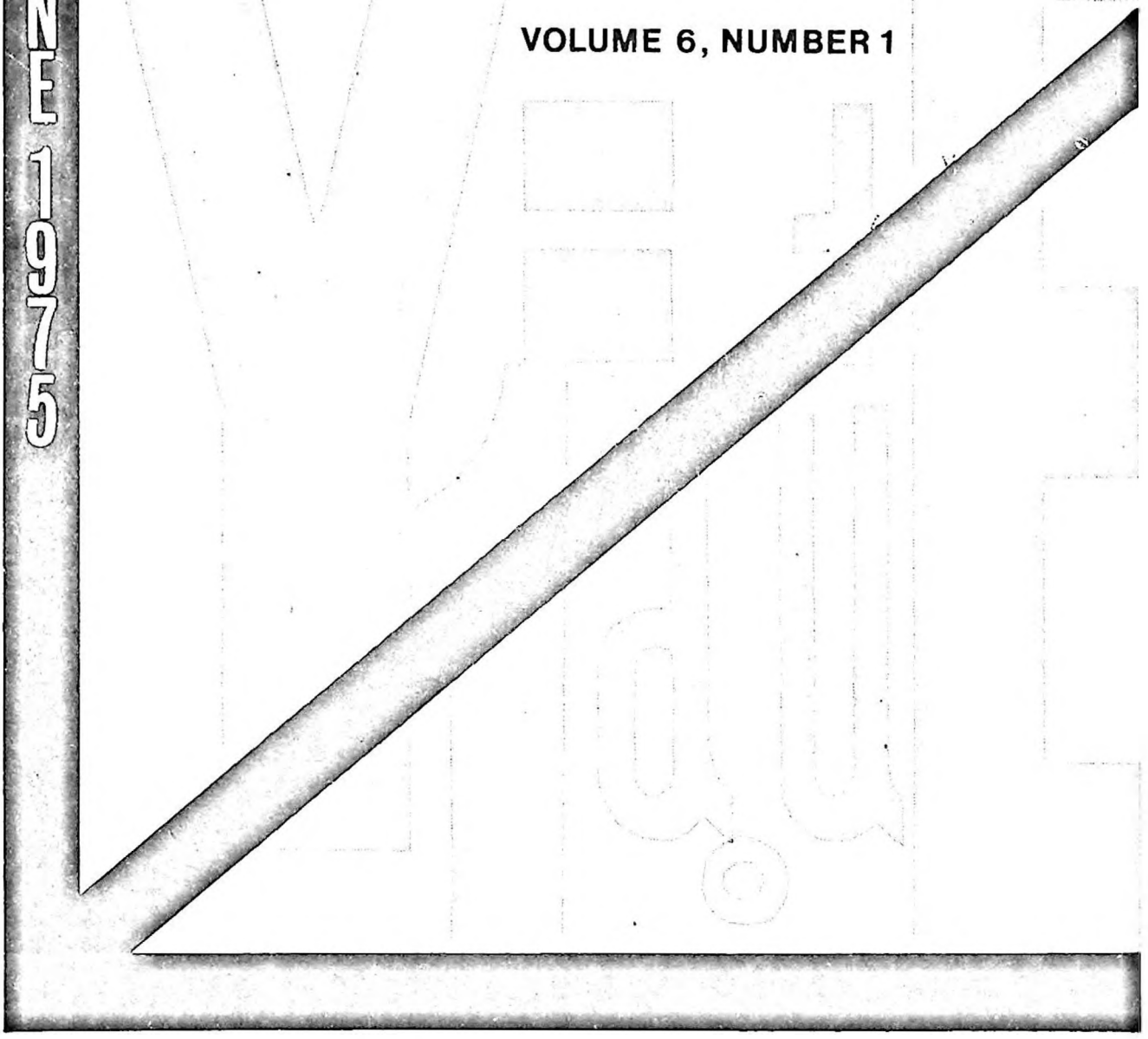
JUNE 1975

VOLUME 6, NUMBER 1



MULTIPLE LINEAR REGRESSION VIEWPOINTS

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

MULTIPLE LINEAR REGRESSION VIEWPOINTS

Chairman ..... John T. Pohlman, Testing Center, Southern Illinois University, Carbondale, IL 62901

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

Secretary and Chairman-elect ..... Michael McShane, Student Affairs Center, Southern Illinois University, Carbondale, IL 62901

Cover by ..... David G. Barr

TABLE OF CONTENTS

Dear SIG Members:

A NEW EDITORIAL POLICY

As you may or may not know, the special interest group on Multiple Linear Regression at the AERA meeting in Washington voted to initiate a new editorial policy on an experimental basis for next year. Since we feel that "Viewpoints" should allow our members to express their positions or ideas freely, an article will be published as is, without any editorial review, if the author so desires. This article will then be placed in a non refereed section of the journal.

The other option an author may exercise is to request that his paper be refereed. The editorial comments will be sent to the author for his benefit and he will be asked to rewrite in accordance with the suggestions before resubmitting his article. The rewritten article will then be published in the refereed section of the journal. The major drawback of this procedure is that it will take a little longer, but the author should benefit from the comments of the reviewers.

When you submit a paper, I would appreciate your informing me whether or not you would like your article reviewed. I would also like to remind you that a cost of $1 per page should be submitted with the paper. Reprints are also available from the editor if they are ordered when the paper is submitted. Twenty reprints will cost $.50 per page of manuscript.

Respectfully,

Isadore Newman

Isadore Newman, Ph.D.
Editor, Multiple Linear
Regression Viewpoints

IN/kaj

Multiple Linear Regression Viewpoints
Vol. 6, No. 1, 1975

Logical Steps in the Creation and Manipulation of Fixed X Linear Models

Earl Jennings
University of Texas at Austin

Joe H. Ward, Jr.
Air Force Human Resources Laboratory

The remarks in this paper were prepared for a meeting of a special interest group of AERA at the annual meeting in Washington, 1975. Presumably the views expressed are intelligible, although not necessarily agreeable, to members of the special interest group. Others who are interested will find these steps and the concepts that serve as a foundation for them elaborated and exemplified in Introduction to Linear Models, Prentice-Hall, 1973 by the authors.

Over a period of some years, the authors have been approached for advice on the formulation of linear models by individuals with varying degrees of expertise. It has been our experience that many of the difficulties encountered by such individuals can be traced to a lack of consideration of certain concepts and ideas that we have found helpful in model creation, manipulation, and interpretation. In the list which follows an effort has been made to identify a series of steps in the approximate order in which they ordinarily occur. Although all problem situations may not require detailed attention to each step, we are of the opinion that investigators who reach a point where they do not know what to do next generally have failed to consider one or more of the following steps:

1. State the question in "natural language." The question may contain technical terms in substantive areas but the use of technical terms of a statistical or mathematical nature should be avoided at this stage. For example, the investigator should not ask, "Are the regression slopes for methods A and B homogenous?" A preferable phrasing is, "Does the difference in performance between method A and B depend on aptitude," or perhaps, "Is the difference between method A and B the same at all levels of aptitude?"

2. Accept the idea that model creation and manipulation is a way of formalizing an argument. Identify the entities and the attributes associated with these entities in the problem situation. If at all possible, translate the question in 1 in such a way that specifiable relationships among a set of means (expected values) can be used as evidence concerning the answer to 1. By concentrating on the expected values, one can usually avoid the problems of non-estimability that characterize over parameterized models, and except for possible power considerations, eliminate any concern over unbalanced data.

   It has been our experience that a failure to identify a relationship among a set of expected values leads to a jungle of confusion in terms of undefined or tautologically defined concepts. Our general position is that the process of assigning an attractive name such as "proportion of variance accounted for," "joint contributions" or "linear by linear component of the AB interaction" to a specifiable quantity does not enhance ones understanding. The purpose of a technical vocabulary should be to speed communication, not to serve as a substitute for thought.

3. Translate natural language questions into a symbolic expression that relates expected values (means) in a model independent form. This step forces an investigator to think seriously about his

question before he attempts to create a model. In particular it forces him to identify the expected values that his model must estimate and the relationship that must exist among those expected values to defend a particular inference about 1. It is particularly important at this stage not to get overly concerned with the "name" of the hypothesis. It may be a stimulating intellectual exercise and a profitable educational experience to determine what the models are that produce the test of the AB interaction in a multiple covariable factorial analysis of covariance. We call this process "model recognition" and although it is a desirable skill to possess, it is different from model creation. Step 3 forces the investigator to state in a relatively unambiguous form what his definition of an "effect" or a "relationship" is. In a sense this step represents the translation of the natural language used in step 1 into an operational definition.

4. Write a linear model which will produce "good" estimates of the expected values which are to be compared. Although the superiority of matrix notation is acknowledged for some purposes, we prefer to write the model in vector notation because we believe it facilitates later steps, particularly step 8 and to some extent steps 5, 6, 7 and 9. This model is referred to as the starting, assumed or full model.

5. Investigate the properties of the model so as to identify what assumptions are being made. In connection with this, the researcher should be able to distinguish between a model which is known to be "true" and a model the truth of which depends on certain assumptions. For a large class of fairly simple models step 5 can be carried out by inspection. For more complex models, the researcher may find it necessary to iterate on steps 4 and 5. An understanding of the concept of equivalent models can be invaluable at this stage because a model with specified properties can be parameterized in an infinite number of ways.
6. Substitute for the expected values expressed in 3 their estimates in symbolic form from the model generated in 4.
7. State the restrictions on the parameters of the model which are implied by simplifying the mathematical expressions of 6.
8. Impose the restrictions identified in 7 on the model generated in 4 to produce a restricted or reduced model. For multi-attribute models, steps 7 and 8 can require a considerable degree of skill in algebraic manipulations. A recent issue of Viewpoints contained an article on "contrast coding." In addition to reducing computational labor, contrast coding usually reduces step 8 to a trivial problem. The disadvantage of contrast coding is that the work saved in step 8 is transferred to steps 4 through 6 and the

starting model, although equivalent, is usually not as intuitively appealing.

9. Investigate the properties of the restricted model so as to verify that the estimates of the expected values are related in the manner specified in 3. If not, it may be necessary to iterate on steps 6 through 9.

   This step is frequently omitted with discouraging results. Depending in part on how the starting model was parameterized, the restricted model may look "peculiar" and cause the creator to think a mistake has been made. Failure to verify the restricted model may cause the investigator to give up prematurely or proceed erroneously.

10. Obtain a least squares solution for both the full and restricted model.

11. Compare the full and restricted model by constructing an appropriate test statistic on the basis of which a decision is reached about 3 and by inference about 1.

Steps 1 through 4 are logically the most difficult and intellectually challenging. The decisions reached in these steps are always subject to debate and they generally require a good deal of creative or intuitive insight and substantive expertise in a problem area. Except for steps 10 and 11, the remaining steps generally require only the polishing of relatively routine algebraic skills.

Multiple Linear Regression Viewpoints
Vol. 6, No. 1, 1975

COMPARISON OF REGRESSION COEFFICIENTS IN MULTIVARIATE REGRESSION EQUATIONS

by

Robert A. Karabinus, Northern Illinois University
Charles H. McCormick, Northern Illinois University

The main purpose of this study was to explore methods of comparing regression coefficients in multivariate regression equations containing the same variables but for independent groups. Secondary purposes included the exploration of methods to compare multiple correlation coefficients and canonical correlation coefficients between independent groups.

The uses of multiple regression analysis have been well covered in research literature and in various statistical texts. Problems with numbers and kinds of independent variables to choose, multicollinearity, sampling error, etc., that need careful attention with multiple regression analysis also need special attention with multivariate analysis. Since the study of human behavior is one involving complex and overlapping variables, the use of multivariate regression analyses should and will be playing an ever increasing role as more of us familiarize ourselves with the techniques and the available computer programs that have been developed to handle the rather complicated procedures.

The initial research problem that made this investigation necessary was one involving the relative importance of certain nonintellective variables in the prediction of academic success of school children in three different ethnic groups. Academic success was defined as a compound dependent variable comprising arithmetic and reading subtest scores on the Metropolitan Achievement Tests (1970). Since this dependent variable was really two different sets of scores that were significantly related, the use of multivariate techniques was mandated. The three nonintellective independent variables were the Coopersmith Self-esteem Inventory (SEI), Sarason's Test Anxiety Scale for Children (TASC), and Sarason's Test Anxiety Scale for Children (GASC). In addition to these three independent variables, it was felt necessary to include four other variables: intelligence, sex, ethnic group, and grade in school. Intelligence was measured by the Otis-Lennon Mental Ability Test, sex was boys and girls, ethnic groups were White, Black, and Spanish-surname, and the children were all from grades 4, 5, and 6 from a wide range of socio-economic levels. Over 1200 children were selected from a large suburban mid-western community to participate in this project.

While the detailed results of how all those variables interrelated and what the regression equations looked like would be of general interest (see McCormick, 1975), our concern here was limited to a further exploration of viable statistical techniques to meet the needs of that research as expressed in the purpose statement of this paper. Each of the eight techniques presented was illustrated with data from the McCormick study.

Comparison of two independent multiple correlation coefficients:

One of the secondary purposes of this paper was to find a way to test the significance of difference between two independent multiple correlation coefficients generated from the same variables. While the z test for differences of independent correlation coefficients utilizing Zeta transformations was quite appropriate for zero order correlation coefficients having ranges from -1.0 to +1.0, it was not considered for use with multiple correlations having ranges only from 0 to +1.0.

Some consideration was given to an F test formula that was modeled after the simple homogeneity test for variances. It was identified as Formula 1.

Formula 1: $$F = \frac{(SSreg./df)_1}{(SSreg./df)_2} \quad \text{or simply} \quad \frac{(SSreg.)_1}{(SSreg.)_2}$$

Subscripts of 1 & 2: group designation
SSreg.: regression sum of squares
df: degrees of freedom = # of independent variables, or k. This term dropped out of the equation because k was a constant for each pair of groups compared.

One problem was immediately apparent, however. While in the homogeneity test the n of each group was involved in the determination of the df, for regression variances only the # of independent variables or k was involved. Therefore, the variances were weighted equally in this equation when in fact their actual values were a function of the n of each group. This was considered a major fault with this approach, though when the n of the two groups involved were approximately the same, the results were reasonable.

When two independent R^2 were compared for groups of unequal n using Formula 1 (see Table I, Illustration A), the F value of 5.806 was found to be significant. Conclusion: the correlation with arithmetic was significantly higher ($p=.05$) for White than for Spanish-surname 4th grade children. However, because of the large differences in n between the two groups, this conclusion was probably biased in favor of the larger group. Therefore, this formula might not be valid for comparisons of groups with unequal n.

TABLE I

Illustrations of Independent R^2 Comparisons

A: The R^2 for two ethnic groups in the 4th grade were compared.
Dependent variable: arithmetic achievement
Independent variables: I.Q., sex, SEI, GASC, & TASC

Group	n	k	R^2	SSreg.	SSres.
White	301	5	.6199	32636.604	20010.3195
Spanish-surname	74	5	.51996	5621.1195	5189.597

Formula 1: $F = \frac{(SSreg.)_1}{(SSreg.)_2} = \frac{32636.604}{5621.1195} =$ 5.806 df: 5/5 $p < .05$

2: $F = \frac{[SSres./(n-k-1)]_1}{[SSres./(n-k-1)]_2} = \frac{5189.597/68}{20010.3195/295} =$ 1.125 df: 86/295 $p \approx .25$ n.s.

3: $F = \frac{(SSreg./k)_1 - (SSreg./k)_2}{[SSres./(n-k-1)]_1 + [SSres./(n-k-1)]_2}$

$= \frac{32636.604/5 - 5621.1195/5}{20010.3195/295 + 5189.597/68} =$ 37.483 df: 5/363 $p < .001$

4: $?= \frac{F_1}{F_2} = \left[\frac{R^2/k}{(1-R^2)/(n-k-1)}\right]_1 \div \left[\frac{R^2/k}{(1-R^2)/(n-k-1)}\right]_2$

$= \frac{.6199/5}{.3801/295} \div \frac{.51996/5}{.48004/68} =$ 6.532 ? ?

B: The R^2 for 5th grade Boys and Girls were compared.
Dependent variable: reading achievement
Independent variables: I.Q. sex, SEI, GASC, & TASC

Group	n	k	R^2	SSreg.	SSres.
Boys	189	5	.700	45119.837	19314.809
Girls	190	5	.636	35642.762	20365.916

Formula 1: $F = \frac{45119.837}{35642.762} =$ 1.266 df: 5/5 $p > .25$ n.s.

2: $F = \frac{20365.916/184}{19314.809/183} =$ 1.049 df: 184/183 $p > .25$ n.s.

3: $F = \frac{45119.837/5 - 35642.762/5}{19314.809/183 + 20365.916/184} =$ 8.766 df: 5/367 $p < .001$

4: $?= \frac{.700/5}{.30/183} \div \frac{.636/5}{.364/184} =$ 1.328 ? ?

When two independent R^2 were compared for groups of nearly equal n using Formula 1 (see Table I, Illustration B), the F value of 1.266 was not significant. Conclusion: there was no significant difference between the two R^2. If the formula was valid for equal n groups, then this conclusion could be accepted.

Because of the problem with the n of each group not being considered in Formula 1, we thought we might still be able to use the same basic model but instead of comparing the SS regression, comparing the SS residual. By comparing the SS residual, the n would be involved in the determination of the df for each group. Then, if significance were found between the two independent residual variances, it could be inferred that the two regression variances were also significantly different (since they are hypothetically complementary).

Formula 2: $$F = \frac{(SSres./df)_1}{(SSres./df)_2}$$

Subscripts of 1 & 2: group designation
SSres.: residual sum of squares
df: degrees of freedom = n-k-1

When two independent R^2 were compared for groups of unequal n using Formula 2 (see Table I, Illustration A), the F of 1.125 was not significant. Conclusion inferred: since the residual variances were not significantly different, there was no significant difference between the two R^2. The validity of this conclusion rests with the validity of the formula and the inference. Concerning the latter, e.g., we recognize the fact that because of the df involvement, the regression variances are not directly complementary (even though the residual and regression sum of squares are).

This same formula was then applied to Illustration B, Table 1, and again no significant difference was found between the R^2 of two groups of nearly equal size. Though it would appear that this formula gave rather conservative results, it did seem to be fairer in its treatment of the data than did Formula 1.

Another exploratory method used a combination of models: one, the typical inferential test of mean differences; the other, the format for the test of significance for R^2. It was identified below and in Table I as Formula 3.

Formula 3: $$F = \frac{(SSreg./k)_1 - (SSreg./k)_2}{[SSres./(n-k-1)]_1 + [SSres./(n-k-1)]_2}$$

Symbols used are as defined previously.

On the surface, this Formula seemed plausible, but it generated such large F values that we were not too confident with it. For example, when two independent R^2 were

compared for groups of unequal n using Formula 3 (see Table I, Illustration A), we found the F value of 37.483 to be extremely significant. Conclusion: the R^2 for White 4th grade children was significantly higher ($p < .001$) than the R^2 for Spanish-surname 4th grade children. Even considering the fact that both the individual R^2 for each group were highly significant, it seemed doubtful that the difference between the two could be so highly significant. Though we thought this formula took the bias out of favoring one group over the other because of sample size, it might well have other problems inherent in its logic.

A second example with nearly equal n groups was given in Table I, Illustration B. Here an F of 8.766 was found to be highly significant, which is in sharp contrast to the results from Formulas 1 and 2. Again, the two separate R^2 were each highly significant, so the significant finding here of a difference between them rests with the validity of the formula itself.

Our final exploratory method involved an unknown and therefore unnamed distribution. It was the ratio of F values representing the significance levels of each of the two R^2. We did not know how to interpret it, but believed it was worth further study. It was identified below and in Table I as Formula 4.

Formula 4:
$$? = \frac{F_1}{F_2} = \left[\frac{SS\,reg./k}{SS\,res./(n-k-1)}\right]_1 \div \left[\frac{SS\,reg./k}{SS\,res./(n-k-1)}\right]_2$$
$$= \left[\frac{R^2/k}{(1-R^2)/(n-k-1)}\right]_1 \div \left[\frac{R^2/k}{(1-R^2)/(n-k-1)}\right]_2$$

Using the same data as with the previous Formulas, we computed ? statistics as noted in Table I, Illustration A & B, Formula 4. The ratios of ratios seemed to be in the right proportion for the two different Illustrations, but how to interpret them was unknown.

Comparison of two independent canonical correlation coefficients:

The other secondary purpose of this paper was to explore methods of comparing two independent canonical correlation coefficients. Since a common method to measure the significance of a Rc is the chi square method proposed by Bartlett (1947) to test the significance of Lambda (Λ), it seemed logical to use the chi squares for two independent Lambdas in a ratio so the F distribution could be used. That formula was identified as Formula 5 below and in Table II, Illustration C & D.

Formula 5:
$$F = \frac{\chi_1^2}{\chi_2^2}$$

The F table would be entered with the same df applied to each chi square.

TABLE II

Illustrations of Independent Rc Comparisons

C: The Rc for two ethnic groups of 5th grade children were compared.
Dependent variables: arithmetic and reading achievement
Independent variables: I.Q., sex, SEI, TASC, & GASC

Group	n	k	Rc_1	Λ_1	χ^2	t	m	s	v
White	252	5	.88251	.221176	374.18	2	247.5	2	3
Spanish-surname	58	5	.82644	.316997	62.039	2	53.5	2	3

Formula 5: $F = \frac{\chi^2_1}{\chi^2_2} = \frac{374.18}{62.039} = 6.031$ df: 6/6 $p < .025$

6: $? = \frac{F_1}{F_2} = \left[\frac{(1-\Lambda^{\frac{1}{s}})/t(k-1)}{\Lambda^{\frac{1}{s}}/(ms-v)}\right]_1 \div \left[\frac{(1-\Lambda^{\frac{1}{s}})/t(k-1)}{\Lambda^{\frac{1}{s}}/(ms-v)}\right]_2$

$= \frac{(1 - .221176^{\frac{1}{2}})/8}{.221176^{\frac{1}{2}}/492} \div \frac{(1 - .316997^{\frac{1}{2}})/8}{.316997^{\frac{1}{2}}/104}$

$= \frac{69.275266}{10.089587} = 6.87$? ?

D: The Rc for Boys and Girls in the 5th grade were compared.
Dependent variables: arithmetic and reading achievement
Independent variables: I.Q., sex, SEI, TASC, & GASC

Group	n	k	Rc_1	Λ_1	χ^2	t	m	s	v
Boys	189	5	.891	.2061	292.521	2	184.5	2	3
Girls	190	5	.866	.250	257.798	2	185.5	2	3

Formula 5: $F = \frac{292.521}{257.798} = 1.135$ df: 6/6 $p > .25$ n.s.

6: $? = \frac{(1 - .2061^{\frac{1}{2}})/8}{.2061^{\frac{1}{2}}/366} \div \frac{(1 - .25^{\frac{1}{2}})/8}{.25^{\frac{1}{2}}/368}$

$= \frac{55.028783}{46.003238} = 1.1962$? ?

We limited our examples to the first canonical coefficients, which in all cases were significant. In our first example (see Table II, Illustration C), we used data from two 5th grade ethnic groups, which had unequal n sizes. Since the same variables were used with both groups, the df of 6 (# of independent and dependent variables - 1) was constant. The F value of 6.031 was found to be significant at the .025 level. Conclusion: the Rc for White 5th graders was significantly higher ($p < .025$) than the Rc for Spanish-surname 5th graders.

The use of this formula might be questionable when the n was different for each group. Each chi square was determined in part from the n of its respective group, but the df for each chi square when going into the F table was not. Therefore, the F test could be biased in favor of the larger group in much the same way Formula 1 was for the comparison of independent R^2.

A second example was given in Table II, Illustration D, where the n sizes were nearly equal. Here the chi squares derived from the Lambda for each Rc for the 5th grade Boys and Girls were compared, and the F value of 1.135 was not found to be significant. Whether this lack of significance was a result of no bias with n sizes and actually no difference between the Lambdas, or the difference between the two Lambdas was just not large enough to be significantly different, regardless of n sizes, was not known.

With this approach, it is interesting to note that the chi square is a significance test for Lambda, the residual variance. Therefore, the F formula presented was really a ratio of residual variances, as was Formula 2 with R^2 comparisons, and since the n of each group is built into the determination of the chi square value, this approach might very well be considered valid.

Another approach to the problem of testing the significance of difference between canonicals was explored, this time using the model from Formula 4. We took the ratio of estimated F values to obtain a factor indicating how many more times one F value was larger than the other. The F value for testing the significance of each canonical separately was approximated using a model developed by Rao (Rulon & Brooks, 1968). The resulting formula was identified as Formula 6, below and in Table II.

Formula 6: $$? = \frac{F_1}{F_2} = \left[\frac{(1-\Lambda^{\frac{1}{s}})/t(k-1)}{\Lambda^{\frac{1}{s}}/(ms-v)}\right]_1 \div \left[\frac{(1-\Lambda^{\frac{1}{s}})/t(k-1)}{\Lambda^{\frac{1}{s}}/(ms-v)}\right]_2$$

Λ_1 = Wilks' Lambda, derived from the first canonical correlation $(\Lambda_1 = 1 - Rc^2)$

t = # of dependent variables

k = # of independent variables

$$m = \frac{2n - t - k - 2}{2}$$

$$s = \sqrt{\frac{t^2(k-1)^2 - 4}{t^2 + (k-1)^2 - 5}}$$

$$v = \frac{t(k-1) - 2}{2}$$

When this formula was applied to Illustrations C & D (see Table II), we found ratios that seemed to be proportional to what Formula 5 gave us for the two Illustrations. But precisely how to interpret the results was, as with Formula 4, unknown. This formula was believed to be worthy of further study.

A third and last approach which was considered had so many serious problems that we did not include it Table II. It followed the model of Formula 3, which we initially thought could easily be adjusted to handle the canonical data. Itwould have involved the basic ingredients of Formula 6, with the two fractions made into a compound fraction by removing the division operation and replacing it in the numerator with a subtraction operation and in the denominator with an addition operation - plus the inclusion of total sum of squares for the composite dependent variable. The attempt to calculate the latter was the main trouble, for in reality it exists as a SS-SCP matrix and not as a single value. Though Determinates could have been calculated, we still would not have had a reasonable value to plug into the formula. Therefore, it was not presented.

Comparison of two independent canonical beta weights:

The primary purpose of this paper was to develop a statistical method to compare canonical beta weights for two independent groups on the same variables. The need for such comparisons was made clear in McCormick's study (1975), but since a technique was not available then, comparisons in that study were limited to beta weights in multiple regression equations on the same variables for two independent groups.

The following formula, which was identified as Formula 7, below and in Table III, utilized the t distribution and was considered an exploratory attempt to meet the need of independent canonical beta weight comparisons.

Formula 7: $$t = \frac{(b_{y_1.2..k})_1 - (b_{y_2.13..k})_2}{\sqrt{\left[s^2 b_{y_1.2..k}\right]_1 + \left[s^2 b_{y_2.13..k}\right]_2}}$$

This is basically the same formula used to compare zero order beta weights in independent regression equations (Edwards, 1967).

Adapting the formula to use with canonical beta weights presented two serious problems. The first one was the conversion of normal beta weights to raw score beta weights required in this formula. We naively used the same procedure followed with other partial beta weight conversions, and that was to multiple the normal beta weight by the ratio of standard deviations, the composite dependent variable standard deviation over the selected independent variable standard deviation. The problem, of course, was to figure or estimate the composite dependent variable standard deviation. After much deliberation, it was decided that the square root of the average cross products term of the two dependent variables would be a reasonable estimate of the hypothetical composite standard deviation to use in ratio with the independent variable's standard deviation.

The second problem was the estimating of the $SE^2est.$ term in the formula. The formula for the error variance of each partial beta was taken from Kerlinger and Pedhazur (1973).

$$s^2 b_{y_1.2..k} = \frac{SE^2est.}{SS_1(1 - R^2_{1.2..k})}$$

$SE^2est.$: residual variance

SS_1 : SS of variable whose partial beta is being compared

R^2 : multiple R^2 of that selected independent variable with all other independent variables in the equation

Though Lambda gave us the proportion of residual variance, it did not in any way help us determine the actual amount of residual variance. Again we were faced with the need for a value that existed only in a matrix of within SS-SCP, and not as a single value. We resolved the problem in this way. First we estimated the total sum of squares of the composite using the canonical weights of the two dependent variables. Then we took that portion of the estimated composite sum of squares attributed to error (which we knew to be Lambda), and then divided by the df to obtain the residual variance. To illustrate how this was done, we have calculated below the total sum of squares for the 5th grade White children in Illustration E, Table III.

SS_t for Reading variable in multiple regression table was 77371.853.

SS_t for Arithmetic variable in multiple regression table was 49811.853.

Canonical weights for the composite dependent variable were .515 for Reading and and .569 for Arithmetic.

Squaring these weights gives us the proportion that each carried in the composite.

Therefore, $.515^2$ times 77371.853 = 20520.949 (proportion SS for Reading)

$.569^2$ times 49811.853 = 16127.135 (proportion SS for Arithmetic)

36648.084 (weighted SS_t for composite)

Multiplying this composite SS_t by Lambda should give us that portion attributed to error.

Lambda times weighted SS_t = .221176 · 36648.084 = 8105.6766 (SSresidual)

Dividing this by the appropriate df, which in this case is 246, gives us residual variance.

$$\frac{8105.6766}{246} = 32.95 \text{ (estimated residual variance)}$$

In Table III, this Formula 7 was demonstrated with Illustrations E & F. The first of these, Illustration E, concerned the significance of difference between two independent TASC beta weights for 5th grade White and Spanish-surname children. In this example the beta weight for the Spanish-surname group was significant while the beta weight for the White group was not. The former group, though it was considerably smaller in size, proved to have a significantly larger beta ($p < .001$) than did the White children. Considering the difference between the two beta weights, this finding was not surprising. Therefore, the formula seemed to do a reasonable job in spite of all the estimating of values that was involved.

A similar significant difference was found between two independent GASC beta weights for 4th grade White and Black children. In this case, Illustration F, Table III, the beta weight for the White children (larger group) was found to be significantly larger than the beta weight for Black children.

In spite of the fact that estimated values were used in this formula, because it seemed to work well in spite of n differences between groups, we think it is worth further examination. It was the only formula that was considered plausible to meet the need of testing the significance of independent canonical beta weight differences. However, we would hope that there could be developed a way to compare the normal beta weights that would also not involve estimated residual variance calculations.

TABLE III

Illustrations of Canonical Beta Weight Comparisons

E: The beta weights of the TASC variable for 5th grade White and Spanish-surname children were compared.

Dependent variables: arithmetic and reading achievement

Independent variables: I.Q., sex, SEI, GASC, & TASC

Group	n	TASC β	Rc_1	Λ_1	Beta Weights Rdg.	Beta Weights Arith.	SS Total Rdg.	SS Total Arith.
White	252	.075	.88251	.221176	.515	.569	77371.853	49811.853
Spanish-	58	- .285	.82644	.316997	.655	.402	11664.345	9680.7759

Formula 7:

$$t = \frac{(b_{y_1 .a..k})_1 - (b_{y_2 .13..k})_2}{\sqrt{\left[s^2 b_{y_1 .a..k}\right]_1 + \left[s^2 b_{y_2 .13..k}\right]_2}}$$

$$= \frac{.3328 - (-.8351201)}{\sqrt{\frac{32.95}{40893(1-.755)} + \frac{40.98}{15259(1-.843)}}}$$

$$= \frac{1.1679201}{.1427935} = 8.18 \quad \text{df: } 298 \quad p < .001$$

F: The beta weights of the GASC variable for 4th grade White and Black children were compared.

Dependent variables: arithmetic and reading achievement

Independent variables: I.Q., sex, SEI, GASC & TASC

Group	n	GASC β	Rc_1	Λ_1	Beta Weights Rdg.	Beta Weights Arith.	SS Total Rdg.	SS Total Arith.
White	301	-.127	.851	.2758	.577	.493	97320.578	52646.924
Black	66	-.030	.813	.339	.644	.416	6832.3182	7912.1212

Formula 7:

$$t = \frac{-.347359 - (-.0516364)}{\sqrt{\frac{42.255}{96034(1-.855)} + \frac{23.7461}{29235(1-.95)}}}$$

$$= \frac{-.2957226}{.1415835} = -2.09 \quad \text{df: } 355 \quad p < .05$$

Summary:

Obviously, the kind of research that McCormick dealt with in his dissertation raised a considerable number of technical and statistical questions, of which only a few were considered in this paper. Each one of the three purposes of this paper could well have comprised a paper in itself.

Of the four exploratory formulas presented for independent R^2 comparisons, we believe that Formula 2 makes the best sense, even though it appeared to give very conservative results. We feel much less confident with Formulas 1 and 3, but Formula 4 might prove to be the best of them if a distribution and interpretation were developed.

Of the two exploratory formulas presented for independent Rc comparisons, we believe that Formula 5 makes good sense. Formula 6 has potential if Formula 4 makes any sense and proves to be feasible.

The only exploratory formula presented for independent Rc beta weight comparisons was Formula 7. However, we do not consider it satisfactory in its present form. Hopefully, there will be better ways to estimate the significance of difference between these beta weights.

We would encourage others to look at these models and give serious thought to the development of sound approaches to meet the significance testing needs of multivariate analyses.

REFERENCES

Bartlett, M.S. Multivariate Analysis Supplement to the Journal of the Royal Statistical Society, 1947, 9, 176-197.

Edwards, A.L. Statistical Methods for the Behavioral Sciences, Holt, Rinehart, and Winston, 2nd ed., 1967

Kerlinger, F.N. & Pedhazur, E. J. Multiple Regression in Behavioral Research, Holt, Rinehart, and Winston, Inc., 1973.

McCormick, C.H. The Prediction of Arithmetic and Reading Achievement from Selected Affective Measures, Unpublished Doctoral Dissertation, Northern Illinois University, 1975.

Rulon, P.J. & Brooks, W.D. On Statistical Tests of Group Differences, in Handbook of Measurement and Assessment in Behavioral Sciences, (ed., Whitla), Addison-Wesley, 1968.

Multiple Linear Regression Viewpoints
Vol. 6, No. 1, 1975

Paper Presented to a Meeting of the
American Educational Research Association
Special Interest Group/Multiple Linear
Regression.

Washington D.C.

March 1975

The Effect of Multicollinearity and the Violation of the Assumption of Normality on the Testing of Hypotheses in Regression Analysis

Ellen S. Vasu
University of North Carolina at Chapel Hill

Patricia B. Elmore
Southern Illinois University at Carbondale

Abstract

This study investigated the effects of the violation of the assumption of normality coupled with the condition of multicollinearity upon the outcome of testing the hypothesis $\beta' = 0$ in the two-predictor regression equation. A monte carlo approach was utilized in which three different distributions were sampled for two sample sizes over thirty-four population correlation matrices. The preliminary results indicate that the violation of the assumption of normality has no significant effect upon the outcome of the hypothesis testing procedure. As was expected, however, the population correlation matrices with extremely high collinearity between the independent variables resulted in large standard errors in the sampling distributions of the standardized regression coefficients. Also, these same population correlation matrices revealed a larger probability of committing a type II error. Many researchers rely on beta weights to measure the importance of predictor variables in a regression equation. With the presence of multicollinearity, however, these estimates of population standardized regression weights will be subject to extreme fluctuation and should be interpreted with caution, especially when the sample size involved is relatively small.

The Effect of Multicollinearity and the Violation of the Assumption of Normality on the Testing of Hypotheses in Regression Analysis

One of the goals of applied research is to define functional relationships among variables of interest. If such relationships can be found, then this knowledge can be used for prediction purposes. For example given a subject's scores on selected X variables, the mathematical relationship can be utilized to predict that same subject's score on the associated Y variable. If the relationship is not a stable one, then perfect prediction is not possible. This is generally the situation that exists in social science research. The best that a prediction rule can do is to provide a 'good' fit to the data. Nevertheless, knowledge of such a rule can greatly decrease the errors in prediction and can be of practical utility in behavioral research (Hays, 1963).

Multiple linear regression is one mathematical approach to the problem of prediction. Given a set of independent variables and a criterion variable, least squares regression weights can be calculated which will maximize the squared multiple correlation between the criterion vector and the predicted criterion vector (Kerlinger, 1973). If the variables used in the determination of the regression weights are transformed into z score form, then the resulting weights are standardized regression coefficients and sometimes are referred to

as beta coefficients (McNemar, 1969). In the remainder of this paper the symbol β' will be used to refer to the population standardized regression coefficient and the symbol b' will represent the sample weight which estimates it.

These b' weights have been interpreted by some researchers to reflect the strength and direction of the relationship between an independent variable and the criterion. However, b' weights in most cases are not a useful measure of the importance of a predictor variable when the independent variables are highly intercorrelated (Darlington, 1968). There is no requirement in multiple regression analysis that the predictor variables used in the regression equation be uncorrelated or orthogonal (Johnston, 1963). From a linear algebra perspective this is reasonable since a criterion vector (dependent variable) can fit perfectly into a common vector space spanned by basis vectors (independent variables) which are not orthogonal. (The criterion vector can be a linear combination of these basis elements). Therefore, situations may occur in regression analysis in which the independent variables are highly intercorrelated. The presence of such highly intercorrelated predictors is termed multicollinearity. These predictor variables are, in fact, measuring approximately the same thing which makes the determination of the relative influence of each independent variable upon the criterion virtually impossible to disentangle (Goldberger, 1968). Also, the presence of multicollinearity increases the standard error of b' values which results in a statistically less consistent estimator of β' (Goldberger, 1968).

When exact multicollinearity occurs, one of the independent variables becomes a multiple of another. In the case of two predictor variables this would mean that the best fitting function which should be represented by a plane (see Figure 1) can instead be represented by a line. Again visualizing this situation from the perspective of linear algebra, it is evident that since linear dependencies cannot exist among basis elements which span a common vector space, the dimensionality of the vector space would in this case be reduced to two and the best fitting function would degenerate to one of a line. Exact multicollinearity is rare in applied research but multicollinearity is a rather common occurrance.

Statistical tests of significance can be run to determine whether or not a specific β' value is different from zero in the population. In order to test hypotheses such as these, an assumption of normality must be made in the distribution of the criterion measures (Draper & Smith, 1966). This assumption is rarely met in psychological or social science research. Many variables of interest to psychologists and educators are extremely skewed in the population making such an assumption invalid.

One of the goals of this study was to examine the effect of the violation of this assumption upon the probability of committing a type II error in the testing of hypotheses based upon b' coefficients. In order to answer this research question and the others which will be explained in turn, a monte carlo approach was taken. Extremely

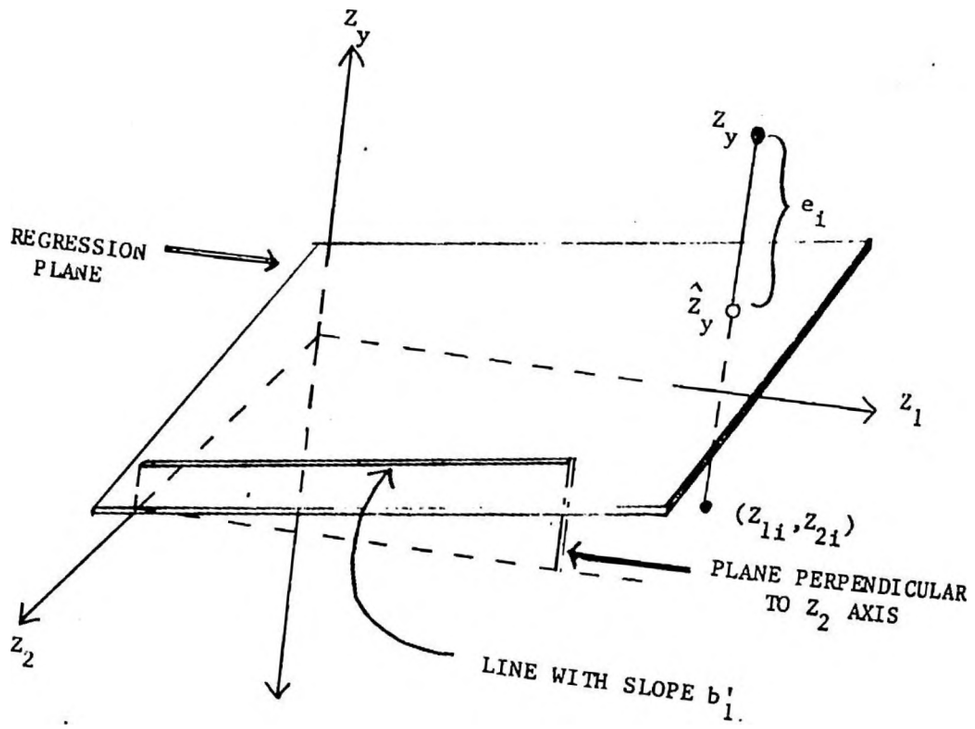

z_y
z_y
e_i
REGRESSION PLANE
$\hat{z}_y$
z_1
(z_{1i}, z_{2i})
PLANE PERPENDICULAR TO z_2 AXIS
z_2
LINE WITH SLOPE b_1'

Figure 1

skewed distributions were included in the distributions of the variables in the populations for the purpose of making the research more meaningful.

Turning once again to the problem of multicollinearity, one might consider the effect of highly correlated variables upon the outcome of the testing of hypotheses such as $H_0:\beta' = 0$ for each independent variable involved in the regression equation. Ostle (1963) states that the F tests used in testing these hypotheses are not all independent since the predictor variables themselves may be correlated. This was another goal of the study, to investigate the effects of multicollinearity upon the probability of committing a type II error in the testing of these hypotheses.

In review the main focus of the authors was the effect of multicollinearity coupled with the violation of the assumption of normality in the criterion measures upon the outcome of the testing of hypotheses concerning population regression coefficients in the two-predictor regression equation. Answers were sought to the following specific research question:

1. What effect does the violation of the assumption of normality have upon the probability of committing a type II error for alpha .05 in the testing of the null hypothesis $H_0:\beta_i' = 0$ $(i = 1,2)$ for both small and large sample sizes?
2. What effect does the presence of multicollinearity have upon the probability of committing a type II

error in the testing of these hypotheses for small and large samples?

3. Does this effect (if any) change as the distribution sampled becomes more skewed?

The mathematical model under investigation may be written as:

$$Z_y = \beta_1' Z_1 + \beta_2' Z_2 + e$$

or equivalently:

$$Z_y = Z_x \beta'' + e$$

where Z_y is an (n x 1) vector of observations in z score form

Z_x is an (n x 2) matrix of known form whose elements are also standardized

β'' is a (2 x 1) vector of parameters

e is an (n x 1) vector of errors

and where the e_i are independently and normally distributed (Draper & Smith, 1966). This last statement is needed in order to test the significance of β'. We must also make the important assumption that the linear model defines the best functional fit to the data in the population. This assumption can be met by sampling from a multivariate normal distribution (Blalock, 1972) which was accomplished through the monte carlo program.

The test of the null hypothesis that a specific β' value was different from zero was determined from the following test statistic (McNemar, 1969):

$$F = \frac{(R_1^2 - R_2^2)/(m_1 - m_2)}{(1 - R_1^2)/(N - m_1 - 1)}$$

where R_1 is the multiple correlation coefficient based upon m_1 of the predictor variables and R_2 is the multiple correlation coefficient based upon m_2 of the remaining variables where $m_2 = m_1 - 1$. Sample b' values were calculated using the following formulae (McNemar, 1969):

$$b_1' = \frac{r_{y1} - r_{y2}r_{12}}{(1 - r_{12}^2)} \qquad b_2' = \frac{r_{y2} - r_{y1}r_{12}}{(1 - r_{12}^2)}$$

The population correlation matrices, sample sizes and population distributions chosen will be outlined in the next section.

Method

In order to answer the research questions it seemed necessary to construct approximate sampling distributions of b_1' and b_2' values from the sample regression equation:

$$z_y = b_1'z_1 + b_2'z_2 + e$$

The hypotheses dealt with the violation of the assumption of normality, level of collinearity between the independent variables, sample size and the effect of these upon the hypothesis testing of β'. Three different distributions were chosen from which to generate random samples of z scores: the multivariate normal, χ^2 with 5 degrees of freedom and χ^2 with 20 degrees of freedom. Three different levels of intercorrelation between the predictor variables were chosen: ρ_{12} = .95, .70 and .45. In addition two different sample sizes were selected: n = 25 and n = 100.

The basic element in the monte carlo procedure was the intercorrelation between the independent variables in the population. At one level of intercorrelation between Z_1 and Z_2 different levels of correlation between Z_y and Z_1 were selected as were different levels of correlation between Z_y and Z_2. Thirty-four different triplets of population intercorrelations among Z_y, Z_1 and Z_2 were selected and are displayed in Table 1. Five cases involved a ρ_{12} value of .95, fourteen cases involved a ρ_{12} value of .70 and fifteen cases involved a ρ_{12} value of .45. These triplets of population Pearson product-moment correlation coefficients were transformed into factor structure matrices which were then used as input into a monte carlo program written by the main author and based upon a previously developed Fortran program (Wherry, 1965). By focusing on one of the population correlation matrices, the logic behind the monte carlo technique can be more easily explained and comprehended.

For one set of fixed ρ_{y1}, ρ_{y2} and ρ_{12} values a factor structure matrix was calculated and a distribution and sample size were chosen for generating sample r_{y1}, r_{y2} and r_{12} values. Because the authors were interested in examining standardized regression coefficients which are based upon z score values, these sample r coefficients were all that was needed in order to calculate b_1' and b_2' coefficients for a sample regression equation. Five-hundred sample correlation matrices were produced for each selected distribution and sample size, therefore five-hundred sample regression equations in z score form were developed

for the population regression equation. The five-hundred b_1' coefficients were then used to form an approximate sampling distribution for b_1'. The same procedure was followed for b_2'.

As each sample b' value was produced, an F test was used to determine if the regression weight was significantly different from zero at the .05 level of significance. This information was tabulated and used in the calculation of the empirical probability of committing a type II error: which was estimated by taking the proportion of b' values which were retained in the hypothesis testing procedure. All the population β' values present in this study (see Table 1) were different from zero. Therefore, the only kind of error which could be examined was type II error; the probability of retaining a false hypothesis.

For each factor structure matrix six approximate sampling distributions for b_1' were developed and six approximate sampling distributions for b_2' were simultaneously developed. One was formed for each combination of distribution and n size: multivariate normal, χ^2_5 and χ^2_{20}; n=100 and n=25. Since there were thirty-four factor structures in total, two-hundred and four approximate sampling distributions were formed for each b' coefficient.

Characteristics of the sampling distributions, population ρ values, distributional type, sample size and population β' values were examined for the presence of relationships in accordance with the research hypotheses. Table 2 through Table 7a contain the summary statistics of the sampling distributions of each b'.

Results

Table 2 and Table 3 consist of calculations based upon the bias involved in each sampling distribution. Since the model employed in the regression procedure was fixed, the mean of each sampling distribution of b' should equal the population β' value. In Table 2 and Table 3, however, there is evidence of bias. The average bias, whether mean or median, is slight: the largest negative bias is -.056 while the largest positive bias is .051. Since each sampling distribution involved a finite number of b' values and was, therefore, only approximate, it would seem logical to attribute the presence of bias to the approximation technique. By scanning each table across distributional shape, (Dist. Type), there appears to be little difference in the reported statistics and no consistent pattern appears as the deviation from normality becomes more marked. A Spearman correlation coefficient was calculated between bias and distribution shape and was found to be non significant in all cases. (see Table 8). Likewise, by scanning the columns of Table 2 and Table 3 there appears to be little difference in the reported statistics. A Spearman correlation coefficient was calculated between bias and level of intercorrelation between predictors in the population. This coefficient was also found to be nonsignificant in all but one case. (see Table 8).

Tables 4 and 5 contain statistics on the standard deviations of the sampling distributions of the b' values. Scanning across each table from left to right there appears to be little change in the average of the standard errors for the b' coefficients. The Spearman

correlation coefficient calculated between empirical standard error, (S_{e_1} and S_{e_2}), and distributional type was not found to be significant. As was expected, however, there is a significant correlation between the standard error of each b' sampling distribution and the level of intercorrelation present between the independent variables in the population. (see Table 8). By examination of Tables 4 and 5 one can see a decrease in the average standard error of the sampling distributions of the b' values as the ρ_{12} value decreases from .95 to .45. This decrease is consistent for a sample size of 25 and a sample size of 100 regardless of the distribution sampled. As the ρ_{12} value decreases, the spread of the standard error values for the distributions also decreases as indicated by the standard deviation statistics.

In Table 6 and Table 6a there appear statistics calculated on difference values obtained by subtracting the theoretical probability of committing a type II error from the empirical proportion of false hypotheses which were retained. Again, there seems to be little change among the average of the difference values as the shape of the distributions sampled becomes more skewed. However, as ρ_{12} decreases, the average difference between empirical and theoretical probability of committing a type II error also decreases/ The maximum difference for n = 25. appears when ρ_{12} equals .95; the maximum difference at this level is .502. As ρ_{12} decreases to .45, the maximum difference is found to be .192. The spread of the difference values decreases as the ρ_{12} value

for n = 25.
decreases from .95 to .45/ A significant correlation was found to exist between the difference values, $(Diff_1(25), Diff_2(25))$, and ρ_{12} for a sample size of 25. As the sample size increased to 100, the correlation was found to be non-significant. These difference values can be attributed to the approximation of the monte carlo technique. When ρ_{12} was relatively low and the sample size was large, the approximation technique was much more accurate.

Table 7 and Table 7a contain the proportion of times the null hypothesis was falsely retained; an approximation of the probability of committing a type II error. As the distribution becomes more skewed, there is no significant change in the average proportion of times a false hypothesis was retained regardless of sample size. The Spearman correlation coefficient calculated between empirical proportion of type II errors committed and distributional shape was found to be non-significant regardless of sample size. The largest Spearman found was .03.

from .95 to .70 or from .95 to .45,
As the ρ_{12} value decreases/ the probability of committing a type II error also decreases as would be expected. This finding is consistent for all distributions sampled for both sample sizes. The average type II error for sample sizes of 100 within a level of ρ_{12} is smaller for a sample size of 100 than for one of 25.

Conclusions and Implications

The results illustrate that a departure from normality in the distribution from which random samples are selected for inclusion in a regression equation with two predictors does not significantly influence the probability of committing a type II error in the testing of the null hypothesis $H_0:\beta_1' = 0$; (i = 1,2). Because the assumption of normality can rarely be met in the distribution of psychological and educational variables, and if it seems plausible to generalize beyond two independent variables, the results indicate that this violation should not be of great concern to a researcher.

Level of intercorrelation confounded with a departure from normality did not significantly influence the probability of committing a type II error either.

As was expected, multicollinearity does have an effect upon the sampling distribution of b' values. This fact is consistent with the theory behind the effects of multicollinearity upon distributions of standardized regression coefficients. The more highly the predictor variables are correlated, the larger the standard error of the b' values. This implies that a confidence interval around a b' value for the purpose of estimating β' would have to be much larger in the case of a regression equation with an r_{12} value which is exceedingly high. In general /the smaller the amount of collinearity between two predictors and the larger the sample size, the more statistically consistent the b' values are: in other words the probability that the b' value is close to the β' value of the population regression equation is increased.

Based upon the findings of this research report it would seem that researchers dealing with variables selected from populations with extremely skewed distributions do not have to be concerned with any detrimental effects upon the probability of committing a type II error. However, with small sample sizes and highly correlated (.95) predictors, generalizations about the contribution of an independent variable to any regression equation should be made with caution. Sample b' values in situations such as these are subject to extreme fluctuation and, although they are unbiased in the long run, most researchers are dealing with only one regression equation and, therefore, only one estimate of any population β' value.

Selected References

Blalock, H.M., Jr. Social Statistics. New York: McGraw-Hill Book Company, 1972.

Cooley, W., & Lohnes, P.R. Multivariate Data Analyses. New York: John Wiley and Sons, Inc., 1971.

Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

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

Ezekiel, M., & Fox, K. A. Methods of Correlation and Regression Analysis. New York: John Wiley and Sons, Inc., 1967.

Freund, J. E. Mathematical Statistics. New Jersey: Prentice-Hall, Inc., 1971.

Goldberger, A. S. Topics in Regression Analysis. New York: The Macmillan Company, 1968.

Hays, W. L. Statistics for Psychologists. New York: Holt, Rinehart & Winston, 1973.

Johnston, J. Econometric Methods. New York: McGraw-Hill Book Company, 1963.

Kelley, F. J. Beggs, D. L. & McNeil, K. A. Research Design in the Behavioral Sciences: Multiple Regression Approach. Illinois: Southern Illinois University Press, 1969.

Kerlinger, F. N. Foundations of Behavioral Research. New York: Holt, Rinehart & Winston, Inc., 1973.

McNemar, Q. Psychological Statistics. New York: John Wiley and Sons, Inc., 1962.

Mood, A. M., & Graybill, F. A. Introduction to the Theory of Statistics. New York: McGraw-Hill Book Company, 1963.

Ostle, B. Statistics in Research, Basic Concepts and Techniques for Research Workers. Iowa: Iowa State University Press, 1963.

Rummel, R. J. Applied Factor Analysis. Evanston: Northwestern University Press, 1970.

Tatsuoka, M. M. Multivariate Data Analysis: Techniques for Educational and Psychological Research. New York: John Wiley and Sons, Inc., 1971.

Walker, H., & Lev, J. Statistical Inference. New York: Henry Holt and Company, 1953.

Wherry, R. J., Sr. Naylor, J. C., Wherry, R. J., Jr., & Fallis, R. F. Generating multiple samples of multivariate data with arbitrary population parameters. Psychometrika, 1965, 30, 303-313.

Winer, B. J. Statistical Principles in Experimental Design (2nd ed.). New York: McGraw-Hill Book Company, 1971.

Wonnacott, T. H. & Wonnacott, R. J. Introductory Statistics. New York: John Wiley and Sons, Inc., 1969.

Table 1

Population Intercorrelations[a] Specified in Monte Carlo Procedure and Accompanying Theoretical Standardized Regression Weights

ρ_{y1}	ρ_{y2}	ρ_{12}	β_1	β_2
.95	.95	.95	.4872	.4872
.70	.70	.95	.3590	.3590
.45	.70	.95	-2.2051	2.7949
.70	.45	.95	2.7949	-2.2051
.45	.45	.95	.2308	.2308
.70	.95	.70	.0606	.9020
.45	.95	.70	-.4216	1.2451
.95	.70	.70	.9020	.0686
.70	.70	.70	.4118	.4118
.45	.70	.70	-.0784	.7549
.00	.70	.70	-.9608	1.3725
.95	.45	.70	1.2451	-.4216
.70	.45	.70	.7549	-.0784
.45	.45	.70	.2647	.2647
.00	.45	.70	-.6176	.8824
.70	.00	.70	1.3725	-.9608
.45	.00	.70	.8824	-.6176
-.45	.00	.70	-.8824	.6176
-.70	.00	.70	-1.3725	.9608
.70	.95	.45	.3417	.7962
.45	.95	.45	.0282	.9373
.95	.70	.45	.7962	.3420
.70	.70	.45	.4828	.4828
.45	.70	.45	.1693	.6238
.00	.70	.45	-.3950	.8777
.95	.45	.45	.9373	.0282
.70	.45	.45	.6238	.1693
.45	.45	.45	.3103	.3103
.00	.45	.45	-.2539	.5643
-.45	.45	.45	-.8182	.8182
.70	.00	.45	.8777	-.3950
.45	.00	.45	.5643	-.2539
-.45	.00	.45	-.5643	.2539
-.70	.00	.45	-.8777	.3950

[a] ρ_{y1} is the population correlation between the criterion variable, z_y, and the predictor variable, z_1. ρ_{y2} is the population correlation between the criterion variable, z_y, and the predictor variable, z_2. ρ_{12} is the population correlation between the independent variables, z_1 and z_2. These population correlations were utilized in the determination of factor structure matrices for input into the Monte Carlo technique. There are five factor structure matrices which have a ρ_{12} value of .95, fourteen which have a ρ_{12} value of .70 and fifteen which have a ρ_{12} value of .45.

Table 2

Calculations of the Bias[a] Present in Estimating β_1'

Pop. Correlation Bet. Ind. Variables	Normal	χ^2_5	χ^2_{20}	Normal	χ^2_5	χ^2_{20}
$\rho_{12} = .95$						
Mean	-.004	.010	-.003	-.003	-.009	-.015
Standard Deviation	.026	.025	.028	.023	.010	.028
Median	-.001	.002	-.007	-.004	-.009	-.017
Mode	.001	-.004	.002	.001	-.010	-.002
Skewness	-.524	1.090	.619	.357	-.098	.224
Minimum	-.044	-.010	-.032	-.032	-.024	-.050
Maximum	.027	.051	.039	.032	.004	.024
$\rho_{12} = .70$						
Mean	-.000	-.002	-.004	-.003	-.005	-.002
Standard Deviation	.010	.011	.013	.012	.012	.011
Median	.002	-.002	-.005	-.002	-.006	-.001
Mode	.015	.019	.022	.005	.013	.006
Skewness	-.556	.081	.180	-.505	-.851	-.258
Minimum	-.023	-.024	-.026	-.029	-.035	-.022
Maximum	.015	.019	.022	.016	.013	.019
$\rho_{12} = .45$						
Mean	-.004	-.001	.001	-.003	-.000	.001
Standard Deviation	.007	.013	.011	.008	.013	.008
Median	-.003	.003	.001	-.000	.001	.001
Mode	-.003	.003	-.004	.001	.000	.001
Skewness	-.390	-.806	.152	-1.748	-.004	1.568
Minimum	-.017	-.031	-.020	-.025	-.019	-.012
Maximum	.006	.018	.020	.005	.018	.027
N =		(25)			(100)	

[a]Bias was determined by the following formula: $(\bar{b}_1' - \beta_1')$ where $\bar{b}_1'$ is the mean of the sampling distribution of five-hundred standardized regression weights and β_1' is the corresponding theoretical standardized regression weight.

Table 3

Calculations of the Bias[a] Present in Estimating β_2'

Pop. Correlation Bet. Ind. Variables	Normal	χ^2_5	χ^2_{20}	Normal	χ^2_5	χ^2_{20}
$\rho_{12} = .95$						
Mean	-.004	-.020	-.010	-.003	-.004	.006
Standard Deviation	.025	.019	.029	.029	.013	.026
Median	-.003	-.016	-.005	.002	-.005	.008
Mode	-.001	-.003	-.005	-.002	.007	.001
Skewness	-.029	-1.120	-.726	-.714	.126	-.346
Minimum	-.039	-.052	-.056	-.049	-.017	-.033
Maximum	.030	-.003	.024	.030	.011	.040
$\rho_{12} = .70$						
Mean	-.003	-.005	-.001	-.002	-.000	-.003
Standard Deviation	.015	.015	.012	.008	.104	.013
Median	-.003	-.004	-.002	-.001	-.003	-.003
Mode	.004	.000	-.020	-.010	.003	-.004
Skewness	.254	.476	.764	.392	.748	-.468
Minimum	-.029	-.028	-.020	-.013	-.017	-.032
Maximum	.026	.027	.028	.014	.029	.020
$\rho_{12} = .45$						
Mean	-.004	-.005	-.002	-.001	-.004	-.002
Standard Deviation	.007	.008	.008	.006	.009	.007
Median	-.005	-.006	-.002	-.001	-.001	-.001
Mode	-.002	-.006	-.001	-.002	-.005	-.002
Skewness	1.853	.059	.247	-.072	-.773	-.677
Minimum	-.011	-.018	-.016	-.011	-.022	-.019
Maximum	.019	.010	.015	.010	.008	.009
N =		(25)			(100)	

[a]Bias was determined by the following formula: $(\bar{b}_2' - \beta_2')$ where $\bar{b}_2'$ is the mean of the sampling distribution of five-hundred standardized weights and β_2' is the corresponding theoretical standardized weight.

Table 4

Standard Deviation of the Empirical Sampling Distribution of b_1' [a]

Pop. Correlation Bet. Ind. Variables	Normal	χ^2_5	χ^2_{20}	Normal	χ^2_5	χ^2_{20}
$\rho_{12} = .95$						
Mean	.428	.431	.445	.417	.436	.420
Standard Deviation	.163	.138	.159	.147	.141	.146
Median	.440	.452	.457	.437	.450	.445
Mode	.179	.214	.202	.187	.214	.184
Skewness	-.483	-.662	-.435	-.593	-.624	-.820
Minimum	.179	.214	.202	.187	.214	.184
Maximum	.623	.588	.643	.589	.600	.570
$\rho_{12} = .70$						
Mean	.175	.195	.182	.177	.194	.184
Standard Deviation	.062	.073	.067	.064	.073	.068
Median	.202	.222	.216	.202	.221	.207
Mode	.215	.260	.221	.213	.238	.225
Skewness	-.731	-.729	-.800	-.629	-.753	-.677
Minimum	.073	.079	.077	.078	.078	.080
Maximum	.256	.265	.252	.258	.273	.272
$\rho_{12} = .45$						
Mean	.135	.150	.137	.134	.151	.139
Standard Deviation	.055	.064	.055	.056	.063	.060
Median	.144	.159	.150	.145	.161	.145
Mode	.078	.079	.081	.077	.077	.072
Skewness	-.405	-.233	-.436	-.345	-.437	-.327
Minimum	.043	.049	.045	.043	.049	.043
Maximum	.212	.248	.214	.221	.226	.219
N =		(25)			(100)	

[a] Where b_1' is a standardized regression coefficient corresponding to z_1 in the prediction equation.

Table 5

Standard Deviation of the Empirical Sampling Distribution of b_2' [a]

Pop. Correlation Bet. Ind. Variables	Normal	χ^2_5	χ^2_{20}	Normal	χ^2_5	χ^2_{20}
$\rho_{12} = .95$						
Mean	.428	.435	.449	.414	.432	.420
Standard Deviation	.164	.143	.163	.145	.141	.147
Median	.441	.453	.459	.437	.448	.446
Mode	.179	.211	.202	.188	.212	.182
Skewness	-.445	-.629	-.409	-.623	-.605	-.851
Minimum	.179	.211	.202	.188	.212	.182
Maximum	.628	.600	.653	.582	.597	.567
$\rho_{12} = .70$						
Mean	.175	.197	.183	.179	.197	.185
Standard Deviation	.063	.075	.069	.064	.076	.069
Median	.197	.215	.207	.206	.224	.203
Mode	.236	.285	.244	.235	.251	.248
Skewness	-.672	-.559	-.681	-.692	-.727	-.510
Minimum	.074	.081	.070	.075	.074	.077
Maximum	.256	.285	.264	.260	.276	.289
$\rho_{12} = .45$						
Mean	.136	.155	.142	.137	.153	.141
Standard Deviation	.055	.067	.059	.056	.065	.059
Median	.154	.173	.157	.149	.171	.155
Mode	.047	.048	.047	.047	.046	.043
Skewness	-.490	-.406	-.508	-.453	-.512	-.459
Minimum	.047	.048	.047	.047	.046	.043
Maximum	.206	.249	.214	.210	.234	.219
N =		(25)			(100)	

[a] Where b_2' is a standardized regression coefficient corresponding to z_2 in the prediction equation.

Table 6

Calculations of the Difference[a] Between Empirical Type II Error and Theoretical Type II Error in the Testing of the Hypothesis[b], $H_0: \beta_1' = 0$

Pop. Correlation Bet. Ind. Variables	Normal	χ^2_5	χ^2_{20}	Normal	χ^2_5	χ^2_{20}
$\rho_{12} = .95$						
Mean	.211	.220	.218	.124	.136	.126
Standard Deviation	.240	.245	.241	.150	.149	.152
Median	.173	.184	.186	.089	.113	.093
Mode	.000	.000	.000	.000	.000	.000
Skewness	.300	.273	.254	.434	.304	.370
Minimum	.000	.000	.000	.000	.000	.000
Maximum	.479	.502	.485	.320	.329	.303
$\rho_{12} = .70$						
Mean	.053	.053	.058	.016	.016	.014
Standard Deviation	.060	.058	.065	.078	.093	.072
Median	.046	.051	.053	-.005	.002	-.002
Mode	.000	.000	.000	.000	.000	.000
Skewness	.728	.841	.718	.876	-.260	.301
Minimum	.000	.000	.000	-.127	-.209	-.139
Maximum	.165	.181	.193	.211	.199	.157
$\rho_{12} = .45$						
Mean	.012	.023	.019	.019	.022	.023
Standard Deviation	.030	.037	.038	.085	.101	.097
Median	.004	.006	.008	.013	.005	.007
Mode	.000	.000	.000	.000	.000	.000
Skewness	.364	1.463	.664	-1.641	-1.280	-.965
Minimum	-.051	-.021	-.057	-.238	-.268	-.247
Maximum	.080	.124	.102	.140	.188	.192
N =		(25)			(100)	

[a]The differences were determined by subtracting the theoretical probability of committing a type II error from the resulting empirical proportions of type II errors committed. Theoretical probabilities were calculated under the assumption of normality in the sampling distribution of standardized regression coefficients.

[b]The hypothesis, $H_0: \beta_1' = 0$ involved an F test at $\alpha = .05$.

Table 6a

Calculations of the Difference[a] Between Empirical Type II Error and Theoretical Type II Error in the Testing of the Hypothesis[b], $H_0: \beta_2 = 0$

Pop. Correlation Bet. Ind. Variables	Normal	χ^2_5	χ^2_{20}	Normal	χ^2_5	χ^2_{20}
$\rho_{12} = .95$						
Mean	.213	.213	.212	.124	.122	.135
Standard Deviation	.238	.237	.233	.154	.143	.173
Median	.180	.180	.180	.084	.090	.088
Mode	.000	.000	.000	.000	.000	.000
Skewness	.270	.265	.251	.478	.401	.444
Minimum	.000	.000	.000	.000	.000	.000
Maximum	.475	.481	.481	.331	.307	.355
$\rho_{12} = .70$						
Mean	.060	.067	.067	.020	.015	.019
Standard Deviation	.080	.076	.081	.066	.082	.064
Median	.028	.044	.043	-.003	.000	-.000
Mode	.000	.000	.000	.000	.000	.000
Skewness	.691	.692	.914	1.168	-.562	.726
Minimum	.000	.000	.000	-.098	-.187	-.097
Maximum	.199	.205	.231	.195	.161	.167
$\rho_{12} = .45$						
Mean	.029	.027	.032	.031	.038	.032
Standard Deviation	.054	.045	.050	.097	.107	.105
Median	.013	.012	.007	.018	.028	.014
Mode	.000	.000	.000	.000	.000	.000
Skewness	.514	1.846	1.377	-1.025	-1.451	-1.123
Minimum	-.065	-.017	-.025	-.238	-.272	-.264
Maximum	.142	.140	.162	.170	.172	.184
N =		(25)			(100)	

[a]The differences were determined by subtracting the theoretical probability of committing a type II error from the resulting empirical proportions of type II errors committed. Theoretical probabilities were calculated under the assumption of normality in the sampling distribution of standardized regression coefficients.

[b]The hypothesis, $H_0: \beta_2 = 0$ involved an F test at $\alpha = .05$.

Table 7

Proportion of False Hypotheses Retained in Testing $H_0: \beta_1' = 0$

Pop. Correlation Bet. Ind. Variables	Normal	χ^2_5	χ^2_{20}	Normal	χ^2_5	χ^2_{20}
$\rho_{12} = .95$						
Mean	.422	.431	.429	.238	.249	.239
Standard Deviation	.464	.468	.465	.303	.300	.310
Median	.364	.375	.377	.156	.181	.149
Mode	.000	.000	.000	.000	.000	.000
Skewness	.242	.225	.215	.432	.392	.480
Minimum	.000	.000	.000	.000	.000	.000
Maximum	.934	.958	.940	.616	.620	.646
$\rho_{12} = .70$						
Mean	.259	.259	.264	.125	.125	.123
Standard Deviation	.362	.355	.360	.215	.202	.209
Median	.082	.089	.090	.004	.015	.007
Mode	.000	.000	.000	.000	.000	.000
Skewness	1.042	1.016	.987	1.343	1.258	1.341
Minimum	.000	.000	.000	.000	.000	.000
Maximum	.926	.890	.912	.574	.510	.562
$\rho_{12} = .45$						
Mean	.277	.288	.284	.122	.126	.127
Standard Deviation	.342	.335	.347	.197	.185	.196
Median	.132	.176	.139	.024	.032	.026
Mode	.000	.000	.000	.000	.000	.000
Skewness	.880	.810	.856	1.500	1.348	1.358
Minimum	.000	.000	.000	.000	.000	.000
Maximum	.940	.928	.926	.614	.584	.606
N =		(25)			(100)	

Table 7a

Proportion of False Hypotheses Retained in Testing $H_0: \beta_2 = 0$

Pop. Correlation Bet. Ind. Variables	Normal	χ^2_5	χ^2_{20}	Normal	χ^2_5	χ^2_{20}
$\rho_{12} = .95$						
Mean	.424	.424	.423	.238	.235	.248
Standard Deviation	.462	.461	.457	.306	.298	.328
Median	.371	.371	.371	.155	.156	.154
Mode	.000	.000	.000	.000	.000	.000
Skewness	.224	.221	.211	.427	.436	.443
Minimum	.000	.000	.000	.000	.000	.000
Maximum	.930	.936	.936	.614	.610	.658
$\rho_{12} = .70$						
Mean	.277	.279	.278	.129	.125	.129
Standard Deviation	.363	.350	.359	.221	.198	.216
Median	.109	.111	.081	.008	.017	.010
Mode	.000	.000	.000	.000	.000	.000
Skewness	.925	.863	.917	1.403	1.302	1.397
Minimum	.000	.000	.000	.000	.000	.000
Maximum	.952	.880	.942	.612	.514	.604
$\rho_{12} = .45$						
Mean	.336	.334	.339	.148	.155	.149
Standard Deviation	.364	.334	.352	.206	.200	.201
Median	.220	.241	.244	.035	.061	.040
Mode	.000	.000	.000	.000	.000	.000
Skewness	.521	.523	.482	1.071	.970	.983
Minimum	.000	.000	.000	.000	.000	.000
Maximum	.932	.944	.932	.614	.580	.588
N =		(25)			(100)	

Table 8

Spearman Correlation Coefficients[a]

	Dist. Type[c]	ρ_{12}[d]		Dist. Type	ρ_{12}
$Bias_1$[b]	-.02	-.02	$Bias_1$	.04	-.19
	p<.43	p<.41		p<.34	*p<.03
$Bias_2$	.05	-.10	$Bias_2$	-.00	.04
	p<.31	p<.16		p<.49	p<.36
$Diff_1$ (25)[e]	.05	.28	$Diff_1$ (100)	-.00	.06
	p<.31	*p<.00		p<.48	p<.26
$Diff_2$ (25)	.02	.24	$Diff_2$ (100)	.00	-.00
	p<.41	*p<.01		p<.50	p<.49
S_{e_1}[f]	.05	.60	S_{e_1}	.03	.59
	p<.32	*p<.00		p<.38	*p<.00
S_{e_2}	.06	.59	S_{e_2}	.02	.58
	p<.29	*p<.00		p<.41	*p<.00
N =	(25)			(100)	

[a]Some of the correlation coefficients tabled were calculated on variables whose elements involve statistics of sampling distributions. These statistics were tabulated from regression equations originally involving a sample size of 25 or a sample size of 100. The number of cases upon which the significance was determined was 102: the number of factor structures (34) multiplied by the number of distributions sampled (3), which equals the number of sampling distributions examined.

[b]See notes tables 2 and 3.

[c]Dist. Type refers to the shape of the population from which the z scores were generated for input into the regression equations for the purpose of constructing sampling distributions. Three distributions were involved: normal, χ^2_5 and χ^2_{20}.

[d]ρ_{12} is the population correlation between the predictor variables. Three levels were examined: .95, .70 and .45.

[e]$Diff_1$ (25) can vary between zero and one and was calculated by subtracting the theoretical probability of committing a type II error from the empirical proportion of type II errors committed in the testing of the hypothesis $H_0:\beta_1 = 0$, at $\alpha = .05$. $Diff_2$ (25) was determined in the same manner for the hypothesis $H_0:\beta_2 = 0$, as was $Diff_1$(100) and $Diff_2$(100).

[f]S_{e_1} is the empirical standard deviation of the sampling distribution of b_1 values.

*Significant at $\alpha = .05$.

Table 9

Pearson Correlation Coefficients[a]

	$Bias_1$[b]	$Bias_2$	$Diff_1$ (25)[e]	$Diff_2$ (25)	S_{e_1}[f]	S_{e_2}
$Bias_1$	1.00	-.65 *p .00	.13 p .19	.20 *p .05	.07 p<.49	.08 p<.40
$Bias_2$	-.65 *p<.00	1.00	-.30 *p<.00	-.29 *p<.00	-.29 *p<.00	-.28 *p<.01
$Diff_1$ (25)	.13 p<.19	-.30 *p<.00	1.00	.86 *p<.00	.72 *p<.00	.70 *p<.00
$Diff_2$ (25)	.20 *p<.05	-.29 *p<.00	.86 *p<.00	1.00	.66 *p<.00	.70 *p<.00
S_{e_1}	.07 p<.49	-.29 *p<.00	.72 *p<.00	.66 *p<.00	1.00	.99 *p<.00
S_{e_2}	.08 p<.40	-.28 *p<.01	.70 *p.<00	.70 *p<.00	.99 *p<.00	1.00
N -			(25)			

(See notes table 8)

*Significant at $\alpha = .05$.

Table 10

Pearson Correlation Coefficients[a]

	$Bias_1$[b]	$Bias_2$	$Diff_1$ (100)[e]	$Diff_2$ (100)	S_{e_1}[f]	S_{e_2}
$Bias_1$	1.00	-.65 *p<.00	-.12 p<.21	-.06 p<.56	-.22 *p<.03	-.22 *p<.02
$Bias_2$	-.65 *p<.00	1.00	-.06 p<.57	-.10 p<.34	-.06 p<.57	-.02 p<.81
$Diff_1$ (100)	-.12 p<.21	-.06 p<.57	1.00	.65 *p<.00	.57 *p<.00	.58 *p<.00
$Diff_2$ (100)	-.06 p<.56	-.10 p<.34	.65 *p<.00	1.00	.59 *p<.00	.57 *p<.00
S_{e_1}	-.22 *p<.03	-.06 p<.57	.57 *p<.00	.59 *p<.00	1.00	.99 *p<.00
S_{e_2}	-.22 *p<.02	-.02 p<.81	.58 *p<.00	.57 *p<.00	.99 *p<.00	1.00
N =			(100)			

(See notes table 8)

*Significant at $\alpha = .05$.

Multiple Linear Regression Viewpoints
Vol. 6, No. 1, 1975

REGRESSION ANALYSIS FOR REPEATED MEASURES DESIGNS: Dealing with Missing Data and the Use of Covariates as an Alternative to Person Vectors

Judy T. McNeil
High/Scope Educational Research Foundation
Ypsilanti, Michigan

Funds for the writing and typing of this manuscript were provided by Educational Monitoring Systems

A paper presented at the annual meeting of the American Educational Research Association's Multiple Linear Regression Special Interest Group

Washington, D. C.
April 1975

REGRESSION ANALYSIS FOR REPEATED MEASURES DESIGNS: Dealing with Missing Data and the Use of Covariates as an Alternative to Person Vectors

Judy T. McNeil
High/Scope Educational Research Foundation

When subjects are measured on a particular criterion variable at more than one point in time, a "repeated measures" analysis is generally used to test hypotheses about the data. One of these hypotheses may be that the group's mean score increased (or decreased) over time. Since it is generally expected that there is a correlation between early and later scores for the same subjects, it is desirable to take this correlation into account, which is in effect controlling for each individual's mean score on the criterion variable. In fact, some writers argue that failure to extract variance attributable to subject differences results in a violation of the assumption of independence of errors (Glass, Peckham, and Sanders, 1973). (Some authors including this one [see also Dixon and Massey, 1969, and Downie and Heath, 1970] would argue instead that the crucial aspect in deciding to use a repeated measures analysis is "accounting for variance," not "meeting assumptions." If a source of variance is known, it is usually wise--from an heuristic point of view as well as for the power of the test--to include that source of variance in the analysis. In those cases where there is no expected correlation between pre and post scores--as would be the case with some criterion-referenced tests with little or no variance on pre or post--then a repeated measures analysis would not be beneficial.)

In regression analysis, controlling for each subject's mean score on the criterion variable (repeated measures analysis) is accomplished by using person vectors, each of which identifies a subject and contains a "1" if the criterion score belongs to that subject and a "0" if not. Using this regression approach, one can accomplish analyses identical to the "correlated" t test or the repeated measures ANOVA; in addition, this approach allows for greater flexibility in the analysis application.

The application of the general linear model to repeated measures problems was discussed in detail in a presentation to the Special Interest Group in Multiple Linear Regression at the AERA annual meeting in 1974 (Pohlmann and McShane, 1974). Therefore, this paper will describe only briefly the method of using person vectors in regression analysis to test research hypotheses regarding repeated measures, for a one group--two trial situation.

Based on these analyses with person vectors there are two particular focuses of this paper: (1) a proposed solution to the problem of missing data, and (2) the use of covariates as an alternative to person vectors in controlling for differences between individuals.

A One Group--Two Trial Research Hypothesis

Although there are many research hypotheses that a researcher may wish to test in this situation, the most typical one would be the following: "The mean score of the group increases from timepoint A to timepoint B, over and above differences between individual mean scores." For example, a researcher may have measured four infants' ability to focus on and visually follow a moving human face at two ages, say 10 days and 30 days. According to developmental theory, the infants' scores should increase from the first timepoint to the second timepoint (if the measure is well constructed). Since she expects a correlation between pre and post scores on this measure at these ages, the researcher wishes to covary these individual differences. The researcher therefore wishes to test the hypothesis that "the mean score on visual following will increase from 10 days to 30 days, over and above differences between individual mean scores." The resulting full model for testing the hypothesis is thus:

Model 1:

$$X_1 = a_0U + a_1T_1 + a_2T_2 + a_3P_1 + a_4P_2 + a_5P_3 + a_6P_4 + E_1$$

where: X_1 = criterion vector of both 10-day and 30-day scores on visual following;

U = the unit vector;

T_1 = 1 if score on X_1 is a 10-day score, 0 otherwise;

T_2 = 1 if score on X_1 is a 30-day score, 0 otherwise;

P_i (i=1 to 4) = a vector containing a 1 if the score is from person i, 0 otherwise;

$a_0, a_1 \ldots a_6$ = a set of least squares weights derived to minimize the sum of the squared elements in E_1; and

E_1 = the error vector.

Sample data that will be used to demonstrate this analysis is shown in Table 1. The data would be organized into vectors as in Figure 1.

Person	10-day Score on Visual Following	30-day Score on Visual Following
1	4	8
2	3	5
3	6	6
4	2	7

Table 1. Sample Data for a One Group--Two Trial Hypothesis

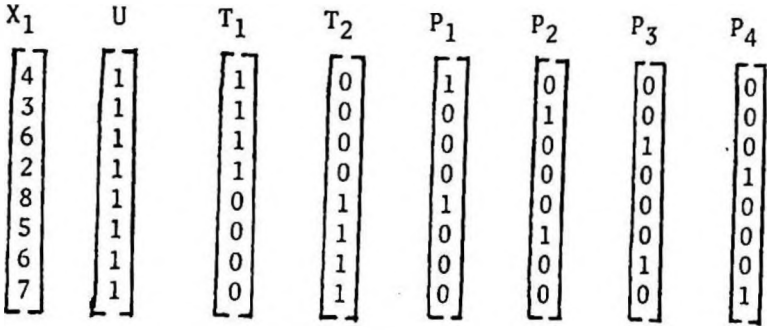


Figure 1. Vectors for Model 1.

The weights a_1 and a_2 in Model 1 will take on values which will reflect the difference between the 10-day and 30-day means. The restriction placed on Model 1 to test the hypothesis would be $a_1=a_2$, which results in the restricted model, Model 2, which would be compared to the full model using an F test. The full model contains 5 linearly independent vectors, and the restricted model contains 4. The degrees of freedom for the F test are therefore 1 and 3.

Model 2:

$$X_1 = a_0U + a_3P_1 + a_4P_2 + a_5P_3 + a_6P_4 + E_2$$

The type of analysis just described is the basis for the following discussions of missing data and the use of covariates.

Missing Data

When collecting longitudinal data (repeated measures), it is unfortunately all too common to find that some subjects are not available for at least one measurement timepoint. Several possible approaches to handling the problem of missing data, not restricted

to repeated measures, are presented by McNeil, Kelly, and McNeil (1975), and three of them are discussed below. A fourth approach is proposed here as a suggested alternative when dealing with repeated measures.

(1) Insert in place of the missing score the mean value of the variable.

(2) Insert in place of the missing score a random value which is within the range of the variable.

(3) Eliminate subjects for whom any data is missing.

(4) Eliminate a subject for only that timepoint on which data is missing; utilize other data for that subject.

(1) Inserting the mean. Using this approach, the researcher makes the assumption that the person with missing data is like the average subject with data. If, in the example on infant visual following used above, subject number 1 was missing the 10-day score, the researcher would assign the mean of the other three subjects' 10-day scores (3.67) to subject number 1. Sometimes, however, the researcher may not be willing to make the assumption that persons with missing data are like persons with data--it may be that they are absent from testing because they are less healthy or less willing to cooperate or different on some other variable which may be relevant to the construct under investigation. In addition, the insertion of mean values reduces the variance of the predictor variable, resulting generally in a variable with lowered predictive value.

(2) Inserting a random value. One way of getting around the problem of reduced variance due to inserting a mean value is to insert a random score (from the range of observed scores) into the missing data locations. Thus, in the example above, if subject number 1 was missing the 10-day score, the researcher would assign a random value (say 6) to subject 1. However, this procedure tends to decrease the relationship between that predictor variable and any other variable because it adds random variance. It should be pointed out that in a repeated measures analysis, inserting either random scores or mean scores tends to reduce the correlation between pre and post scores--a correlation which is expected to be high and on which the justification of a "repeated measures" or person-vector analysis is based.

(3) Eliminating subjects. Using this approach, all scores for a subject would be eliminated if the subject was missing any score. In repeated measures studies, this method can result in the discarding of much data. If, in the example given above, subject number 1 was missing the 10-day score, the 30-day score for subject number 1 would be eliminated as well. This would eliminate the P_1

vector from Models 1 and 2, and the elements of the remaining vectors would be as shown in Figure 2.

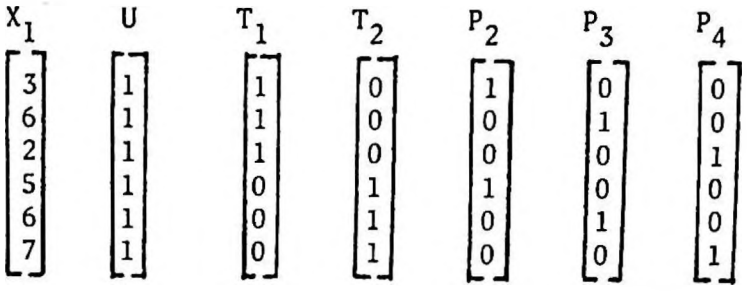


Figure 2. Elements of vectors when subject 1 is eliminated.

A drawback to this approach is that eliminating subjects because of missing data most likely redefines the population from which one has sampled and hence to which one can generalize. Infants who could not be tested at one timepoint may be the least healthy ones or may come from less organized families who could not arrange to keep the testing appointment. Therefore, if one continually uses only complete data, the population to which one can generalize may be restricted to healthy infants from organized families. This restriction may be unavoidable in some studies. But when one has some data on these differing subjects, it is unfortunate and possibly unnecessary to ignore that data.

(4) Eliminating only missing timepoints. I would propose that in most longitudinal or repeated measures studies the researcher would neither want to reduce the relationship between pre and post scores by inserting mean or random values nor want to give up the use of a subject's data on all timepoints just because the subject is missing a score at one timepoint. I would therefore recommend that a subject be included in the analysis if scores were obtained for him at the earlier timepoint, or at the later timepoint, or at both timepoints. This would mean that all obtained data would be utilized and that for some subjects only pre or only post data would be in the analysis. For example, if infant subject number 1 was missing the 10-day score on visual following, that score would be missing from the analysis but his score at 30 days would be included. The elements of the vectors for the repeated measures analysis would be as shown in Figure 3.

When there are only two timepoints in the analysis, there is no gain in degrees of freedom in using this approach rather than approach number 3 above (eliminating all scores for the subject with missing data). Note that in Figure 2 there are six observations; the full

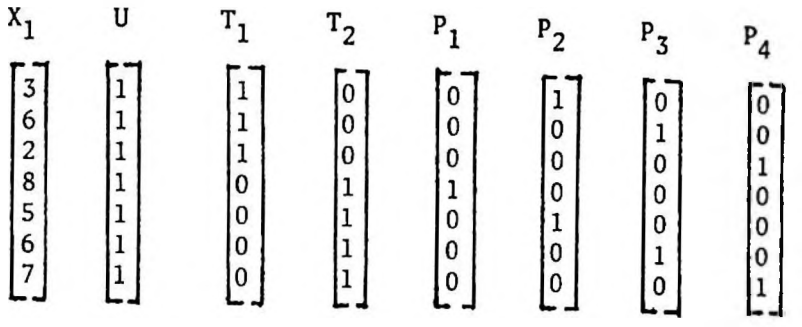


Figure 3. Elements of vectors when the 10-day score for subject 1 is eliminated, but the 30-day score for the same subject is retained.

model would contain four linearly independent vectors and the restricted model would contain three. The degrees of freedom when all scores for subject number 1 were eliminated would thus be 1 and 2. In Figure 3 there are seven observations; the full model would contain five linearly independent vectors (because P_1 appears) and the restricted model would contain four. The degrees of freedom when only the 10-day score is eliminated would thus be 1 and 2, the same as for Figure 2. The fact that the degrees of freedom are the same in both instances makes sense when you consider that, by including the P_1 vector along with the 30-day score for subject 1, the 30-day score for this subject is completely accounted for. Note that when there are more than two timepoints in the full model, the elimination of only one timepoint for a subject--rather than all timepoints for that subject--will result in a gain in denominator degrees of freedom. This makes sense because now the person vector for that subject does not completely account for that subject's variance.

The advantage in using this approach when there are only two timepoints is therefore not to be found in degrees of freedom. When there are more than two timepoints, it will yield a gain in df. But an advantage in both cases is that the increased number of observations yields a more stable estimate of the population mean on the criterion score, and this estimate is more representative of the population of subjects that the researcher set out to measure--not just subjects with complete data.

Covariates as an Alternative to Person Vectors

The above discussion has assumed that the use of person vectors is the path a researcher would wish to take when there is an expected correlation between individuals' scores at two timepoints.

But there is another approach that a researcher may wish to consider when dealing with measures that are repeated over time on the same subjects, and that approach is the use of a particular kind of covariate in place of the person vectors.

Using person vectors is essentially covarying for each person's uniqueness. It is acknowledging the expected correlation between pre and post scores by assuming that the reason a particular individual tends to score high (or low) at both timepoints in relation to other subjects is simply because he is that unique individual. But the researcher may know (or have reason to suspect) that there are measurable dimensions along which her subjects vary and which relate to the criterion behavior in explaining why individuals tend to score high (or low) at both timepoints. For example, our researcher may have evidence from prior studies that an infant's ability to focus on and follow a moving human face at both 10 and 30 days is related to the sex and the birth weight of the infant. Using scores on these two variables to predict the criterion score will probably yield a lower R^2 (greater errors of prediction) than predicting on the basis of which scores belong to which individuals (i.e., using person vectors), but it will be more valuable in a theory-building attempt as well as more generalizable to another group of subjects. It will also be a more parsimonious model. While using person vectors is a powerful technique for testing the hypothesis of differences between timepoints, it does not tell us anything about how individuals' differences on other variables causes them to be different from one another and yet consistent with themselves on the criterion measure. And since person vectors represent each person in the sample individually, they cannot be used to generalize beyond the sample.

Covariates which do not change across timepoints. It is not possible to use both person vectors and one or more covariates which do not differ for a subject across timepoints. An infant's sex and birth weight are this type of potential covariate. Whether one is considering a 10-day or 30-day criterion score, the sex of the infant would be the same, as would the birth weight of that infant. If a variable of this type were used in a regression model in conjunction with person vectors, a linear dependency would be generated. This is illustrated in Figure 4, where it is shown that a set of weights can be found such that the birth weight variable is a linear combination of the person vectors. A choice must be made, then, between the covariate and the person vectors. Another way of looking at this choice is to consider the hypotheses which represent the two choices. For the hypothesis, "the mean score on visual following will increase from 10 days to 30 days, over and above the effect of birth weight," the researcher would use the birth weight variable as a covariate. If the researcher chose instead to state the hypothesis, "the mean score on visual following will increase from 10 days to 30 days, over and above differences between individual mean scores on the criterion," person vectors would be used. This author would argue that, if the researcher has sufficient evidence to expect that a certain covariate

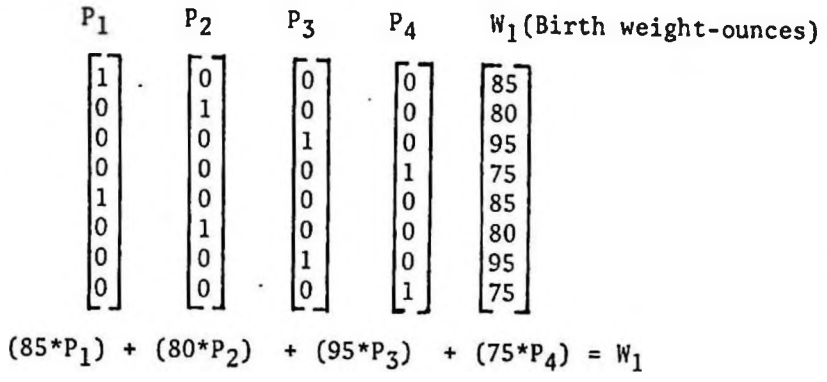


Figure 4. Linear dependency generated by the use of person vectors and a covariate which does not differ across timepoints.

or set of covariates is related to the criterion in a repeated measures design, it would be beneficial to state the hypothesis in terms of the covariate. The benefits are: (1) the covariate model is more parsimonious than the person vector model, (2) the findings are more generalizable to a new sample when the covariate is used than when person vectors are used, and (3) an advance can be made in theory building because the analysis containing the covariate gives a more refined estimate (than person vectors) of what specifically enters into individual differences on the criterion. In short, the researcher may wish to state the hypothesis in terms of a covariate rather than person vectors and thus will give up some of the power of the statistical test of differences over time in return for gaining generalizability of findings that will also aid in theory building. (Even if the researcher chooses to use person vectors, I would recommend that some thought and possibly additional analyses be devoted to accumulating evidence regarding variables which do relate to individual differences on the criterion--for possible use in the future.)

Covariates which change across timepoints. The above discussion has focused on covariates which do not change across timepoints. A short comment should be made about the use in a repeated measures design of covariates which do change across timepoints. It is possible to use in the same analysis both person vectors and covariates which differ across timepoints. If, for example, the covariate of interest were the weight of the infant at the time of testing, this would probably be different at 10 days and 30 days. The researcher might then state the following hypothesis: "the mean score on visual following will increase from 10 days to 30 days, over and above the effects of weight at the time of testing and differences between individual mean scores on the criterion." Examples of

vectors that would be used to test this hypothesis are shown in Figure 5. (A changing-over-time covariate might be useful in a case like this, in which the researcher may wish to know if the infants' scores increase over time, beyond the increase that could be predicted by weight gain alone. It should be pointed out, however, that in a repeated measures design which includes treatment as a factor, one would not wish to include a covariate on which change over time is a possible function of that treatment because the covariate may account for some of the variance in the criterion which is due to treatment.)

X_1	U	T_1	T_2	P_1	P_2	P_3	P_4	W_2 (Ounces at testing)
4	1	1	0	1	0	0	0	85
3	1	1	0	0	1	0	0	84
6	1	1	0	0	0	1	0	97
2	1	1	0	0	0	0	1	75
8	1	0	1	1	0	0	0	90
5	1	0	1	0	1	0	0	95
6	1	0	1	0	0	1	0	99
7	1	0	1	0	0	0	1	85

Figure 5. Example vectors for a situation in which the researcher wishes to use both person vectors and a covariate which changes over time.

Discussion

Two issues regarding the analysis of longitudinal or repeated measures data have been presented--the handling of missing data and the use of covariates. These concepts need not be applied separately in practice; it would be perfectly reasonable to use a covariate instead of person vectors with a set of data containing missing scores. Some additional comments on missing data, covariates, and the basic choice of a repeated measures analysis are presented in this section.

In handling the problem of missing data, the researcher needs to consider how many subjects are missing data on how many timepoints--this is a data-based question. It has been pointed out above that, when there are only two timepoints in an analysis, the use of subjects who have data for one timepoint in addition to those with complete data results in no additional degrees of freedom. The mean obtained is a better estimate of the population mean, but are the total results more generalizable to the population? This writer is not at all certain--my guess is that they are not. But when there are more than two timepoints, it would appear to me

that the use of subjects who have data for one or more timepoints would result in better estimates of both the population mean and the population variance and would yield results that are more generalizable to the population one initially intended to sample.

If one can refer to a question of the "power" of the test when discussing the use of a "repeated measures" analysis versus a "regular" analysis (person vectors versus no person vectors), then it seems appropriate to refer to "power" when the decision is between person vectors and a covariate. This writer would say, then, that the decision regarding the use of a covariate rather than person vectors in a repeated measures analysis is one of weighing theory and statistical power. But if the covariate or set of covariates account for nearly as much variance in the criterion as do the person vectors, and if there are far fewer covariates than person vectors, the covariate analysis may actually be more powerful (rather than less) because it is more parsimonious and thus generates greater degrees of freedom.

It seems appropriate at this point to develop one final issue: when should one use a "repeated measures" analysis. The argument was made early in this paper that, even though data is obtained on the same subjects at two timepoints, one may not necessarily wish to covary for differences in individuals' mean scores (i.e., use a repeated measures analysis). Just as it is not beneficial to covary on any variable which is not related to the criterion, it is not beneficial to control for a correlation between pre and post scores if no such correlation exists. The example was given earlier of a criterion-referenced test--one may expect all students to achieve criterion on the post test. In this case there would be no correlation between pre and post scores. However, the researcher in this case would probably not be interested in a hypothesis regarding an increase in scores from pre to post so the question of repeated measures would not apply. Consider, though, the situation in which a researcher is developing a measure of, say, mathematics concepts. In validating the measure, she wants to show that scores increase from pre to post and also that prescores are correlated with post-scores. She is interested in both questions--increasing scores and correlations between scores. The procedure she follows might be this: First she tests a number of subjects at two timepoints. Then she inspects the correlation between pre and post scores. This answers her question regarding the correlation. If there is a substantial correlation, she proceeds with a repeated measures analysis on the difference (increase) between pre and post scores. If on the other hand there is no correlation, she must decide if she is still interested in the question of increasing scores. In the validation of the instrument, this may still be a question of interest. She would therefore test the difference between pre and post scores--without a repeated measures design (it would be an "uncorrelated" t test).

This paper has presented the repeated measures type of analysis as it is formulated in regression models. Issues regarding its use

with missing data and covariates have been presented. Before making decisions on these issues and applying those decisions to data analysis, the researcher must first decide (1) what hypothesis is being asked, and (2) how a lack of correlation between timepoints will affect that hypothesis.

NOTE--added after presentation of paper at AERA, 1975:

As a result of discussion following the presentation of the above paper, the following note is added.

In the section on missing data above, it was not stated what the restriction on the full model and the resultant restricted model would be when partial data is used. One would not necessarily want to assume that the restriction is the same as in the example case with complete data.

If one decides to use the restriction used in the example case ($a_1=a_2$), I believe the research hypothesis being tested would be, "The mean score at one timepoint is greater than the mean score at another timepoint, assuming that the person with the missing score is conceptually average--that is, assuming (1) that the missing score is equal to the average of the obtained scores at that timepoint and (2) that the score(s) which was obtained for that individual has a deviation from its timepoint mean which is equal to the average deviation of all other individuals' scores from the grand mean.

If one did not wish to make this assumption, the logic to be followed in making the restriction on the full model would be as follows: First, return to the case with complete data (Model 1 and Figure 1). The restriction to test the stated hypothesis--with complete data--is actually:

$$\frac{(a_1+a_3)+(a_1+a_4)+(a_1+a_5)+(a_1+a_6)}{4} = \frac{(a_2+a_3)+(a_2+a_4)+(a_2+a_5)+(a_2+a_6)}{4}$$

which simplifies to:

$$\frac{4a_1 + (a_3+a_4+a_5+a_6)}{4} = \frac{4a_2 + (a_3+a_4+a_5+a_6)}{4}$$

$$\frac{4a_1}{4} = \frac{4a_2}{4}$$

$$a_1 = a_2$$

This restriction on Model 1 results in restricted Model 2 above.

Applying the above sequence to the case in Figure 3 where there is no prescore for subject 1, the restriction would be:

$$\frac{(a_1+a_4)+(a_1+a_5)+(a_1+a_6)}{3} = \frac{(a_2+a_3)+(a_2+a_4)+(a_2+a_5)+(a_2+a_6)}{4}$$

which simplifies to:

$$\frac{3a_1 + (a_4+a_5+a_6)}{3} = \frac{4a_2 + (a_3+a_4+a_5+a_6)}{4}$$

$$a_1 + \frac{(a_4+a_5+a_6)}{3} = a_2 + \frac{(a_3+a_4+a_5+a_6)}{4}$$

$$a_1 = a_2 + \frac{(a_3+a_4+a_5+a_6)}{4} - \frac{(a_4+a_5+a_6)}{3}$$

This restriction is complicated but could be algebraically applied to the full model (Model 1) and would result in a model conceptually similar to but looking very different from Model 2.

REFERENCES

Dixon, W.J. and Massey, F.J. Introduction to Statistical Analysis, Third Edition. New York: McGraw-Hill, 1969, (p. 120).

Downie, N.M. and Heath, R.W. Basic Statistical Methods, Third Edition. New York: Harper and Row, 1970, (p. 177).

Glass, G., Peckham, P., and Sanders, J. Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. Review of Educational Research, 1972, 42(3), 237-288.

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

Pohlmann, J.T. and McShane, M.G. Applying the general linear model to repeated measures problems. Paper presented to the Annual Meeting of the American Educational Research Association, Chicago, 1974.

Multiple Linear Regression Viewpoints
Vol. 6, No. 1, 1975

STWMULTR; A COMPUTER PROGRAM TO EXPEDITE THE RETRIEVAL OF RESIDUAL SCORES

Carl E. Edeburn and David P. Ochsner

South Dakota State University

Summary - Residual gain analysis was described in general terms and a new computer program, STWMULTR, designed to retrieve and punch residual scores was described. Samples of input and output data cards were included.

One of the most difficult tasks in psycho-educational research investigation has been the measurement of change. When pre and post-testing models have been implemented, the most prevalent application in most analyses has been the use of raw gain scores. Due to the inadequacies of this approach several different solutions, including residual gain analysis, have been proposed.

The residual gain analysis approach has been discussed by Dubois (1957, 1970) and Bakan (1970). Others, (Williams and Maresh, 1972, Buzzahora and Williams, 1973 and Edeburn and Landry, 1974) have applied this technique in test-retest situations using elementary school students' scores as elicited by various cognitive and affective measures.

Essentially, the residual gain method can be concieved as a partial correlation between the group membership variable and the residuals in the posttest data using the pretest as a predictor.

One of the physical limitations of the residual gain application has been the amount of clerical time spent in extracting the residual scores for each student from the computer printout, and repunching them

on appropriately identified cards. The present effort was aimed at overcoming this limitation.

To a common version of a stepwise multiple linear regression program (STWMULT) originally adapted from the Scientific Subroutine Package (IBM, 1972), the present authors have added an optional feature which stores and then punches the residual scores for each subject on a new data card. This new program is identified in the S.D.S.U. Computing Center as STWMULTR. The only restriction in the STWMULTR version is that the first card for each subject in the original data set must include 1-8 columns of Alphanumeric I.D. (see Figure 1.).

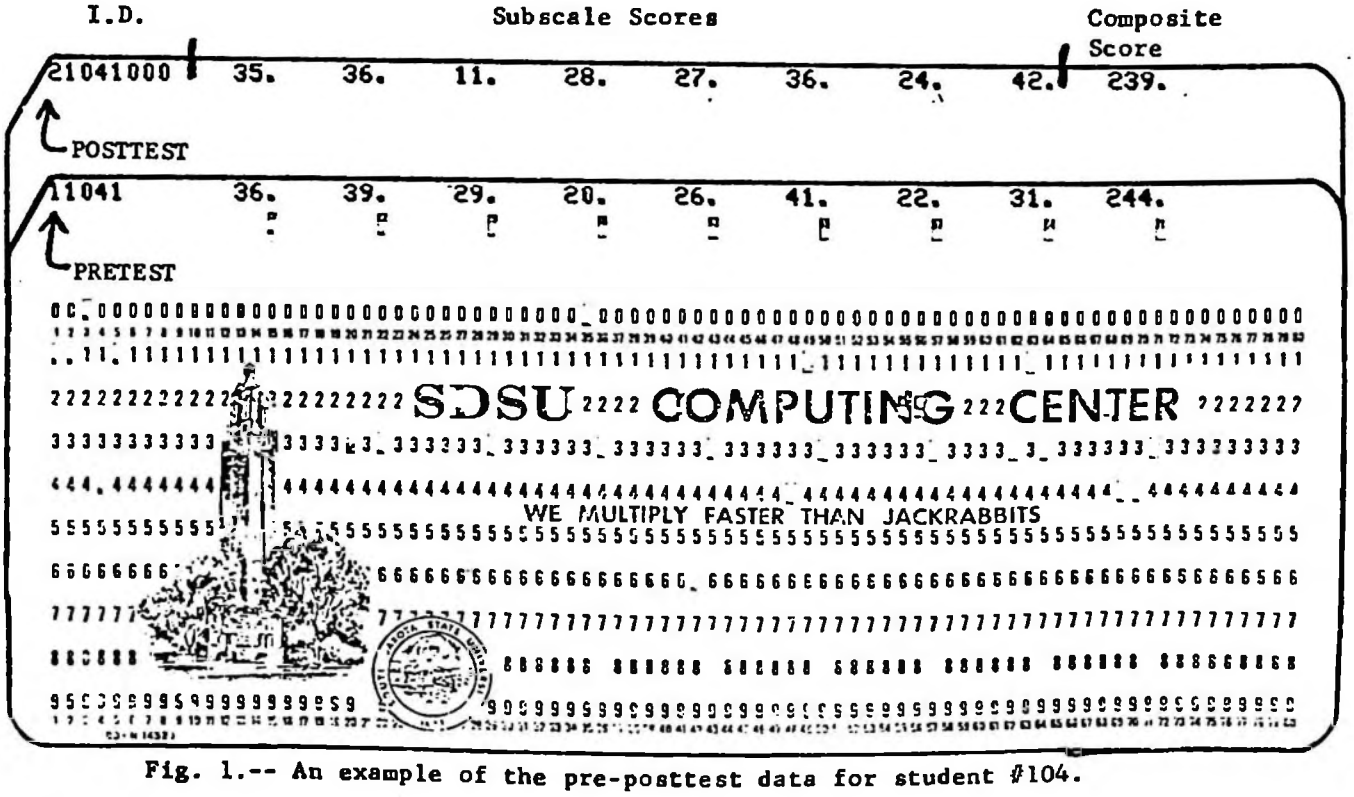


Fig. 1.-- An example of the pre-posttest data for student #104.

By using individual selection cards (pretest predicting posttest on each subscale) to determine the residuals, and calling for the punch option, the researcher is able to obtain a new set of data cards containing up to 8 columns of the original alphanumeric I.D., and the residual gain scores for each subject. As will be noted in Figure 2., these scores are punched in fields of F 7.3, and there is room for 10 scores on the card. Should the research in question re-

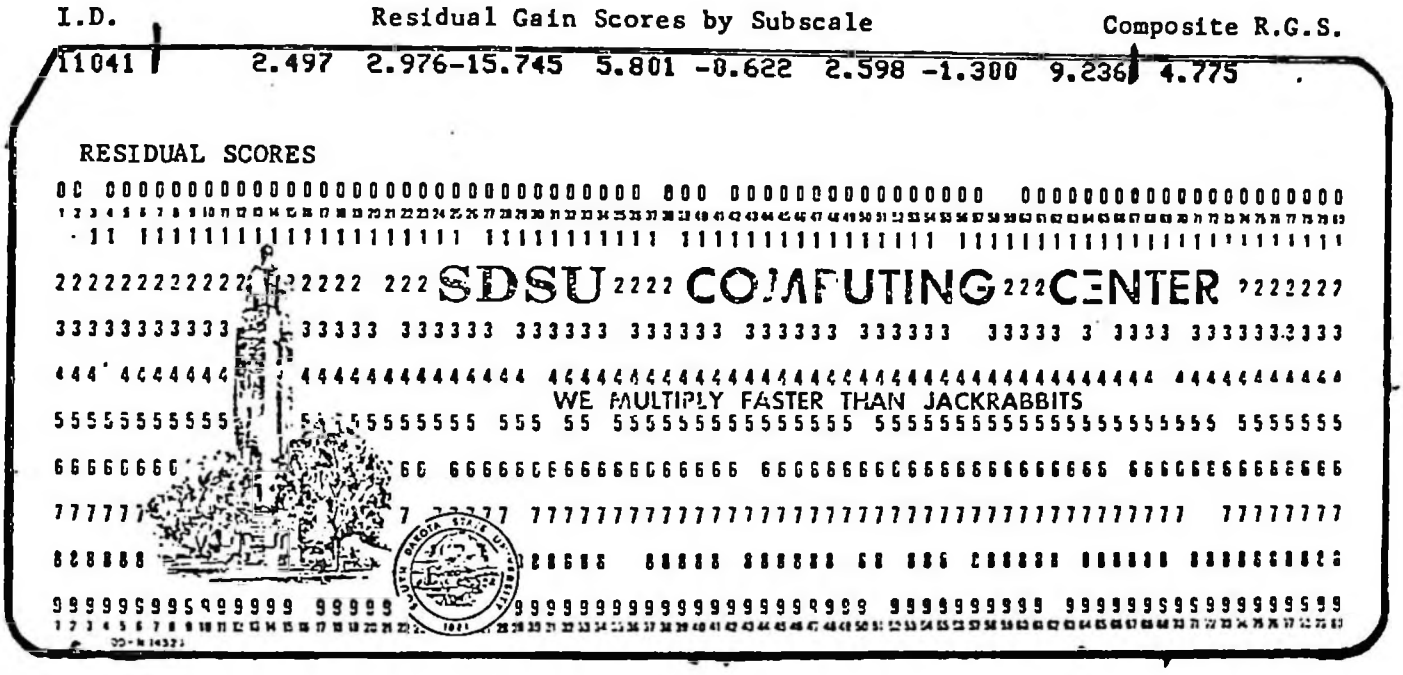


Fig. 2.--An example of the residual gain data for student #104.

quire more than 10 scores, the program will continue to punch on additional cards in the same format, retaining the I.D. for each new card.

After obtaining the new cards, normal regression procedures can be easily implemented.

Needless to say, STWMULTR has enhanced the present investigators' abilities in pursuing residual gain problems. Interested users may contact either of the authors for further details.

REFERENCES

Bakan D. Analysis of covariance in training research. In P.H. Dubois and G.D. Mayo (Eds.), Research strategies for evaluation training. Chicago: Rand McNalley, 1970.

Buzzahara R.C. & Williams J.D. An empirical comparison of residual gain analysis and the analysis of covariance. Multiple Linear Regression Viewpoints, 1973, 4, 10-17.

Dubois P.H. Multivariate correlational analysis. New York: Harper, 1957.

Dubois P.H. Correlational analysis in training research. In P.H. Dubois and G.D. Mayo (Eds.), Research strategies for evaluation training. Chicago: Rand McNalley, 1970.

Edeburn C.E. & Landry R.G. Teacher self concept and student self concept. Paper presented at the annual convention of the American Educational Research Association, April 16, 1974.

IBM Concepts and applications of regressions analysis 3rd. Ed., White Plains, New York: Author, 1972.

Williams, J.D. & Maresh, R.T. A comparison of raw gain scores, residual gain scores, and the analysis of covariance with two modes of teaching reading. Multiple Linear Regression Viewpoints 1972, 2, 2-16.

Multiple Linear Regression Viewpoints
Vol. 6, No. 1, 1975

THE ANALYSIS OF COVARIANCE WITH RANDOMIZED BLOCKS DESIGNS BY REGRESSION

John D. Williams and John G. Watson
The University of North Dakota

A regression solution is given for a research situation that includes both the analysis of covariance and randomized blocks. Basically, the solution includes the successive use of three linear models. The first model uses the covariate as the predictor while the second model uses both the covariate and the group membership variables; the difference (in R^2 units) between these two models is the proportion of the variance that is attributable to the group membership variables independent of the covariate. The third model includes the covariate, the group membership variables and the blocks. The difference (in R^2 units) between the third model and the second model is the proportion of the variance due to the blocks independent of both the group variables and the covariate.

Over the last several years, several writers have shown regression analogs to analysis of variance designs; the usual designs (t test, one way analysis of variance, two way analysis of variance, the analysis of covariance or randomized blocks designs) have found their way into recently written texts from a regression viewpoint (Ward and Jennings, 1973; Williams, 1974, and McNeil, Kelly, and McNeil, 1975). More complex designs have been explicated in a regression approach as well. Pyle (1974) showed a regression solution to the split-plot (AxBxS) design. Other repeated measures designs have been considered by Olson (1973), Fanning and Newman (1974) and Pohlman and McShane (1974). Higher order factoral analysis designs with disproportionate cell frequencies has been considered by Williams (1974a) and Deitchman, Newman, Burkholder and Sanders (1974).

Need for the Design

In extending regression to a solution to a randomized blocks design with an analysis of a covariance, necessity, as usual, is the "mother of invention". In a substantive application, the following problem presented itself. An analysis of the difference in motivation of black and white managers was being investigated. Black and white managers (N = 64 pairs) were chosen such that one black and one white manager from the same participating organization were included in the sample; in two companies, two pairs were chosen, so that 62 different business organizations were selected (four commercial banking companies, four life insurance companies, four transportation companies, eight utilities, 36 industrials, and six assorted companies). The Thematic Apperception Test (TAT) was mailed to each participant; detailed directions for self administration of the test were included.

As the TAT scoring utilizes count information for measuring the strength of various motives (need for achievement, need for affiliation and need for power), it was felt, in absence of administering the instruments to each participant and thus keep control over time, that the use of number of words in the TAT stories might serve as a useful covariate to help achieve statistical control. Thus, the necessary statistical analysis was seen to be randomized block design (each pair of a black and white managers serves as a block) with an analysis of covariance imposed upon the design (to control for word length). While the scores on each of the 128 subjects is not duplicated here, a prototype of the solution is given.

An Example

To see how a solution can be formulated by regression for a randomized block design with the analysis of covariance, a hypothetical example with 12 blocks, three groups of subjects and a covariate are given. First, several variables can be defined:

Y = the criterion variables;

X_1= the covariate;

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

X_3= 1 if a score from a member of the second group, 0 otherwise;

X_4= 1 if a score from block 1, 0 otherwise, and

X_5-X_{13}= 1 if a score respectively from blocks 2-11, 0 otherwise.

Table 1 contains the information for a regression solution.

To accomplish this analysis, it might first be recalled that the analysis of covariance attemps to assess the independent contribution of the group membership variables; on the other hand, the randomized block design is usually employed to increase the precision (alternatively, reduce the error variance) in an experiment.

With this distinction in mind, three linear models can be defined:

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

TABLE 1

Data for Randomized Block Design with the Analysis of Covariance

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
27	18	1	0	1	0	0	0	0	0	0	0	0	0	0
29	20	0	1	1	0	0	0	0	0	0	0	0	0	0
22	16	0	0	1	0	0	0	0	0	0	0	0	0	0
27	22	1	0	0	1	0	0	0	0	0	0	0	0	0
29	27	0	1	0	1	0	0	0	0	0	0	0	0	0
25	20	0	0	0	1	0	0	0	0	0	0	0	0	0
24	24	1	0	0	0	1	0	0	0	0	0	0	0	0
25	28	0	1	0	0	1	0	0	0	0	0	0	0	0
28	25	0	0	0	0	1	0	0	0	0	0	0	0	0
23	21	1	0	0	0	0	1	0	0	0	0	0	0	0
28	27	0	1	0	0	0	1	0	0	0	0	0	0	0
17	20	0	0	0	0	0	1	0	0	0	0	0	0	0
17	18	1	0	0	0	0	0	1	0	0	0	0	0	0
25	20	0	1	0	0	0	0	1	0	0	0	0	0	0
22	19	0	0	0	0	0	0	1	0	0	0	0	0	0
23	23	1	0	0	0	0	0	0	1	0	0	0	0	0
28	21	0	1	0	0	0	0	0	1	0	0	0	0	0
27	22	0	0	0	0	0	0	0	1	0	0	0	0	0
21	20	1	0	0	0	0	0	0	0	1	0	0	0	0
27	20	0	1	0	0	0	0	0	0	1	0	0	0	0
19	20	0	0	0	0	0	0	0	0	1	0	0	0	0
26	25	1	0	0	0	0	0	0	0	0	1	0	0	0
27	23	0	1	0	0	0	0	0	0	0	1	0	0	0
23	24	0	0	0	0	0	0	0	0	0	1	0	0	0
23	24	1	0	0	0	0	0	0	0	0	0	1	0	0
25	22	0	1	0	0	0	0	0	0	0	0	1	0	0
22	21	0	0	0	0	0	0	0	0	0	0	1	0	0
31	26	1	0	0	0	0	0	0	0	0	0	0	1	0
28	20	0	1	0	0	0	0	0	0	0	0	0	1	0
24	25	0	0	0	0	0	0	0	0	0	0	0	1	0
33	30	1	0	0	0	0	0	0	0	0	0	0	0	1
27	20	0	1	0	0	0	0	0	0	0	0	0	0	1
22	25	0	0	0	0	0	0	0	0	0	0	0	0	1
34	31	1	0	0	0	0	0	0	0	0	0	0	0	0
38	27	0	1	0	0	0	0	0	0	0	0	0	0	0
31	27	0	0	0	0	0	0	0	0	0	0	0	0	0

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e_2, \text{ and} \quad (2)$$

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + \ldots + b_{14}X_{14} + e_3. \quad (3)$$

In general, b_0 will be different for equations 1-3. Similarly, b_1-b_3 will be respectively different for equations 1-3. Equation 1 might be thought of as a model wherein the covariate is used as a predictor of the criterion; the error variance for equation 1 is the total variation for the remaining sources of variation (groups, blocks, and error variance). The second equation allows the assessment of the contribution of the group membership variables independent of the covariate. In turn, the error variance of equation 2 can be thought of as due to two sources: blocks and error variance.

The sum of squares for the four sources of variation can be easily determined through the use of squared multiple correlation (R^2) terms. The sum of squares for the covariate (SS_C) will equal the sum of squares total (SS_T) times the R^2 term for equation 1:

$$SS_C = (SS_T)\ R_1^2.$$

The sum of squares for the group membership variables will be

$$SS_G = (SS_T)\ (R_1^2 - R_2^2).$$

The sum for the blocks is given by

$$SS_B = (SS_T)\ (R_3^2 - R_2^2).$$

Finally, the error sum of squares is given by

$$SS_E = (SS_T)\ (1 - R_3^2).$$

Table 2 contains a summary table with this information.

TABLE 2

Summary Table for the Analysis of Covariance with Randomized Blocks Design

Source of variation	df	SS	MS	F
Covariate	1	254.125		
Groups	2	92.279	46.140	6.446*
Blocks	11	196.033		
Error	21	150.313	7.158	
Total	35	692.750		

* significant at .01 level.

Some readers might recognize that the process described here is actually a hierachical model with the following ordering: covariate, groups and blocks. In that this is a hierachical model, some researchers might prefer the following ordering: covariates, blocks and groups, finding that part of the variation with the group membership variables independent of both the covariate and blocks. While these writers have no quarrel with those who prefer the second ordering if it fits their research purpose, it should be pointed out that the first ordering utilized the blocks as a way to increase the precision of the experiment, whereas the second ordering could be concievably construed as an analysis of covariance where both the covariate and the blocks are covaried!

Finding The Adjusted Means

If there is interest in finding the adjusted means, then the following equation can be utilized:

$$\overline{Y}_i adj = \overline{Y}_i - b(\overline{X}_i - \overline{X}_T) \qquad (4)$$

where Y = the regression coefficient for the covariate in equation 2, $\overline{X}_i$ = the mean of the ith group on the covariate, and the $\overline{X}_T$ = the overall mean of the covariate.

For the data in Table 1, b = .72667, $\overline{Y}_1$ = 25.75, $\overline{Y}_2$= 28.00, $\overline{Y}_3$= 23.50, $\overline{X}_1$ = 23.50, $\overline{X}_2$ = 22.92, $\overline{X}_3$= 22.00 and $\overline{X}_T$ = 22.81. Then $\overline{Y}_1 adj$= 25.75-.72667(23.50-22.81)= 25.25. In a similar manner, $\overline{Y}_2 adj$= 27.92 and $\overline{X}_3 adj$ = 24.09. If more than one covariate is desired, it can be easily included. The adjusted means are found in a manner analogous to that described here. Complete details regarding the use of multiple covariates are given in Williams (1974b, pp. 109-112, 115.)

Study On Black and White Managers

In the comparison of black and white managers, Watson and Williams (1975) used the models as described here. Their result for need for power is shown in Table 3. White managers were found to have a significantly higher need for power. Interestingly, the use of randomized blocks slightly decreased the efficiency of the experiment, in that had the blocking not taken place, the mean square for error would have been slightly smaller. Given the manner in which the blocks were formed, (a paired black and white manager from a given company), it is still intuitively a "cleaner" study to include the blocks.

TABLE 3

Analysis of Covariance with Randomized Blocks Design for Comparing Black and White Managers On Need for Power

Source of variation	df	SS	MS	F
Words	1	122.200		
Group	1	32.164	32.164	4.710*
Subjects	63	358.257		
Error	62	423.357	6.828	
Total	127	935.978		

* significant at .05 level

References

Deitchman, R., Newman, I., Burkholder, J., and Sanders, R.E. A application of the higher order factorial analysis design with disproportionality of cells. Multiple Linear Regression Viewpoints, 1974, 5, No. 2, 45-57.

Fanning, F.W. and Newman, I. The development and demonstration of multiple regression models for operant conditioning questions. Multiple Linear Regression Viewpoints, 1974, 4, No. 4, 65-87.

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

Olson, G.H. A general least squares approach to the analysis of repeated measures designs. Multiple Linear Regression Viewpoints, 1973, 4, No. 3, 60-79.

Pohlman. J.T. and McShane, M.G. Applying the general linear model to repeated measure problems. Presented at the American Educational Research Association Convention, Chicago, April 17, 1974.

Pyle, T.W. Classical analysis of variance of completely within-subject factorial designs using regression techniques. Multiple Linear Regression Viewpoints, 1974, 5, No. 1, 23-32.

Ward, J.H. and Jennings, E. Introduction to Linear Models. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Watson, J.G. and Williams, J.D. A study of the achievement, power, and affiliation motivation of black and white managers. Proceedings of the 18th Annual Conference of the Midwest Division of the Academy of Management, 1975, 189-196.

Williams, J.D. Four-way disproportionate hierarchical models. Multiple Linear Regression Viewpoints, 1974a, 5, No. 2, 32-40.

Williams, J.D. Regression Analysis in Educational Research. New York: MSS Information Corp., 1974b.

Multiple Linear Regression Viewpoints
Vol. 6, No. 1, 1975

A REGRESSION FORMULATION OF DUNN'S AND SCHEFFÉ'S TESTS

John D. Williams
University of North Dakota

Perhaps one of the most widely disseminated multiple comparison procedures is Scheffé's (1953) S-method for judging all possible contrasts. As Scheffé's test has the property that it can be used on an a posteriori basis after the rejection of the overall F test in the analysis of variance and still maintain an experimentwise error rate, applied researchers find the test to be particularly useful in investigating unplanned contrasts that are suggested by the data.

On the other hand, Dunn (1961) devised a multiple comparison technique that allows the researcher to state the hypotheses to be tested on an a priori basis; if the number of comparisons to be made can be severely limited, Dunn's test can be the most powerful multiple comparison method under these circumstances, and retains a per experiment error rate. Dunn's test would be particularly appropriate in those theory testing situations that require less than all possible simple comparisons of means.

Both Tukey's (1953) and Dunnett's (1955) tests have been explored in a regression formulation. Both tests have been shown to be rather simple applications of using a regression approach, being a by-product of testing each regression coefficient for significance (Williams, 1971, 1974).

An Example

In that Dunn's test requires a severe limitation of the number of comparisons of interest, suppose the comparisons of interest when four groups are being included in the analysis are

$\bar{X}_1$ to $\bar{X}_3$

$\bar{X}_1$ to $\bar{X}_4$

$\bar{X}_2$ to $\bar{X}_4$.

Before completing the analysis, comments on other multiple comparison procedures with these specific hypotheses should be made. First, Dunnett's test would not apply, as there is no single group being compared to other groups. Orthogonal comparisons would not apply, as the stated comparisons do not transform to an orthogonal set of comparisons. Both Tukey's and Scheffé's tests would appropriately contain these comparisons (among a much larger set), but with considerably less power. Thus, the kind of experimental situation implied in the posited comparisons fits very well the test proposed by Dunn.

TABLE 1

Data From Four Hypothetical Groups

Group One	Group Two	Group Three	Group Four
9	8	13	15
8	7	10	12
6	8	12	10
3	6	11	17
4	6	14	11

$\overline{X}_1 = 6.0$, $\overline{X}_2 = 7.0$, $\overline{X}_3 = 12.0$ and $\overline{X}_4 = 13.0$

In that the three posited comparisons are "simple" in that no more than two means are being compared at any one time, a methodology very similar to that given in Williams (1971, 1974) can be employed. The process involves employing a pseudo-replication of Dunnett's test until all comparisons can be made. The comparisons of $\overline{X}_1$ to $\overline{X}_4$ and $\overline{X}_1$ to $\overline{X}_3$ can be found from using the linear model

$$Y = b_0 + b_2X_2 + b_3X_3 + b_4X_4 + e_1. \qquad (1)$$

The comparison of $\overline{X}_2$ to $\overline{X}_4$ can be found from using the linear model

$$Y = b_0 + b_1X_1 + b_3X_3 + b_4X_4 + e_1. \qquad (2)$$

For equations 1 and 2,

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

X_2 = 1 if the score is from a member of Group Two;
0 otherwise,
X_3 = 1 if the score is from a member of Group Three;
0 otherwise,
X_4 = 1 if the score is from a member of Group Four:
0 otherwise,
b_0 = the Y-intercept,

b_1 = the regression coefficient for Group One,

b_2 = the regression coefficient for Group Two,

b_3 = the regression coefficient for Group Three,

b_4 = the regression coefficient for Group Four; and

e_1 = the error involved in prediction.

It should be noted that the value for b_0 will in general be different for each equation. In fact, it has been shown (Williams, 1971) that b_0 will equal the mean of the group that is being compared to the other groups. For example, in equation 1, $b_0 = \bar{X}_1$. Similarly, b_1, b_2, b_3 and b_4 will differ for the two equations. For example, $b_3 = 6$ for equation 1, and $b_3 = 5$ for equation 2. It can also be noted that the regression coefficients are the difference between the means of each group to the group "left out" in a particular analysis. Thus, $X_3 - X_1 = 12 - 6 = 6$, which of course is b_3 for equation 1. The value for e_1 will remain the same for each of the equations, as each of these equations can be conceptualized as a reparametrization of an analysis of variance model.

The information necessary for a regression solution to Dunn's and Scheffé's tests is given in Table 2.

TABLE 2
Regression Formulation for Dunn's and Scheffé's Tests

Y	X_1	X_2	X_3	X_4
9	1	0	0	0
8	1	0	0	0
6	1	0	0	0
3	1	0	0	0
4	1	0	0	0
8	0	1	0	0
7	0	1	0	0
8	0	1	0	0
6	0	1	0	0
6	0	1	0	0
13	0	0	1	0
10	0	0	1	0
12	0	0	1	0
11	0	0	1	0
14	0	0	1	0
15	0	0	0	1
12	0	0	0	1
10	0	0	0	1
17	0	0	0	1
11	0	0	0	1

Using the two linear models, most computer programs include a "computed t value" for each regression coefficient. Each linear model generates three computed t values; however, only three of the six computed t values are of interest.

The computed t values for equation 1 are

t = .735 for comparing $\overline{X}_1$ to $\overline{X}_2$ (not a comparison of interest),
t = 4.411 for comparing $\overline{X}_1$ to $\overline{X}_3$, and

t = 5.147 for comparing $\overline{X}_1$ to $\overline{X}_4$.

The computed t values for equation 2 are

t = -.735 for comparing $\overline{X}_2$ to $\overline{X}_1$ (not a comparison of interest),
t = 3.676 for comparing $\overline{X}_2$ to $\overline{X}_3$ (not a comparison of interest),
t = 4.411 for comparing $\overline{X}_2$ to $\overline{X}_4$.

To evaluate these t values, Dunn's (1961) table can be used. With 3 comparisons and $df_w = 16$, the critical value is 3.45, with $\alpha = .01$.

TABLE 3

Computed t Values for Dunn's and Scheffé's Tests in a Regression Formulation

Comparison	t
$\bar{X}_1$ to $\bar{X}_3$	4.411*
$\bar{X}_1$ to $\bar{X}_4$	5.147*
$\bar{X}_2$ to $\bar{X}_4$	4.411*

*Significant at the .01 level

The formulation given for Dunn's test is identical for completing Scheffé's test. The only difference in the interpretation lies in determining the critical value; for Scheffé's test,

$S = (k-1)_{\alpha}F_{k-1,n-k}$, which, for $\alpha = .01$, $k = 4$, $n = 20$, is $S = 3.98$. The corresponding critical value for Tukey's test is $\underline{q}$ or 3.67. That Dunn's test is $\sqrt{2}$ more powerful than Scheffé's test for this set of posited simple comparisons is due to the limited number of comparisons. For the application given, then, Dunn's test is the most powerful and Scheffé's test the least powerful, with Tukey's test being intermediate. For the three comparisons, given, however, significance at the .01 level for all three comparisons by all three tests was attained.

Complex Comparisons

Both Dunn's test and Scheffé's test allow for more complex comparisons (more typically called contrasts) of the means than is indicated by the three comparisons already shown. The more complex comparisons typically will require a more complex solution than was demonstrated for the simple comparisons.

Suppose the following four comparisons were of interest:

$\frac{1}{2}\overline{X}_1 + \frac{1}{2}\overline{X}_2$ to $\frac{1}{2}\overline{X}_3$ to $\frac{1}{2}\overline{X}_4$

$\frac{1}{2}\overline{X}_1 + \frac{1}{2}\overline{X}_2$ to $\overline{X}_4$

$1/3\overline{X}_1 + 2/3\overline{X}_2$ to $3/7\overline{X}_3 + 4/7\overline{X}_4$

$\overline{X}_2$ to $\overline{X}_3$.

There are several approaches to solving for these comparisons. One approach is to define full and restricted models, following the methodology of Ward and Jennings (1973). To consider the first comparison, the full model is identical to any one of the previously given equations 1 or 2, as they all yield the same R^2 value of .71429. Actually, the full model can be thought of as $Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_1$. (3)

In that most computer programs automatically generate a constant term (b_0), equation 3 will typically not produce a desired result; if, for example, b_4X_4 is thought of as being b_0, then one solution would be to use equation 1, where b_0 will equal $\overline{X}_4$. The restricted model could then be formed from the restriction implied in the comparison $\frac{1}{2}b_1 + \frac{1}{2}b_2 = \frac{1}{2}b_3 + \frac{1}{2}b_4$. If this restriction is shown in terms of b_1, it is given as $b_1 = b_3 + b_4 - b_2$. If this restriction is made on the full model (equation 3)

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

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

or

$$Y = b_2V_1 + b_3V_2 + b_4V_3 + e_2, \text{ where}$$

V_1 = 1 if a member of Group Two, -1 if a member of Group One, 0 otherwise;
V_2 = 1 if a member of Group Three, 1 if a member of Group One, 0 otherwise;
V_3 = 1 if a member of Group Four, 1 if a member of Group One, 0 otherwise.

Then, a restricted model could be used which uses any <u>two</u> of the variables of V_1, V_2 and V_3. Using V_1 and V_2, the restricted model is

$$Y = b_0 + b_5V_1 + b_6V_2 + e_2. \qquad (4)$$

The resulting R^2 for equation 4 is .01931.

To evaluate the significance of this restriction, the equation

$$F = \frac{(R^2_{FM} - R^2_{RM})/1}{(1 - R^2_{FM})/df_w}, \qquad (5)$$

can be used, where R^2_{FM} refers to the R^2 value from the full model, R^2_{RM} refers to the R^2 value from the restricted model and df_w refers to the degrees of freedom for the within term in the full model.

For the comparison $\frac{1}{2}X_1 + \frac{1}{2}X_2$ to $\frac{1}{2}X_3 + \frac{1}{2}X_4$,

$$F = \frac{(.71429 - .01931)/1}{(1 - .71429)/16} = 38.925.$$

If $\sqrt{F}$ is found, $\sqrt{F} = 6.239$, which can be interpreted as a t value for Dunn's test (or as the value of S in Scheffé's test) for this comparison.

A simpler approach can be used to gain the same result. If a new predictor V_4 is defined as $V_4 = \frac{1}{2}$ for a member of Group One, $\frac{1}{2}$ for a member of Group Two, $-\frac{1}{2}$ for a member of Group Three or $-\frac{1}{2}$ for a member of Group Four, so that V_4 is a direct utilization of the comparison of interest,

then the R^2 value found from using V_4 as a predictor is .69498, which is equal to $R^2_{FM} - R^2_{RM}$: .71429 - .01931 = .69498. The value for R^2 for the use of V_4 can be directly utilized as equal to $R^2_{FM} - R^2_{RM}$ in equation 5. Similar solutions can be given for the comparisons

$$\tfrac{1}{2}\overline{X}_1 + \tfrac{1}{2}\overline{X}_2 \text{ to } \overline{X}_4 \quad (F = 30.463,\ t = 5.519)$$

$$1/3\overline{X}_1 + 2/3\overline{X}_2 \text{ to } 3/7\overline{X}_3 + 4/7\overline{X}_4 \quad (F = 35.380,\ t = 5.948)$$

and

$$\overline{X}_2 \text{ to } \overline{X}_3 \quad (F = 13.513,\ t = 3.676).$$

The last comparison is of particular interest in that it is of the type given earlier for simple comparisons and is in fact identically equal to the earlier obtained result from the value given from the use of equation 2, although it was not then a comparison of interest.

To evaluate these comparisons for significance, the tabled value for Dunn's test at the .01 level for four comparisons is 3.59. Each of the four comparisons are significant. The critical value for Scheffé's test remains as 3.98. While Tukey's test also yields significance on the four posited tests, it is of interest to note that on each of the three complex comparisons, Scheffé's test has a shorter interval than does Tukey's test.

As was noticed, including four comparisons rather than three reduced somewhat the advantages of Dunn's test over Scheffé's test. For $\alpha = .01$, $k = 4$, $df_w = 16$, Dunn's test will be more powerful than Scheffé's test for any number of comparisons up to 9 comparisons. Beyond 9 comparisons, Scheffé's test is more powerful. As both Dunn's and Scheffé's tests employ the same standard error for a contrast, they are easily comparable. On the other hand, comparisons to Tukey's test are more difficult.

Tukey's test maximizes its power with simple comparisons of the form $\bar{X}_i - \bar{X}_j$; for complex comparisons Scheffé's test tends to be more powerful.

If all seven of the previously given comparisons had been posited a priori, then Dunn's test has a critical value of 3.86, with Scheffé's test critical value remaining at 3.98. For such a situation, Tukey's test would be more powerful for the four simple comparisons, but both Dunn's and Scheffé's tests are more powerful on the complex comparisons, with Dunn's test being slightly the more powerful of the two.

References

Dunn, O. J. "Multiple Comparisons Among Means." Journal of the American Statistical Association. 1961, 56: 52-64.

Dunnett, C.W. "A Multiple Comparison Procedure for Comparing Several Treatments with a Control." Journal of the American Statistical Association, 1955, 50: 1096-1121.

Scheffé, H. "A Method for Judging all Contrasts in the Analysis of Variance." Biometrics, 1953, 40: 87-104.

Tukey, J.W. "The Problem of Multiple Comparisons." Dittoed, Princeton University, 1953.

Ward, J.H., and Jennings, E. Introduction to Linear Models. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Williams, J.D. "A Multiple Regression Approach to Multiple Comparisons for Comparing Several Treatments with a Control." Journal of Experimental Education, 1971, 39: No. 3, 93-96.

Williams, J.D. "A Simplified Regression Formulation of Tukey's Test." Journal of Experimental Education, 1974.

Multiple Linear Regression Viewpoints
Vol. 6, No. 1, 1975

THREE REASONS WHY PERCENT VARIANCE ACCOUNTED FOR IS IMPORTANT TO THE DEVELOPMENT OF THEORY*

by
James Gillham and Darlene Napady
THE UNIVERSITY OF AKRON

Abstract

Percent variance accounted for describes the degree of ambiguity in a test of a theory. This percentage is a parsimonious statement of the relative success of each attempt to solve a particular puzzle; it is also a guide to forming still better solutions.

At professional meetings discussants can find themselves confronted with a group of papers, nominally related in subject matter but so divergent in theoretical approach that substantive comparisons are difficult. In a larger sense, of course, this dilemma faces anyone wishing to evaluate the goodness of one theory against that of another. Although it has received little attention in this regard, the percent of variance accounted for can serve as a highly practical tool for such comparisons of theories (Klein and Newman, 1974; Byrne and McNeil, 1975).

Three papers presented at one session of the American Sociological Association's 1974 annual meeting illustrate the enormous differences that can be uncovered by this approach. Among these three papers, the

---

*We are indebted to Keith McNeil for his very helpful comments on an earlier version.

percentage of variance accounted for ranged from an astounding 96 percent down to a miniscule 11 percent.

The paper (Segal, 1974) with the astonishingly high percentage of variance figure reported a Spearman's rho of .98. It provided rather unambiguous support for the relationship between spatial propinquity and interpersonal attraction.

The second paper (Moschetti, 1974) accounted for 64 percent of the variance between written instructions and continuance of the criterion behavior. Although this figure also seems high, the operationalization left ambiguous the effects of verbal instructions that accompanied the written instructions (for a summary of this evidence, see McGuire, 1969).

The third paper (Sykes and Fox, 1974) accounted for only 11 percent of variance in its argument that persons tend to interact more with those with whom they share certain characteristics rather than with those with different characteristics. A long line of research (summarized by Sears and Abeles, 1969) already had thoroughly repudiated several versions of this homophily hypothesis, so this small figure is to be expected. But the authors made no mention of this literature. They also failed to take into account the tendency of some people to be more gregarious than others or the characteristics that make some people more attractive than others (Smith and Freedman, 1972; Tomeh, 1973; Lindzey and Byrne, 1969). No wonder the percent variance accounted for was so small!

Drafts of these papers, sent to the discussant prior to the meeting, were all suspiciously strong. Each reported results significant at $p < .001$, a rather high--but, in this instance, deceptive--figure. In any case, significance tests are not always helpful because they only

indicate how much the theory's results surpass complete randomness, which is sheer chaos. Data can be extremely non-random without even approaching a meaningful relationship between observed and predicted values.

None of these papers, however, actually reported the percentage of variance accounted for. For the first paper, of course, this readily could be calculated by squaring the Spearman's rho figure. The second paper, however, presented an analysis of variance, which required the reader to apply Hayes' (1963:407) formula to obtain the percentage of variance for which it accounted.

Originally, the third paper also presented an analysis of variance, which, through Hayes' formula, yielded the 11 percent variance-accounted-for figure. In a revised version, however, the authors substituted Goodman's log-linear model (1971), a statistical procedure not related to multiple regression. Although they apparently acted out of concern over violating the classical assumptions underlying multiple regression, they admitted that they had made no systematic effort to determine if these assumptions, indeed, had been violated. Goodman's model has been put to fruitful use elsewhere, but it provides no means of even calculating a percentage-of-variance-accounted-for figure. The authors, therefore, perpetrated an even greater disservice to the development and evaluation of theory.

As a tool in theory development and evaluation, the comparison of percent variance accounted for has several unique strengths. First, it is parsimonious. Expressed numerically, it is much more succinct than verbal evaluations of ambiguity that involve criticizing either concepts

or operationalization. Since it is derived mathematically, percent variance accounted for also incorporates formal logic to a greater degree than do most verbal approaches.

Secondly, percent variance accounted for provides a record of the relative success of each past attempt to solve a particular puzzle. Science, as Kuhn (1970) points out, can best be viewed as a series of such attempts to solve commonly agreed upon puzzles. Each new theory, therefore, contributes to the advancement of science only if its solution to such a puzzle is better than those of its predecessors. Here again, percent variance accounted for permits a direct comparison. Unless the new theory accounts for a higher percentage of variance than established theories, it should not be allowed to supercede them, regardless of its intuitive appeal.*

Thirdly, percentage of variance accounted for can guide subsequent investigators in formulating new theories that can provide still better solutions to the puzzle. In selecting variables for a study, an investigator initially can look to previous works that accounted for relatively high percentage of variance. Then, of course, the investigator must carry the matter further, incorporating new variables or combining traditional ones in novel ways. Here the process virtually becomes a form of art, demanding creativity and insight on the part of the investigator. Even such a sophisticated and productive statistical technique as multiple regression provides no assistance. The error figure reveals

*Although some may prefer to judge theories by their generalizability, most social science theories have yet to account for enough variance to make generalizations feasible (Phillips, 1971; Gibbs, 1972: Chapter 3).

only that something has been left out or that at least some of the included variables were not measured appropriately. But it provides no indication of what, specifically, was omitted or even whether the problem stemmed from omission or poor measurement. The investigator must look elsewhere for answers to these questions. Unfortunately, many investigators probably do not pay close enough attention to what actually happens to the persons in their observations; consequently, they tend to overlook valuable clues to the true state of nature.

The authors of the third paper presented at the ASA session not only failed to carry their inquiry to the art form stage but also, by neglecting previous research, overlooked a valuable source of scientific guidance. They then compounded the damage by reporting their results in a way that made direct comparisons with other work difficult, to say the least. But if behavior sciences are to make any progress in the accumulation of knowledge, its practitioners must fight off this temptation to obscure weak results. Investigators, understandably, may be reluctant to report percentage of variance accounted for; in most cases this figure is embarrassingly small. Although it may be ego-deflating, investigators should report their results in a manner that permits their fellow scholars at least to calculate this figure. When investigators fail to do so, their colleagues should demand this information, for, as a statistical tool, percentage of variance accounted for provides a direct measure of how far a discipline has come in its efforts to reduce ambiguity.

REFERENCES

Byrne, J. F. and McNeil, K. A. A further discussion of issues related to causal inference. Multiple Regression Viewpoints, 1975, 5, 1-10.

Gibbs, J. P. Sociological Theory Construction. Hinsdale, Illinois: Dryden Press, 1973.

Goodman, L. A. Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. Journal of the American Statistical Association, 1971, 66, 339-344.

Hayes, W. H. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963.

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

Kuhn, T. S. The Structure of Scientific Revolutions. Chicago: University of Chicago Press, 1970.

Lindzey, G. and Byrne, O. Measurement of social choice and interpersonal attractiveness. In G. Lindzey and E. Aronson (Eds.), The Handbook of Social Psychology. Reading, Massachusetts: Addison Wesley, 1969, pp. 452-525.

Maola, J. F. Causality and prediction, similar but not synonymous. Multiple Linear Regression Viewpoints, 1974, 5, 1-2.

McGuire, W. J. The nature of attitudes and attitude change. In G. Lindzey and E. Aronson (Eds.), The Handbook of Social Psychology. Reading, Massachusetts: Addison Wesley, 1969, pp. 136-314.

Moschetti, G. J. The role of the individual in the tradition perpetuation problem. Paper read at the 1974 annual meeting of the American Sociological Association.

Newman, I. and Newman, C. A discussion of component analysis; its intended purpose, strengths, and weaknesses. Multiple Regression Viewpoints, 1975, 5, 35-48.

Phillips, D. Knowledge from What. Chicago: Rand McNally, 1971.

Sears, D. O. and Abeles, R. P. Attitudes and opinions. In P. H. Mussen and M. R. Rosenburg (Eds.), Annual Review of Psychology, Volume 20. Palo Alto, California: Annual Reviews, Inc., 1969, pp. 267-270.

Segal, M. W. Alphabet and attraction: an unobtrusive measure of the effect of propinquity in a field setting. Paper read at the 1974 annual meeting of the American Sociological Association.

Smith, C. and Freedman, A. Voluntary Associations. Cambridge, Massachusetts: Harvard University Press, 1972.

Sykes, R. E. and Fox, J. C. Proximity and similarity effects on frequency of interaction in a class of naval recruits. Paper read at the 1974 annual meeting of the American Sociological Association.

Tomeh, A. K. Formal voluntary organizations: participation, correlates and interrelationships. Sociological Inquiry, 1973, 43, 89-122.

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

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

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

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

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

TABLE OF CONTENTS