN D.
BEHAVIORAL STUDIES
UNIV OF MO-ST LOUIS
ST LOUIS, MISSOURI 63121

TABLE OF CONTENTS