NOVEMBER 1979

VOLUME 10 NUMBER 1

MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group on Multiple Linear Regression

MLRV Abstracts appear in CIJE, the ERIC System, and microform copies are available from University Microfilms International
MLRV is listed in EBSCO Librarians Handbook.

ISSN 0195-7171

MULTIPLE LINEAR REGRESSION VIEWPOINTS

EDITORIAL BOARD

TABLE OF CONTENTS

Multiple Linear Regression Viewpoints
Vol. 10, No. 1, 1979

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 10 No. 1 1979

MULTIVARIATE TECHNIQUES FOR MEETING FEDERAL REQUIREMENTS CONCERNING VALIDATION

Francis D. Bertram, Marquette University
Kathleen M. Roblee, University of Wisconsin-Oshkosh
Glenn E. Tagatz, Marquette University

ABSTRACT

The purpose of this study was to use multivariate techniques in a federally-regulated validation study, and to compare the results obtained from zero-order correlations and multiple correlations with the results obtained using factor scores and canonical correlation. The subjects consisted of fifty-one individuals who were selected from five hundred thirty-two applicants for the position of patrolman. The investigation revealed that multivariate techniques yield higher correlation coefficients than zero-order correlation or multiple correlation in a number of instances, and thus multivariate techniques may be the method of choice for certain federally-regulated validation studies.

INTRODUCTION

Tests and other assessment instruments that possess appropriate psychometric characteristics can reduce discrimination in the selection of employees (Anastasi, 1976). Yet, discrimination in hiring can be perpetuated by inappropriate tests (Griggs vs. Duke Power Company, 1971; Morrow vs. Crisler, 1971; Castro et al. vs. Beecher et al., 1971). To prevent this from occurring, the Equal Employment Opportunity Commission (EEOC) Guidelines (1970) state that criterion-related validity is required of all selection procedures.

Traditionally, zero-order correlation and multiple correlation are the statistical procedures used in determining criterion-related validity. However, beginning with Peck and Stephens (1964) who used factor analysis to obtain composite predictor and criterion variables in a study of the future vocational success of male retardates, various researchers have found that certain multivariate techniques are applicable in validation studies. Factor scores derived from factor analysis were used as predictors or criteria by Riccobono and Cunningham (1971), Chissom (1971), Logan and Palmer (1972), and Goldstein and Barrows (1972). Factor scores and canonical correlation were used by Mayerske et al. (1969) and Friedman (1972). Yet, the application of multivariate techniques to employee selection procedures for the purpose of meeting federal requirements (EEOC Guidelines, 1970) concerning validation has not been attempted.

The purpose of this study is two-fold: 1) to apply multivariate techniques in a validation study of employee selection procedures, and 2) to compare the validation results obtained from the traditional procedures of zero-order correlation and multiple correlation with the results obtained from the multivariate techniques of factor scores and canonical correlation. The latter purpose of this study perhaps is of more significance for future applied research, for if multivariate techniques are found to yield results comparable to traditional procedures, two advantages are obtained through their use: 1) the set of variables is reduced, thereby enabling a more simple and logical explanation of the relationships found, and 2) the fundamental dimensions underlying the predictor and criterion variables are delineated.

METHOD

Subjects

The subjects utilized in this study consisted of fifty-one individuals who were selected from a group of five hundred thirty-two applicants for the position of police patrolman. Due to the imposition of an affirmative action quota, forty-one of the subjects were White males, while ten of the subjects were Black males. They were chosen on the basis of high scores on the written test in relation to other applicants from their racial group, coupled with passing scores on the physical agilities test and passing ratings on the oral interview. Since these subjects were selected on the basis of their scores on the written test, physical agilities test and oral interview, which are the assessment instruments being investigated in this study, the validity coefficients obtained in this study are affected by this preselection and probably underestimate the relationships among predictors and criteria.

Data

Data utilized as predictors in this study consisted of the subjects' scores on a written test and on a physical agilities test, as well as their average rating on a six category, one hundred points per category, oral interview rating scale. These three instruments were used to select police recruits (who constituted the subjects in this study) from a pool of applicants. Data utilized as both predictors and criteria in this study consisted of the subjects' ten ratings on a policy academy field training rating scale. This scale was used to measure the subjects' performance while at the police academy in the following ten areas: 1) appearance, 2) ability to learn, 3) attitude, 4) self-confidence, 5) willingness to work, 6) knowledge of job, 7) quality of work, 8) reliability, 9) quantity of work, and 10) eagerness to learn. The scale contained five rating categories for each of the ten areas, and was completed for all the subjects by the sergeant in charge of recruits.

Data utilized as criteria in this study consisted of the subjects' eight ratings on a monthly follow-up performance rating scale. This scale was used to measure the subjects' on-the-job performance as patrolmen in the following eight areas: 1) aggressiveness and initiative, 2) ability, 3) conduct, 4) judgment, 5) temperament, 6) appearance, 7) physical condition, and 8) reliability. This scale contained five rating categories for each of the eight areas, and was completed by the sergeant responsible for supervising their work as patrolmen.

Data Analysis

The statistical techniques of simple correlation, multiple correlation, multiple correlation with factor scores as criteria, and canonical correlation with factor scores as criteria, were used to investigate the relationships of: 1) the three screening measures with the academy measures, 2) the three screening measures with the monthly follow-up measure, 3) the academy measures with the monthly follow-up measures, 4) combination of the three screening measures and the academy measures with the follow-up measures, and 5) the screening measures with a combination of the academy measures and the monthly follow-up measures. The Factor Analysis subprogram, the Multiple Regression subprogram, and the Canonical Correlation subprogram of the Statistical Package for the Social Sciences (Nie et al., 1970) were used to make the statistical analyses. The factor scores used as criterion measures were derived through Alpha Factor Analysis and Incomplete Image Analysis, the results of both being subjected to both orthogonal and oblique rotations.

RESULTS

Simple Correlations

Table 1 contains the 134 coefficients obtained through correlating: 1) the three screening measures with the eight monthly follow-up measures, 2) the three screening measures with the ten academy measures, and 3) the ten academy measures with the eight monthly follow-up measures. Each of the three screening measures correlated significantly with one or more of the monthly follow-up measures, and each of the monthly follow-up measures correlated significantly with one or more of the screening measures. Each of the screening measures correlated significantly with two or more of the academy measures, and nine of the ten academy measures correlated significantly with one or more of the screening measures. Seven of the ten academy measures correlated significantly with one or more of the monthly follow-up measures, and four of the eight monthly follow-up measures correlated significantly with one or more of the academy measures.

The nine significant correlations involving the physical agilities test were negative, thus indicating that the lower a patrolman scored on the physical agilities test, the better his academy and job performance. Since this finding suggested that the physical agilities test was measuring a trait different from the cognitive and affective traits measured by the other two screening measures and the academy and job performance ratings, the remaining statistical analyses were performed both with and without the physical agilities data. (See Conclusion)

TABLE 1

SIMPLE CORRELATION COEFFICIENTS

MONTHLY FOLLOW-UP MEASURES	SCREENING MEASURES		
	WRITTEN TEST	PHYSICAL AGILITIES TEST	ORAL INTERVIEW
Aggressiveness & Initiative	-.05	-.27*	.11
Ability	.35**	-.13	.15
Conduct	.48**	-.04	.18
Judgement	.29*	-.22	.06
Temperament	.29*	-.22	.06
Appearance	.15	.04	.26*
Physical Condition	.13	.09	.27*
Reliability	.29*	.00	.10
ACADEMY MEASURES			
Appearance	.17	-.02	.09
Attitude	.31*	-.40**	.14
Ability to Learn	.46**	-.39**	.35**
Self-Confidence	.17	-.44**	.07
Willingness	.26*	-.33**	.15
Knowledge of Job	.36**	-.36**	.17
Quality of Work	.42**	-.22	.30*
Reliability	.23	-.39**	.17
Quantity of Work	.39**	-.46**	.17
Eagerness to Learn	.19	-.33**	.14

TABLE 1 (Cont.)

SIMPLE CORRELATION COEFFICIENTS

ACADEMY MEASURES	MONTHLY FOLLOW-UP MEASURES			
	Aggressiveness & Initiative	Ability	Conduct	Judgement
Appearance	-.17	.18	.33**	.11
Attitude	.13	.33**	.38**	.24*
Ability to Learn	-.03	.18	.31*	.19
Self-Confidence	-.01	.13	.16	.12
Willingness	-.06	.12	.10	.10
Knowledge of Job	-.00	.22	.24*	.31*
Quality of Work	-.05	.31*	.32**	.22
Reliability	.00	.27*	.35**	.14
Quantity of Work	.06	.33**	.39**	.21
Eagerness	-.10	.15	.18	.05
	Temperament	Appearance	Physical Condition	Reliability
Appearance	.07	.08	.08	.11
Attitude	.15	.16	.18	.27*
Ability to Learn	.06	.04	.04	.16
Self-Confidence	-.00	-.01	-.08	-.01
Willingness	-.10	-.05	-.02	.07
Knowledge of Job	-.04	.03	.00	.16
Quality of Work	.05	.06	.03	.18
Reliability	.02	.02	.05	.25*
Quantity of Work	.09	.04	.07	.24*
Eagerness to Learn	-.14	-.13	.02	.20

**$p \leq .01$ level of significance

*$p \leq .05$ level of significance

Multiple Correlations

Table 2 contains the significant multiple correlations obtained. The three screening measures were used as predictors of both the ten academy measures and the eight monthly follow-up measures. Six of the eight monthly follow-up measures were predicted significantly by the screening measures. Five of these six predictions involved only the written test. The sixth involved the physical agilities test with its negative relationship to performance; when the physical agilities test was removed from this analysis, no significant prediction was found, leaving five significant predictions based on the written test. Nine of the ten academy measures were predicted significantly by the screening measures. Eight of the nine predictions involved the physical agilities test. When the physical agilities test was removed from the analysis, five of the ten academy measures were predicted by the written test.

The ten academy measures were used to predict the eight monthly follow-up measures. Three of the eight monthly follow-up measures were predicted by the academy measures using multiple correlation.

The combined set of screening measures, plus the academy measures were used as predictors of the monthly follow-up measures. Significant predictions were found for five of the eight monthly follow-up measures. The written test alone was involved in three of these predictions, the academy measure of knowledge of job was involved in one of these predictions, and the written test and the academy measure of appearance were involved in the remaining prediction.

TABLE 2

MULTIPLE CORRELATIONS

SCREENING MEASURES WITH MONTHLY FOLLOW-UP MEASURES	R	Significance Level
Physical Agilities with Aggressiveness and Initiative	.27	.05
Written Test with Ability	.35	.05
Written Test with Conduct	.48	.01
Written Test with Judgement	.29	.05
Written Test with Temperament	.29	.05
Written Test with Reliability	.29	.05
SCREENING MEASURES WITH ACADEMY MEASURES		
Physical Agilities with Attitude	.40	.05
Written Test, Physical Agilities, and Oral Interview with Ability to Learn	.61	.05
Physical Agilities with Self-Confidence	.44	.01
Physical Agilities with Willingness to Work	.33	.05
Written Test and Physical Agilities with Knowledge of Job	.46	.05
Written Test with Quality of Work	.42	.01
Physical Agilities with Reliability	.39	.01
Written Test and Physical Agilities with Quantity of Work	.56	.05
Physical Agilities with Eagerness to Learn	.33	.05

TABLE 2 (Cont.)

MULTIPLE CORRELATIONS

SCREENING MEASURES WITH ACADEMY MEASURES AFTER PHYSICAL AGILITIES TEST WAS REMOVED	R	Significance Level
Written Test with Attitude	.31	.05
Written Test with Ability to Learn	.46	.01
Written Test with Knowledge of Job	.36	.05
Written Test with Quality of Work	.42	.01
Written Test with Quantity of Work	.39	.05
ACADEMY MEASURES WITH MONTHLY FOLLOW-UP MEASURES		
Attitude with Ability	.33	.05
Quantity of Work with Conduct	.39	.01
Knowledge of Job with Judgement	.31	.05
SCREENING MEASURES AND ACADEMY MEASURES WITH MONTHLY FOLLOW-UP MEASURES		
Written Test with Ability	.35	.05
Written Test and Appearance with Conduct	.55	.05
Knowledge of Job with Judgement	.31	.05
Written Test with Temperament	.29	.05
Written Test with Reliability	.29	.05

Multiple Correlations Predicting Factor Scores

Table 3 contains the significant multiple correlations obtained between the three predictor variables and the three Alpha and the ten Incomplete Image factor scores computed from the criterion variables. The three screening measures were used to predict the factor scores from factor analyses of the monthly follow-up measures. Of the two factors obtained using Alpha Factor analysis, the written test was predictive of the scores of Factor II using the orthogonal solution, and of both the scores of Factor I and Factor II using the oblique rotation. Of the four factors obtained using Image factor analysis, the written test was predictive of the scores of Factor I using the orthogonal solution, and the scores of Factor I and Factor II using the oblique solution. The written test and the oral interview were predictive of the scores of Factor IV using the oblique rotation.

The three screening measures were also used to predict the factor scores from factor analysis of the combined academy and monthly follow-up measures. Of the three factors obtained using Alpha factor analysis: 1) the written test and the physical agilities test were predictive of the scores of Factor I using both the orthogonal and oblique solutions, and 2) the written test was predictive of the scores of Factor II and Factor III using the oblique solution. The data was re-analyzed without the physical agilities test, and the written test alone was found to be predictive of the scores of Factor I using both solutions. Of the ten factors obtained using Image Factor analysis: 1) the written test and physical agilities test were predictive of the scores of Factor I using the oblique solution, 2) the written test alone was predictive of the scores of Factor II using the orthogonal solution, and the scores of Factor II, Factor IV, Factor VIII, and Factor IX using the oblique solution, and 3) the physical agilities test alone was predictive of the scores of Factor I and Factor IV using the orthogonal solution, and the scores of Factor IV, Factor VII and Factor X using the oblique solution. Upon removing the physical agilities test, the data was re-analyzed and no significant predictors were found for the scores of Factor IV using either solution, or for the scores of Factor VII using the oblique solution. The written test did significantly predict the three remaining factor scores (the scores of Factor I using both solutions, and the scores of Factor X using the oblique solution).

Canonical Correlations

Canonical correlations were computed between the screening measures and the monthly follow-up measure Alpha and Imcomplete Image factor scores. Using the factor scores from the Image factor analysis, canonical correlations of .500 were found for both rotational procedures. These correlations were significant at the .05 level. Upon re-computation of the correlations after removing the physical agilities test, a negligible reduction occurred in both the orthogonal solution (r=.498) and the oblique solution (r=.499). The canonical correlations computed using Alpha factor scores were not significant.

Canonical correlations were also computed between the screening measures and the combined sets of academy measures and monthly follow-up measures. Using the factor scores obtained from the Alpha factor analysis, canonical correlations of .624 were found for both rotational procedures, which were significant at the .01 level. When the correlations were re-computed without the physical agilities test, a substantial reduction occurred (r=.501;p=.05). Using the factor scores from the Image factor analysis, canonical correlations of .678 were found for both rotational procedures, which were significant at the .05 level. When the correlations were re-computed without the physical agilities test, no significant correlations were obtained.

TABLE 3

MULTIPLE CORRELATIONS USING FACTOR SCORES AS CRITERIA

FACTOR SCORES DERIVED FROM MONTHLY FOLLOW-UP MEASURES	R	Significance Level
Written Test with Factor II Scores of the Alpha Factor Analysis (Orthogonal Rotation)	.27	.05
Written Test with Factor I Factor Scores of the Alpha Factor Analysis (Oblique Rotation)	.31	.05
Written Test with Factor II Factor Scores of the Alpha Factor Analysis (Oblique Rotation)	.29	.05
Written Test with Factor I Factor Scores of the Image Factor Analysis (Orthogonal Rotation)	.45	.01
Written Test with Factor I Factor Scores of the Image Factor Analysis (Oblique Rotation)	.41	.01
Written Test with Factor II Factor Scores of the Image Factor Analysis (Oblique Rotation)	.27	.05
Written Test and Oral Interview with Factor IV Factor Scores of the Image Factor Analysis (Oblique Rotation)	.42	.05
FACTOR SCORES DERIVED FROM ACADEMY MEASURES COMBINED WITH MONTHLY FOLLOW-UP MEASURES		
Physical Agilities Test and Written Test with Factor I Factor Scores of the Alpha Factor Analysis (Orthogonal Rotation)	.44	.05
Physical Agilities Test and Written Test with Factor I Factor Scores of the Alpha Factor Analysis (Oblique Rotation)	.54	.05
Written Test with Factor II Factor Scores of the Alpha Factor Analysis (Oblique Rotation)	.32	.05
Written Test with Factor III Factor Scores of the Alpha Factor Analysis (Oblique Rotation)	.27	.05

TABLE 3 (Cont.)

MULTIPLE CORRELATIONS USING FACTOR SCORES AS CRITERIA

FACTOR SCORES DERIVED FROM ACADEMY MEASURES COMBINED WITH MONTHLY FOLLOW-UP MEASURES	R	Significance Level
Physical Agilities Test and Written Test with Factor I Factor Scores of the Image Factor Analysis (Oblique Rotation)	.51	.05
Written Test with Factor II Factor Scores of the Image Factor Analysis (Orthogonal Rotation)	.39	.05
Written Test with Factor II Factor Scores of the Image Factor Analysis (Oblique Rotation)	.35	.05
Written Test with Factor IV Factor Scores of the Image Factor Analysis (Oblique Rotation)	.43	.01
Written Test with Factor VIII Factor Scores of the Image Factor Analysis (Oblique Rotation)	.50	.01
Written Test with Factor IX Factor Scores of the Image Factor Analysis (Oblique Rotation)	.41	.01
Physical Agilities Test with Factor I Factor Scores of the Image Factor Analysis (Orthogonal Rotation)	.46	.01
Physical Agilities Test with Factor IV Factor Scores of the Image Factor Analysis (Orthogonal Rotation)	.33	.05
Physical Agilities Test with Factor IV Factor Scores of the Image Factor Analysis (Oblique Rotation)	.46	.01
Physical Agilities Test with Factor VII Factor Scores of the Image Factor Analysis (Oblique Rotation)	.29	.05
Physical Agilities Test with Factor X Factor Scores of the Image Factor Analysis (Oblique Rotation)	.47	.01
FACTOR SCORES DERIVED FROM ACADEMY MEASURES COMBINED WITH MONTHLY FOLLOW-UP MEASURES WITH PHYSICAL AGILITIES TEST REMOVED AS A PREDICTOR		
Written Test with Factor I Factor Scores of the Alpha Factor Analysis (Orthogonal Rotation)	.34	.05
Written Test with Factor I Factor Scores of the Alpha Factor Analysis (Oblique Rotation)	.37	.05
Written Test with Factor I Factor Scores of the Image Factor Analysis (Orthogonal Rotation)	.31	.05
Written Test with Factor I Factor Scores of the Image Factor Analysis (Oblique Rotation)	.39	.05
Written Test with Factor X Factor Scores of the Image Factor Analysis (Oblique Rotation)	.29	.05

CONCLUSION

The two purposes of this study were: 1) to establish the validity of a set of employee selection procedures using multivariate techniques, and 2) to compare the results obtained through traditional validation procedures with the results obtained through using multivariate procedures. With respect to the first purpose, the written test was found to be an excellent predictor of on-the-job performance. It met the federal regulations (EEOC Guidelines, 1970) for criterion-related validity, and its continued use was recommended. The physical agilities test, although showing a substantial relationship with many performance measures, was inversely related to the performance measures. Possible explanations of this finding include: 1) Individuals with high physical agility prefer physical methods of approaching the tasks required of a patrolman rather than the cognitive or affective methods used as this study's criterion measures; 2) Individuals rating the performance of the subjects are biased against subjects displaying high physical agility; or 3) The sample was composed of individuals who were either high in physical agilities or high in cognitive - affective ability, but not high in both. Since it was not possible to ascertain the reason why the physical agilities test demonstrated an inverse relationship with performance measures, it was recommended that the items composing the physical agilities test be evaluated with respect to their pertinence for a patrolman and/or that criterion measures be designed to measure the physical component of the patrolman's job. The oral interview, although demonstrating some correlation with performance, failed to add significantly to any of the predictions. Its discontinuance as a screening device was recommended, although its retention as a means of maintaining personal contact with applicants was suggested.

With respect to the second purpose, the comparison of multiple correlation using factor scores as the criteria with multiple correlation using the original performance measures as criteria, yielded mixed results. When the screening measures were correlated with the monthly follow-up measures, the multiple correlations using Alpha factor scores were somewhat lower than those obtained using the original variables, while the multiple correlations using Image factor scores tended to be somewhat higher than those obtained using the original criterion variables. One possible explanation of this result is that Image factor analysis uses the estimated communality of the variables as the diagonal elements in the correlation matrix, whereas Alpha factor analysis does not. When the screening measures were correlated with the combined academy and monthly follow-up measures, the multiple correlations using the Alpha factor scores were in the same range as those obtained using the original criteria variables, while the multiple correlations using Image factor scores were somewhat higher than those obtained using the original criteria variables. In general, multiple correlation using factor scores as the criteria appears to be a productive method of approaching predictive validity, although further research is indicated.

The multivariate techniques of canonical correlation between predictor variables and scores derived through factor analysis also appears to have promising possibilities for predictive validity studies, when compared to zero-order correlation and multiple correlation. The canonical correlation obtained between the screening measures and the factor scores derived from the monthly follow-up measures in this study, was higher than the obtained zero-order and multiple correlations. After removal of the physical agilities measure, these canonical correlations still maintained their greater predictive ability. The canonical correlation obtained between the screening measures and the factor scores obtained from the combined academy and monthly follow-up measures also demonstrated superior predictive power.

In conclusion, this study demonstrates the applicability of multivariate techniques to validity studies. The results obtained suggest that these techniques yield higher correlation coefficients than zero-order correlations or multiple correlations in a number of instances. This result is probably due to the exclusion of a greater proportion of error variance than true variance in the factor analysis portion of the analyses. In addition, the use of multivariate techniques has logical and theoretical advantages in that: 1) the set of variables is reduced, thereby enabling a more simple and logical explanation of the relationships found, and 2) the fundamental dimensions underlying the predictor and criterion variables are delineated. Thus, while this study does not suggest that these results can be generalized to all validity studies using composite measures, it does suggest that: 1) multivariate techniques may be appropriate in some validity studies, and 2) further delineation concerning the conditions under which multivariate techniques are applicable to validation efforts is needed.

REFERENCES

Anastasi, A. Psychological Testing. (3rd. Ed.) New York: Macmillan, 1976.

Castro, Cuffs, Darby, Franklin, Garcia, Green, O'Bryant, and Upshaw vs. Beecher, Duffy, Feinberg, LaFlamme, and Mitchell, Civil Action No. 70-1220-W, U. S. District Court, District of Massachusetts, Nov. 17, 1971.

Chissom, B.S. A factor-analytic study of the relationship of motor factors to academic criteria for first and third grade boys. Child Development, 1971, 42, 4, 1133-1143.

Equal employment opportunity commission guidelines on employee selection procedures. Federal Register, 35, 149, 12333-12336, 1970.

Friedman, D. "The use of pattern analysis for the prediction of achievement criteria." Paper presented to the annual meeting of the American Educational Research Association (Chicago, Illinios, April, 1972).

Goldstein, L.S., and Barrows, T.S. The structure of three instruments intended for police selection. Educational Testing Service: Report No. PR-72-14, August, 1972.

Griggs vs. Duke Power Co. 30, L.W. 4317, 401 U.S. 424, March 8, 1971.

Logan, W.L., and Palmer, M. "Identification and description of the intrinsic sources of individual differences in concept learning." Final report, Northern Colorado University, Greeley: Office of Education, U. S. Department of Health, Education and Welfare, August, 1972.

Mayerske, G.W., Wisler, C.E., Beaton Jr., A.E., Cohen, W.M., Okada, T., Proshek, J.M., and Tabler, K.A. A Study of our nation's schools. Office of Education, U. S. Department of Health, Education and Welfare, 1969.

Morrow vs. Crisler, Civil Action No. 4716, U. S. District Court, Southern District of Mississippi, Jackson Division, Sept. 29, 1971.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H. Statistical Package for the Social Sciences. New York: McGraw-Hill, 1970.

Peck, J., and Stephens, R. "Personality and success profiles characteristic of young adult male retardates." Texas University, Austin: Cooperative Research Program of the Office of Education, U. S. Department of Health, Education and Welfare, Report No. CRP-S-116, 1964.

Riccobono, J.A., and Cunningham, J.W. "Work dimensions derived through systematic job analysis: A study of the occupational analysis inventory." North Carolina State University, Raleigh: Center for Occupation Education for the National Center for Education Research and Development, Office of Education, U. S. Department of Health, Education and Welfare, 1971.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 10 No. 1 1979

MULTIPLE COMPARISONS IN THE ANALYSIS OF COVARIANCE USING MULTIPLE LINEAR REGRESSION

John T. Williams
University of North Dakota

ABSTRACT

A process is described for multiple comparisons when covariates are involved in the analysis. The method can be accomplished with considerable ease whenever pairwise comparisons are involved. More complex contrasts require the use of full and restricted models.

While many explications regarding multiple comparisons have been made for the usual one-way analysis of variance, most authors on the subject of multiple comparisons are silent regarding the analysis of covariance. The silence is understandable; each separate comparison will have its own standard error of estimate even if equal N occur in each cell. The equation for the standard error of estimate for a comparison in the analysis of covariance is given by Winer (1971, p. 772),

$$s_{\overline{Y}_i adj - \overline{Y}_j adj} = \sqrt{MS'_W \left[\frac{1}{n_i} + \frac{1}{n_j} + \frac{(\overline{X}_i - \overline{X}_j)^2}{E_{XX}}\right]} \qquad (1)$$

where

$\overline{Y}_i adj$ = the adjusted mean for group i;

$\overline{Y}_j adj$ = the adjusted mean for group j;

MS'_W = the error term in the analysis of covariance;

n_i, n_j = respectively cell frequencies for the i^{th} and j^{th} groups;

$\overline{X}_i$, $\overline{X}_j$ = respectively the means on the covariate for the i^{th} and j^{th} groups; and

E_{XX} = SS_W for the covariate.

While researchers may feel justifiably ill at ease in attempting to use equation 1, the use of regression can eliminate the tedious calculations. Further, more than one covariate can easily be accommodated.

An Example

Table 1 is taken from Williams (1974, p. 104 and p. 109). In Table 1, X_1 is a binary variable for membership in group 1, X_2 is a binary variable for membership in group 2 and X_3 is similarly a binary variable for membership in group 3. Also, X_4 represents a pretest score and X_5 represents a measure of intelligence; the Y value represents a posttest score. Only the pretest is considered as a covariate in this section; both the pretest and intelligence are considered as covariates in section under multiple covariates.

TABLE 1

Data for the Analysis of Covariance

Y	X_1	X_2	X_3	X_4	X_5
35	1	0	0	12	120
27	1	0	0	17	98
32	1	0	0	13	102
29	1	0	0	10	106
27	1	0	0	8	94
38	0	1	0	29	123
25	0	1	0	12	96
36	0	1	0	17	108
35	0	1	0	22	115
31	0	1	0	15	128
27	0	0	1	17	90
35	0	0	1	22	110
19	0	0	1	10	94
17	0	0	1	8	95
32	0	0	1	13	116

Under the assumption of a single regression line on the covariate (the pretest, X_4) an analysis of covariance can be accomplished with two linear models:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_4X_4 + e_1, \qquad (2)$$

and

$$Y = b_0 + b_4X_4 + e_2. \qquad (3)$$

In that a large part of the print-out regarding equation 2 is useful, the print-out is reproduced in Table 2.

Print-Out for Equation 2

VARIABLE	MEAN	STAND. DEV.	CORRELATION X VS Y	REG. COEF.	STD. ERROR OF REG. COEF.	COMPUTED T VALUE
4	15.00	5.85	0.689	0.76	0.22783	3.33582
1	0.33	0.48	0.039	5.52	2.73396	2.01905
2	0.33	0.48	0.398	3.20	2.92653	1.09345
DEPENDENT						
Y	29.66	6.12				
INTERCEPT	15.36					

MULTIPLE CORRELATION	0.78714
STD. ERROR OF ESTIMATE	4.26230
MULTIPLE CORRELATION SQUARED	0.61959
ONE MINUS MULTIPLE CORRELATION SQD	0.38041

Analysis of Variance for the Regression

SOURCE OF VARIATION	df	SS	MS	F
ATTRIBUTABLE TO REGRESSION	3	325.49	108.497	5.972
DEVIATION FROM REGRESSION	11	199.84	18.167	
TOTAL	14	525.33		

The usual analysis of covariance can be completed by using

$$F = \frac{(R_2^2 - R_3^2)/(g - 1)}{(1 - R_2^2)/dfw_2} = \frac{(.61959 - .47476)/2}{(1 - .61959)/11} = 2.09,$$

which for df = 2, 11, $p > .05$.

In equation 2, the X_3 variable has been omitted. Thus $b_1 = \overline{Y}_1\text{adj} - \overline{Y}_3\text{adj}$ and $b_2 = \overline{Y}_2\text{adj} - \overline{Y}_3\text{adj}$. To find the adjusted means, the following equations can be used:

$$\overline{Y}_3\text{adj} = b_0 + b_4\overline{X}_4 = 15.36 + .76(15) = 26.76;$$

$$\overline{Y}_1\text{adj} = b_1 + \overline{Y}_3\text{adj} = 5.52 + 26.76 = 32.28; \text{ and}$$

$$\overline{Y}_2\text{adj} = b_2 + \overline{Y}_3\text{adj} = 3.20 + 26.76 = 29.96.$$

The adjusted values agree with those originally given by Williams (1974, p. 106), though the method shown here is simplified somewhat.

More importantly, the standard error of the regression coefficients corresponding to X_1 and X_2 are respectively equal to the standard errors from equation 1 for comparing $\overline{Y}_1$adj to $\overline{Y}_3$adj and $\overline{Y}_2$adj to $\overline{Y}_3$adj. Thus, the computed t values given in Table 2 are directly usable in whichever multiple comparison procedure the researcher prefers. The use of Dunnett's (1955), Tukey's (1953), Dunn's (1961) and Scheffé's (1959) tests are described in a regression format using computed t values in Williams (1976, 1979). Were there interest in comparing $\overline{Y}_1$adj to $\overline{Y}_2$adj, a model of the form

$$Y = b_0 + b_1X_1 + b_3X_3 + b_4X_4 + e_1 \qquad (4)$$

could be used, with focus on the computed t value for the X_1 variable.

Complex Comparisons

Complex comparisons, or contrasts, can be completed in a regression analysis for the analysis of covariance as well. Suppose a contrast of the form

$$\Psi = \overline{Y}_3\text{adj} - \tfrac{1}{2}\overline{Y}_1\text{adj} - \tfrac{1}{2}\overline{Y}_3\text{adj} \qquad (5)$$

is contemplated. First, equation 2 is reparametized as

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_1. \qquad (6)$$

Then a restriction corresponding to Ψ_1, $b_3 = \tfrac{1}{2}b_1 + \tfrac{1}{2}b_2$ is placed on equation 6:

$$Y = b_1X_1 + b_2X_2 + (\tfrac{1}{2}b_1+\tfrac{1}{2}b_2)X_3 + b_4X_4 + e_3.$$

Or,

$$Y = b_1(X_1+\tfrac{1}{2}X_3) + b_2(X_2+\tfrac{1}{2}X_3) + b_4X_4 + e_3. \qquad (7)$$

Two new variables can be constructed such that V_1 = 1 if a member of group 1, ½ if a member of group 3, 0 if a member of group 2; and V_2 = 1 if a member of group 2, ½ if a member of group 3, 0 if a member of group 1. Then equation 7 can be rewritten as

$$Y = b_1V_1 + b_2V_2 + b_4X_4 + e_3. \qquad (8)$$

Equation 8 (and also equation 6) could be processed using a program such as Ward and Jennings' (1973) DATRAN or McNeil et al.'s (1975) LINEAR.

However, equation 8 can also be reparametized back into a form using a unit vector as was done earlier. This can be accomplished by setting either b_1 or b_2 equal to zero. Setting b_2 = 0 yields

$$Y = b_0 + b_1V_1 + b_4X_4 + e_3. \qquad (9)$$

Then R^2_9 = .50151.

To test Ψ, $t = \sqrt{\dfrac{R_2^2 - R_9^2}{(1-R_2^2)/11}} = \dfrac{.61959 - .50151}{(1-.61959)/11}$;

$t = 1.85$, $p > .05$.

Concerns of Homogeneity of Variance

To this point, the assumption of a single regression line for the covariate has been made. A test can be made of this assumption; three new variables are defined such that

$X_6 = X_1 \cdot X_4$;

$X_7 = X_2 \cdot X_4$; and

$X_8 = X_3 \cdot X_4$.

Then a model can be written as

$$Y = b_0 + b_1X_1 + b_2X_2 + b_6X_6 + b_7X_7 + b_8X_8 + e_4. \qquad (10)$$

$R_{10}^2 = .71825$. To test this for significance,

$$F = \frac{(R_{10}^2 - R_2^2)/(g - 1)}{(1 - R_{10}^2)/(N - 2g)},$$

$$F = \frac{(.71825 - .61959)/2}{(1 - .71825)/9} = 1.58;\ p > .05.$$

Had the F value been significant, some researchers would prefer to abandon the analysis given earlier; however, there are no real alternatives short of abandonment. It would be inappropriate to attempt to use the computed t values for testing b_1 and b_2 in equation 10. The "adjusted means" would occur where separate regression lines occur for each group on the covariate. Since the covariance process is occurring separately for each group, differences in the adjusted means would not test any meaningful hypotheses regarding group differences on the criterion score. Table 3 should help show why this is so.

TABLE 3

Regression Output With Separate Regression Lines for Each Group on the Covariate

VARIABLE	MEAN	STAND. DEV.	CORRELATION X VS Y	REG. COEF.	STD. ERROR OF REG. COEF.	COMPUTED T VALUE
6	4.00	6.12	0.039	0.02	0.59792	0.03623
7	6.33	9.93	0.505	0.64	0.30396	2.10702
8	4.67	7.46	-0.165	1.20	0.36127	3.31721
1	0.33	0.48	0.039	20.52	9.14573	2.24344
2	0.33	0.48	0.398	11.61	8.09397	1.43434
DEPENDENT						
Y	29.67	6.12				
INTERCEPT	9.22					

MULTIPLE CORRELATION	0.84750
STD. ERROR OF ESTIMATE	4.05527
MULTIPLE CORRELATION SQUARED	0.71826
ONE MINUS MULTIPLE CORRELATION SQD	0.28174

Analysis of Variance for the Regression

SOURCE OF VARIATION	df	SS	MS	F
ATTRIBUTABLE TO REGRESSION	5	377.32	75.465	4.588
DEVIATION FROM REGRESSION	9	148.01	16.445	

Does the information given in Table 3 suggest that $\overline{Y}_1$adj - $\overline{Y}_3$adj = 20.51791? The answer is a qualified "no". Only under the condition that each group has its separate regression on the covariate, and its separate mean on the covariate would b_1 = $\overline{Y}_1$adj - $\overline{Y}_3$adj. However, that condition is very different than most users of the analysis of covariance would wish to use.

It is clearly quite different from asking, "If the groups were equal (on the covariate) at the beginning of the experiment, how do they compare at the end?" Even if all groups are "adjusted" by using a covariate mean of $\overline{X}_4 = 15$, the difference in the regression coefficients preclude interpreting b_1 as "a treatment difference after covariate adjustment between groups 1 and 3". The analysis of covariance is usually enlisted to test treatment differences in groups whose members were unable to be randomly assigned to a treatment group, so that a statistical control is used. While a test of significance on b_1 can be legitimately done, it does not address questions usually asked by researchers using the analysis of covariance.

Multiple Covariates

Extensions to more than one covariate can easily be accommodated both for the analysis of covariance and for multiple comparisons. The intelligence score, X_5, could be used together with the pretest as covariates. Assuming single regression lines for all three groups on the two covariates, the model can be given as

$$Y = b_0 + b_1X_1 + b_2X_2 + b_4X_4 + b_5X_5 + e_5. \qquad (11)$$

The use of the computed t values for b_1 and b_2 allow a test regarding differences among the adjusted means for comparing groups 1 and 2 with group 3 respectively; $t_1 = 1.95059$ and $t_2 = .38191$. To test the differences between the adjusted means of group 1 to group 2, a model such as

$$Y = b_0 + b_1X_1 + b_3X_3 + b_4X_4 + b_5X_5 + e_5 \qquad (12)$$

can be used.

Here, $t_1 = 1.33421$; also, $t_3 = -.38191$, reaffirming the t value for the difference between the adjusted means of groups 2 and 3. The sign is changed because the direction of the comparison has changed; for b_2 in equation 11, t_2 addresses $\overline{Y}_2\text{adj} - \overline{Y}_3\text{adj}$, for b_3 in equation 12, t_3 addresses $\overline{Y}_3\text{adj} - \overline{Y}_2\text{adj}$.

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.

McNeil, K. A., Kelly, F. J. and McNeil, J. T. Testing research hypotheses using multiple linear regression. Carbondale, Ill.: Southern Ill. Univ. Press, 1975.

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

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

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

Williams, J. D. Regression analysis in educational research. New York: MSS Information Corp., 1974.

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

Williams, J. D. Contrasts with unequal N by multiple linear regression. Multiple Linear Regression Viewpoints, 1979, 9, No. 3, 1-7.

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

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 10 No. 1 1979

A DEMONSTRATION OF A TYPE VI ERROR: AN APPLIED RESEARCH PROBLEM

Steve Roll
Ken Hoedt
Isadore Newman
The University of Akron

Introduction

The inappropriate fitting of a research question with a research design may result in the costly loss of power necessary to reject the false null hypothesis. This type of error has been labeled a Type VI error (inconsistency between the research question and the statistical design) by Newman, Deitchman, Burkholder, Sanders and Ervin (1976). Those most apt to fall prey to this error are researchers dependent on standard research designs which they use in a cookbook manner. Compounding the problem is an over reliance on packaged statistical programs that are appropriate for use with traditional designs, but which unfortunately may not reflect the substantive question.

Particularly vulnerable to the Type VI error are research designs applied by the novice researcher in which the statistical analysis and research design exist in a "symbiotic" relationship. Such a situation may occur when the test of significance is a traditional analysis of variance (ANOVA). In an attempt to maintain the symmetry necessary to use ANOVA, the researcher may include in his or her analysis, combinations of treatments and levels which make little theoretical or practical sense.

The Type VI error may also occur when the application of the ANOVA procedure forces a continuous variable into an artificially categorized variable. This results in the additional loss of degrees of freedom and power.

Newman et al. (1976) point out that the Type VI error results from "the inconsistency between the researchers' question of interest and the statistical procedures employed to analyze the data." It is suggested that the study of a problem should begin by asking a substantive question that reflects an area of interest. A design should then be developed to answer the research question followed by the selection of an appropriate statistical model that fits both the research question and the research design.

Unfortunately many researchers become overly concerned with selecting a statistical package with which they are familiar. Seemingly, these researchers are not cognizant of the fact that they may be allowing their statistical procedures to dictate their research questions. To demonstrate possible problems resulting from a Type VI error, the discussion which follows will include:

1) a presentation of a Type VI error resulting from the inappropriate use of ANOVA,

2) a power analysis based on the inappropriately applied ANOVA,

3) a presentation of an alternative and more appropriate design utilizing multiple linear regression, and

4) a power analysis based on the more appropriate design.

Applied Example of a Type VI Error

The Type VI error reported below may be obvious, but it does clearly illustrate how a research design and statistical model can control the substantive questions being investigated. The illustration comes from an initial research design considered by Roll (1979) in order to compare three behavioral treatments and a no treatment control group. The target of the treatment was the alleviation of test anxiety in undergraduate college students. The three active treatments were systematic desensitization (Wolpe, 1958), self-control desensitization (Goldfried, 1971) and a modified version of covert positive reinforcement (Cautela, 1970).

The initial direction in selecting a research design came from a review of literature related to the behavioral treatments of concern. Previous researchers had utilized ANOVA in their comparisons. Denney and Rupert (1977), for instance, had compared self control desensitization to systematic desensitization in the treatment of test anxiety utilizing a 2 X 2 factorial design and ANOVA to test for statistical significance. Factor A was instructions; one group received instructions which described the treatment as teaching a self-control procedure, the second group received instructions which described the treatment in passive reciprocal inhibition terms. Factor B was the method of therapy: Wolpe's or Goldfried's method.

The new problem to be studied extended this research to compare the effectiveness of Wolpe's, Goldfried's and

Cautela's behavioral therapies which differ in terms of instruction given to the client, reinforcement or no reinforcement to insure the instructions are followed, and method of stimulus presentation. The comparisons involved three levels of instruction, two levels of reinforcement and two levels of stimulus presentation. Initially it seems obvious that an appropriate research design to use would be to expand Denney's and Rupert's 2 X 2 ANOVA to a 3 X 2 X 3 ANOVA. The design was set up as illustrated in Table I. Upon examination, however, this extension was abandoned as it would have resulted in a Type VI error. The research design was determining the substantive questions with a resultant loss of power for a fixed "n" size. The theoretical and practical problems that the application of a symetrical factorial design using ANOVA would have caused are presented below.

From a practical standpoint a clinician would not be interested in applying any of the main effect treatments except in combination with each other (see Table I). In other words, the only clinical interest would be in the simple effects. Further, the only simple effects of interest are those labeled in Table I T_2, T_8 and T_{11}: T_2 is Wolpe's systematic desensitization procedure; T_8 is Goldfried's self-control desensitization procedure; T_{11} is a modified form of Cautela's covert positive reinforcement procedure.

The reasons for not including the remaining nine simple effects are as follows:

T_1: It would not make sense to instruct a client that his part in the treatment is passive in nature, as Wolpe would do, and then covertly reinforce the client for doing nothing. Further, there was no theoretical interest in combining covert reinforcement and systematic desensitization.

T_3 and T_4: These treatments make no practical sense since they combined Wolpe's instructions with Goldfried's method of scene presentation. Wolpe's instructions stress that desensitization occurs due to the incompatibility of anxiety and relaxation and that effective treatment results from the pairing of relaxation with an anxiety provoking stimuli. On the other hand, Goldfried's method of scene presentation requires anxiety to occur as a result of exposure to anxiety provoking stimuli.

T_5 and T_6: These treatments make no practical sense as they combine Goldfried's instructions with Wolpe's method of scene presentation. Goldfried's instructions stress that effective treatment results from practicing relaxing away anxiety produced by exposure to an anxiety provoking stimuli. On the other hand, Wolpe's method of scene presentation does not allow anxiety to be evoked, hence, instruction to practice relaxing it away, which would be given by Goldfried, is not possible.

T_7: This treatment combination makes the most sense of the nine excluded treatments. It is very similar to T_{11} which was included. The instructions for T_{11} inform the client that he will covertly be reinforced while the instructions for T_7 do not. This treatment combination was not of interest to the researcher as it has not been advocated by the behaviorists of concern.

T_9 and T_{10}: It would make little sense to instruct subjects that they should practice relaxing anxiety away and then never allow them to experience any anxiety due to Wolpe's method of scene presentation.

T_{12}: It would not make sense to tell a subject that he will be covertly reinforced, as would happen in treatment C, and then never covertly reinforce him.

Power Analysis for ANOVA

In addition to the problems outlined above, the Type VI error would result in a loss of power. A power analysis for an overall F following Cohen's procedure (1977) based on 3 X 2 X 2 ANOVA with an alpha level set at 0.05, a medium effect size ($f^2 = 0.15$) and a sample size of 60 is presented below.

Given,

$N = 60$
Alpha $= 0.05$
$f^2 = 0.15$ (medium effect size)
$K = 13$ (twelve treatment groups and a no treatment control group)
$U = K-1 = 13-1 = 12$
$V = N-U-1 = 60-12-1 = 47$
and $L = f^2V = 0.15(47) = 7.05$

With these parameters Cohen's (1977) power tables indicates an estimated power of 0.35. This level of power is low.

Hypotheses of Substantive Interest

Since a factorial design using ANOVA resulted in various theoretical and practical problems including unacceptably low power, it was abandoned in favor of a completely randomized post-test-only design (Campbell and Stanley, 1963) using multiple linear regression for the statistical analysis.

The research hypotheses to be tested in the afore mentioned designs were directional and are presented below.

Hypothesis I $T_{11} < T_{13}$

The modified covert positive reinforcement group will have a significantly lower mean anxiety score than the control group.

Hypothesis II $T_{11} < T_{2}$

The modified covert positive reinforcement group will have a significantly lower mean anxiety score than the systematic desensitization group.

Hypothesis III $T_{11} < T_{8}$

The modified covert positive reinforcement group will have a significantly lower mean anxiety score than the self control desensitization group.

Hypothesis IV $T_{8} < T_{13}$

The self control desensitization group will have a significantly lower mean anxiety score than the control group.

Hypothesis V $T_{8} < T_{2}$

The self control desensitization group will have a significantly lower mean anxiety score than the systematic desensitization group.

Hypothesis VI $T_{2} < T_{13}$

The systematic desensitization group will have a significantly lower mean anxiety score than the control group.

The MLR models developed to test the hypotheses were:

Hypothesis One
Full model: $Y_1 = a_0U + a_1T_{11} + a_2T_{13} + E$
Restricted model: $Y_1 = a_0U + E$

Hypothesis Two

$Y_1 = a_0U + a_1T_{11} + a_2T_2 + E$
$Y_1 = a_0U + E$

Hypothesis Three

$Y_1 = a_0U + a_1T_8 + a_2T_{11} + E$
$Y_1 = a_0U + E$

Hypothesis Four $Y_1 = a_0U + a_1T_8 + a_2T_{13} + E$
$Y_1 = a_0U + E$

Hypothesis Five $Y_1 = a_0U + a_1T_2 + a_2T_8 + E$
$Y_1 = a_0U + E$

Hypothesis Six $Y_1 = a_0U + a_1T_2 + a_2T_{13} + E$
$Y_1 = a_0U + E$

Y_1 = test anxiety scale score
T_2 = membership in the systematic desensitization group (Wolpe's procedure)
T_8 = membership in the self-control desensitization group (Goldfried's procedure)
T_{11}= membership in the covert positive reinforcement group (Modified Cautela procedure)
T_{13}= Control Group
E = Error term

Power Based Upon Substantive Questions

Since specific hypotheses could be written to compare the target treatments, it was not necessary to include in the analysis inappropriate comparisons. Also, by limiting the number of comparisons a higher subject to treatment group ratio was available which resulted, as shown below, in a more acceptable power level.

Power analysis for an overall F using MLR

N	=	60
Alpha	=	0.05
f^2	=	0.15 (medium effect size)
K	=	4 (3 treatment groups and a no treatment control group)
U	=	K-1 = 4-1 = 3
V	=	N-U-1 = 60-3-1 = 56
L	=	f^2V = .15(56) = 8.40

With these values Cohen's (1977) power tables indicate an estimated power of 0.68 as compared to 0.35 for the originally considered design.

Summary Conclusion

The research question of interest presented in this paper was concerned with only the four groups which make logical sense from an applied therapeutic research framework. If a traditional analysis of variance approach had been taken, one would most likely have conceptualized a 3 X 2 X 2 design. As pointed out in the paper the ANOVA approach would have resulted in testing hypotheses which were either illogical or which were not of interest to the researcher, and in a loss of power.

The importance of the applied researcher utilizing a research design that does not compromise the substantive questions of concern, especially with respect to it fitting into a logical and/or theoretical frame, cannot be overemphasized. This is necessary in order to facilitate meaningful research.

References

Campbell, E. T. and Stanley, J. C. Experimental and Quasi-Experimental Designs for Research. Chicago, Ill: Rand McNally College Publishing Co., 1963.

Cautela, J. R. Covert Reinforcement. Behavior Therapy, 1970, 1, 33-50.

Cohen, J. Statistical Power Analysis For the Behavioral Sciences. New York: Academic Press, 1977.

Denney, D. R. and Rupert, P. A. Desensitization and self-control. Journal of Consulting and Clinical Psychology, 1977, 24, 272-280.

Goldfried, M. R. Systematic desensitization as training in self-control. Journal of Consulting and clinical psychology, 1971, 37, 228-234.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R. and Ervin, L. Type VI Error: Inconsistency between the statistical procedure and the research question. Multiple Linear Regression Viewpoints, 1976, 6, 1-19.

Roll, S. A Comparison of Three Behavioral Self-control Procedures in the Treatment of Test Anxiety. Doctoral Dissertation in Progress, 1979.

TABLE I

Treatment Conditions Possible with Instruction (3 levels) by Scene Presentation Procedure (2 levels) by Reinforcement Design (2 levels).

	A Instructions for Wolpe's Procedure		B Instructions for Goldfried's Procedure		C Instructions for Covertly Reinforced Procedure		D Control Group
	Covert Reinforce	No Covert Reinforce	Covert Reinforce	No Covert Reinforce	Covert Reinforce	No Covert Reinforce	
Wolpe's Method of Scene Presentation	T_1	T_2	T_5	T_6	T_9	T_{10}	
Goldfried's Method of Scene Presentation	T_3	T_4	T_7	T_8	T_{11}	T_{12}	T_{13}

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 10 No. 1 1979

USING MULTIPLE REGRESSION TO INTERPRET CHI-SQUARE CONTINGENCY TABLE ANALYSIS

Dennis W. Leitner
Southern Illinois University at Carbondale

ABSTRACT

In the analysis of bivariate categorical data, the most common statistical test is the chi-square test of independence. A significant chi-square value leads the researcher to reject the null hypothesis of no relationship between the categorical variables. But the size of the chi-square statistic is a function of its degrees of freedom. This leaves the researcher with no indication of how large (or small) the relationship is. The purpose of this paper is to demonstrate multiple regression analyses of the 2 x 2, R x 2, and R x C contingency tables using "dummy" coding. The multiple correlation coefficient (whose square has a well-known interpretation) will be shown to equal Pearson's r, Cramer's V, and a function of the chi-square statistic.

This paper was presented to the Special Interest Group on Multiple Linear Regression program session on methodological advances and applications during the 1979 annual meeting of the American Educational Research Association, San Francisco, April, 1979.

A reinactment

Student: I have found it! A significant relationship between sex and political party affiliation. Look at this 2 x 2 contingency table and the chi-square value is significant at the .01 level.

Table 1. Frequencies of political party affiliation by sex of respondent.

	Democrat	Republican	
Male	60	40	$X_1^2 = 8.0$
Female	40	60	

Faculty Member (while scratching on a pad of paper and punching on his calculator): But are you sure that you have found a meaningful relationship?

S: What do you mean? It's significant and chi-squares with one deg of freedom seldom get beyond 6.

F (while fumbling through a file cabinet to get a scatter diagram): But what if I show you a bivariate scatterplot illustrating the strength of the relationship you have shown. (See Figure 1.)

S (crestfallen): But there is no relationship there.

F: Yes there is. A significant one at the .05 level of significance for the 200 points, r = .2.

S: But that means only 4% of the variance is accounted for, 96% is unexplained.

F: Yes, and that is the strength of the relationship you have found with the chi-square analysis.

(The scene continues and ends with the student and faculty member commiserating at a local hangout.)

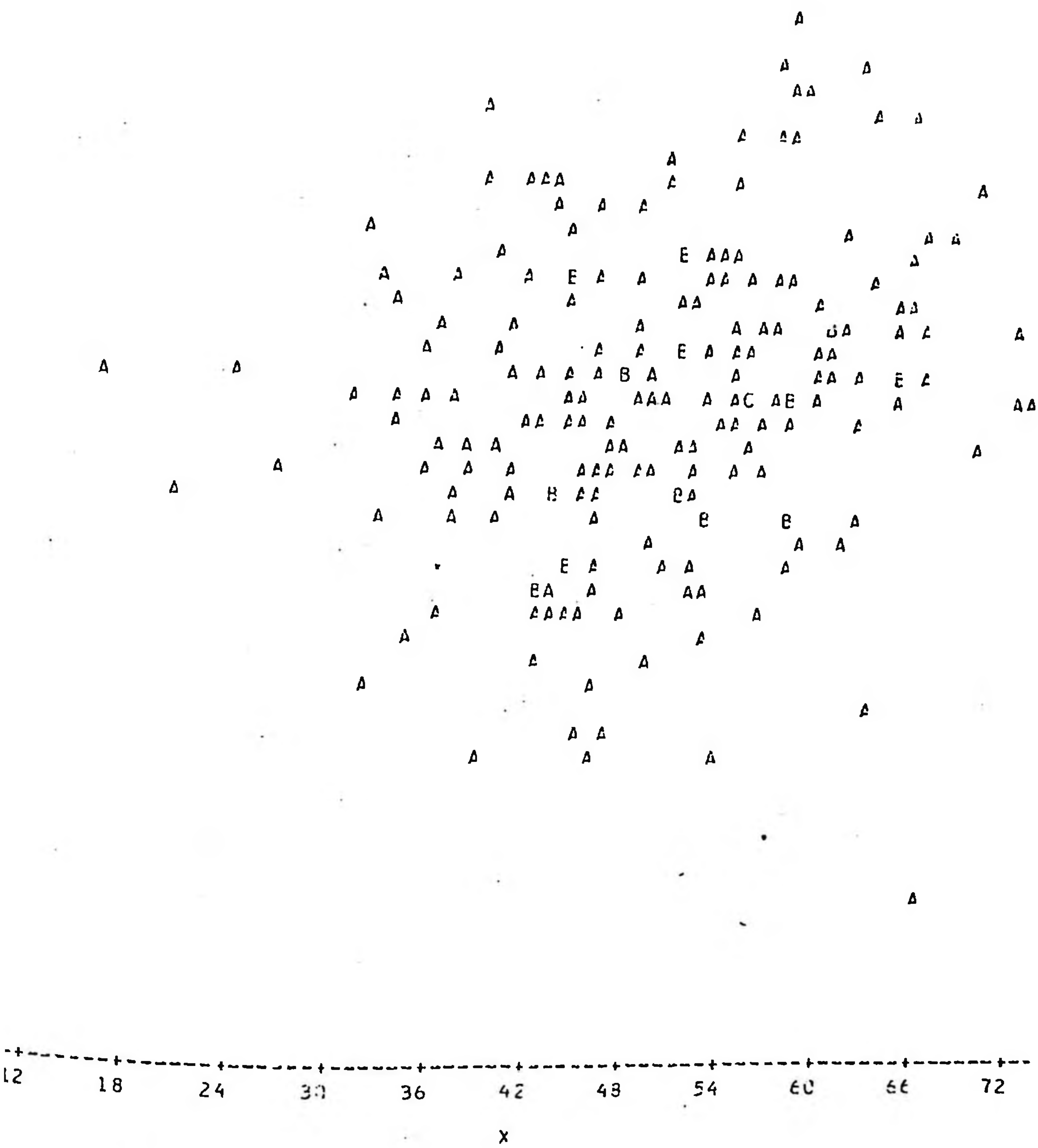

Figure 1. A scatterdiagram variables X and Y where $\mu_x = \mu_y = 50$, $\sigma_x = \sigma_y = 10$, $N = 200$, $r = .2$

The purpose of this paper is to emphasize the interpretation of R^2 in multiple regression that carries over to chi-square contingency table analysis.

2 x 2 Contingency Tables

Multiple regression aficionados would not have found themselves in the role of the student in the above scenario. By "dummy coding" sex and political party affiliation, and regressing one on the other, an R^2 (actually r^2) value is obtained which is numerically equal to χ^2/N, where N is the number of subjects on which the two variables are measured (McNeil, Kelly, McNeil, 1975, pp. 246-248). In general, if row variable A has two levels and column variable B has two levels, let

$$X = \begin{cases} 1 & \text{if observation is from level 1 of A} \\ 0 & \text{otherwise} \end{cases}$$
$$Y = \begin{cases} 1 & \text{if observation is from level 1 of B} \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

then, $r^2_{xy} = \chi^2/N$. (See proof in Bishop, Fienberg, and Holland, 1975, p. 382).

Another related statistic is the phi (ϕ) coefficient developed by Karl Pearson. If a, b, c, and d denote cell frequencies as indicated by the table at the left, ϕ can be computed directly using the formula on the right.

		B	
		1	2
A	1	a	b
	2	c	d

$$\phi = \frac{bc - ad}{\sqrt{(a+b)\ (c+d)\ (a+c)\ (b+d)}}$$

But the formula for ϕ can be derived mathematically from the formula for the Pearson product-moment correlation coefficient. (See Glass and Stanley, 1970, pp. 158-160.) So we have

$$\phi^2 = r^2 = \chi^2/N.$$

Unlike the bivariate scatterplot of two continuously distributed variables as in Figure 1, the plot of X and Y in (1) does not show much. But the interpretation of r^2 (the coefficient of determination) as the proportion of variance in one variable explained by variation in the other holds for the categorical as well as the continuous case. (With a computer package like SAS (Barr, et al, 1976), it is easy to demonstrate this by calculating and printing predicted and residual scores, and computing their variances and σ_y^2.)

R x 2 Contingency Table

If the row variable has more than two categories, an R x 2 contingency table can be constructed, and the coding method in (1) can be extended. Code Y as before, and extend the coding of X as follows:

$$X_1 = \begin{cases} 1 & \text{if observation is from level 1 of A} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$X_2 = \begin{cases} 1 & \text{if observation is from level 2 of A} \\ 0 & \text{otherwise} \end{cases}$$

$$\cdots\cdots$$

$$X_r = \begin{cases} 1 & \text{if observation is from level R-1 of A} \\ 0 & \text{otherwise} \end{cases}$$

Regression Y on X_1, X_2, ..., X_r yields an R^2 which equals χ^2/N, where χ^2 is the test statistic for independence. Again, we can use our notions of R^2 to add meaning to the relationship between A and B tested using the value of χ^2.

A modification of the phi coefficient is made for contingency tables larger than 2 x 2: Cramer's V is given by

$$V = \left\{\frac{\phi^2}{\min\{(R-1),\ (C-1)\}}\right\}^{\frac{1}{2}} .$$

(The denominator is the maximum that ϕ^2 attains, so that V ranges from 0 when no relationship is presented to a value of 1.) Substituting R^2 for ϕ^2 and solving for R^2 gives

$$R^2 = V^2 \times \min\{(R-1),\ (C-1)\} \qquad (3)$$

So from Cramer's V or the χ^2 value, we can compute a proportion of variance of one variable accounted for by the other.

R x C Contingency Table

The most general form of the contingency table has R rows for variable A and C columns for variable B. Variable A may be coded X_i as in (2). But, the coding of Y needs to be an orthogonal partition of the variability in B. This is not difficult, using orthogonal polynominals, providing the frequencies in each level of A are equal. Assuming variable B to have four levels, code Y as follows:

$$Y_1 = \begin{cases} 1 & \text{if observation is from level 1 of B} \\ -1 & \text{if observation is from level 2 of B} \\ 0 & \text{otherwise} \end{cases}$$

$$Y_2 = \begin{cases} 1 & \text{if observation is from level 1 or 2 of B} \\ -2 & \text{if observation is from level 3 of B} \\ 0 & \text{otherwise} \end{cases}$$

$$Y_3 = \begin{cases} 1 & \text{if observation is from levels 1, 2, or 3 of B} \\ -3 & \text{otherwise} \end{cases}$$

Then regress each Y_i $(i = 1, 2, \ldots, C-1)$ on $X_1, X_2, \ldots, X_r$, denoting the respective multiple correlation coefficients by R_i.

Then

$$R^2 = \sum_{i=1}^{C-1} R_i^2$$

is equal to χ^2/N from the R x C table. Also, Cramer's V computed on the table using equation (3) yields an R^2 equal to χ^2/N.

Conclusion

The purpose of this paper was to relate common statistics from contingency table analysis to the more familiar R^2 terminology in order to better understand the strength of the relation implied. The method of coding contingency tables in order to compute R^2's was shown, as well as how R^2 relates to ϕ, V, and χ^2. It is not implied that all contingency tables be recoded so that multiple regression can be performed, but it is hoped that proportion of variance interpretations be done in addition to tests of significance.

References

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

Bishop, Y. M. M., Fienberg, S. D., & Holland, P. W. Discrete Multivariate analysis: Theory and Practice. Cambridge: MIT Press, 1975.

Glass, G. V., & Stanley, J. C. Statistical Methods in Education and Psychology. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1970.

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

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 10 No. 1 1979

CONTROLLING THE TYPE I ERROR RATE IN STEPWISE REGRESSION ANALYSIS

John T. Pohlmann
Southern Illinois University

Stepwise regression has become a widely used technique for selecting a subset of potential predictors for some dependent variable. Three procedures have been used under the rubric of stepwise regression analysis: Forward selection, backward elimination, and true stepwise (Draper and Smith, 1966).

The forward selection procedure forms a model of the dependent variable by first selecting the best single predictor, then the second predictor is chosen which makes the strongest contribution to the prediction of Y, controlling for the effects of the first predictor. The process continues so that at each step, the variable selected for inclusion in the model increases the prediction of Y more than any other predictor. The selection process stops when the remaining variables fail to contribute significantly to the prediction of Y. The backward elimination procedure begins with a model containing all potential predictors, and then at each step a variable is eliminated if its removal from the model results in the smallest reduction in the model's effectiveness. The elimination process continues until the removal of any variable results in a significant reduction in the model's R^2. The true stepwise procedure is a variant of the forward selection technique. It differs from the forward selection procedure in that at each step, a variable that has been previously included in the model may be deleted if a partial F-test shows that variable to be an insignificant predictor.

In most of the computer statistical packages that have stepwise regression procedures, the criterion used for variable selection is an F-test formed as follows:

$$F = \frac{R_F^2 - R_R^2}{(1-R_F^2)/(N-p-1)}, \tag{1}$$

where:

R_F^2 = the coefficient of determination for the model containing all predictors included at previous steps, plus the variable under test.

R_R^2 = the coefficient of determination for the model containing all predictors except the variable under test.

N = the number of observations.

p = the number of predictors used in the model that produced R_F^2.

As with any statistical test, two kinds of inferential errors can be made. A type I error would occur if a variable was selected, using the F ratio criterion, when that variable's population regression weight was zero. A type II error occurs when a variable is not selected, using the F-test criterion, when that variable has a non-zero population weight.

Most users of stepwise regression adopt one of the traditional significance levels (.05 or .01) when evaluating the F-test in (1). This significance level will determine the type I error rate for each test. However, another perspective can be taken when considering the type I error rate, the problem-wide error rate.

The problem-wide error rate is the probability of selecting any variable when all variables have population regression weights of zero. In other words, the problem-wide error rate is the probability of forming a sample regression model, when none should be formed. The rest of this paper addresses this error rate, and a procedure will be presented that allows researchers to control its value.

The problem-wide error rate is comparable to the family-wide error rate commonly encountered in the context of <u>post hoc</u> tests conducted after a significant effect has been found in an ANOVA. For example, the probability of making one or more type I errors in a family of orthogonal tests is:

$$\alpha_F = 1 - \prod_{i=1}^{k} (1-\alpha_i) \qquad (2)$$

where α_F = the family-wide error rate.

k = the number of orthogonal tests.

α_i = the significance level on test i.

When the α_i's are all equal to α_T,

$$\alpha_F = 1 - (1-\alpha_T)^k. \qquad (3)$$

If a researcher wished to control α_F by reducing α_T, (3) could be solved for α_T:

$$\alpha_T = 1 - \sqrt[k]{1-\alpha_F}\ . \tag{4}$$

Alternately, the researcher could conservatively approximate α_T using the Bonnferoni inequality,

$$\alpha_T \approx \alpha_F/k \tag{5}$$

When the members of the family of tests are not orthogonal, formulae (4) and (5) yield conservative values of α_T. That is, the use of α_T from (4) or (5) will result in an α_F less than the desired value. The solution for α_T is considerably more complex when the tests are not orthogonal. The solution for a critical F that will maintain α_F at a desired value should be done using the correlated F distributon (Pope and Webster, 1972). Unfortunately the integration of the correlated F distribution is an extremely tedious process, and only limited tables of critical values derived from it are available. Consequently, an approximate solution was sought using Monte Carlo methods.

METHOD

A Monte Carlo program, written in FORTRAN IV, was prepared by the author for this project. The program incorporated subroutines supplied in the International Mathematical and Statistical Library (1975). The IMSL subroutines were selected because of their proven accuracy and efficiency. A copy of the program is supplied in the Appendix of this paper.

The program generated sample data matrices (cases by variables) sampled with a given population dispersion matrix. Subroutine GGNRM was used for this purpose. Various population correlation matrices were supplied to GGNRM and a sample data matrix of standard normal deviates was produced. All population correlations between the predictors and the criterion variable were set equal to zero. The inter-predictor correlations were all set equal to a common value, and for the various replications examined in this study, the inter-predictor correlations were 0, .3, .5, .7, and .9. In addition, the numbers of predictors used were 2, 3, 4, 5, 7, 10, and 20. For every combination of the number of predictors and the average inter-predictor correlation (35 in all), a thousand sample data sets were generated.

Each data set thus generated was then subjected to a stepwise regression analysis using IMSL subroutine RLSTEP. Subroutine RLSTEP uses a true stepwise procedure. Variable selection is governed by a significance testing process. When, at any step, no F-test is significant, the selection process ceases.

For the purposes of this study, an error occured when a model, other than the null model, was formed by subroutine RLSTEP. The proportion of analyses resulting in a model was treated as an empirical estimate of the probability of erroneously forming a model using stepwise regression analysis.

RESULTS

Table 1 shows the results obtained when a variable selection significance level of .05 is used. The table entries in Table 1 are the proportion of 1000 stepwise regression analyses that produced a sample model when none should have been produced. For example, when a researcher has ten potential predictors that have correlations with each other equal to .50, the probability of erroneously forming a model is approximately .308.

Since the values in Table 1 are empirical estimates of the actual probabilities of making an error, there is some sampling error. The magnitude of the sampling error can be conservatively estimated by using the standard error of a proportion when $p = .5$. Since 1000 replications were used to derive each table entry, the standard error of a sample proportion will be less than or equal to .016. Consequently, a conservative 68% confidence interval for the true probability of making an error will be: tabled value ± .016.

The figures in Table 1 support two conclusions: (1) The probability of erroneously forming a regression model increases dramatically as a function of the number of predictors, and (2), as the inter-predictor correlation increases, the probability of making an error decreases. Consequently, any solution to the error rate problem must take into consideration the number of predictors and the inter-predictor correlation.

After Table 1 was prepared, an attempt was made to develop an algorithm that could be used to select a significance level for variable selection that would control the problem-wide error rate.

The rationale for the algorithm presented here was based on the formula that gives the family-wide error rate in k independent tests. Formula (3) is reproduced here for this purpose:

$$\alpha_F = 1 - (1-\alpha_T)^k , \tag{6}$$

All terms are defined in (3). If α_T and α_F are known, k can be solved for as follows:

$$k = \frac{\ln(1-\alpha_F)}{\ln(1-\alpha_T)} \tag{7}$$

Formula (7) was applied to each entry in Table 1, and the resulting k values are given in Table 2. In producing Table 2, α_T was .05 and α_F was taken as the corresponding value in Table 1. The k values in Table 2 were then plotted as a function of various measures of the inter-predictor correlation. Figure 1 shows one of these plots for the 10 predictor variable case. The k values were observed to be an inverse linear function of ρ^2_{xx}, the inter-predictor correlation. The following function was considered to be a reasonable approximation:

$$k = p - (p-1)\rho^2_{xx} \tag{8}$$

where p = the number of predictors

ρ^2_{xx} = the inter-predictor correlation.

This function seemed suitable since for the extreme cases of ρ^2_{xx}, 0 and 1.0, (8) produced k values of p and 1 respectively. When ρ^2_{xx} is equal to 0, the problem-wide error rate should equal the α_F value given by (6). Under this condition ($\rho^2_{xx} = 0$) the error rate is directly analogous to the family-wide error rate for a family of orthogonal tests. When ρ^2_{xx} is equal to 1, every predictor is linearly dependent on the other predictors, hence there is in fact only one predictor. Formula (8) yields a k value of 1, when ρ^2_{xx} equals 1. In addition, inspection of plots, such as Figure 1, suggested that (8) was also accurate for estimating k for values of ρ^2_{xx} between 0 and 1.

Unfortunately, a researcher using stepwise regression never knows ρ^2_{xx}, so it must be estimated. A less biased estimate of the squared correlation coefficient can be obtained using the shrinkage formula (McNumar, 1969):

$$\hat{\rho}^2 = 1 - (1-r^2)\frac{N-1}{N-2} \quad . \tag{9}$$

The estimate of ρ^2_{xx} used for this study was obtained as follows:

Let R_{pp} = the inter-predictor correlation matrix.

Define each element in $\hat{R}_{pp}$ as

$$\hat{r}^2_{ij} = 1 - (1-r^2_{ij})\,\frac{N-1}{N-2} \,, \tag{10}$$

where r^2_{ij} = the square of the ijth element of R_{pp}, and

N = the number of observations.

$$\text{Let } \bar{r}^2 = \frac{\sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \hat{r}^2_{ij}}{\frac{1}{2}(p^2-p)} \,, \tag{11}$$

which is the mean of the off diagonal elements of $\hat{R}_{pp}$.

The sample estimate of ρ^2_{xx} is then substituted into (8) to obtain

$$k = p - (p-1)\bar{r}^2 \,. \tag{12}$$

After k has been obtained via (12), α_T is obtained.

$$\alpha_T = 1 - \sqrt[k]{1-\alpha_F} \quad , \tag{13}$$

where α_F is the desired problem-wide error rate. A concise worked example is given in the Appendix of this paper.

The validity of the proposed algorithm was then tested by modifying the Monte Carlo program, used to produce Table 1, to use (13) to select an α_T. The results of this validation study are presented in Table 3. As can be noted in Table 3, the probability of erroneously forming a model, using (13) to determine α_T, approaches the desired value of .05. There is a slight tendency for this procedure to produce conservative values of α_T. The average value of α_F in Table 3 is .045, and the conservative nature of the procedure is most apparent for problems with large numbers of predictors and high inter-predictor correlations.

DISCUSSION

The type I error rate in stepwise regression analysis deserves serious consideration by researchers. The literature is replete with "significant" findings that fail the ultimate test of replication. One possible explanation for this state of affairs might lie in the increasing problem-wide error rate that can occur in stepwise regression analysis.

If a researcher considers the problem wide error rate important, he or she should take some corrective action. Three possibilities exist, depending on the kind of analysis contemplated. They are: (1) Prior to the stepwise analysis conduct an omnibus test of the model containing all potential predictors, (2) use the backward elimination procedure and use an α_T obtained by substituting the number of predictors for k in (13), or (3) use the algorithm for obtaining α_T presented here, if a forward selection or true stepwise procedure is used.

The Omnibus Test

The analysis begins by forming a full model containing all predictors. The R^2 for this model is tested for significance at the α_F level. The F is obtained as follows:

$$F = \frac{R^2/p}{(1-R^2)/(N-p-1)} \quad , \tag{14}$$

where R^2 = the coefficient of determination for the model containing all potential predictors,

p = the number of predictors,

N = the number of cases.

This F ratio yields a simultaneous test of significance for all weights in a model. Proceed with the analysis only if a significant F using (14) is obtained.

The Backward Elimination Procedure

The backward elimination procedure is comparable to testing a family of orthogonal hypotheses. At each step, the variance accounted for in the dependent variable that is tested for each predictor is independent of all other sources of variation. Consequently, the use of

$$\alpha_T = 1 - \sqrt[p]{1-\alpha_F} \quad , \tag{15}$$

will maintain α_F at its desired value.

Finally, the algorithm developed in this paper is recommended if a forward selection or true stepwise procedure is used. Since the value of α_T obtained using (13) will be greater than that obtained using (15), when some covariance among the predictors is present, the use of (13) will produce a more powerful analysis.

References

Draper, N. R., & Smith, H. Applied regression analysis. New York: John Wiley & Sons, 1966.

International mathematical and statistical library (5th Ed.). Houston: International Mathematical and Statistical Libraries, Inc., 1975.

McNemar, Q. Psychological statistics. New York: John Wiley & Sons, 1969.

Pope, P. T., & Webster, J. T. The use of an F-statistic in stepwise regression procedures. Technometrics, 1972, 14(2), 327-340.

Table 1

Monte Carlo Estimates of the Probability of

Erroneously Forming a Sample Model Using

Stepwise Regression Analysis with

a Variable Selection Significance Level of .05

Inter-Predictor	Number of Predictors						
Correlation	2	3	4	5	7	10	20
.0	.102	.130	.184	.216	.304	.410	.653
.3	.101	.130	.178	.213	.275	.367	.552
.5	.097	.128	.171	.196	.235	.308	.417
.7	.085	.125	.140	.153	.185	.225	.314
.9	.073	.094	.101	.111	.122	.126	.169

Table 2

k Values Derived Using Formula (7)

on the Values from Table 1

Inter-Predictor	Number of Predictors						
Correlation	2	3	4	5	7	10	20
.0	2.10	2.72	3.96	4.74	7.06	10.29	20.63
.3	2.08	2.72	3.82	4.67	6.27	8.92	15.70
.5	1.99	2.67	3.66	4.25	5.22	7.18	10.52
.7	1.73	2.60	2.94	3.24	3.98	4.97	7.35
.9	1.48	1.92	2.08	2.28	2.54	2.63	3.61

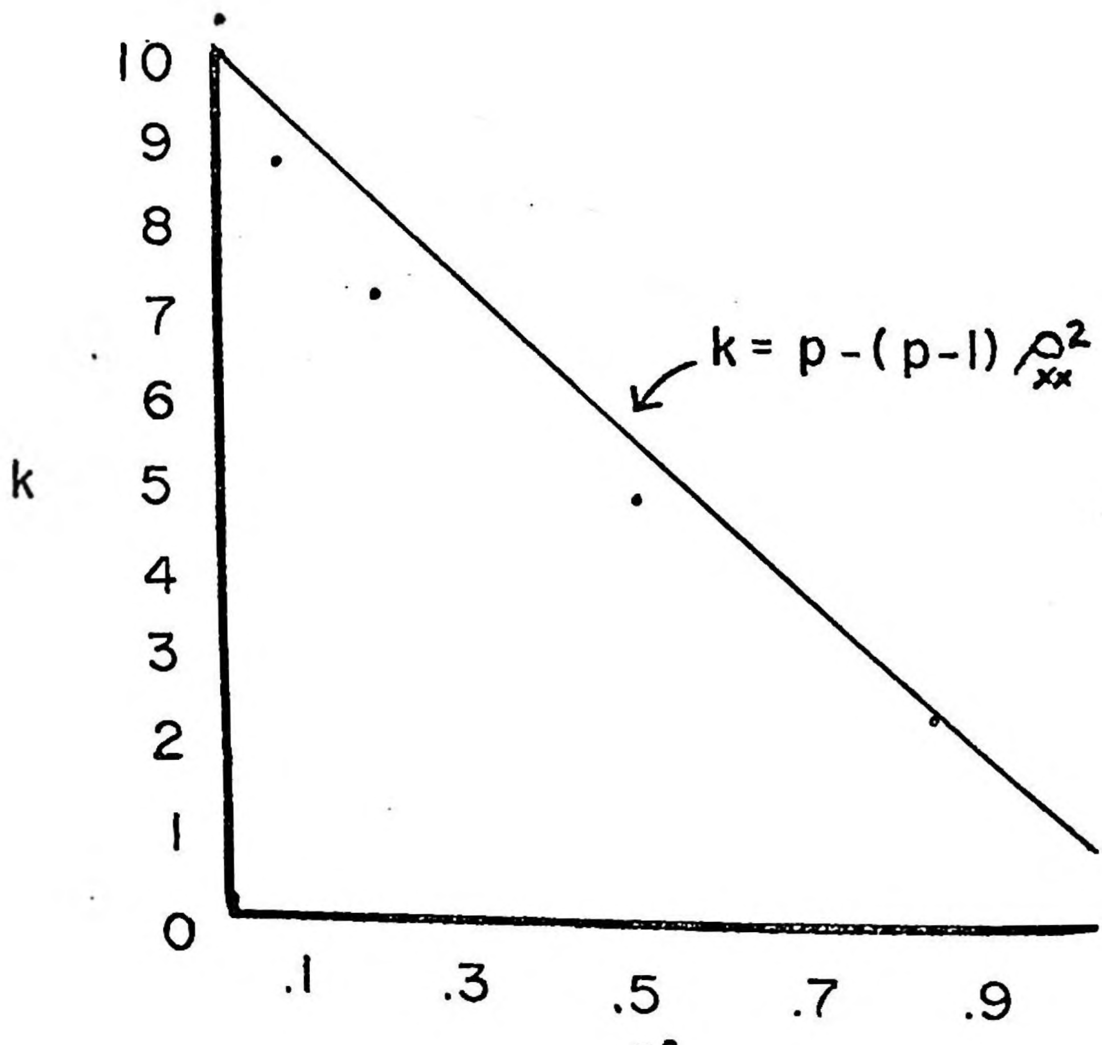

Figure 1. Plot of k as a function of ρ^2_{xx} for 10 predictors.

Table 3

Monte Carlo Estimates of the Probability of Erroneously Forming a Sample Model Using Stepwise Regression Analysis with a Variable Selection Significance Level Obtained Using Formula 13. The Desired α_F was .05

Inter-Predictor	Number of Predictors						
Correlation	2	3	4	5	7	10	20
.0	.052	.044	.058	.044	.048	.045	.055
.3	.050	.045	.044	.055	.046	.047	.038
.5	.060	.044	.041	.063	.041	.044	.042
.7	.059	.050	.041	.046	.037	.031	.032
.9	.045	.054	.056	.050	.033	.027	.011

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 10 No. 1 1979

TWO METHODS OF COMPUTING MATRICES OF WITHIN-GROUP CORRELATIONS USING FULL MODEL DUMMY VARIABLES

Gary J. Coles
American Institutes for Research

Abstract

Using matrices of pooled within-group correlations in identifying and defining multi-item indices on survey instruments permits the researcher to create indices that will not, for methodolgical reasons alone, be confounded with those group differences. This paper discusses how full model dummy variables can be used with partial correlation or multiple regression procedures to compute such correlation matrices.

Introduction

Because multi-item indices, if constructed properly, are more reliable than a single-item measure of the same construct, researchers usually attempt to develop internally consistent multi-item indices from available data. However, the large amount of data required for sound index development typically is not collected in a pilot test of the instrument and its items. One of the reasons for this is the availability of adequate resources for thorough pilot testing of instruments. Another is that in research efforts sponsored by the federal government, special clearance (taking anywhere between six and 12 weeks, or longer) must be obtained before any instruments can be administered by a contractor or grantee to more than a small number of individuals. Thus, a random sample of the cases in one's final data base is the most frequent source for data on which to conduct index development. The remainder of the cases in the data base can then be used to cross-validate the internal consistency of the indices identified and to provide index reliability estimates.

One frequently overlooked issue inherent with such an approach, however, is that index development may be affected by the very same group differences which one hopes to describe by means of the indices being identified and defined. This is because if one has drawn a random sample of one's data for index development purposes, cases from the various groups of interest should be selected (within the limits of sampling error) from each group in proportion to the number of cases in that group relative to the total number of cases in the data base. Thus, the procedure of using a random sample of cases in one's data base for index development and accessing the remaining cases for purposes of index cross-validation will not guarantee that among-group effects have been adequately controlled. Subsequent use of such indices to describe group differences could be misleading (at best) or could represent "bootstrapping" (at worst). Is there, however, a practical solution to this issue?

Discussion

Perhaps the most sound approach is to identify indices on a priori grounds and/or to ascertain if the items that are to belong to a given index are associated with one another, independently of systematic group differences among the items. Assuming that one is using simple correlations and factor analysis or principal components analysis to examine inter-item covariation, the proposed approach can be accomplished by analyzing a matrix of correlations from which group differences have been covaried out (i.e., analyze a matrix of pooled within-group correlations).

It can be shown that any correlation based upon data from individuals in different groups can be broken down into among-group and within-group components (Coles, 1976).[1] The within-group component is, in fact, a pooled within-group correlation coefficient or a pooled within-group multiple correlation coefficient in the case of a single criterion variable and multiple predictor variables. Because this is true, it is possible and practical to compute matrices of such coefficients using currently existing computer software.

Due to the widespread availability of multiple regression computer programs and, to a lesser extent, partial correlation computer programs, the most feasible methods to compute matrices of pooled within-group correlations are to do so via partial or multiple correlations. In both cases, the researcher must associate with each data record a series of k-1 dummy variables that encode that case's membership in one of the *k* mutually exclusive groups of concern in the study. In regression analysis terminology, this series of *k*-1 vectors are the variables included in the full model-equation for one's particular design. (And, because they are, these variables can be used in one's later analyses of group differences.) If one is using a partial correlation program, the matrix of partial correlations will, in fact, be a pooled within-group correlation matrix if the *k*-1 vectors are declared as covariables. This is because the full model equation's *k*-1 dummy-coded vectors describe all group differences and because the partial correlation program in effect will covary these from each of the items prior to their being intercorrelated.

The use of a multiple correlation program is somewhat more laborious because it is necessary to treat each item as a criterion variable, to compute the regression of each criterion on the full model equation's *k*-1 variables, to compute the residuals for each item given the results obtained (sometimes an optional form of output with some multiple regression/correlation computer programs), and to intercorrelate the residuals. The variance of the residuals is, obviously, the pooled within-groups variance since the full-model predictor variables encoded all group differences. And, intercorrelations among variables that have no among-group variance will be pooled within-group correlations since it can be shown that without among-group variance in two or more variables there can be no among-groups covariance (Coles, 1976).

There, of course, is no strict requirement that one's covariables in this particular approach to adjusting correlations be variables merely describing group differences. The full model equation could, for example, contain variables measuring attributes of individuals. In this case, one could covary out of each and every item, all among-group variance, as well as all variance (among-group and within-group) for the individual level variable(s). Such might be the case, for example, if one wished to identify and define multi-item indices that were independent of group/treatment membership and of the socioeconomic status of the individual. As in all such analyses, however, the particular covariates used should be chosen because there is a clearcut reason for their being used.

Technical Notes

[1]Within-group and among-groups covariance and variance components can be shown derivationally by partitioning the deviation of a given score from that variable's overall or grand mean into (1) the deviation of the score from the mean of all scores in the group of which it is a member plus (2) the deviation of that group's mean from the overall or grand mean. When this is done for two variables being correlated the formula for the Pearson product-moment correlation coefficient (r_{xy}) becomes

$$r_{xy} = \frac{Cov_{xy}(W) + Cov_{xy}(A)}{\sqrt{Var_x(W) + Var_x(A)}\ \sqrt{Var_y(W) + Var_y(A)}}$$

where

$Cov_{xy}(W)$ = weighted and pooled within-groups covariance

$Cov_{xy}(A)$ = weighted and pooled among-groups covariance

$Var_x(W)$, $Var_y(W)$ = weighted and pooled within-group variances

$Var_x(A)$, $Var_y(A)$ = weighted and pooled among-group variances

References

Coles, G. J. Guidelines for the use of multiple regression, commonality analyses and levels analysis with multi-level predictor variables. Appendix IV-E. In G. J. Coles, et al., Impact of educational innovation on student performance: Project methods and findings for three cohorts - Volume I Appendices. Palo Alto, California: American Institutes for Research, 1976. (ERIC Document No. ED-132-178)

MULTIPLE LINEAR REGRESSION VIEWPOINTS
Vol. 10 No. 1 1979

RECIPROCAL CAUSATION IN REGRESSION ANALYSIS

Lee M. Wolfle
Virginia Polytechnic and State University

With even the simplest bivariate regression, least-squares solutions are inappropriate unless one assumes *a priori* that reciprocal effects are absent, or at least implausible. While the discussion to follow is limited to bivariate regression, the issues apply equally to multivariate regression, including stepwise regression. McNeil (1976: 49) has written that, "most stepwise applications are based on one-shot studies that are not based on a priori hypotheses." This paper will demonstrate that *any* regression estimate is based on *a priori* assumptions. Furthermore, while the discussion is framed in the context of regression analysis, the issues to be raised apply to all forms of data analysis, including all analysis of variance designs.

A situation commonly faced by researchers is to have a set of random variables, each of which may be considered a dependent variable, to be explained by a different set of independent variables. Thus, each X_i could be regressed on all of the other variables, X_j $(j \neq i)$. In the simplest case, consider two variables, X_1 and X_2. With no *a priori* reason to select only one variable as dependent, we have:

$$X_1 = a_1 + b_{12}X_2 + e_1$$

$$X_2 = a_2 + b_{21}X_1 + e_2,$$

where X_1 and X_2 are random variables; a_1, a_2, b_{12}, and b_{21} are constants, their values to be estimated from the data; and e_1 and e_2

are disturbance terms, or residuals.

However, these two equations *taken together* have no unique solution. To demonstrate, take the regression:

$$X_1 = a_1 + b_{12}X_2 + e_1.$$

Letting $x_1 = (X_1 - \overline{X}_1)$ and $x_2 = (X_2 - \overline{X}_2)$, the equation may be transformed to:

$$x_1 = b_{12}x_2 + e_1.$$

The constant b_{12} is equivalent in both the original and transformed equations, as are the individual values of e_1. In order to obtain an unbiased least-squares estimate of the population value, β_{12}, one must assume,

$$E(x_2e_1) = 0,$$

where the E-notation indicates the expected value. The solution for b_{12} would then proceed in the usual way (see, for example, Kerlinger and Pedhazur, 1973). The equation,

$$x_1 = b_{12}x_2 + e_1,$$

however, may be rewritten,

$$x_2 = \frac{1}{b_{12}} x_1 - \frac{1}{b_{12}} e_1.$$

Letting $\frac{1}{b_{12}} = b_{21}$ and $\frac{-e_1}{b_{12}} = e_2$, we have,

$$x_2 = b_{21}x_1 + e_2,$$

which is a simple transformation of,

$$X_2 = a_2 + b_{21}X_1 + e_2.$$

In order to obtain a least-squares estimate of b_{21}, the assumption that $E(x_1e_2) = 0$ is required; but on the assumption that the first equation is true, including the specification that $E(x_2e_1) = 0$, it can be shown that $E(x_1e_2) \neq 0$. To do so, take the equation,

$$x_2 = b_{21}x_1 + e_2;$$

it may be multiplied by e_2, obtaining,

$$x_2e_2 = b_{21}x_1e_2 + e_2^2.$$

Taking the expected value of this equation, we have,

$$E(x_2e_2) = b_{21}E(x_1e_2) + E(e_2^2).$$

And by transposition and substitution, we obtain:

$$\begin{aligned} E(x_1e_2) &= \frac{1}{b_{21}} E(x_2e_2) - \frac{1}{b_{21}} E(e_2^2) \\ &= b_{12}E(x_2e_2) - b_{12}E(e_2^2) \\ &= b_{12}E(x_2[\frac{-e_1}{b_{12}}]) - b_{12}E([\frac{-e_1}{b_{12}}]^2) \\ &= \frac{-b_{12}}{b_{12}} E(x_2e_1) - b_{12} E(\frac{e_1^2}{b_{12}^2}) \\ &= -E(x_2e_1) - \frac{1}{b_{12}} E(e_1^2). \end{aligned}$$

Since $E(x_2e_1) = 0$ by assumption, we have,

$$E(x_1e_2) = \frac{-1}{b_{12}} E(e_1^2).$$

Thus, $E(x_1e_2)$ will be zero only when $E(e_1^2) = 0$; that is, when every data point falls on a straight line -- a rare occurrence, indeed.

In similar fashion, it can be shown that the usual formula for computing b_{21} is inappropriate in situations where we cannot by assumption eliminate reciprocal causation. Taking,

$$x_2 = b_{21}x_1 + e_2,$$

the equation may be multiplied by x_1, obtaining,

$$x_1x_2 = b_{21}x_1^2 + x_1e_2.$$

Taking the expected value of this equation, we have,

$$E(x_1x_2) = b_{21}E(x_1^2) + E(x_1e_2).$$

Each of these variables has a variance, and it is convenient to adopt the following notation for the variances and covariances:

$$\sigma_{x_1x_1} = E(x_1^2)$$

$$\sigma_{x_1x_2} = E(x_1x_2)$$

$$\sigma_{x_1e_2} = E(x_1e_2).$$

We may therefore write the covariance of x_1x_2 as,

$$\sigma_{x_1x_2} = b_{21}\sigma_{x_1x_1} + \sigma_{x_1e_2}.$$

If we were able to assume $E(x_1e_2) = \sigma_{x_1e_2} = 0$, then

$$b_{21} = \frac{\sigma_{x_1x_2}}{\sigma_{x_1x_1}},$$

the usual formula. But $E(x_1e_2) \neq 0$ when one assumes the first equation to be true. Therefore, b_{21} may not be solved with the least-squares formula used in the solution of b_{12}.

How does this conclusion square with the well-known fact that there are two regressions, X_1 on X_2 and X_2 on X_1? The regression of X_1 on X_2, as we have seen, requires the assumption that $E(x_2e_1) = 0$. This assumption is equivalent to saying that there is no reciprocal causation; that is, in order to estimate b_{12} one must assume $\beta_{21} = 0$. At the same time, regressing X_2 on X_1 requires the assumption that $E(x_1 e_2) = \beta_{12} = 0$. In order to obtain unbiased least-squares estimates of b_{12} and b_{21} *together*, one must assume $\beta_{12} = \beta_{21} = 0$. In other words, although nonzero numeric values may be attached to b_{12} and b_{21}, there can be no plausible interpretation of b_{12} and b_{21} *taken together*, when consideration is taken of the underlying assumptions that $\beta_{12} = \beta_{21} = 0$.

Furthermore, the two statistics, b_{12} and b_{21}, are constrained by the fact that,

$$b_{12} b_{21} = r_{12}^2.$$

That is, they must always be zero or of the same sign, and their product may never exceed unity. In situations in which reciprocity exists, this relation clearly leads to the conclusion that something is wrong. If, for example, one "knows" that the association of X_1 on X_2 is positive, but that the association of X_2 on X_1 is negative, the statistics, b_{12} and b_{21}, must nevertheless have the same sign. To continue the example, if the price of a certain commodity (X_1) was regressed on real income (X_2), one would expect b_{12} to be positive, since increases in real income would increase demand, thereby increasing the price of the commodity. On the other hand, an increase in the price of the commodity should produce a decrease in real income; nonetheless, b_{21} would be positive even when one "knows" it should be negative.

CONCLUSION

In sum, ordinary least squares is inappropriate when one cannot eliminate by assumption the possibility of reciprocal causation. To obtain least-squares estimates for the association of only two variables, one must "argue away" the existence of what is variously called simultaneity, feedback loops, or reciprocal causation. That is, if one decides that X_1 is the dependent variable, one must also assume at the same time that $\beta_{21} = 0$. In some cases this may be done with little ambiguity; if X_1 is a performance score and X_2 a dummy variable indicating the respondent's sex, clearly the former could not in any realistic way be said to have caused the latter. In other situations, the matter is not nearly as clear-cut. For example, in the case of the two variables, "educational plans" (X_1) and "best friend's educational plans," (X_2) educational plans can be thought of as deriving in part from the influence of the plans of one's best friend, and educational plans is regressed on best friend's plans. However, if one stipulates that one's educational plans are influenced by one's best friend, it must also be the case that from the perspective of the best friend, the respondent's plans influenced his or her educational plans. These variables are reciprocal causes of each other, and regressing one on the other would violate the assumption required of least-squares regression that the independent variable is uncorrelated with the error term.

Thus, in order to perform even the simplest bivariate regression, one has to make assumptions about the real relationship between variables. Specifically, one has to postulate the absence of reciprocal causation. In doing so, reference should be made to the models and theories which synthesize the area under study. In a few words, the researcher must have a firm grasp on reality in order to proceed with the analysis of his data. Data analysis thus becomes a two-way process; the researcher must have some basis for ordering variables before analyzing the data, which when finished will further illuminate the real world.

Do not conclude from this that regression models are bad. Most variables can be plausibly ordered in terms of their dependence on one another, and least-squares solutions are clearly appropriate. However, the researcher should be sensitive to the possibility of reciprocal effects, because when present the regression estimates are biased.

ACKNOWLEDGEMENT

The author would like to express his thanks and appreciation to two anonymous reviewers for *Multiple Linear Regression Viewpoints*. The errors that remain are mine.

REFERENCES

Kerlinger, Fred N., & Pedhazur, Elazar J. *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart and Winston, 1973.

McNeil, Keith. "Position Statement on the Roles and Relationships between Stepwise Regression and Hypothesis Testing Regression." *Multiple Linear Regression Viewpoints*, 1976, *6*, pp. 46-49.

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

Title
Author and affiliation
Indented abstract (entire manuscript should be single spaced)
Introduction (purpose—short review of literature, etc.)
Method
Results
Discussion (conclusion)
References

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

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

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

"Membership in the *Multiple Linear Regression Special Interest Group* is renewed yearly at the time of the American Educational Research Association Convention. Membership dues pay for a subscription to the *Viewpoints* and are divided into two categories: individual = $5.00; and institutional (libraries and other agencies) = $18.00. Membership dues and subscription requests should be sent to the Executive Secretary of the MLRSIG."

THE UNIVERSITY OF AKRON
AKRON, OH 44325

AKRON OHIO U.S. POSTAGE 53 P.B. METER 672269

122MCNEO
MCNEIL, KEITH A.
NTS RESEARCH CORP.
2634 CHAPEL HILL BLVD
DURHAM, NORTH CAROLINA 27707

TABLE OF CONTENTS

Multiple Linear Regression Viewpoints
Vol. 10, No. 1, 1979

ISSN 0195-7171