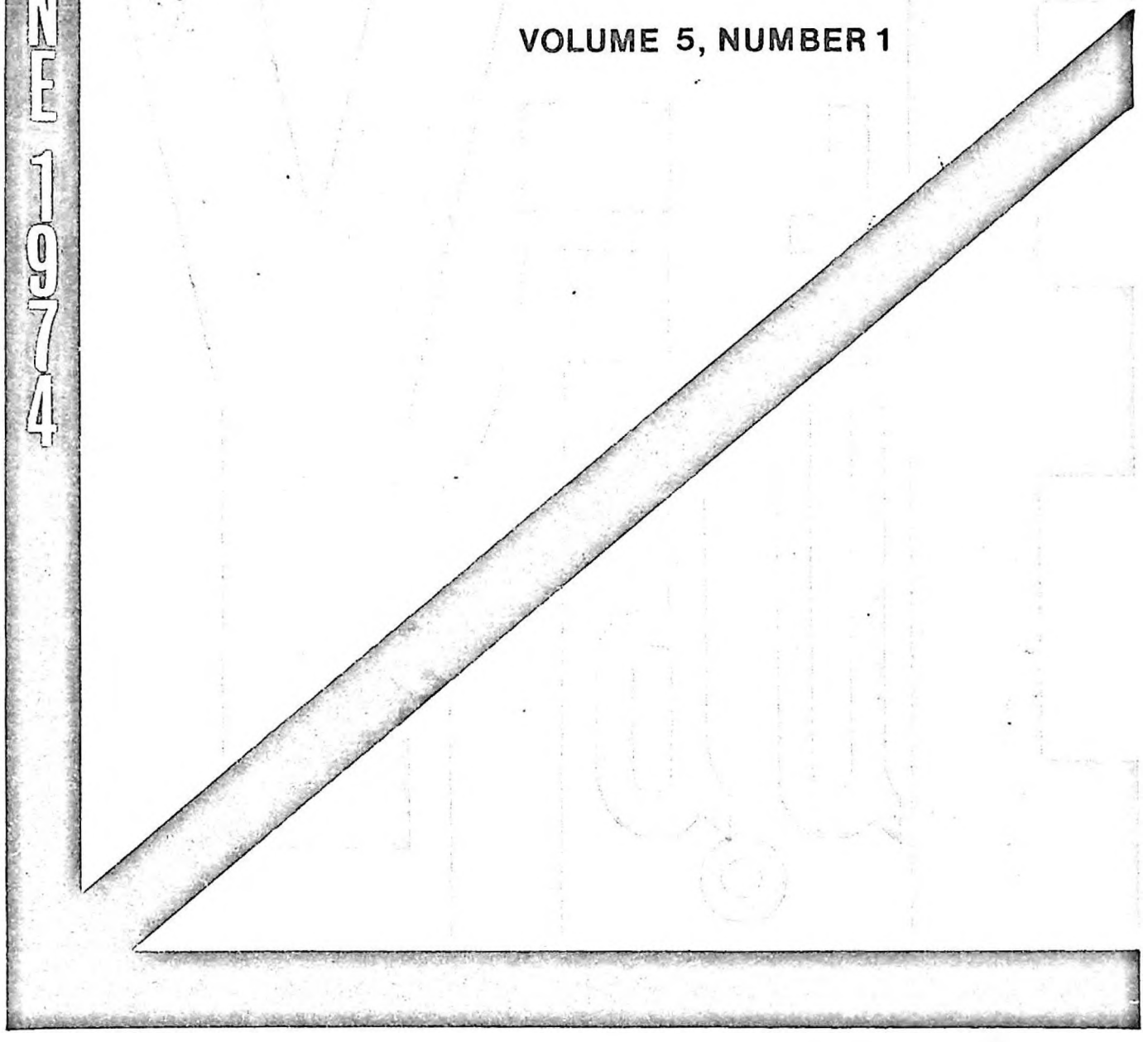
JUNE 1974

VOLUME 5, NUMBER 1



MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group
on Multiple Linear Regression
of The American Educational Research Association.

IF YOU ARE SUBMITTING A RESEARCH ATRICLE OTHER THAN NOTES OR COMMENTS, I WOULD LIKE TO SUGGEST THAT YOU USE THE FOLLOWING FORMAT, AS MUCH AS POSSIBLE:

TITLE
AUTHOR AND AFFILIATION
SINGLE-SPACED INDENTED ABSTRACT (THE REST OF THE MANUSCRIPT SHOULD BE DOUBLE-SPACED)
INTRODUCTION (PURPOSE--SHORT REVIEW OF LITERATURE, ETC.)
METHOD
RESULTS
DISCUSSION (CONCLUSION)
REFERENCES

ALL MANUSCRIPTS SHOULD BE SENT TO THE EDITOR AT THE ABOVE ADDRESS. (ALL MANUSCRIPTS SHOULD BE CAMERA-READY COPY)

TABLE OF CONTENTS

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 5, No. 1, 1974

METHOD FOR COMPARISON OF NON-INDEPENDENT MULTIPLE CORRELATIONS

Jack R. Haynes and Ronald G. Swanson
North Texas State University

In the last several years, multiple linear regression and multiple correlation have been utilized in a wide variety of research situations. Multiple linear regression models have been found applicable to curve fitting problems (Christal, 1967) and classical analysis of variance problems (Bottenberg & Ward, 1963) as well as problems requiring prediction equations, correlational values, and communality estimates.

A typical situation is one in which an investigator is trying to measure the relationship between a criterion and a set of predictor variables. Usually an R is computed and a procedure such as stepwise multiple regression is employed to determine if some subset of the predictor variables is as efficient in prediction as all of the predictors. Since the amount of predictable variance is readily obtained from R^2, the decision can be made via an F test comparing the magnitude of difference between two R^2 values. Since the form of the distribution of R is known, the significance of a given R is also available by use of the F test.

Types of Problems

No technique has been presented for the comparison of two or more non-independent correlations. A search of the literature, plus queries to a number of eminent statistical theorists and

practitioners, made evident the fact that this particular type of comparison had not been dealt with previously. Such a comparison could be useful in solving two types of research problems.

Problem one is to determine if the multiple correlation between a criterion and a set of predictors is significantly different from an R between the same criterion and a different set of predictors within the same sample. For example, an investigator may wish to find out if a set of personality measures correlated higher with job performance than a set of demographic variables in the same group of workers.

Problem two deals with the question of the differential predictive ability of a set of predictors from population to population. In this situation the criterion measures on the subjects would be constant, but the measures on the predictors would be obtained from different groups of subjects. An example would be where different groups such as subordinates, peers, and superiors were all evaluating the same subjects on a set of measures. If the multiple correlations were significantly different from each other, there would be differential predictive ability among the three groups. This problem differs from the situation where the Rs are computed on totally different samples, for in problem two the same subjects and criteria are used in the computation of each R. If the Rs are computed on independent samples, the multiple correlations can be treated as zero order correlations and a test of significance can be

made with an r to z transformation (Kendall & Stuart, 1961). The r to z transformation cannot, however, be used where non-independent samples are involved.

Rationale for a Comparison Procedure

Although problems one and two differ in some respects, they are similar in that the subjects' measures on the criterion are the same in the computation of all the multiple correlations. Thus estimates of any differences must consider the interrelationships among the measures. In problems one and two, the basic consideration is the degree of accuracy in predicting the criterion which is reflected in the multiple correlation and the amount of residual variance. Since R is a correlation between a criterion and a linearly weighted combination of variables, the whole matrix of correlations between the variables does not need to be considered (Fisher, 1928). Thus, the differential contribution of the individual predictor variables is not a consideration; only the sum of their linear combination, for different weighted combinations of sets of variables, could produce equal predictability. Therefore, the situation having the highest multiple correlation would logically be the one which produced the best prediction. The best prediction would also yield the lowest residual variance and the smallest difference between the actual and predicted criterion scores. The analysis of difference scores would indicate differences in predictability and, thus, differences between non-independent multiple correlations.

Procedure for Comparison

The technique for the comparison of related Rs is based on the residual criterion scores rather than directly on the R or R^2 values. The procedural steps are: (a) generate predicted criterion scores for each set of predictor variables; (b) for each set of predicted scores, obtain the absolute difference between each predicted and corresponding actual criterion scores; (c) analyze these difference scores by a single classification analysis of variance for repeated measures. Absolute difference scores are used in the interest of maintaining a relatively conservative test of the error residuals. Squaring large difference scores would result in biasing the test in favor of lesser deviation values, i.e., an error residual of two when squared equals four, but an error score of ten squared equals one-hundred. Hence, an inordinate amount of supposed error for larger error scores would be subject to the scrutiny of the ANOVA, biasing the test of differences. Each set of difference scores are thus treated as a sample of dependent measures from different treatments, and since the criterion measures are the same for the computation of all difference scores, the residual scores would be considered correlated or non-independent. For problem one the treatments would be different predictor sets, and for problem two the treatments would be the different sources of prediction, i.e., subordinates, peers, and superiors. In addition, the mean of each treatment condition is the mean of the deviation values for that condition, If a significant F is obtained from the ANOVA, some or all of the multiple correlations would be considered different. Multiple comparisons to determine which Rs differ

from each other are then conducted using techniques such as the Newman-Keuls studentized range statistical test.

The assumptions necessary to use such a comparison technique would involve the already established assumptions inherent in the use of any multiple linear regression model or single-classification repeated measures ANCVA. This proposed technique is seen as a simple and logical approach to an important mode of comparison not previously dealt with.

References

Bottenberg, R. A. & Ward, J. H. Applied multiple linear regression. Technical Documentary Report PRL - TDR - 63 - 6, March, 1963, Personnel Research Laboratory, Lackland Air Force Base.

Christal, R. E. Selecting a harem and other applications of the policy-capturing model. Technical Report PRL - TR - 67 - 1, March, 1967, Personnel Research Laboratory, Lackland Air Force Base.

Fisher, R. A. The general sampling distribution of the multiple correlation coefficient. Proceedings of the Royal Society of London, Series A., Vol. 121, 1928, 654-673.

Kendall, M. G. & Stuart, A. The advanced theory of statistics. Vol. 2. Inference and Relationship. London: Charles Griffin, 1961.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 5, No. 1, 1974

THE MULTIPLE REGRESSION APPROACH FOR ANALYZING DIFFERENTIAL TREATMENT EFFECTS – THE REVERSED GESTALT MODEL

Joseph F. Maola
Graduate Assistant in Guidance and Counseling
The University of Akron

Many times in educational and psychological research the same experimental treatment is provided by different people. An example of this effect would be a comparison of Behavioral vs Gestalt group counseling where there may be three different counselors providing each treatment. Since the individuals providing the treatments are unique then each treatment, no matter how objectively defined, becomes for the most part a different treatment effect. The three counselors providing the Gestalt group counseling all present the treatment to the subjects in a way which is uniquely theirs while the subjects will perceive the treatment differently from each counselor. The purpose of this paper is to present a statistical procedure for examining differences in criteria above independent differences. The theoretical position underlying the approach is that "the sum of the parts is greater than the whole." Since this position is directly opposite that of the Gestaltists, the model will be referred to as the "Reversed Gestalt Model." Before the reader begins dismembering, hoping his body will function more effectively, keep in mind that this is a statistical position for examining differences.

For the sake of explanation, the following model will be used: A researcher is trying to determine whether the Gestalt method or the

Behavioral method of group counseling is more effective for increasing the self concept of a certain population. The models for testing this hypothesis are:

Full Model 1: $Y_1 = a_0U + a_1(X_1) + a_2(X_2) + E$

restriction: $a_1 = a_2$

Restricted Model 99: $Y_1 = a_0U + E$

where: Y_1 = self-concept score

X_1 = Gestalt Treatment (1 if Gestalt, 0 otherwise)

X_2 = Behavioral Treatment (1 if Behavioral, 0 otherwise)

Whether or not significant differences are found may lead the researcher to ask questions whether the effects can be maximized by covariance of differential effects of the treatment.

In order to begin the comparison the researcher would first determine if there are significantly different treatment effects. The model for testing the hypothesis of differences among counselors for the Gestalt group treatment would be:

Full Model 2: $Y_1 = a_0U + a_1(X_3) + a_2(X_4) + a_3(X_5) + E$

restriction: $a_1 = a_2 = a_3$

Restricted Model 99: $Y_1 = a_0U + E$

where: Y_1 = self-concept score

$X_1 = X_3 + X_4 + X_5$ (1 if Gestalt, 0 otherwise)

X_3 = Gestalt Counselor A's group (1 if Gestalt Counselor A, 0 otherwise)

X_4 = Gestalt Counselor B's group (1 if Gestalt Counselor B, 0 otherwise)

X_5 = Gestalt Counselor C's group (1 if Gestalt Counselor C, 0 otherwise)

If significant differences are found the researcher will now want to determine whether a significant amount of variance is accounted for by the differential effects above and beyond treatment differences. The model (Reversed Gestalt Model) would be:

Full Model 3: $Y_1 = a_0U + a_1(X_2) + a_2(X_3) + a_3(X_4) + a_4(X_5) + E$

restriction: $a_2 = a_3 = a_4$

Restricted Model 4: $Y_1 = a_0U + a_1(X_1) + a_2(X_2) + E$

where: Y_1 = self-concept score

$X_1 = X_3 + X_4 + X_5$ (1 if Gestalt, 0 otherwise)

X_2 = Behavioral Treatment (1 if Behavioral, 0 otherwise)

X_3 = Gestalt Counselor A's group (1 if Gestalt Counselor A, 0 otherwise)

X_4 = Gestalt Counselor B's group (1 if Gestalt Counselor B, 0 otherwise)

X_5 = Gestalt Counselor C's group (1 if Gestalt Counselor C, 0 otherwise)

This model allows the researcher to look at the independent Gestalt Counselor differences above and beyond the knowledge of being in the Gestalt or Behavioral group. Since the restriction implies that all the regression weights are equal, then if they are not, there will be significantly more variance accounted for by the counselor differences.

Testing Full Model 3 against Restricted Model 4 permits the researcher to determine whether a significant amount of variance is accounted for by the differential effects than from the total Gestalt treatment. The reversed Gestalt model, therefore, allows the researcher to determine whether a significant difference exists between the different applications of a treatment (the 3 groups) above and beyond the subject knowledge of the overall group treatment. The method should also be repeated for the behavioral group counseled treatment to determine whether differential effects are also significant.

After determining whether significantly greater variances are accounted for by the differential treatment effects as compared to the total treatment effect, then there are two applications the researcher may make. One application would be using the procedure for allowing the researcher an opportunity of generalizing not only to the criteria but also to the effect of independent differences of providing treatment. Another application of the procedures may be using the Full Model with the differential effects, accounting for more variance, as the Gestalt Counseling criteria in order to maximize the Gestalt treatment effect.

This paper was written in order to demonstrate a method of using multiple regression for determining the effect of independent differences of providing treatment. The procedures involves a model of comparing differential treatment effects to the total treatment and was, therefore, named the "Reversed Gestalt Model" because of its theoretical base. The writer is hopeful that the reader will find this procedure useful for future research.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 5, No. 1, 1974

THE USE OF REGRESSION EQUATIONS TO DEMONSTRATE CAUSALITY

John F. Byrne
Behavioral Research Department
Westinghouse Electric Corporation
Pittsburgh, Pennsylvania 15235

A universal objective of scientists involved with explanations of behavior or phenomenon is to demonstrate their knowledge of what is causing it. Explanations of causality usually entail knowledge of a number of elements or underlying variates that interrelate to produce the phenomenon.

Causality may exist at three levels, at least. It is important that a brief mention of the philosophy of knowledge be made since the value of regression equations to demonstrate causality can be lost in the shuffle between types of causal explanation.

Science generally acknowledges that a causal explanation can exist independent of human perception and that the purpose of research is to develop a human picture or understanding of what exists "out there" which comes as close as possible to what is really out there. For our purposes we can accept three levels of causality, -- the first being what is really causing the phenomenon, the second being what we think is causing the phenomenon, and the third being what our measurements and analysis of variance state is causing the phenomenon. This can be pursued further with Simon (1971).

An assertion that causality is understood must be defended at the second level by means of reason, interpretation, consensus and subjective reliability among scientists. An assertion that causality is understood at the third level can only be defended with a high R Square, low probability of chance occurrence, and replication.

Of the three terms introduced to defend the assertion of causality at the statistical level, R Square is the one which requires the greatest amount of explanation. A number of formal statistics texts written in the language of regression analysis such as Kelly et.al (1969) can take the reader through basic mathematical operations and other operational aspects of R Square. The purpose of the present digression is to focus on the meaning of R Square in a larger than mathematical sense, e.g. once it has been computed, what does it mean to the research project?

R Square (R2) is a term which summarizes the effectiveness of the experimental search for causality. A high R2 indicates that the interrelationships among the measurements can be statistically connected to changes that take place in the phenomenon being studied.

R2 values range from 1.00 to 0.0. An R2 of 1.00 means that 100% of the change that takes place in the phenomenon has been numerically connected to changes that take place among the variables and experimental manipulations controled by the scientist. As the R2 value approaches 0.0, it means that more and more things outside of the comprehension of the scientist are appearing which cause the phenomenon. At 0.0 R2, the scientist is farthest from a statistical explanation of causality. Still, one may only guess, and not too emphatically, as to his grasp of level one or level two causality from this statistical fact.

Unfortunately it is not unusual to find statistical analyses which omit R2 from the report. Many researchers mistakenly believe it meaningful to report that some manipulation of variables was shown to have a significant effect upon a behavior or phenomenon. It must be underlined that a significant effect of unspecified magnitude communicates very little to the reader or to the scientist himself. It has been argued that the purpose of science is not to demonstrate that effects of unknown magnitude can be caused, but rather the purpose of science is to demonstrate that what causes a phenomenon is completely understood or completely connected to a set of measurable variables which are subject to the knowledge and/or manipulation of the scientist. When you consider a research report which reports only the alpha level significance of an effect, you must realize neither you nor the author understands (level 3) what he does and does not know about the behavior he is affecting. It is well known that an R2 of .05 can be associated with a t or F ratio which could only have occurred one time in a thousand by random sampling error or chance. In such a case the scientist can be relatively certain that he produced an effect, but he must admit he knows very little about the thing he is affecting. However, it is also likely that the short course in statistics and short-cut formulae which have been overly popular have simply omitted the role of R2 since the typical scientist is foremost concerned with causality as it appears at the substantive (level 2) or knowledge level. The fact is, however, that when the scientist puts his human grasp of causality to an experimental test involving measurements

and the analysis of variance among those measurements, there is a statistical term in the statistical realm of causality which he must regard, e.g. R2.

Friedman (1969) provides a table for converting variance ratios such as t,Z,X2, and F to R2. The following equations from McNeil, Kelly and McNeil (1974) can be used to calculate R2 when the reader happens upon an incomplete statistical report.

$$R2 = \frac{F\frac{(dfn)}{(dfd)}}{1 + F\frac{(dfn)}{(dfd)}} \qquad (1.1)$$

$$R2 = \frac{t2}{df + t2} \qquad (1.2)$$

Where

dfn = degrees of freedom numerator

dfd = degrees of freedom denominator

df = degrees of freedom

Note that the main point to this presentation is that R2 must be taken into consideration when one is interpreting the practical significance an experimental undertaking. For example, a recent study of the effect of Scanned Illumination upon perceptual motor behavior by Ramage (1974) reported a t value of 4.05 with 6 degrees of freedom and

an alpha level less than .01 to summarize the statistical analysis of variance between two types of illumination and one behavior. That statistic found it highly probable that a correlation existed between a change in illumination and a change in behavior.

The reader who wants to know what the experimental finding was must still ask another question, "What was the magnitude of the correlation between the change in treatment condition and the change in the dependent variable?". Somehow many researchers do not seem to realize that a t test is also a test of the significance of the correlation between treatment change and change in the dependent variable. Of course, correlation also refers to shared variance. R2, the percent of variation which is shared between the independent variables and the dependent variable happens to be the mathematical square of the correlation or multiple correlation as the case may be. It must be emphasized that analysis of variance, a procedure which sets out to separate the variation which is shared between the independent and dependent variables, from the variation in these variables which is not, is definitely concerned with determining the amount of shared, explained, or caused variation, e.g. R2. A researcher who attempts to communicate the practical significance of his experiment, but does not understand how close he is to understanding causality at the experimental level, must experience an uncomfortable moment while searching for the words (level 2).

A comparison of the researcher's interpretation without knowledge of R2, with an interpretation based on knowledge of R2 will clearly point

to the value of R2. The author states in conclusion that "The results of this experiment tend to confirm previous work with scanning light sources in that motor responses are better sustained under scanning light than under steady light." I submit this is an incomplete statement. What does "better sustained" mean? Let's look at the R2 associated with the correlation between the change in illumination source and the change in motor response. Using Equation 1.1

$$R2 = \frac{4.05^2}{6 + 4.05^2} = .73217$$

R2 for this experiment is .73. Seventy-three percent of the variation in motor response that occurred during the experiment is attributable to the experimental effect of scanned light! The reader now knows the magnitude of the experimental effect, and he knows that 27% of the variation was due to causes still unmeasured by the researcher. R2 has also given the reader a means of evaluating the significance of the researcher's knowledge. With this knowledge the reader might even be interested in studying the experimental methodology apparatus, etc., and even funding further development of scanned lighting!

R2 has been described as the statistical description of causality which is at the heart of the goals of research. To understand R2 more fully, one must also realize that it is an index expressing the degree to which a regression equation has determined relationships between a dependent variable and one or more independent

variables. The statistical language of regression equations enables the modern statistician to write specific analysis of variation equations for specific research questions.

Again, the purpose of statistics is related to determining how much variation that occurs among the independent variables and the dependent variables is correlated as well as the alpha level significance of the correlation. This purpose can be concealed in short cut formulae applications such as the t test where the basic regression equation has been concealed. To illustrate this, we can again look at the statistics which were used in the scanning light project. This project used a t test for repeated observations on the same subjects, but if we look at the author's definition of research design, we see that this ready-made technique only partly reflects his design. It is possible that relationships within his experimental design, but which remained outside of his statistical analysis could change the statistical finding. Unless the full experimental model is reflected by the statistical treatment, a conservative reader could not feel secure in interpreting the statistical analysis.

The researcher expressed his research design as follows:

FIGURE 1

Conducted Research Design

	Steady Light	Scanning Light
Subjects group A	First	Second
Subjects group B	Second	First

The above design was described as a counterbalanced design in which each participant experienced both light levels. The two levels or types of illumination were experienced by each group in a counterbalanced order so that learning or practice effects would be theoretically cancelled out. By pooling the data in a statistical simplification such as the t test which was used, the researcher analyzes his data as if he conducted a different experiment. The experiment which he reports in his analysis does not map the Figure 1 design. The reported experiment is presented in Figure 2.

FIGURE 2

Reported Experimental Design

Steady Light	Scanning Light
Subject 1	Subject 1
etc	etc
Subject n	Subject n

t = 4.05, df = 6, alpha = .01

(R2 = .73)

A complete analysis of the actual experiment looks very different from the crunched analysis presented in Figure 2. A full statistical equation written for the Figure 1 design produced the following results.

FIGURE 3

Summary of Analysis of Variance

Source	SS	DF	MS	R2	F	ALPHA
Factor A Down	.00	1	.00	.00	.00	1.0000
Factor B Across	11.52	1	11.52	.72	26.59	.0007
Interaction	.10	1	.10	.01	.22	.6523
Within cell	4.33	10	.43			
Full Model	11.62	3		.73	8.94	.0039
TOTAL	15.95	13				

Note that Figure 3 is a standard ANOVA Table with the exception that R2 has been added and the facts associated by the concealed regression equation full model are presented.

Figure 3 reports an R2 of .73 just as we computed with Equation 1.2 for the t test version of the scanning light experiment. But we can answer from knowledge several questions which were muted by the t test. Questions such as:

1) Was there a practice effect or learning effect?

   Answer: Factor A (Down) found no correlation between the exposure sequence and behavior. There was no practice effect for the counterbalanced design to cancel out. If there had been a practice effect, would it really have cancelled out with small numbers of very complicated subjects? How would a real practice effect with unequal amounts have affected the R2 reported in the t test? The point is, why should the reader be forced to guess? Why not do the statistics?

2) What percent of the variation in the Full Model was attributable to the correlation between change in treatment and behavior change?

   Answer: 72 percent. 1 percent of the 73 per in the Full Model was attributed to another source.

3) What was the alpha level associated with the R2?

   Answer: Not only was it less than .01 as reported with the t test, it was .0007.

The particular model presented in Figure 3 is fully documented by Winer (1962) Chapter 7. Because the research design presented in Figure 1 is a standard design, it is unnecessary to write out the regression equation behind it. However, it is important to note that had more knowledge been involved in the experimental design, a unique equation could be provided so that when the researcher got to the point of communicating what his experiment accomplished, he could statistically talk about what happened rather than point to mean differences.

As far as reporting mean differences, which seems to be some kind of ritualistic zoomorph, you could as well report any points on parallel regression lines that pass over the same point on the x coordinate. A two group model is only a special case where there is only one point on the x coordinate to pass over, e.g. the group membership point. The problem is, that by reporting on the significance of the distance between such points to the exclusion of R2, it is obvious that the researcher has missed the point.

The magnitude of the experimental effect produced in the technically excellent Scanning Light experiment has exciting implications. The excitement generated by an R2 of .72 will encourage thought given to industrial and educational applications and further research in the perceptual area. At the statistical level of causality, the R2 associated with the Scanning Light experiment implies that the researcher knows a great deal about what is causing the behavior he is investigating. The excitement of an experiment is conveyed through the R2 associated with each hypothesis the researcher tests.

In attempting to convert the front line scientist to a belief in R2, statistical causality, and an awareness that there is a regression equation behind every great experiment, the statistical analysis of the Scanning Light experiment was unmercifully used as an example of the many technically excellent experiments which stand to lose if their statistical counterpart is not complete. I owe a debt of gratitude to Mr. W. Ramage of the Westinghouse Behavioral Research Department for the good humored acceptance of the review of the statistical aspects of his exemplar project. Hopefully, R2 has found a home at Westinghouse. Who knows, perhaps one day we will say, "You can be sure, if it's 1.00".

REFERENCES

Friedman, H., Magnitude of Experimental Effect and Table for its Rapid Estimation. Psychological Bulletin, 1968, 70, 245-251.

McNeil, K. A., Kelly, J. and McNeil, J., Testing Research Hypotheses Using Multiple Linear Regression, Unpublished text.

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

Ramage, W. W., Scanned Illumination and Driving. Westinghouse Research Laboratories, Research Report 73-1Y1-FRLIT-R3, 1974.

Simon, H. A., Spurious Correlation: A Causal Interpretation, In Causal Models in the Social Sciences, edited by Blalock, H. R., Aldine-Atherton Press, Chicago, 1971.

Winer, B. J., Statistical Principles in Experimental Design, McGraw-Hill, 1962.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 6, No. 1, 1974

CLASSICAL ANALYSIS OF VARIANCE OF COMPLETELY WITHIN-SUBJECT FACTORIAL DESIGNS USING REGRESSION TECHNIQUES

Thomas W. Pyle
Eastern Washington State College

The use of regression analysis in analyzing data obtained from Between-subject experimental designs is well-documented and easily obtainable (e.g., Bottenberg and Ward, 1963; Jennings, 1967; Williams, 1970; Ward and Jennings, 1973). In addition some texts cover the classical analysis of variance (ANOVA) of single-factor repeated measurements designs (e.g., Kelly, Beggs, McNeil, Eichelberger, and Lyon, 1969), and an identification has also been made of regression models which yield F ratios (for Main and Interaction effects) equivalent to those obtained by standard computing formulas for split-plot factorial and simple hierarchical designs (Pyle, 1973). However, documentation for the standard ANOVA of completely Within-subject factorial designs using regression analysis is not available as far as this author is aware. The purpose of this article is to provide such documentation by means of an example. The regression solution for such problems is not the most efficient way to generate the desired F ratios, at least not the F ratio for interaction. However, the information will be valuable for the didactic purpose of understanding the models, and some of the assumptions underlying the analysis of this type of design will become apparent.

The following example assumes that the dependent variable of interest was collected for three independent Ss under two levels each of independent variables A and B. Each S receives all levels of all independent variables and therefore contributes four scores for a total of 12 observations. Let Y_{ijk} refer to the score made by the i^{th} individual at the j^{th} level of A and the k^{th} level of B. For this example, i = 1,2,3; j = 1,2; and k = 1,2. The design may be represented as follows:

	B_1		B_2	
	A_1	A_2	A_1	A_2
S_1	Y_{111}	Y_{121}	Y_{112}	Y_{122}
S_2	Y_{211}	Y_{221}	Y_{212}	Y_{222}
S_3	Y_{311}	Y_{321}	Y_{312}	Y_{322}

(1)

For the purposes of the example, assume the following scores have been obtained:

	B_1		B_2	
	A_1	A_2	A_1	A_2
S_1	1	3	4	4
S_2	2	2	5	4
S_3	3	1	6	6

In order to obtain the correct F ratios for the A and B Main Effects, two different Full models must be constructed. For the A Main Effect, the elements of the criterion vector, Y_A, will consist of the sum (or any linear transformation on the sum) of each S's scores at each level of B for a given level of A. Note that each S will have two scores in Y_A: one obtained under treatment A_1, and the other obtained under treatment A_2. For example, for S_1 his two scores in Y_A are $Y_{111} + Y_{112}$ (1 + 4 = 5) and $Y_{121} + Y_{122}$ (3 + 4 = 7). Let Equation (2) be the Full Model for the A Main Effect:

$$Y_A = a_1X_1 + a_2X_2 + b_1P_1 + b_2P_2 + b_3P_3 + E_2 \qquad (2)$$

where:

Y_A is the vector of criterion scores (defined above);

X_j(j=1,2) contains a one if the corresponding element in Y_A was made by a person receiving Treatment A_j; zero otherwise;

P_i(i =1,2,3) contains a one if the corresponding element in Y_A was made by person i; zero otherwise;

a_1, a_2, b_1, b_2, and b_3 are weighting coefficients which minimize the length of the residual (error) vector;

E_2 is the error vector.

For this example the R^2 for Model (2) is 0.50943. When the restriction $a_1 = a_2 = a_0$ (a common value) is imposed on (2), the following model obtains:

$$Y_A = a_0U + b_1P_1 + b_2P_2 + b_3P_3 + E_3 \qquad (3)$$

where:

Y_A, P_1, P_2, P_3 are defined as in (2);

U is the unit vector;

a_0, b_1, b_2, b_3 are weighting coefficients which minimize the the length of E_3;

E_3 is the error vector.

The R^2 associated with (3) is 0.49057. The $\underline{F}$ ratio comparing Models (2) and (3) may be obtained from the following formula (Bottenberg and Ward, 1963):

$$F = \frac{(R_F^2 - R_R^2)/df_1}{(1 - R_F^2)/df_2}$$

where:

R_F^2 is the R^2 associated with the Full model [Model (2)];

R_R^2 is the R^2 associated with the Restricted model [Model (3)];

df_1 = the number of linearly independent vectors in the Full Model (m_1) minus the number of linearly independent vectors in the Restricted model (m_2);

df_2 = the dimension of the vectors (N) minus m_1.

Note that $m_1 = 4$ and that $m_2 = 3$. N, for this example, is six, and the obtained value of $\underline{F}$ is 0.077. This is precisely what one obtains for the A main effect using traditional computing formulas (e.g., Myers, 1972, Chap. 7).

The analysis for the B Main Effect is structurally quite similar; however, the values of the criterion vector are not the same as for the Full Model for the A Main Effect and some of the predictor vectors are also changed.

Let the elements of Y_B consist of the sum (or any linear transformation on the sum) of each S's scores at each level of A for a given level of B. Using S_1 as an example again, his two scores in Y_B will be [from (1)] $Y_{111} + Y_{121}$ and $Y_{112} + Y_{122}$ (i.e., 1 + 3 = 4 and 4 + 4 = 8, respectively). Define Model 4 as:

$$Y_B = a_1X_1 + a_2X_2 + b_1P_1 + b_2P_2 + b_3P_3 + E_4 \quad (4)$$

where:

Y_B is defined as above;

X_k(k = 1, 2) contains a one if the corresponding element in Y_B was made by a person receiving Treatment B_k; zero otherwise;

P_i(i = 1, 2, 3) contains a one if the corresponding element in Y_B was made by person i; zero otherwise;

a_1, a_2, b_1, b_2, and b_3 are weighting coefficients which minimize the length of E_4;

E_4 is the error vector.

If the restriction that $a_1 = a_2 = a_0$ (a common value) is imposed on Model (4), Model (5) obtains:

$$Y_B = a_0U + b_1P_1 + b_2P_2 + b_3P_3 + E_5 \quad (5)$$

All of the predictors in (5) have been defined previously. The R^2 for Model (4) is 0.92375; the R^2 for Model (5) is 0.07625; $df_1 = 1$, $df_2 = 2$ and the obtained value of F is 22.23. Again, this is exactly the same value of F one obtains using traditional computing formulas. Notice also that the values of df_1 and df_2 (as defined above) are the same as those obtained by traditional procedures. It can be shown that this will, in general, be true.

A x B Interaction

In order to obtain the df's and F ratio for the AxB interaction which correspond to those obtained by traditional computing formulas, one must create a new Full Model. To obtain this "Full" model, another model must first be created, which will be referred to as the Largest model. A restriction will be imposed on the Largest model that will yield the Full model which allows for AxB interaction. Then another restriction will be placed on the Full model which will not allow for AxB interaction. These last two models are the ones which are compared using the F ratio.

In order to obtain the "Full" model, one must first create the Largest model, which is defined as follows:

$$Y = \sum_{k=1}^{B} \sum_{j=1}^{A} \sum_{i=1}^{n} d_{ijk} X_{ijk} + E_6 \qquad (6)$$

where:

Y is the criterion vector. The elements of Y are the raw scores from the original data matrix [see (1)]. The dimension of Y is 12;

X_{ijk} contains a one if the corresponding element in Y was made by person i while receiving Treatment combination A_jB_k;

The d_{ijk}'s are weighting coefficients which minimize the length of E_6;

E_6 is the error vector.

For the largest model there are 12 linearly independent predictors; since the dimension of Y is also 12, E_6 is null and $R^2 = 1.0$.

At this point on assumption must be made: there is no AxBxS interaction. When this restriction is imposed on the Largest model (Model 6), the Full model for the AxB interaction hypothesis obtains.

In general, there will actually be (A-1)(B-1)(n-1) restrictions imposed on the largest model. For the specific example under consideration, two restrictions will be made; these two restrictions constitute what is meant by the term "no AxBxS interaction":

$$(d_{111} - d_{211})-(d_{121} - d_{221}) = (d_{112} - d_{212})-(d_{122} - d_{222}) \quad (7)$$

$$(d_{111} - d_{311})-(d_{121} - d_{321}) = (d_{112} - d_{312})-(d_{122} - d_{322}) \quad (8)$$

Equation (7) can be solved in terms of d_{111} and Equation (8) can be solved for d_{311}. The actual restrictions imposed on Model (6) are (7') and (8'):

$$d_{111} = d_{112} - d_{212} - d_{122} + d_{222} + d_{211} + d_{121} - d_{221} \quad (7')$$

$$d_{311} = -d_{212} + d_{222} + d_{211} - d_{221} + d_{321} + d_{312} - d_{322} \quad (8')$$

After substituting the expressions on the right side of the equations in (7') and (8') for d_{111} and d_{311} in Model (6), and after collecting vectors that have like coefficients, Model (9) obtains:

$$\begin{aligned} Y = {} & d_{112}(X_{111} + X_{112}) + d_{212}(X_{212} - X_{111} - X_{311}) \\ & + d_{122}(X_{122} - X_{111}) + d_{221}(X_{221} - X_{111} - X_{311}) \\ & + d_{211}(X_{111} + X_{211} + X_{311}) + d_{121}(X_{111} + X_{121}) \\ & + d_{222}(X_{111} + X_{311} + X_{222}) + d_{321}(X_{311} + X_{321}) \\ & + d_{312}(X_{311} + X_{312}) + d_{322}(X_{322} - X_{311}) + E_9 \end{aligned} \quad (9)$$

Model (9) is rather complicated, but its important characteristic is that it does not allow for AxBxS interaction. It <u>does</u> allow for AxB interaction, however.

The restrictions for no AxB interaction can be formulated verbally as follows: given Model (9), the difference between the average predicted values for A_1B_1 and A_2B_1 equals the difference between the average predicted values for A_1B_2 and A_2B_2. For example, from (9) the average predicted value for cell A_1B_1 is as follows:

$$\left[(d_{112} - d_{212} - d_{122} - d_{221} + d_{211} + d_{121} + d_{222}) + (d_{211}) + (- d_{212} - d_{221} + d_{211} + d_{222} + d_{321} + d_{312} - d_{322})\right]/3.$$

The expression in the first set of parentheses is the predicted value for Person 1 at A_1B_1. The expression in the second set of parentheses is the predicted value for Person 2 at A_1B_1; and the last expression is the predicted value for Person 3 at A_1B_1. The predicted values for the other cells are obtained in an analogous manner.

Although the initial restriction is rather involved, it eventually reduces to the following expression:

$$d_{211} = d_{212} - d_{222} + d_{221}$$

When the above restriction is imposed on Model (9), we obtain the model which does not allow for AxB interaction, Model (10):

$$\begin{aligned} Y = {} & d_{112}(X_{111} + X_{112}) + d_{212}(X_{212} + X_{211}) + d_{122}(X_{122} - X_{111}) \\ & + d_{221}(X_{221} + X_{211}) + d_{121}(X_{111} + X_{121}) + d_{222}(X_{222} - X_{211}) \\ & + d_{321}(X_{311} + X_{321}) + d_{312}(X_{311} + X_{312}) + d_{322}(X_{322} - X_{311}) \\ & + E_{10} \end{aligned} \qquad (10)$$

The R^2 for Model (9) is 0.93418; for Model (10), R^2 = 0.93165. For Model (9), m_1 = 10; for Model (10), m_2 = 9. Therefore, df_1 = 1, and since the dimension of Y is 12, df_2 = 2. The value of $\underline{F}$ is 0.077. If one calculates the $\underline{F}$ ratio for the AxB interaction effect using traditional computing formulas, exactly the same value is obtained (within round off error).

For the purpose of generality it will be shown that the denominator degrees of freedom (df_2), obtained by the procedure illustrated above, is equal to (A-1)(B-1)(n-1), which is the general formula for df_2 (e.g., see Myers, 1972, p. 188). Recall that $df_2 = N-m_1$, where N is the dimension of the vectors and m_1 is the number of linearly independent predictors in the Full Model. For the Full model used to analyze the AxB interaction, N will, in general,

be equal to the product nAB. The Largest model will have nAB linearly independent predictors. When the restriction of no AxBxS interaction is imposed on the Largest model, the Full model for AxB interaction obtains. When one imposes the restriction of no AxBxS interaction on the Largest model, the number of independent restrictions imposed is (A-1)(B-1)(n-1), and m_1, for the Full model, is [nAB - (A-1)(B-1)(n-1)].

Then,

$$df_2 = N - [nAB - (A-1)(B-1)(n-1)];$$

$$df_2 = nAB - [nAB - (A-1)(B-1)(n-1)];$$

$$df_2 = (A-1)(B-1)(n-1).$$

Discussion

In conclusion, several points can be made. (a) Completely Within-subject factorial designs <u>can</u> be analyzed in the multiple regression framework. The obtained dfs and <u>F</u> ratios are equivalent to those obtained by traditional computing formulas. (b) Three different "Full" models are necessary to obtain the correct <u>F</u> ratios. It should be noted that this is a distinctly different state of affairs in comparison to a completely Between-subjects factorial design (see Jennings, 1967), but is somewhat similar to the analysis for the split-plot design, i.e., a design that has one Between-subjects variable and one Within-subjects variable (see Pyle, 1973). The regression approach makes it quite clear that different "Full" models are associated with different aspects of the analysis. This should be of distinct pedagogical interest. (c) It is particularly obvious that in order to obtain the correct <u>F</u> ratio for the AxB interaction one must assume that there is no AxBxS interaction. Similar restrictions are also made for the analysis of the main effects. For example, Model (2) does not allow for AxS interaction. (d) An analysis of the AxB interaction using the regression approach is rather impractical in the sense that a

great deal of algebraic manipulation is required to obtain both the Full and Restricted models. It should be noted that the complexity increases as the number of subjects increases.

References

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

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

Kelley, F. J., Beggs, D. L., McNeil, K. A., Eichelberger, T., & Lyon, J. Research design in the behavioral sciences: Multiple regression approach. Carbondale, Illinois: Southern Illinois University Press, 1969.

Myers, J. L. Fundamentals of experimental design. (2nd ed.) Boston: Allyn and Bacon, 1972.

Pyle, T. W. The analysis of split-plot and simple hierarchical designs using multiple linear regression. Multiple Linear Regression Viewpoints, 1973, 4(3), 1-9.

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

Williams, J. D. A regression approach to experimental design. Journal of Experimental Education, 1970, 39(1), 83-90.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 5, No. 1, 1974

USE OF MULTIPLE LINEAR REGRESSION IN ANALYSIS OF INTELLIGENCE TEST SCORE CHANGES FOR VISUALLY HANDICAPPED ADULTS

Dr. Theodore L. Gloeckler

Multiple linear regression was used to analyze complex data assessing longitudinal changes in IQ test performance of visually handicapped adults. Results indicated: (1) patterns of performance similar to those found in sighted populations, and (2) no influence on IQ changes by a variety of ontological factors.

Although it has been traditionally assumed that patterns of intelligence test performance of visually handicapped persons are not different from those of the sighted (Newland, 1963), there is little supporting evidence in the literature. Longitudinal studies with the blind or the visually handicapped have been limited in number and these few have in turn been limited by population size, by instrumentation, and by subjects' (Ss) chronological age spans. Therefore this study was undertaken to examine longitudinal pre-post intelligence test performances of a group of visually handicapped adults.

Six major hypotheses were posited. Additional hypotheses related to subtest performance and other variables are not reported here. The six central hypotheses were:

Hypothesis 2: Changes in pre-post intelligence test scores are not significantly related to chronological age at initial testing.

Hypothesis 3: Changes in pre-post intelligence test scores are not significantly related to chronological age at the time to posttest administration.

Hypothesis 1: For the total sample of visually handicapped adults there will be a significant relationship between initial test scores and subsequent score changes.

Hypothesis 4: There is no significant relationship between changes in pre-post intelligence test scores and the length of the time interval between the two test administrations.

Hypothesis 5: There is no significant relationship between the age of onset of the visual impairment and changes in intelligence test performance as measured by the pre-post verbal IQ scores.

Hypothesis 6: For the total sample of visually handicapped adults there will be a significant relationship between the degree of the disability and pre-post verbal IQ score differences.

Method

The study was an extension of an earlier work on vocational success of the visually handicapped by Scholl, Bauman and Crissey (1969). Biographical and historical data had been gathered on 929 visually handicapped Ss in six states and approximately one-third of these were retested with instruments which had been originally used earlier when these Ss were completing their public and/or private school education. These Ss had taken one of the Wechsler scales: Wechsler-Bellevue I, Wechsler-Bellevue II, or the Wechsler Adult Intelligence Scale. From this population 159 Ss were selected. Criteria included the availability of data central to the investigation and that Ss had been administered the same form of the Wechsler scale at both pre and post testings.

The study was limited by several factors. First were the usual restrictions resulting from ex post facto research. In addition, there was wide variability in the available data. These included the following: (1) Whereas each S was retested at approximately the same time (Spring, 1968), the time of

initial testing varied widely. As a result, test-retest time intervals ranged between 4 and 19 years. (2) Chronological ages of the Ss ranged from 16 to 22 years on the pretest and from 23 to 40 years on the posttest. (3) IQ ranges were from 57 to 141 on the pretest and from 53 to 145 on the posttest. (4) Age on onset of the visual impairment ranged from birth (congenital blindness) to 17 years for the adventitiously blinded. The degree of visual acuity of the Ss ranged from 20/30 in the better eye to total blindness in both eyes.

In addition there were wide ranges within other variables including the amount and kind of educational experience, socioeconomic indeces of Ss and their parents, measures of job success, amount of occupational training, general health, presence of additional disabilities, financial and counseling assistance received from agencies, and travel ability of the Ss.

This wide variability of the data plus the fact that the study dealt with a series of continuous variables resulted in the selection of multiple regression analysis as the main statistica method used in the study. This method, as described by Kelly Beggs, McNeil and Eichelberger (1969) provided the necessary flexibility in dealing with these complex data. In addition, the method had the capability of providing analyses of variance and analyses of covariance at a more robust level than the traditional F test. Of these six major hypotheses only No. 1 was not tested with multiple regression analysis. For this initial hypothesis Person Product Moment correlation and a Chi-Square were used.

Results

Table 1 presents the chi-square results used for Hypothesis 1. With the Ss stratified into subgroups on the basis of initial IQ scores, an inspection was made to determine how many Ss moved up or down from one IQ group to another.

The chi-square test was made on these frequencies. As indicated in the table, an increasingly larger percentage of Ss were able to improve their scores sufficiently to move to the next higher IO range as initial IO scores increased. In addition, a test-retest correlation resulted in a .88. On the bases of these two indicators Hypothesis 1 was accepted to show subsequent score changes significantly related to initial IO.

The remaining hypotheses were all tested with multiple linear regression and Table 2 gives these results. Hypothesis 2 was rejected and it was found that there was a significant relationship between chronological age at the time of initial testing and subsequent score changes. As a result of these two findings both initial IO and initial CA were covaried on subsequent models.

Hypotheses 3, 4, and 5 were accepted. No significant relationships were found between longitudinal IO score changes and either chronological age at the time of retest, time interval between the two test administrations, or the age of onset of the visual impairment. Hypothesis 6 was rejected indicating no significant relationship between the degree of visual impairment and subsequent score changes.

An additional 30 models were tested although no formal hypotheses were made for them in the study. Results of these tests ruled out any significant relationships between score changes and socioeconomic indices of Ss and their parents, various educational factors, additional training, health, travel ability, measures of occupational success, and counseling with Ss. Only one variable (Table 2, Hypothesis A), money spent by rehabilitation agencies on the Ss, was found to be related to IO score change.

Discussion

In general, the results indicated that changes in the verbal intelligence scores, over time, of visually handicapped adults are unrelated to a wide variety of ontological data. The relationship of such changes to financial assistance was probably an artifact of the selection process of rehabilitation agencies which tend to select those with most promising abilities (e.g. high IQ scores) to receive financial assistance (Scott, 1969). Thus financial assistance is probably related to initial IQ which, as the data have indicated, is in turn related to score changes.

In summary, changes in the IQ test performance of these visually handicapped adults, over time, appear to be similar to those changes observed in other, sighted populations.

References

Kelly, F. J., Beggs, D. L., McNeil, K. A., Eicherberger, T., and Lyon, J. Multiple regression approach. Carbondale, Ill., Southern Illinois University Press, 1969.

Newland, T. E. Psychological assessment of exceptional children and youth. In W. M. Cruickshank (Ed.) Psychology of Exceptional Children and Youth. Englewood Cliffs, N. J.: Prentice Hall, 1963.

Scholl, G. T., Bauman, M. K., and Crissey, M. S. A Study of Vocational Success of Groups of the Visually Handicapped. Ann Arbor, The University of Michigan, 1969.

Scott, R. A. The Making of Blind Men. New York: Russell Sage Foundation, 1969.

TABLE 1

Frequencies of Subjects Moving
Between IQ Score Ranges

IQ Group	Pretest N	Moved Up N	Moved Up %	Moved Down N	Moved Down %	Chi-Square df	Chi-Square Mean	Chi-Square Sig.
50-84	24	3	12.5	1	.04	1	.66	NS
85-99	42	25	58.5	2	.02	1	13.6	.001
100-114	54	42	77.7	1	.03	1	16.3	.001
115-plus	38	0	0.0	1	.02	-	-	-

TABLE 2

Models, F ratios, and R^2 for Predicting Posttest Scores
Predictor Variabled

Models and Explanations	Models	R^2	df	F	p
Hypothesis 2: An analysis of covariance (Initial IQ covaried) to determine if initial chronological age is related to IQ score change.					
Model 1: $Y_2 = a_0u + a_1X_1 + a_3X_3 + E$	Full	.80	1/157	8.47	.04
Model 2: $Y = a_0u + a_1X_1 + E$	Rest.	.80			
Hypothesis 3: An analysis of covariance (Initial age covaried) to determine if retest chronological age is related to IQ score change.					
Model 1: $Y_2 = a_0u + a_1X_1 + a_4 X_4 + E$	Full	.80	1/157	8.47	.04
Model 2: $Y = a_0u + a_1X_1 + E$	Rest.	.80			

TABLE 2 (CONTINUED)

Models and Explanations	Models	R^2	df	F	p
Hypothesis 4: An analysis of covariance (Initial age and IQ covaried) to determine if test-retest time interval is related to IQ score change.					
Model 1: $Y_2 = a_0u + a_1x_1 + a_3x_3 + a_4x_4 + E$	Full	.80	1/157	1.34	.24
Model 2: $Y_2 = a_0u + a_1x_1 + a_3x_3 + E$	Rest.	.80			
Hypothesis 5: An analysis of covariance (Initial age and IQ covaried) to determine if age of onset in the better eye is related to IQ score change.					
Model 1: $Y_2 = a_0u + a_1x_1 + a_3x_3 + a_5x_5 + E$	Full	.80	1/157	0.02	.86
Model 2: $Y_2 = a_0u + a_1x_1 + a_3x_3 + E$	Rest.	.80			
Hypothesis 6: An analysis of variance (Initial age and IQ covaried) to determine if degree of disability in the better eye is related to IQ score change.					
Model 1: $Y_2 = a_0u + a_1x_1 + a_3x_3 + a_6x_6 + E$	Full	.80	1/157	0.15	.69
Model 2: $Y_2 = a_0u + a_1x_1 + a_3x_3 + E$	Rest.	.80			
Hypothesis A: An analysis of covariance (Initial age and IQ covaried) to determine if amount of money spent by agency is related to IQ score change.					
Model 1: $Y_2 \neq a_0u + a_1x_1 + a_3x_3 + a_7x_7 + E$	Full	.80	1/103	4.32	.04
Model 2: $Y_2 = a_0u + a_1x_1 + a_3x_3 + E$					

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 6, No. 1, 1974

DINNER WITH DR. EARL JENNINGS

Isadore Newman
The University of Akron

At the last A.E.R.A. convention, I had the opportunity to meet and have dinner with Dr. Earl Jennings. During dinner, a variety of topics were discussed, which may be of interest to the Viewpoint audience. Some of the ideas are not fully developed since this was an informal meeting, however, they might provide fuel for thought. Just a few of these ideas will be presented here.

Since one of my interests is shrinkage, we discussed its appropriateness. Earl's response was somewhat surprising. "Shrinkage is O.K. if you're interested in interpreting R^2." He then went on to say that he believes it is more useful to interpret "error mean square." This value can be calculated for any model by

$$\frac{\text{Error Sum of Squares}}{n - k}$$

where n is the dimension of the vectors and k is the number of linearly independent predictor vectors. Under certain assumptions the error mean square is an unbiased estimate of common variance. Even if the assumptions are not met the quantity has a considerable intuitive appeal. Notice that for any model which has a unit vector in the space of the predictors the quantity

$$\frac{\text{Error Sum of Squares}}{n}$$

is the variance of the observed errors.

---

* I would like to publicly thank Earl Jennings for his very helpful comments.

$$[\text{Error Sum of Squares} = (1-R^2)n\sigma^2 y]$$

The difficulty with this quantity and R^2 is that R^2 approaches 1 and ESS approaches zero as k approaches n. However the error mean square does not necessarily approach zero as k is increased. The primary disadvantage of the error mean square is that its value depends on how Y is measured. For example if Y contains annual budget deficits for the last 20 years the error mena square for a particular model might be 50 million dollars. Although the value sounds enormous the model probably would predict next year's deficit more accurately than is currently the case.

Another example Earl used was to assume one was interested in predicting grade point average (y) from a variety of predictor variables. Also assume that the R^2 was about .4. One could increase the R^2 to .6 by taking out the middle scores in the distribution. This would increase the variability, thereby increasing the correlation. Earl said that he actually tried this with some data and found that the weights for the predictor variables did not change and neither did the error mean square, even though the R^2 changed considerably.

For this reason and others, he feels that the error mean square is more stable and at least as good an estimate of the equation's accuracy of prediction, as is the R^2. (This argument is further elaborated on in appendix B in Ward & Jennings Regression Textbook.) I found this rather thought provoking.

Another subject that was briefly discussed was which loss functions should be minimized in the predictor equation. In other words, the equations are calculated to minimize E

$$E=y-\hat{y}$$

The loss functions talked about were:

(1) Mean Square Error

(2) Absolute Error

(3) Hit-Miss (one-zero) Error

The most commonly used is the squared error loss function. For this function the assumption is made that the innacuracy in prediction is equal to the squared error of prediction. The purpose is to minimize

$$(y-\hat{y})^2.$$

Most of the statistical literature uses this function in their calculations. The major reason given for this is that the mathematics that results from such a function are easier to deal with in more sophisticated calculations.

The major problem with using the mean-square error function is that it penalizes large errors considerably more since it squares the errors. However, as long as scores are not very extreme, this makes little difference, and most data in educational psychology is bounded, so that enormous differences are generally not encountered (i.e., test scores generally range from 0 to 100, etc.)

The absolute error (/E/) loss function attempts to minimize the magnitude of the absolute difference between $y - \hat{y}$. The major advantage of using this as a loss function is that large error differences are not overly weighted as they are in the mean-square loss function. A disadvantage is that when using this error function, the computations that are required for fitting a regression line are more complicated. This probably is the reason that it is not used very often.

The 0-1 loss function is appropriate to use when one is interested in minimizing the error of any miss in prediction, for example, it may be most appropriate to use if one is interested in predicting the accuracy of artillary fire, and all misses regardless of how close or far are considered a miss and any hit is considered a hit.

The point is that the least-square solution (minimizing one mean square loss function) is not the only solution that can be used in model solving. However, it is probably so frequently used because it is very robust, especially when the data is bounded. And the mathematics are much easier to manipulate.

One of the major things that I personally derived from the conversation is reinforcement of the idea that one should not become functionally fixed by particular techniques.

The research question should dictate the procedures, whether it be R^2, error mean square, a particular loss function, etc.

P.S. Negative variance accounted for was also discussed, but my recollection of it was not clear enough for presentation in this note.

The workshop on regression being presented by Joe Ward is described below. This workshop has the advantage of being based on an individual approach. There will be a list of objectives and participants will be able to proceed at their own pace. This makes it appropriate for both beginning and advanced work in regression.

Isadore Newman

Bureau of Conferences and Institutes

Division of Continuing Education
University of Colorado
130 Academy Building, 970 Aurora Avenue
Boulder, Colorado 80302
Telephone (303) 443-2211, extension 6485

TO: Educational Researchers, et. al.

FROM: Bureau of Conferences and Institutes

SUBJECT: Seminar on Applied Linear Models in Educational Research, July 8 - 12, 1974, College Inn, Boulder, Colorado

DATE: March 8, 1974

The Seminar on Applied Linear Models in Educational Research will be under the direction of Joe H. Ward, Jr. from the Air Force Human Resource Laboratory. Mr. Ward is well known for his work in the field of linear models, having presented similar programs throughout the nation.

This seminar is designed to develop the capability to use the general linear model as an approach to the formulation and analysis of research problems. Because of the rapid improvement of computer techniques for the systematic organization and analysis of data, researchers are presently unable to formulate problems for computer analyses that will yield answers.

The activities will be divided evenly between lecture-discussion and laboratory exercises related to the objectives. Concepts and exercises will be introduced systematically as they are required in the logical development of the materials.

Specifically, this seminar will enable participants to: Express a question of interest in a natural language form; formulate and defend a linear model containing parameters relevant to the question without conforming to experimental designs for which prescribed computational procedures are available; state the relations among the parameters of a model required to defend specified conclusions about the question; test hypotheses about the relations among the parameters of a linear model

The seminar will be held at the College Inn Center, 1729 Athens Street, Boulder Colorado 80302. The College Inn Center provides an opportunity for the participant to "live in" during the seminar, and also offers free-time recreation.

Rates are $13.00 plus tax per person per day for either single or double accommodations, including room and board. Do not send housing payment in advance.

The registration fee will be $125.00 payable to the University of Colorado. Complete the enclosed form and return it by July 1 to: Bureau of Conferences and Institutes, Academy 217, 970 Aurora, University of Colorado, Boulder, Colorado 80302 For further information call (303) 443-2211, extension 6485 at the above address.

William E. Wright
Conference Coordinator

WEW:hm
Enclosure

REGISTRATION APPLICATION

Please register me for the Seminar on Applied Linear Models in Educational Research, July 8 - 12, 1974:

Name:_____

Address:_____

City:_____ State:_____ Zip:_____

Institution:_____

_____

Please Reserve ☐ Single
☐ Double

Rates are $13.00 per person per day including room and board. Do not make housing payment in advance.

☐ Enclosed is a check for $125.00 registration fee.

☐ Please bill the institution shown above.

Mail this form and your check - made payable to the University of Colorado by July 1 to:

Bureau of Conferences and Institutes
University of Colorado
Academy 217
270 Aurora Avenue
Boulder, Colorado 80302

Southern Illinois
University at Carbondale

CARBONDALE, ILLINOIS 62901

Testing Center

May 2, 1974

Dear Multiple Linear Regression SIG Member:

At our recent annual meeting with AERA in Chicago the membership voted to increase the annual membership dues to $2.00. This action was taken in order to help defray the increase in costs realized by our publication, Viewpoints. The cost per page absorbed by the author will still remain at $1.00.

Those of you who have not yet paid your membership dues for the coming year may send a check to me for $2.00 and your name will be placed on our active membership list. I strongly encourage you to continue your membership in the SIG and I also encourage you to solicit memberships from any friends or colleagues that you think would benefit from membership in the SIG.

Send your dues to:

John T. Pohlmann
Student Affairs Research and Evaluation Center
Washington Square C
Southern Illinois University
Carbondale, IL 62901

Sincerely,

John Pohlmann
Secretary-Treasurer
Multiple Linear Regression SIG

JTP:mjc

Attention SIG Members

In order to be able to continue bringing you Viewpoints at the present minimal cost, we are asking that SIG members request libraries they are affiliated with to subscribe to Multiple Linear Regression Viewpoints at a subscription rate of $12.00 per year. Please send requests for subscriptions to either

John Pohlman
Student Affairs Research
& Evaluation Center
Washington Square C
Southern Illinois University
Carbondale, Ill. 62901

or

Isadore Newman
Research & Design Consultant
University of Akron
Akron, Ohio 44325

This will be of great help, both financially and in spreading interest in regression.

IT IS THE POLICY OF THE SIG=MULTIPLE LINEAR REGRESSION AND OF VIEWPOINTS TO CONSIDER FOR PUBLICATION ARTICLES DEALING WITH THE THEORY AND THE APPLICATION OF MULTIPLE LINEAR REGRESSION. MANUSCRIPTS SHOULD BE SUBMITTED TO THE EDITOR AS AN ORIGINAL, DOUBLE-SPACED TYPED COPY. A COST OF $1 PER PAGE SHOULD BE SENT WITH THE SUBMITTED PAPER. REPRINTS ARE AVAILABLE TO THE AUTHORS FROM THE EDITOR. REPRINTS SHOULD BE ORDERED AT THE TIME THE PAPER IS SUBMITTED AND 20 REPRINTS WILL COST $.50 PER PAGE OF MANUSCRIPT. PRICES MAY BE ADJUSTED AS NECESSARY IN THE FUTURE.

The University of Akron
Akron, Ohio 44325

BOOKS - SPECIAL 4th CLASS RATE

TABLE OF CONTENTS