VOLUME 11, NUMBER [illegible]

SUMMER 1982

MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group on Multiple Linear Regression

MLRV Abstracts appear in
microform and are available from
University Microfilms International.
MLRV is listed in EBSCO Librarians' Handbook.

ISSN 0195-7171

Library of Congress Catalog Card #80-648729

MULTIPLE LINEAR REGRESSION VIEWPOINTS

TABLE OF CONTENTS

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 11, NUMBER 3 SUMMER 1982

ON USING THE AVERAGE INTERCORRELATION AMONG PREDICTOR VARIABLES AND EIGENVALUE ORIENTATION TO CHOOSE A REGRESSION SOLUTION

Beverly Mugrage
The University of Akron

Jesse Marquette
The University of Akron

Isadore Newman
The University of Akron

Background

Since its introduction in 1970 by Arthur Hoerl, the efficacy of ridge regression has been vigorously debated by statisticians. Notable are the debates in the Journal of the American Statistical Association, JASA, in 1980 (Smith and Campbell) and in Technometrics in 1979 (Draper and Van Nostrand). Much research among proponents of ridge regression concentrated on comparisons of various ridge regression solutions. Dempster, Schatzoff, and Wermuth (1977) compared 57 varietios of ridge regression; Galarneau-Gibbons (1981) compared ten of the most promising ridge algorithma. Both were simulation studies.

Presented at AERA 1982, MLR Special Interest Group
Not refereed by editorial staff

Since the introduction of the Monte Carlo method in 1949 by von Neumann and Ulam, simulation studies have been frequently used in statistics to solve problems otherwise difficult or expensive to solve. Monte Carlo simulation can be adapted to any situation for which a model representing reality can be designed and for which a mechanism to simulate this model can be effected.

Analysis of the recent literature of ridge regression reveals essential agreement that ridge regression is an appropriate alternative to least squares regression when predictor variables are highly intercorrelated. Another theme is common. Many researchers from Newhouse and Oman in 1971 to Galarneau-Gibbons in 1981 also suggest that the orientation of the beta vector with respect to the eigenvectors corresponding to the largest and the smallest eigenvalue of the X'X matrix determines the relative performance of ordinary least squares estimators and ridge estimators.

Purpose of this Study

The question of the predictive values of the orientation of beta and/or the average absolute intercorrelation among independent variables in guiding an investigator's choice of regression method is interesting and important. The availability of a computer simulation capable of producing data with given R^2 and average absolute intercorrelation made study of this question possible. The simulation was designed for the 1979 comparison of shrinkage formuli by Newman, McNeil, Garver, and Seymour.

Methods

Twelve populations of 1,000 cases were generated representing four different values of intercorrelation among predictor variables (0.80, 0.50, 0.30, 0.15) and three different values of R^2 (0.50, 0.30, 0.05). From each population 220 samples were drawn with replacement. There were 50 cases per sample.

For each sample generated, Marquette and Du Fala's statistical package ADEPT (1979) was used to calculate the ordinary least squares solution, the principal components solution and three ridge solutions. The ridge solutions chosen were the Lawless-Wang solution, the McDonald-Galarneau solution, and a Hoerl-Kennard-Baldwin solution. The Hoerl-Kennard-Baldwin solution is important historically and because of its good performance in previous studies. The Lawless-Wang solution is a Bayesian solution derived from the assumptions $Y \sim N(X\beta, \sigma^2 I)$ and $\beta \sim N(0, \sigma_\beta^2 I)$ with the ridge parameter $k = \sigma^2/\sigma_\beta^2$ estimated by $k = ps^2/\sum \lambda_i \gamma_i^2$. The McDonald-Galarneau solution is an iterative solution which estimates the true length of the beta vector by $Q = \hat{\beta}'\hat{\beta} - s^2 \sum \lambda_i^{-1}$ and then picks k to minimize $\left|\hat{\beta}'(k)\,\hat{\beta}(k) - Q\right|$. This procedure defaults to ordinary least squares if Q is negative. These three methods of determining k were different enough in derivation to be interesting to compare.

The study was a 3 x 4 x 5 factorial design. There were three values for R^2, four for average absolute intercorrelation and five regression methods.

The various regression solutions were ranked on four criteria:

1. Average variance of regression coefficients.
2. Error in regression coefficients as measured by $(\beta-\hat{\beta})'(\beta-\hat{\beta})$.

3. Mean square error.

4. Shrinkage of R^2 upon cross-validation.

For each sample, solutions were ranked from one to five with smaller rank indicating more desireable solution. Ranks were then summed for each solution on all criteria to give an overall measure of quality of solution.

The orientation of the coefficient vector, beta, with respect to the eigenvector associated with the largest eigenvalue of the $X'X$ matrix was calculated for each sample. For some populations the range of values for the orientation was small enough to cause computational difficulty in the computer packages used in this study. For this reason, the orientation of beta was categorized and interaction between regression method and the orientation of beta was determined using two-way analysis of variance. The decision to categorize the orientation of beta is discussed further in the results section.

Results

Since this study was exploratory, a significance level of $\alpha = .05$ was used. When multiple comparisons were made, the correction suggested by Newman and Fry, $\alpha = .05/n$, was applied (Newman and Fry, 1972). All tests were two-tailed.

Error in Beta

For all populations with high average absolute intercorrelation, $\overline{|r|} = .80$, the error in beta as measured by $(\beta-\hat{\beta})'(\beta-\hat{\beta})$ was significantly different for ordinary least squares regression and each of the ridge solutions tested. For high multicollinearity, the error in beta for each ridge solution was significantly different from that of every other ridge solution with only one exception: Lawless-Wang error in coefficients was not significantly different from that of Hoerl-Kennard-Baldwin for the population with R^2=.50 and $\overline{r}$ =.80. For each of the populations with high multicollinearity, Lawless-Wang regression produced the smallest error in coefficients while ordinary least squares and principal components regression accounting for 100 percent of the trace produced the largest error in coefficients.

For moderate multicollinearity (0.50 and 0 .30), there was always a significant difference between the error in beta for ordinary least squares and each ridge solution's error in beta. The error for the complete principal components solution also was significantly different from that of each of the ridge solutions. Error in beta did not differ significantly for OLS and complete principal components solutions.

For low multicollinearity ($\overline{r}$ =.15), ordinary least squares regression and complete principal components regression produced significantly different error of beta from each other as well as from each ridge solution.

For graphic representation of these results, see Figure 1.

Variance of Betas

For each population, for any given method, the coefficients of each independent variable formed a distribution. Thus if beta 1 is the coefficient of the first independent variable, a distribution for the ordinary least squares beta would exist, as well as one for the Lawless-Wang beta 1, the Hoerl-Kennard-Baldwin beta 1, and the McDonald-

FIGURE 1

Error in Regression Coefficients as a Function of Solution Type, R^2, and $\overline{|r|}$

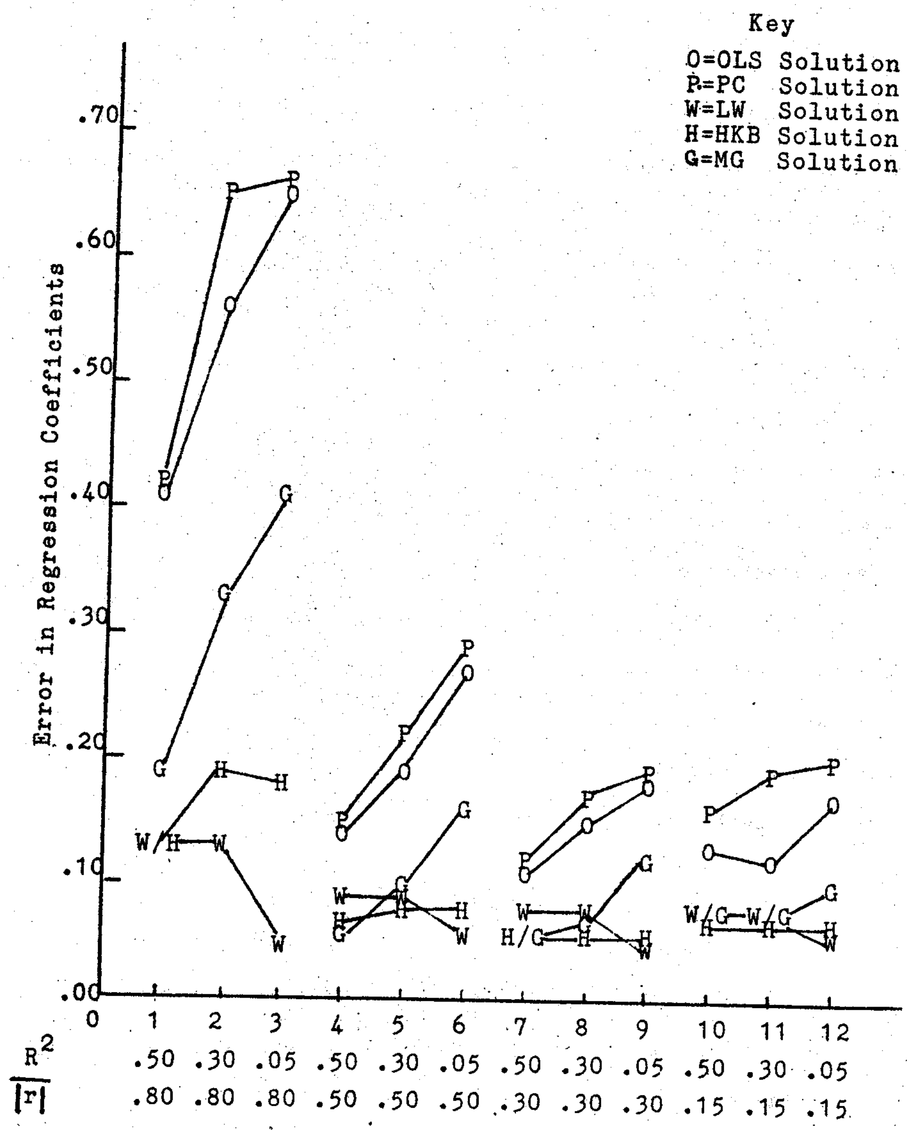


TABLE 1

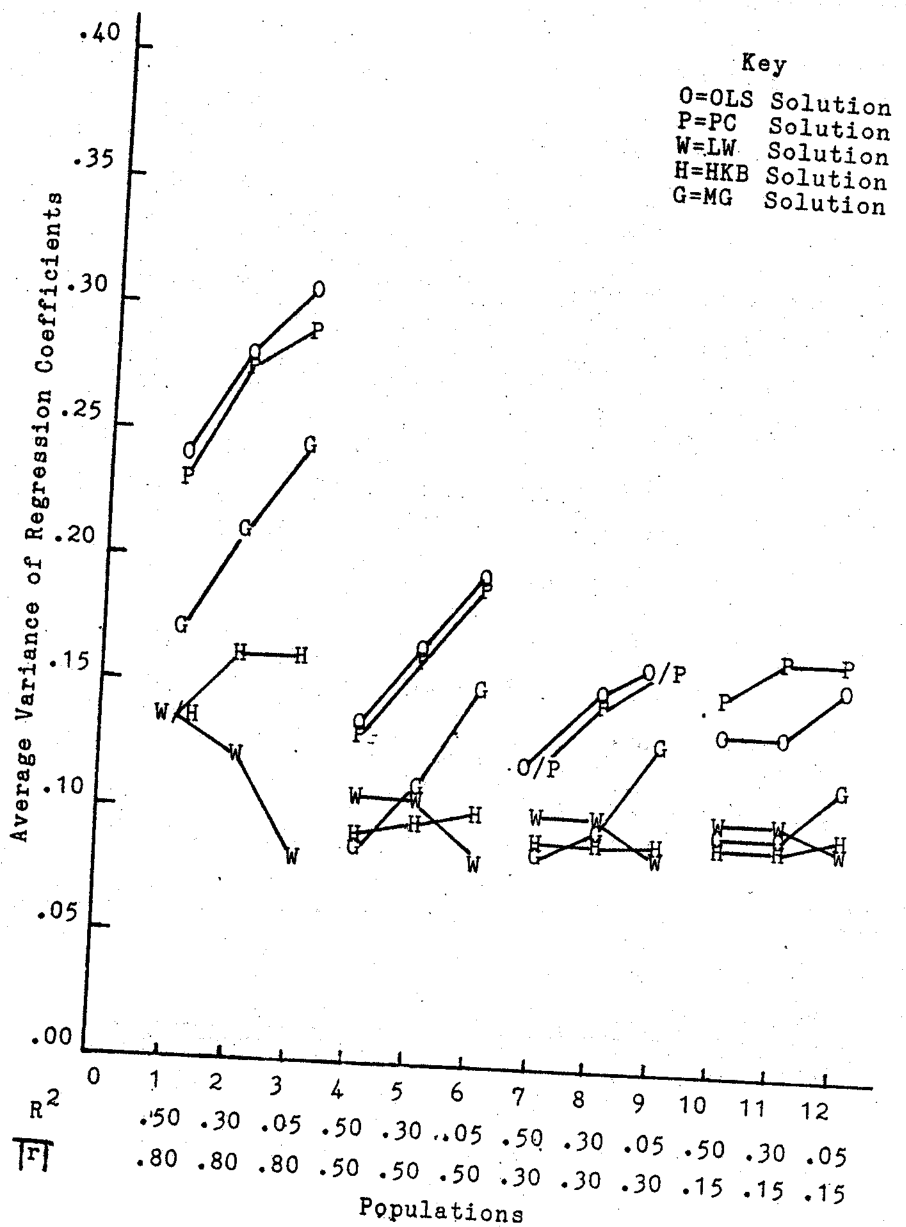
Summary of Results of Cochran's Test for Variance of Betas

Population Parameters $R^2/\|r\|$	Cochran's G for						
	β_1	β_2	β_3	β_4	β_5	β_6	β_7
.50/.80	.3480	.2661	.2979	.2785	.2827	.2825	.2695
.30/.80	.3537	.2881	.2509	.3113	.2950	.3081	.2671
.05/.80	.3300	.2850	.3134	.3038	.2982	.2842	.3021
.50/.50	.3052	.2512	.2686	.2969	.2581	.2491	.2741
.30/.50	.3221	.2587	.2689	.2981	.2767	.2749	.2726
.05/.50	.3322	.2759	.3006	.3049	.3196	.2753	.2841
.50/.30	.2785	.2485	.2718	.2638	.2512	.2415	.2621
.30/.30	.2870	.2534	.2805	.2710	.2535	.2706	.2697
.05/.30	.2983	.3099	.2903	.2862	.2644	.2666	.2656
.50/.15	.2506	.2831	.2873	.2738	.2571	.2728	.2473
.30/.15	.3308	.2753	.2904	.2639	.2771	.2921	.2377
.05/.15	.2748	.2806	.2746	.2619	.2639	.2801	.2845

All tests significant

Critical Region: $G > G_{.05} = .2360$

FIGURE 2
Average Variance of Regression Coefficients as a Function of Solution Type, R^2, and $\overline{|r|}$



Galarneau beta 1. Variances of these distributions were compared using Cochran's test, normality having been verified with a chi square test and sample size being equal. The results appear in Table 1. Cochran's test for each estimated beta for every population showed that the four variances compared were not all equal. To examine the relationship among the variances more closely, multiple comparisons $\alpha = .05/n$ was used for .05 significance. This is the correction suggested by Newman and Fry (1972).

For high multicollinearity (0.80) the variance of the ordinary least squares beta was significantly different from that of Lawless-Wang or Hoerl-Kennard-Baldwin beta for each independent variable. The ordinary least squares beta variance was higher than that of any ridge beta variance for each of the betas for the seven independent variables.

For all population (R^2 = 0.50, 0.30, 0.05) with high multicollinearity the Lawless-Wang estimator was always significantly different from that of the McDonald-Galarneau estimator and for R^2 = 0.05, it was significantly different from both of the other two ridge estimators. See Figure 2 for graphic representation of this information.

Shrinkage Upon Cross-Validation

The shrinkage in R^2 upon cross-validation was not significantly different among the various regression solutions for eight of the twelve populations including the population with R^2 = 0.50 and high average absolute intercorrelation (0.80). For the other two populations (R^2 = 0.30, and R^2 = 0.05) with high multicollinearity there was a significant difference in shrinkage of R^2 upon cross-validation

FIGURE 3

Shrinkage in R^2 upon Cross-Validation as a Function of Solution Type, R^2, and $\overline{|r|}$

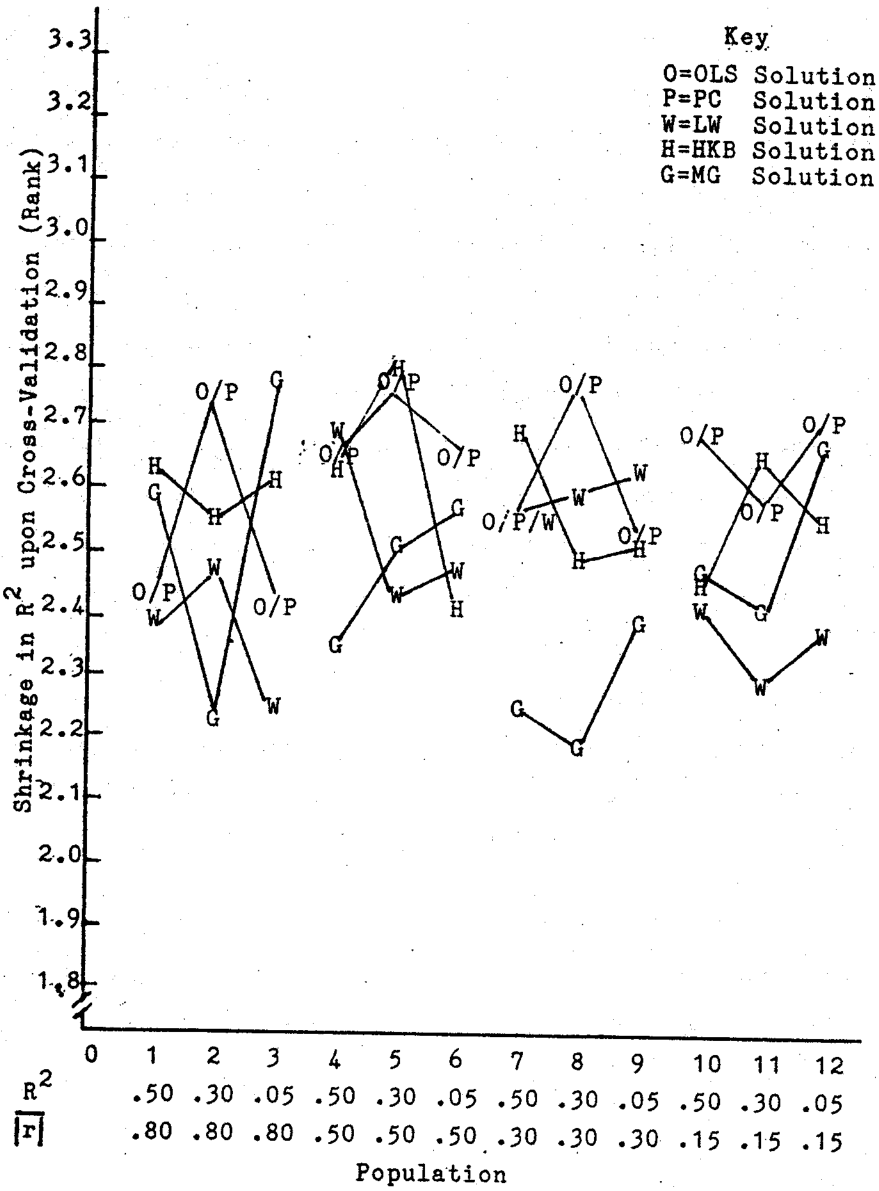
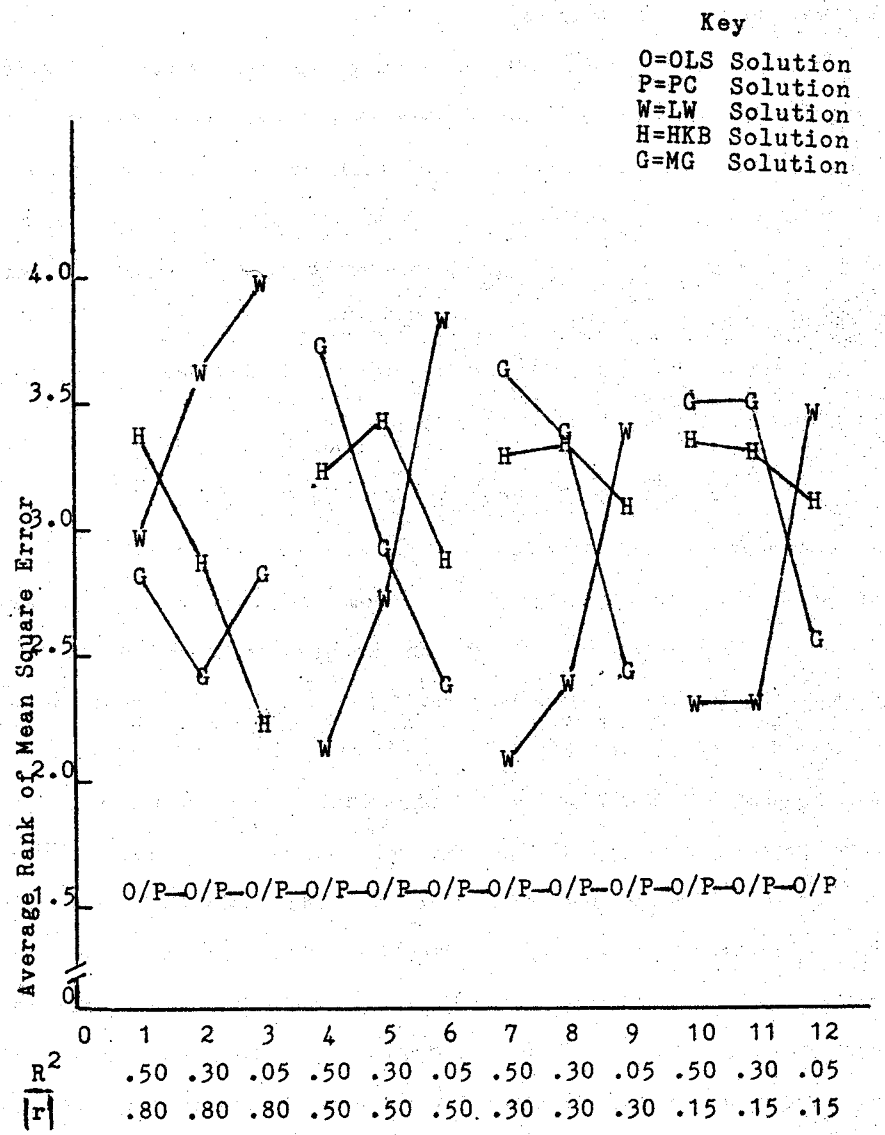


FIGURE 4

Average Rank of Mean Square Error as a Function of Solution Type, R^2, and $\overline{|r|}$



between ordinary least squares and at least some of the ridge solutions. For $R^2 = 0.30$, the ordinary least squares R^2 shrunk more than the ridge solutions and for $R^2 = 0.05$, the ordinary least squares R^2 shrunk less than the other estimators.

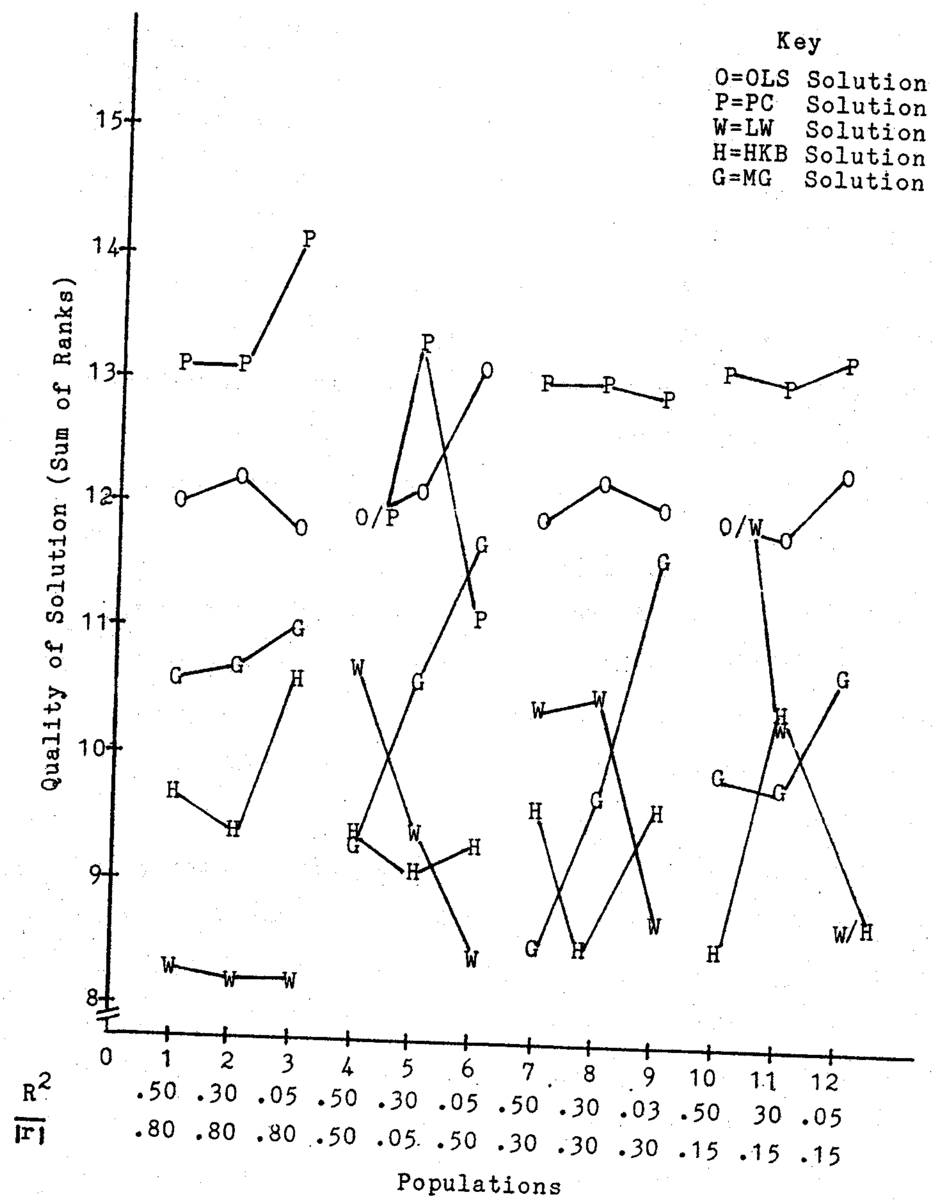
There is no evidence in the results of this study indicating that the ridge regression R^2 shrunk less than the ordinary least squares R^2 for populations with high multicollinearity, a situation in which ridge regression is commonly used. The actual value of the shrunken R^2 may be more useful than the value of the shrinkage of R^2 upon cross-validation.

See Figure 3 for graphic representation of shrinkage for varying R^2 and average absolute intercorrelation.

Means of R^2 Before and After Cross-Validation

Knowledge of the shrinkage in R^2 upon cross-validation may be less valuable than knowledge of the final value of R^2 upon cross-validation. The value of the shrunken R^2 gives a lower bound on R^2. Shrinkage in R^2 is of less interest. For this reason, means of R^2 before and after cross-validation were calculated for the ordinary least squares solution and the ridge solutions for each population. Before cross-validation, R^2 for the ordinary least squares solution was greatest. The values of R^2 for the ridge solutions were only slightly smaller. After cross-validation values for R^2 among the solutions were again close in value. McDonald-Galarneau ridge regression produced the largest R^2 after cross-validation for six of the twelve populations. Ordinary least squares regression and Lawless-

FIGURE 5
Quality of Solution as a Function of Solution Type, R^2, and $\overline{|r|}$



Wang regression produced the highest R^2 for three populations each.

MSE

As expected, the mean square error for ordinary least squares regression was significantly different from that of the ridge solutions for all populations. Generally, the MSE for the ridge solutions were significantly different from each other. For only three populations (R^2=0.50, $\overline{|r|}$ =0.80; R^2=0.30, $\overline{|r|}$ =0.50; R^2=0.30, $\overline{|r|}$ =0.15), was there no significant difference among ridge MSE. Of the ridge solution, the Lawless-Wang solutions had the lowest MSE for six of the twelve populations, McDonald-Galarneau for five, and the Hoerl-Kennard-Baldwin solution for only one of the twelve populations. Graphic representation of MSE for various values of R^2 and r is seen in Figure 4.

Overall Solution Quality

If overall quality is measured by the sum of ranks, analyses of variance indicated a significant F-ratio with a probability of 0.00000 for all populations. Representation of overall quality of solution as a function of R^2 and average absolute intercorrelation occurs in Figure 5.

A good solution was operationally defined as one whose sum of ranks was less than the mean sum of ranks. The number of good solutions for each method for each population are given in Table 2. For average absolute intercorrelation of 0.80 and R^2 = 0.50, 214 of 200 Lawless-Wang solutions were considered good compared with 157 of 220 Hoerl-Kennard-Baldwin solutions and 108 McDonald-Galarneau solutions. For all other highly multicollinear populations, results were similar.

TABLE 2

NUMBER OF GOOD SOLUTIONS

Population 1: $R^2=.50$, $|r|=.80$

Number of Samples: 214

Type of Solution	Number of Good Solutions
Ordinary Least Squares (OLS)	27
Principal Components (PC)	13
Lawless and Wang Ridge (LW)	214
Hoerl, Kennard and Baldwin Ridge (HKB)	157
McDonald and Galarneau (MG)	108

Population 2: $R^2=.30$, $|r|=.80$

Number of Samples: 212

Type of Solution	Number of Good Solutions
OLS	23
PC	9
LW	211
HKB	163
MG	91

Population 3: $R^2=.05$, $|r|=.80$

Number of Samples: 216

Type of Solution	Number of Good Solutions
OLS	22
PC	10
LW	216
HKB	152
MG	135

Population 4: $R^2=.50$, $|r|=.50$

Number of Samples: 215

Type of Solution	Number of Good Solutions
OLS	33
PC	4
LW	81
HKB	177
MG	184

NUMBER OF GOOD SOLUTIONS

Population 5: R^2=.30, $|r|$ =.50

Number of Samples: 215

Type of Solution	Number of Good Solutions
OLS	33
PC	3
LW	165
HKB	214
MG	100

Population 6: R^2= .05, $|r|$ =.50

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	0
PC	69
LW	217
HKB	169
MG	43

Population 7: R^2=.50, $|r|$ =.30

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	38
PC	7
LW	109
HKB	163
MG	204

Population 8: R^2=.30, $|r|$ =.30

Number of Samples: 217

Type of Solution	Number of Good Solutions
OLS	27
PC	5
LW	113
HKB	208
MG	157

TABLE 2

NUMBER OF GOOD SOLUTIONS

Population 9: $R^2 = .05$, $|r| = .30$

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	39
PC	8
LW	217
HKB	166
MG	99

Population 10: $R^2 = .50$, $|r| = .15$

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	40
PC	4
LW	115
HKB	216
MG	153

Population 11: $R^2 = .30$, $|r| = .15$

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	45
PC	7
LW	121
HKB	208
MG	65

Population 12: $R^2 = .05$, $|r| = .15$

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	27
PC	8
LW	174
HKB	217
MG	101

with Lawless-Wang regression producing the largest number of good solutions. For these same populations, in every case, principal components accounting for 100 percent of the trace produced the fewest good solutions followed by ordinary least squares regression.

One must be cautious in interpreting overall quality of solution done as a sum of ranks. In summing ranks, equal weighting is imposed on the criteria for good solution: variance of beta error in beta, shrinkage upon cross-validation, and MSE. This stacks the deck against the OLS solution and the principal components solution accounting for 100 percent of the trace. Theory tells us that ridge should outperform OLS on two of the four criteria used.

Orientation of the Beta Vector

To test for interaction of the orientation of the beta vector and method of regression solution, the orientation of beta was categorized and two-way analyses of variance were run. Categorization of the orientation became necessary because the small range the orientation exhibited in some populations presented serious computational difficulties using the ADEPT model comparison and DPLINEAR. For highly multicollinear data the interaction between the orientation of beta and method was nonsignificant. Significant interaction occurred for R^2=0.05, $\overline{|r|}$ =0.30, and R^2 =0.50, and $\overline{|r|}$ =0.15 only. For these levels of intercorrelation, ridge regression would rarely be considered the method of choice. Orientation of the beta vector appears of little usefulness in choosing among ridge regression methods for highly multicollinear data.

Conclusions

The results of this study indicate that for high degrees of multicollinearity, when stability and interpretability of coefficients is important, ridge regression is an attractive alternative to least squares regression. Low error and small variance of coefficients make ridge regression a useful device for anyone wanting to interpret beta weights for any reason, a device that should prove useful to social science investigators attempting to look at "causation" through correlation as in path analysis. Lawless-Wang ridge regression performed especially well on criteria for stability of coefficients in this study.

The major advantage to ridge regression is not in prediction nor in hypothesis testing but in applications for which the sign or interpretability of coefficients is important.

Principal components using all components was equivalent to the OLS solution in production of R^2, $\hat{Y}$, and MSE. It was not equivalent in variance or error of regression coefficients. For the principal component solution variance of coefficients increased rapidly as components associated with lower eigenvalues were added. Evidence from this experiment supports the use of a cut-off in using principal components regression (Rummel,1970). More work needs to be done concerning appropriate placement of such a cut-off.

Values for R^2 before cross-validation and values for R^2 after cross-validation were close for ordinary least squares and the ridge solutions tested in this study. The value of R^2 after cross-validation seems a more appropriate way of comparing solutions than shrinkage in R^2 upon cross-validation.

The orientation of the eigenvector associated with the largest eigenvalue of the X'X matrix with respect to the population beta vector does not appear to be useful in choosing among ordinary least squares regression, principal components regression accounting for 100 percent of the trace, Lawless-Wang ridge regression, Hoerl-Kennard-Baldwin ridge regression, or McDonald-Galarneau ridge regression.

It is clear from this study that the quality of a solution as determined by error in coefficients, variance of coefficients, MSE or R^2 after cross-validation depends upon the characteristics of the population. There is a strong dependence upon the degree of multicollinearity. Within a given multicollinearity, there is a dependence upon the R^2 of the population.

Ridge regression has a distinct advantage over OLS when stability and interpretibility of coefficients is important but not for purposes of prediction or hypothesis testing.

REFERENCES

Draper, Norman and Van Nostrand, R. Craig Ridge regression and James Stein estimaton: review and comments. Technometrics, 1979, 21, 451-465.

Gibbons, Diane Galarneau A simulation study of some ridge estimators. Journal of the American Statistical Association, 1981, 76, 131-9.

Marquette, Jesse and Du Fala, Margaret ADEPT: Advanced Date Enquiry Package Time-Shared, 1979.

Newhouse, Joseph and Oman, Samuel An Evaluation of Ridge Estimators, Santa Monica, CA: Rand, 1971, (R-716 PR).

Newman, Isadore, and Fry J. A response to "A Note on Multiple Comparisons" and a comment on shrinkage. Multiple Linear Regression Viewpoints, 1972, 2, 36-38.

Newman, Isadore, McNeil, Keith A., Garver, Thomas & Seymour, Gayle A. A Monte Carlo evaluation of estimated parameters of five shrinkage formuli. American Educational Research Association, 1979, 1-18.

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

Smith, Gary and Campbell, Frank A critique of some ridge regression methods. Journal of the American Statistical Association, 1980, 75, 74-81.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 11, NUMBER 3 SUMMER 1982

MULTIVARIATE NONPARAMETRIC ANALYSIS OF VARIANCE THROUGH MULTIPLE REGRESSION - THE TWO GROUP CASE

Bradley E. Huitema
University of Western Michigan

Abstract

The computation of the multivariate nonparametric analysis of variance requires matrix manipulations that are not familiar to many researchers. It is shown that the multivariate test statistic for the two group case can easily be computed with the aid of a conventional multiple linear regression computer program.

Presented at the annual AERA meeting March 19, 1982, New York City.

Introduction

In the randomized two group univariate analysis of variance case, situations arise where the nonparametric Mann-Whitney test is recommended in place of the parametric ANOVA $\underline{F}$ or $\underline{t}$ test or the corresponding regression analog. The choice between these parametric and nonparametric alternatives should generally be based on the nature of the population distributions and the adequacy of the measurement of the response variable. In the case that the population distributions approximate normality and the response measures are known to be

Presented at AERA 1982, MLR Special Interest Group
Not refereed by editorial staff

carefully obtained, the parametric procedures are generally chosen. This is because the relative efficiency (both asymptotic and small sample) of the nonparametric test relative to the parametric test is about .95. That is, if we compute the ratio of the sample sizes associated with the parametric and nonparametric tests having the same power and probability of Type I error, we find that fewer subjects are required for t or F than for the Mann-Whitney. Alternatively, when the sample size is constant, the power of the parametric test is greater. Many data analyzers appear to discount the usefulness of nonparametric alternatives for this reason and because the F test is said to be "robust" or insensitive to departures from distribution assumptions. It turns out, however, that a good case can be made for employing nonparametric statistics in certain situations.

If the population distributions are clearly nonnormal (e.g., exponential, rectangular, two-tailed exponential or long-tailed Cauchy) the parametric test is reasonably robust (using the typical textbook definition of robustness) but this does not mean that the inferences concerning the population means based on the sample means are equally good under all types of nonnormal distributions. The point here is that there is a difference between the effects of different types of nonnormality on a test criterion (such as F) and the effects on inferences made about paramaters. The former has to do with the concept of "criterion robustness" whereas the latter issue is that of "inference robustness". The reader is referred to Box and Tiao (1973) as the basic source on this distinction. The issue here is that the sample arithmetic means associated with a conventional parametric ANOVA may be inappropriate as estimates of the corresponding population means with certain types of nonnormality. The next point has to do with relative efficiency under nonnormality.

It was pointed out earlier that parametric _t_ or _F_ is generally preferable to the Mann-Whitney when normality is present because the relative efficiency of the latter is about .95. But what happens to the relative efficiency or power when the population distributions are clearly not normal?

If the deviation from normality is one of the long-tailed distributions, the Mann-Whitney test is _far_ more efficient. For example, the asymptotic relative efficiency of the Mann-Whitney when the populations are two-tailed exponential is 150%. If the population distributions are Cauchy the asymptotic relative efficiency of the Mann-Whitney is ∞ (infinity) and the efficiency of _t_ or _F_ is zero.

The practical data analyzer should not conclude that there is no use for nonparametric tests such as the Mann-Whitney just because he/she does not encounter extreme nonnormality. There is a second reason why one should consider the use of nonparametrics.

It is not unusual, especially in large studies that involve many variables, to encounter "outliers" or scores that are extreme relative to others in the distribution. Sometimes these extreme scores can be attributed to instrumentation failures or clerical errors. In these situations it makes sense to eliminate the obviously invalid scores from the analysis. But it is frequently the case that we don't know whether an extreme observation is the result of invalid measurement or not. When this happens it is not clear whether the observation should be discarded or left in the sample. A reasonable strategy in this situation is to transform the data in such a way that the extreme score(s) has less influence in the estimation of parameters than when raw data are employed. The ranking transformation, which is a part of the computation of the Mann-Whitney test, is a simple and effective way of decreasing the influence of outliers. Since the chance of encountering an outlier increases with the number of variables analyzed, it is argued here that nonparametric procedures should be given serious consideration in large exploratory studies.

<u>Purpose of Nonparametric Multivariate Analysis of Variance</u>

When multiple dependent variables are employed in a two-group study it is frequently suggested that a multivariate analysis of variance or the mathematically equivalent Hotelling T^2 be computed. These approaches are employed rather than (or in addition to) univariate tests on each dependent variable for two reasons. First, the univariate approach ignors possibly useful information concerning the covariances among the various response measures. Second, the multivariate methods control the probability of Type I error for the whole family of response measures. That is, the probability of making one or more Type I errors in the whole collection of dependent variable tests is equal to or less than the alpha level selected for the analysis. When studies containing multiple dependent variables are analyzed using univariate tests the probability of making a Type I error is greater than the nominal alpha associated with each test. Hence the multivariate approach involves running an overall test that simultaneously considers all dependent variables at once.

In the case of the two-group multivariate nonparametric analysis of variance, the null hypothesis is written as follows:

$$H_o: \quad \underline{\nu}_1 = \underline{\nu}_2 \text{ or } \qquad \begin{bmatrix} \nu_{11} \\ \nu_{21} \\ \cdot \\ \cdot \\ \cdot \\ \nu_{p1} \end{bmatrix} = \begin{bmatrix} \nu_{12} \\ \nu_{22} \\ \cdot \\ \cdot \\ \cdot \\ \nu_{p2} \end{bmatrix}$$

where ν_{ij} is the location parameter associated with the i^{th} dependent variable and the j^{th} population and

$\underline{\nu}_1$ and $\underline{\nu}_2$ are the vectors of the location parameters associated populations 1 and 2.

This is the hypothesis that the two populations are identical with respect to the p response measures. If this overall hypothesis is rejected there are several procedures that are appropriate for the identification of the dependent variable(s) responsible for the overall test conclusion. A simple approach is to run a Mann-Whitney test on each dependent variable. Issues associated with employing tests subsequent to the overall multivariate test are beyond the scope of the present paper.

The nonparametric multivariate techniques are virtually unused at the present time because they have been developed recently and the basic references (e.g., Puri and Sen, 1971) have been written primarily for mathematical statisticians rather than research workers. The purpose of this paper is to describe a simple procedure for computing the two group nonparametric multivariate analysis of variance with the aid of the output of a conventional multiple linear regression computer program.

Conventional Computation

The Puri and Sen nonparametric multivariate ANOVA procedure involves the computation of the test statistic $(N - 1)\mathrm{tr}\underline{BT}^{-1}$

$\underline{B}$ is the between or among group sum of products of ranks matrix and

$\underline{T}^{-1}$ is the inverse of the total sum of products of ranks matrix.

This test statistic* is evaluated as a chi square with $p(J - 1)$ degrees of freedom where p is the number of dependent variables and J is the number of groups.

*While Puri and Sen (1971) have shown that their test statistic $N\mathrm{tr}\underline{BT}^{-1}$ is asymptotically distributed as chi square, the small sample properties are

Regression Procedure

The multiple regression solution requires the following steps:

1. Construct a data matrix that contains a dummy variable to identify subjects in the two groups (column 1), all other columns contain the ranks associated with the p dependent variables included in the design.
2. Regress the group membership dummy variable on the ranks of the dependent variable scores to obtain the multiple rank correlation coefficient R_s
3. Square R_s
4. Multiply N-1 times R_s^2 to obtain the test statistic. That is, $(N-1)R_s^2 = \chi^2$.

It can be seen from a comparison of the conventional and regression approaches that the test statistics are $(N-1)\mathrm{tr}\underline{BT}^{-1}$ and $(N-1)R_s^2$ respectively. It follows that,

$$\mathrm{tr}\underline{BT}^{-1} = R_s^2.$$

A proof is presented in the Appendix.

---

not known (Puri, 1974). I have chosen to define the test statistic as $(N - 1)\mathrm{tr}\underline{BT}^{-1}$ because (a) this statistic is also asymptotically distributed as chi square with p degrees of freedom under the null hypothesis of identical populations and (b) this statistic reduces (exactly) to the Kruskal-Wallis chi square statistic in the case of one dependent variable. Since the small sample properties of the Kruskal-Wallis statistic have been found to differ little from the asymptotic results, it would be suprising if the small sample properties of the multivariate generalization suggested here differ from the theoretical results. There will be almost no difference in the results obtained using these two formulas with respectable sample sizes.

Table 1   Example Raw and Ranked Data from a Two Group Design with Three Dependent Variables

Raw Scores

Group I			Group II		
Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
3	10	12	21	56	11
17	17	7	27	57	10
20	51	5	35	62	6
70	53	0	38	63	1

Ranked Scores

Group I			Group II		
Y_1 ranks	Y_2 ranks	Y_3 ranks	Y_1 ranks	Y_2 ranks	Y_3 ranks
1	1	8	4	5	7
2	2	5	5	6	6
3	3	3	6	7	4
8	4	1	7	8	2

Computational Example

The computation of the multivariate test statistic for the data contained in Table 1 is summarized below for the conventional and regression solutions.

Conventional Solution

$$\underline{B} = \begin{bmatrix} 8.00 & 16.00 & 2.00 \\ 16.00 & 32.00 & 4.00 \\ 2.00 & 4.00 & 0.50 \end{bmatrix}$$

$$\underline{T}^{-1} = \begin{bmatrix} .12877 & -.07241 & .06746 \\ -.07241 & .06857 & -.02732 \\ .06746 & -.02732 & .06319 \end{bmatrix}$$

$$\underline{BT}^{-1} = \begin{bmatrix} .00649 & .46319 & .22886 \\ .01298 & .92638 & .45772 \\ .00162 & .11579 & .05722 \end{bmatrix} \quad \text{and}$$

$$\text{tr } \underline{BT}^{-1} = .00649 + .92638 + .05722 = .99009.$$

The test statistic is $(N - 1)\text{tr}\underline{BT}^{-1} = (7).99009 = 6.93$. Since the critical value of chi square based on $p(J - 1) = 3(1) = 3$ degrees of freedom is 7.81 for alpha = .05, the overall multivariate null hypothesis is retained.

Regression Solution

Step 1 Construct the data matrix as shown below.

(1)	(2)	(3)	(4)
Group Membership	Y_1	Y_2	Y_3
Dummy Variable	Ranks	Ranks	Ranks
1	1	1	8
1	2	2	5
1	3	3	3
1	8	4	1
0	4	5	7
0	5	6	6
0	6	7	4
0	7	8	2

It can be seen that all subjects in the first group have been assigned the dummy score of one and all subjects in the second group have been assigned the dummy score of zero.

Step 2 Regress the group membership dummy variable (column 1) on the ranks of the dependent variable scores (columns 2, 3, and 4). The resulting multiple correlation coefficient (actually the multiple rank correlation coefficient R_s) is .99503.

Step 3 Square R_s. R_s^2 is .99009.

Step 4 Multiply R_s^2 by N-1. (8-1) .99009 = 6.93 = χ^2. Notice that this is the same value obtained with the conventional computation procedure.

Since the obtained chi square does not exceed the critical value of 7.81 the following hypothesis is retained:

$$H_o: \begin{bmatrix} \nu_{11} \\ \nu_{21} \\ \nu_{31} \end{bmatrix} = \begin{bmatrix} \nu_{12} \\ \nu_{22} \\ \nu_{32} \end{bmatrix}$$

There is insufficient data to conclude that the population distributions are not identical. Since the overall hypothesis is not rejected there is no justification for additional tests on the individual dependent variables.

In conclusion, the nonparametric multivariate analysis of variance is a useful method for dealing with long tailed population distributions, possible outliers, and increased probability of Type I error associated with multiple response measures. It is easily computed with the aid of any multiple regression computer program.

Epilog

There is an alternative to the multivariate nonparametric analysis of variance for handling the problem of increased Type I error that is simple, effective and easily understood. This approach is described elsewhere (Huitema, forthcoming).

SUMMARY

There are two situations in which nonparametric procedures such as the Mann-Whitney test should be considered as useful alternatives to the parametric analogs: (1) when the population distributions are of certain nonnormal forms and (2) when the data contain unknown outliers. If responses are obtained on multiple dependent variables both of these problems are more likely to occur than in the univariate case.

An additional problem associated with the multivariate case is an increase in the probability of Type I error; that is, as the number of dependent variables is increased the probability of making a Type I error increases. One method of controlling Type I error is to employ the Puri-Sen nonparametric multivariate analysis of variance. It appears that the Puri-Sen method has virtually never been used. This is so because (a) the original papers presenting this procedure were written for mathematical statisticians (and are inscrutable for the typical research worker), (b) there are no secondary sources that describe the procedure, and (c) there are no widely distributed computer programs available to carry out the analysis.

The Puri-Sen test statistic can easily be computed for the two-group case by regressing a group membership dummy variable on the rank-transformed dependent variables and multiplying the resulting R^2 by N-1.

REFERENCES

Huitema, B. E. Bonferroni Statistics: Quick and Dirty Methods of Simultaneous Statistical Reference. Forthcoming.

Puri, M. L. Personal communication. February 13, 1974.

Puri, M. L. & Sen, P. K. On a class of multivariate multisample rank-order tests, Sankhyā, 1966, 28A, 353-376.

Puri, M. L. & Sen, P. K. Nonparametric Methods in Multivariate Analysis, New York: Wiley, 1971.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 11, NUMBER 3 SUMMER 1982

DISCRIMINANT FUNCTION ANALYSIS AS A POST - HOC OPERATIONS RESEARCH DETERMINANT OF CRITERION STRENGTH AND BINARY DECISIONING RELIABILITY

Dr. Brent E. Wholeben
University of Texas at El Paso
University of Washington

Portions of this study were performed pursuant to a grant from the National Institute of Education, and the Northwest Regional Educational Laboratory. Views presented are those of the author; and no endorsement from either granting institution should be inferred.

INTRODUCTION TO THE STUDY

This investigation sought to evaluate the utility of discriminant functions and their related statistics, in providing a practical post-hoc determinant of criterion strength and decisioning (sic) reliability for decision-making in the multiple alternatives environment (Wholeben, 1980a). Past experience with the use of binary integer programming (operations research) models in the selection of elementary school sites for closure during severe enrollment decline had demonstrated, that discriminant functions could provide a useful tool to the decision modeler -- not only to assist an evaluation of the model's reliability in constructing various solution set vectors (i.e. the schools to be closed versus those to remain open) in the form:

$$[\,1\,0\,0\,1\,1\,1\,0\,0\,0\,\ldots\,0\,]$$

where 1=open and 0=close; but also to provide an accountability framework for the public's understanding of the methodology utilized and the reasonableness of

Presented at AERA 1982, MLR Special Interest Group
Not refereed by editorial staff

the results (solutions) according to the criterion references employed. This current paper seeks to expand upon that 1979 investigation, and provide additional data supporting the use of discriminant functions as an effective post-hoc technique for evaluating not only decisioning reliability but also the relative impact which each of the applied criterion references provided to the construction of the resulting decision (solution set vector formulation).

This paper will proceed to first acquaint the reader briefly with the idea of multiple alternatives modeling (MAM), and present a strong rationale for evaluating and simulating potential alternative decisions via an easily constructable criterion-referenced methodology. Secondly, the reader will be introduced to the "tools" of the MAM evaluator, and the rudiments of a nomenclature which will be utilized within the body of this report. Next, the findings of the 1979 school closure model (SCHCLO) will be summarized as an indication of the utility of discriminant functions in assessing decisioning model reliability for the "complete" matrix model case -- that is, a criterion model with no empty cells due to missing or incomplete (irrelevant) data entries. Finally, the use of discriminant functions for assessing modeling reliability and individual criterion strength associated with each decision will be studied, utilizing the 1981 fiscal deallocation model (ROLBAK) for evaluating budgeting unit alternatives for deallocation during funding roll-backs; and emphasizing the "scant" matrix model case.

The objective of this paper remains to demonstrate the utility of discriminant functions in assessing the relationship between those criterion references designated as providing the rationale underlying the decisions made; that is, to correlate decision sets (solution vectors) with the criteria, and thus measure the relationship of criterion variance in the prediction of solution vector membership. Furthermore as an auxillary objective, the use of discriminant functions will also provide a useful 'at-hand' technique for understanding the weighted value (or strength) for each of the criterion referenced variables entered into the discriminant function formulation. Finally, these results will demonstrate the utility of discriminant functions in the assessment of decisioning reliability and criterion strength for both the "complete" and "scant" criterion matrix of values.

CRITERION STRENGTH AND DECISIONING RELIABILITY

Evaluation and all decision-making resulting therewith, demand a high degree of accountability, visibility and responsibility. Today's complex issues require equally complex methodologies to assess both content and process of such issues, and to provide an understandable environment within which to simulate potential decisions and measure resulting effect or impact. As important moreover, is the secondary demand for providing a means for post-hoc evaluating not only the results of the simulated decisions, but also the influence (singularly as well as collectively) which the criterion references lend in making the original decisions. The clear need for the criterion-referenced decision-maker therefore is to satisfy the following five objectives:

[1] to validate the sophisticated decisioning methodologies which are so necessary for addressing today's complex problems -- yet so often ignored, discounted or feared;

[2] to study criterion effect upon the decisions made, and the impact which the system receives via those decisions; and thereby understand differential criterion weighting and influence -- "what" made a difference in constructing the decisions, and the varying impact resulting;

[3] to provide a high degree of visibility, and therefore accountability, to the public interests served and affected via those decisions -- generating a milieu of trust within which the decisions, no matter how unexpected, can be trusted and accepted;

[4] to simulate the variable impact upon the decisions made by introducing additional criterion influences into the model, and thereby perform a path analysis from solution to solution as different criteria are utilized to construct each decision or solution -- satisfying the innate need of some individuals who must always ask, "... but, what if ...?"; and

[5] to permit easy and quick decisioning replication within an ever changing environment -- knowing the relationships between past successful decisions and the criteria used to construct those solutions, in order to understand the potential of future decisions based upon the new values of more current criterion measures.

This paper demonstrates the superlative ability of a parametrically-based, statistical technique to satisfy each of the five objectives stated above. Relying upon multivariate, linear regression techniques, DISCRIMINANT FUNCTIONS, constructed to relate criterion vectors to a singular 'solution set vector' containing either a binary (1,0) decision representation or the composite entries of a 'selection tally vector' (0,1,2,3,...), provide the basis upon which the required measures of criterion strength and decisioning reliability will be constructed.

Generally, the notion of criterion strength refers to the identification of those measures which in effect constructed the final decision or solution to the modeled problem; and furthermore provide a 'factor' measure of ordinal value or weight within that same group of 'solution-formation' variable measures. Specifically, criterion strength will address three fundamental questions existent within all decisioning evaluation:

[1] which criterion references most clearly defend the decisions made?

[2] to what extent are the criteria individually representative of the decisions made?

[3] how do the most discriminating criteria within this decision setting relate to each other in terms of importance and influence?.

This paper will illustrate the utility of discriminant function(s) formulation for answering these questions of criterion strength, respectively, by evaluating the following rudiments of discriminant analysis:

[1] criteria included within the formation of discriminant functions -- that is, which references were 'entered' into the composition of the prepared functions;

[2] order-of-entry of each of the variables which discriminate the final solution vector; and

[3] weight (or factor strength) relationship between the standardized canonical discriminant coefficients.

Generally, the notion of decisioning reliability refers to the degree of trust which is implicit to the decision model (in this case, the "multiple alternatives model" - MAM); implicit in the sense, that the decision-maker can accept the results of such a criterion-referenced technology, both in terms of content (viz., effect of the criterion references within the model) as well as process (viz., effect of the model upon the criterion references). Specifically, decisioning reliability will address two fundamental questions existent within all decisioning evaluation:

[1] to what extent are the criteria collectively representative of the decisions made?

[2] to what extent can the defined matrix of criterion references re-predict the original binary (include v. exclude) solution?.

This paper will illustrate the utility of discriminant function(s) formulation for answering these questions of decisioning reliability, respectively, by evaluating the following charactistics of discriminant analysis:

[1] canonical correlation coefficients which offer a measure of relationship between the 'set' of discriminating criterion references and the 'set' of dummy variables which are used to represent the solution vector; and

[2] the frequency of mis-inclusions and/or mis-exclusions (or over-estimations and/or under-estimations) discovered when the classification coefficients constructed to predict a solution with the known relationships among the discriminating criteria variables, are utilized to re-predict the original dependent variable (original solution).

DESIGN OF THE MULTIPLE ALTERNATIVES MODELING (MAM) FORMULATION

The complex issue of multiple alternatives decision-making is no stranger to the educational analyst. The selection of some number of schools from a relatively large pool of potential candidates for closure is a MAM problem. Each school site represents varying measures of effectiveness, efficiency, satisfaction and expenditure for each of a number of criterion references (e.g. capacity of building, heating requirements, building age, projected enrollment change over future years, safety factors of neighborhood, and proximity of other schools and their ability to absorb transferees in the event of the first school's closure). Some of these measures will be adjudged satisfactory (or nonsatisfactory) to varying degrees, and will be comparable with other schools across the district. However, to include one site for closure as opposed to another site means, that "good" aspects of a 'to-be-closed' school must be sacrificed in order to keep the other school operational, even though the 'to-be-kept-open' school may have certain unsatisfactory measures on the same criterion variables which the now closed school exhibited as satisfactory. Such modeling of this decisioning situation is known as interactive effects modeling (Wholeben, 1980a), and represents the necessity of constructing solutions sets which will invariably include some form of 'controlled' preference/trade-off mechanics as the various alternatives are evaluated. The issue of complexity is also represented in the statement of the problem: <u>to</u> select some number of schools for closure in order to promote certain defined goals of the district; and thus to determine how many schools will be closed <u>and</u> which ones. Obviously, such a model must in effect be simultaneously performing these two inter-related decisions: "how many?" <u>and</u> "which ones?".

The determination of which program unit budgets will be decisioned for continued funding (versus deallocation) is another example of the multiple alternatives framework, and its superior contribution to the realm of accountable and criterion-referenced evaluation and decision-making (Wholeben and Sullivan, 1981). In the fiscal deallocation model, criteria represent the projected expenditures within each object cost code for each of the units under evaluation; and in addition contain perceptual measures of administrative level of expendability. Once again of course, exists the dual responsibilities for determining how many program budgets will be discontinued, and which ones -- based upon the interactive modeling effects of the various criterion weights across unit alternatives.

The multiple alternatives model is simply a system of simultaneous linear inequalities and equalities which collectively represents the problem to be solved. Such an algebraic linear system is portrayed in ‹Figure 1›. Note how each linear combination represents a vector of values (<u>viz</u>., coefficients) which identifies the total, measureable impact to a system of the alternatives being modeled. Thus there exists a unique (normally) combination of coefficients for each of the criterion references used as input to the decisioning process. The alternatives themselves are further defined as binary variables (that is, taking on the value of either 0 or 1 (to be excluded in the final solution set, or to be included, respectively). Vector formulation for each criterion reference,

$$[\ a_{i1}x_1 \quad a_{i2}x_2 \quad a_{i3}x_3 \quad \ldots \quad a_{ij}x_j\ ]$$

portraying $\underline{i}$ criterion references across $\underline{j}$ alternatives, will then provide a basis for measuring total impact to the system as a whole attributable to the solution set constructed. Bounds (or limits) to what is allowable as a total impact to the system are expressed as vector entries within the conditional vector (or normally named, RHS, the right-hand-side). The RHS-values are the constants of the equations and inequalities modeling the system. ‹Figure 2› presents a listing of the four generic types of criteria to which each model should address content validity; and ‹Figure 3› depicts these criterion entries as members of the modeling framework previously illustrated within Figure 1.

The remainder of the modeling process concerns the use of an additional vector to assist in determining from the potentially hundreds (or millions, in some exercises) of possible alternatives, that one, best mix for which the best, possible solution exists. This process is called the search for optimality, and the vector is known as the objective function (or sometimes, the cost vector). Geometrically, the objective function is a n-1 dimensional figure passing through the n-tuple space (convex) which is feasible (that is, includes all of the constraints postulated through the use of the linear equalities and inequalities) and which seeks a minimum point within the feasible region (if the goal is to minimize the impact of the objective function's values upon the system) or a maximum point within the feasible region (if the goal is to maximize the defined objective function's impact to the system as a whole).

Simply stated, the multiple alternatives model is a technique which seeks to construct a solution set (a vector of 1's and 0's), such that this same solution vector represents the solution of the simultaneous system, constrained by a series of competing criterion measures (vectors), and based upon the optimality demands of the objective function.

Figure 1. Representation of the Augmented Decision Matrix Model as the "Multiple Alternatives Model" (MAM).

(Decision Variables)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8		(RHS)
Constraint #01	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}		b_1
Constraint #02	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	a_{27}	a_{28}	. . .	b_2
Constraint #03	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	a_{37}	a_{38}	$\leq$ = $\geq$	b_3
Constraint #04	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	a_{47}	a_{48}	. . .	b_4
Constraint #05	a_{51}	a_{52}	a_{53}	a_{54}	a_{55}	a_{56}	a_{57}	a_{58}		b_5

c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8

Cost Vector Coefficients

$$\text{Optimize: } \sum_{j=1}^{8} c_j x_j \quad \text{st: } \sum_{j=1}^{8} a_{ij} x_j \ (\leq, =, \geq)\ b_j \quad x \geq 0$$

(If MILP, x is integer; if decisonal, x=0,1 only.)

Figure 2. Representation of a Generic-Criterion Decisioning Model for Analyzing Multiple Competing Alternatives.

Criterion	Foci	Multiple Alternatives A_1 A_2 A_3 A_4 . . . A_n
(Effectiveness Criteria)		
$CRIT_1$	EFFEC-1	a x n sub-matrix
$CRIT_2$	EFFEC-2	effectiveness measures across alternatives
. .	. .	
$CRIT_a$	EFFEC-a	
(Efficiency Criteria)		
$CRIT_{a+1}$	EFFIC-1	b x n sub-matrix
$CRIT_{a+2}$	EFFIC-2	effectiveness measures across alternatives
. .	. .	
$CRIT_{a+b}$	EFFIC-b	
(Satisfaction Criteria)		
$CRIT_{a+b+1}$	SATIS-1	c x n sub-matrix
$CRIT_{a+b+2}$	SATIS-2	satisfaction measures across alternatives
. .	. .	
$CRIT_{a+b+c}$	SATIS-C	
(Expenditure Criteria)		
$CRIT_{a+b+c+1}$	EXPEN-1	d x n sub-matrix
$CRIT_{a+b+c+2}$	EXPEN-2	expenditure increases across alternatives
. .	. .	
$CRIT_{a+b+c+d}$	EXPEN-d	

Figure 3. Fiscal Allocations as a Multiple Alternative Problem, Utilizing the Decision Matrix Framework.

Multiple Alternatives

Criteria		Prog1	Prog2	Prog3	. . .	Progn	
Positive Impact	1.__	+11	+12	+13	. . .	+1n	
	2.__	+21	+22	+23	. . .	+2n	Maximize
	3.__	+31	+32	+33	. . .	+3n	
Negative Impact	1.__	-11	-12	-13	. . .	-1n	
	2.__	-21	-22	-23	. . .	-2n	Minimize
	3.__	-31	-32	-33	. . .	-3n	
Specific Costs	1.__	\$11	\$12	\$13	. . .	\$1n	
	2.__	\$21	\$22	\$23	. . .	\$2n	Sum $\leq$ total budget available
	3.__	\$31	\$32	\$33	. . .	\$3n	

TOOLS OF THE MULTIPLE ALTERNATIVES MODELING MAM FORMULATION

To construct discriminant functions from the relationships between the model just discussed above and the resulting solutions formulated, require the use of linear vectors and combinations of vectors (matrix). Only those vector and matrix formulations most germane to this paper will be discussed below. The reader is invited to be patient until the scheduled publication of the manuscript, "Multiple Alternatives Analysis for Educational Evaluation and Decision-Making" in late summer of 1982, for a detailed illustration of all vectors and matrice pertinent to MAM.

<u>Solution Set Vector.</u> In order to distinguish between alternatives included or excluded as members of the final solution to the system modeled, a vector of binary-decision representations is required, in the form:

$$[\,1\ 0\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ \ldots\ 1\,]$$

where '1' means that the criterion values associated with that particular x(j) will be computed to measure resulting system impact; and '0' means that the underlying criterion values will have <u>no</u> impact upon the system.

<u>Selection Tally Vector.</u> To observe the effect of each criterion reference upon construction of the system solution, a method called cyclic optimization (Wholeben, 1980a; Wholeben and Sullivan, 1981) is used. Under this regimen, the model is executed once for each unique criterion being used to constrain the model, where each unique criterion is cycled through the model as the objective function. For example, during one execution in the case of the school closure model, the intent may be be prepare a solution set whereby existing capacity of the remaining schools will be maximized; in another cycle, the model will be executed such that the schools remaining open within the district will minimize the amount of energy expended for facility heating requirements. The selection tally vector is basically a frequency summation vector, compiling the number of times each alternative was chosen as part of the solution vector, across all cyclic optimizations. Such a vector will be represented as:

showing that the first alternative was selected as solution a total of 3 times, the second alternative a total of 7 times, and so forth. This vector is extremely important when the MAM procedure requires a step-wise decisioning process such as the school closure model -- evaluating a revised database after closing a single school such that the effects of closing each individual site is summarily incorporated into the next decision for determining additional site closures.

Discriminant Criterion Inclusion Vector. This vector simply represents another binary entry vector of 1's and 0's, signifying which particular criterion references were utilized via discriminant functions to develop the canonical classification coefficients, and the standardized canonical discriminant function coefficients.

Discriminant Criterion Entry Vector. This vector contains 1,2,...,k entries, where k criteria were utilized in the development of the discriminant functions, and the 1,2,...,k entries represent their order of entry into the discriminant formulation. Criterion variables not entered into the function(s) receive a value of '0', by convention.

Discriminant Weighting Summary Vector. Applying discriminant procedures to the binary solution vectors will result in the computation of standardized canonical discriminant function coefficients. These coefficients will reflect the utility of entered criterion vectors if those vectors contain standardized measures in lieu of the normal raw scores. By dividing each of the standardized canonical coefficients by the smallest of the standardized canonicals, the quotient will provide a factor of importance for each of the criteria as relative to the other criterion entered in the discriminant formulation. The discriminant weighting summary vector is a linear representation of these factors (quotients), where the minimum entry value is always '1.00' (smallest standardized coefficient divided by itself). Non-entered criterion locations receive a value of '0.00' by convention.

Other 'tools' have been referenced in the proceeding section of this paper: criterion constraint matrix, condition limits vector (RHS), objective function vector, and the cyclic optimization tracking matrix. Other formulations are currently under study by the author (e.g. the optimality weighting matrix) to investigate new relationships which may allow greater accountability and useful reliability of the multiple alternatives modeling framework.

THE "COMPLETE" MATRIX CASE: THE SCHOOL CLOSURE MODEL (SCHCLO)

A total of 32 elementary school sites were measured across 24 relatively independent criteria, resulting from previous factor analyses of an original set of 64 criterion references. The criteria chosen were utilized by the multiple alternatives model for school closures (SCHCLO; Wholeben, 1980a) to evaluate the population of sites for some set of defined closures based upon the characteristics of the data; and the needs of the school district involved. Because the criteria utilized portrayed different value orientations (i.e., positive effects to be maximized; or negative effects to be minimized), the model consisted of a total of 18 cyclic MAXIMIZATIONS, and 6 cyclic MINIMIZATIONS -- for the total 24 optimizations required. The strategy was to operationalize the cyclic model, evaluate the full N=32 sites, analyze the selection tally results, choose a single site for closure, update the database to signify the closure, and then re-evaluate the now reduced N=31 site model for an additional closure. This step-wise closure strategy was considered consistent with the pragmatic reality of deciding school closures due to severe enrollment declines.

‹Figure 4› displays the results ("tracking matrix") of the N=32 cyclic optimization; and in addition, the selection tally vector entries (right column vector). The asterisked (*) vector entries signify those sites considered having the most potential for closure, due to the selection tally entries. These 4 sites were simulated 'closed' (i.e. included as '0' in the solution set vector); and a stepwise discriminant function analysis performed to analyze the relationship between the 24-vector criterion matrix which purportedly constructed the solution set, and the solution set thus constructed.

‹Figure 5› displays the results of the N=32 discriminant analysis. The single discriminant function constructed required a total of 8 criterion vectors to adequately explained the variance found within the binary solution set of 4-0's and 28-1's. The group-correlative relationship between these 8 criteria and the dummy variables formed by the solution set vector, was a canonical of .8512, explaining 72.5 percent of the variance between the criterion and solution sets. Based upon the re-classification coefficients formed, the discriminant function

CRITERION SOURCE(S) FOR SCALING OF THE OPTIMIZATION (COST) VECTOR (1=MAXIMIZATION, 2=MINIMIZATION)

(N=32) SITE	(1) AREA	(1) ENROL	(1) SURVIVE	(1) POTENT	(1) STUD PROX	(2) STUD OL	(2) ZONE OL	(2) SITE AGE	(1) INTER 13	(2) INTER 01	(1) THERM EFF	(2) ENER WAST	(1) CLASS RM	(1) MINOR ITY	(2) SITE AGE	(1) AREA CAPC	(1) AREA HEAT	(1) AREA ELEC	(1) AREA MAIN	(1) AREA REPR	(1) AREA UTIL	(1) ENR HEAT	(1) ENR ELEC	(1) ENR MAIN	(DECISION=CLOSE) FREQUENCY
01			X																						1
02																									--
03						X								X											2
04												X			X										2
05						X										X					X				3
06									X						X					X					3
07											X	X											X		3
08						X				X						X		X							4
09																								X	1
10	X						X	X					X												4
11	X					X			X				X												4
12						X																			1
13																			X						1
14		X									X	X										X	X	X	6 *
15																X									1
16									X	X							X		X		X				5 *
17		X								X	X	X			X							X		X	7 *
18										X					X										2
19										X			X									X	X		4
20			X		X							X								X					4
21																									--
22																	X	X	X		X				4
23					X												X								2
24							X	X							X										3
25					X																				1
26				X										X											2
27			X																						1
28				X			X	X												X					4
29	X																								1
30		X		X			X	X						X											5 *
31																		X							1
32							X	X																	2
Frequency Total: (Site Selection Per Criterion Focus)	3	3	3	3	3	5	5	5	3	5	3	5	3	3	5	3	3	3	3	3	3	3	3	3	

(*=MIP-4A-03 membership)

Frequency of Site Selection for MIP-4A Modeled Closure (N=32) Based Upon Varying Optimality Factors

Figure 4.

Figure 4.

Figure 5.

Summary of Discriminant Function Analysis

Based Upon MIP-4A Results (N=32)

STEP	CRITERION ENTERED	CRITERION REMOVED	WILKS' LAMBDA	SIGNIFICANCE
1	ENROL		.7068	.0014
2	AREAUTIL		.5904	.0005
3	CLASSRM		.4684	.0001
4	AREAREPR		.4124	.0001
5	ENRMAIN		.3628	.0000
6	INTERO1		.3183	.0000
7	SURVIVE		.2921	.0000
8	POTENT		.2755	.0001

Eigenvalue = 2.63000 Canonical Correlation = .8512

Classification Results:

Actual Group	Cases	Close	No Close
Close	4	4 (100.0%)	--
No Close	28	--	28 (100.0%)

Percent of "grouped" cases correctly classified: 100.0

Figure 6.

Summary of Discriminant Function Analysis

Based Upon MIP-4A Sum Results (N=32)

STEP	CRITERION ENTERED	CRITERION REMOVED	WILKS' LAMBDA	SIGNIFICANCE
1	INTERO1		.5745	.0096
2	STUDOL		.3864	.0044
3	AREAREPR		.2320	.0008
4	ENERWAST		.1502	.0003
5	POTENT		.0972	.0001
6	MINORITY		.0666	.0001
7	AREAELEC		.0457	.0001
8	SITEAGE		.0310	.0001
9	THERMEFF		.0208	.0001
10	ENRELEC		.0142	.001
11	ENRHEAT		.0075	.000
12	AREAHEAT		.0036	.000
13	ENRMAIN		.0021	.000
14		THERMEFF	.0027	.000
15	AREACAPC		.0017	.000
16	CLASSRM		.0011	.000
17	ENROL		.0007	.000
18	AREAUTIL		.0005	.000

FUNCTION	EIGENVALUE	PERCENT OF UNIQUE VARIANCE EXPLAINED	CANONICAL CORRELATION
1	24.994	71.67	.9806
2	5.481	15.72	.9200
3	2.300	6.60	.8348
4	1.500	4.30	.7745
5	.600	1.72	.6124

Figure 6. (continued)

Classification Results:

(Predicted Group Membership)

ACTUAL GROUP	CASES	FREQ=0	FREQ=1	FREQ=2	FREQ=3	FREQ=4	FREQ=5
FREQ=0	2	2 (100.0%)	--	--	--	--	--
FREQ=1	9	--	9 (100.0%)	--	--	--	--
FREQ=2	6	--	--	6 (100.0%)	--	--	--
FREQ=3	4	--	--	--	4 (100.0%)	--	--
FREQ=4	7	--	--	--	--	7 (100.0%)	--
FREQ=5	4	--	--	--	--	--	4 (100.0%)

Percent of "grouped" cases correctly classified: 100.0

was able to re-predict group membership for the solution vector (inclusion v. exclusion) with 100.0 percent accuracy.

‹Figure 6› illustrates the results of the discriminant analysis to evaluate the compositional relation between the selection tally vector and the full criterion database. For convenience, any selection frequency ≥ 5 was entered into the discriminant model as a frequency = 5. This was considered necessary in order to provide some control over problems associated with singular frequency tally entries, and a loss therefore of variance potential. To explain the variance existent within the selection tally vector (0,1,...,5), a total of 16 criterion vectors were entered into the final construction of 5 independent discriminant functions. Re-prediction of the original vector entries proceeded with 100.0 accuracy.

Upon the choice of a single school site for closure (j=17, since tally entry = 7), the database was updated to reflect a N=31 base, and the net effect of the student transfers from the closed site. The model was re-executed, and a new tracking matrix constructed, as displayed in ‹Figure 7›. A total of 4 new sites were now simulated as closed (with tally entries ≥ 4); and the discriminant model re-run.

‹Figure 8› displays the discriminant results of analyzing the N=31 solution set. A total of 10 criteria were required to explained the independent variance -- two more than the N=32 analysis. The canonical correlation existed at .8392, or 70.5 percent explained (independent) variance. Re-classification resulted in a 100.0 percent accuracy level. As before, the selection tally vector for the N=31 case was analyzed by discriminant functions; and these results are illustrated in ‹Figure 9›. A total of 4 functions were constructed; and a re-prediction of 87.1 percent accuracy achieved. Within the re-classification, 7 occurrences of 'over-estimation' resulted (viz., an 'expected' tally entry greater than the original 'observed' value); and 1 occurrence of 'under-estimation' (viz., an 'expected' tally entry lesser than the original 'observed' value). Thus, it would seem that reclassification errored on the non-conservative side.

CRITERION SOURCE(S) FOR SCALING OF THE OPTIMIZATION (COST) VECTOR (1= MAXIMIZATION) (2=MINIMIZATION)

(N=31) SITE	(1) AREA	(1) ENROL	(1) SURVIVE	(1) POTENT	(1) STUD PROX	(2) STUD OL	(2) ZONE OL	(2) SITE OL	(1) INTER 01	(2) INTER 13	(1) THERM EFF	(2) ENER WAST	(1) CLASS RM	(1) MINOR ITY	(2) SITE AGE	(1) AREA CAPC	(1) AREA HEAT	(1) AREA ELEC	(1) AREA MAIN	(1) AREA REFR	(1) AREA UTIL	(1) ENR HEAT	(1) ENR ELEC	(1) ENR MAIN	(DECISION=CLOSE) FREQUENCY
01			X																						1
02							X																		1
03						X																			1
04								X				X			X										3
05																X					X				2
06															X										1
07											X	X													2
08						X			X	X						X									4 *
09																								X	1
10																									--
11	X								X	X			X												4 *
12						X																			1
13																			X						1
14		X									X	X										X	X	X	6 *
15																									--
16																	X				X				2
18								X							X										2
19													X									X	X		3
20					X							X								X					3
21																									--
22																	X	X	X						3
23					X																				1
24															X										1
25																									--
26				X										X											2
27			X																						1
28							X													X					2
29	X							X																	2
30		X		X			X							X											4 *
31																		X							1
32							X	X																	2
Frequency Total: (Site selection per criterion focus)	2	2	2	2	2	3	4	4	2	2	2	4	2	2	4	2	2	2	2	2	2	2	2	2	

Frequency of Site Selection for HIP-4B Modeled Closure (N=31)

Based Upon Varying Optimality Factors

Figure 7.

Figure 8.

Summary of Discriminant Function Analysis

Based Upon MIP-4B Results (N=31)

STEP	CRITERION ENTERED	CRITERION REMOVED	WILKS' LAMBDA	SIGNIFICANCE
1	ENROL		.8883	.0661
2	INTER13		.8164	.0585
3	MINORITY		.7238	.0309
4	STUDPROX		.6504	.0206
5		ENROL	.6590	.0095
6	AREAMAIN		.5944	.0073
7	CLASSRM		.5659	.0102
8	ENROL		.5336	.0126
9	ENRHEAT		.4957	.0132
10	ENERWAST		.4330	.0081
11	THERMEFF		.3387	.0020
12	ENRELEC		.2954	.0015

Eigenvalue = 2.38541 Canonical Correlation = .8394

Classification Results:

Actual Group	Cases	Close	No Close
Close	4	4 (100.0%)	--
No Close	27	--	27 (100.0%)

Percent of "grouped" cases correctly classified: 100.0

Figure 9.

Summary of Discriminant Function Analysis

Based Upon MIP-4B-Sum Results (N=31)

STEP	CRITERION ENTERED	CRITERION REMOVED	WILKS' LAMBDA	SIGNIFICANCE
1	INTER01		.6912	.0412
2	INTER13		.5064	.0213
3	MINORITY		.3596	.0091
4	POTENT		.2703	.0070
5	STUDPROX		.2067	.0064
6	SITEOL		.1517	.0047
7	AREAUTIL		.1193	.0057
8	AREAMAIN		.0822	.0034
9	AREA		.0635	.0044

FUNCTION	EIGENVALUE	PERCENT OF UNIQUE VARIANCE EXPLAINED	CANONICAL CORRELATION
1	1.840	42.30	.8049
2	1.522	34.99	.7769
3	.680	15.63	.6363
4	.308	7.08	.4852

Figure 9. (continued)

Classification Results:

ACTUAL GROUP	CASES	FREQ=0	FREQ=1	FREQ=2	FREQ=3	FREQ=4
FREQ=0	4	4 (100.0%)	--	--	--	--
FREQ=1	11	--	10 (90.9%)	1 (9.1%)	--	--
FREQ=2	8	--	1 (12.5%)	5 (62.5%)	1 (12.5%)	1 (12.5%)
FREQ=3	4	--	--	--	4 (100.0%)	4 (100.0%)
FREQ=4	4	--	--	--	--	4 (100.0%)

Percent of "grouped" cases correctly classified: 87.10

THE "SCANT" MATRIX CASE: THE FISCAL DEALLOCATION MODEL (ROLBAK)

A total of 31 program budgeting (unit) alternatives were evaluated for defunding across a total of 10 competing criterion references. In lieu of a step-wise procedure as represented in the school closure modeling framework, the model is further constrained to choose those programs for refunding such that the new operating district budget is not less than 675,000 dollars, but not more than 700,000 dollars for the particular programs under scrutiny. To study the effect of the model's solution generation process, the feasibility region as defined by the constraint matrix and the RHS-values is constructed in two distinct patterns: a highly restricted region in which very stringent controls are defined for the modeling procedure; and a relatively relaxed region in which less stringent controls are modeled. In addition, the ROLBAK formulation is executed both for cyclic maximization of the objective functions, and for cyclic minimization of the objective functions. Thus, a total of 4 tracking matrice containing 10 potential solution sets (each) result.

This particular modeling application represents the "scant" matrix case, in that a high proportion (48.7 percent) of criterion matrix cells contained a 'zero' entry, signifying no cost for that particular alternative within a specific object-expenditure category. For the SCHCLO model, the criterion matrix was "complete" -- all cells contained a value greater than zero.

Under the 'restricted' formulation, the 17 resulting solution sets signify only 2 distinct solution vectors. In contrast under the 'relaxed' formulation, a total of 17 distinct solution vectors result. Under both restricted and relaxed limitations, 3 objective functions were unable to declare optimality due to the inability to find an initial integer-feasible solution.

‹Figure 10› and ‹Figure 11› display the solution sets resulting from optimization within the restricted region environments. The selection tally vector is noted, as well as the impact upon the total budget based upon the simulated cuts (i.e., where X=funded). As can be easily seen, the solutions resulting from optimization within the restricted environment present only two distinct alternatives for later discriminant analyses.

Effect Upon Budget Deallocation Decisions Based Upon the Variable Flows of a Cyclic Objective Function, and the Interaction of a "Maximized, Restricted" Constraint Iterative Problem.

Objective = Maximization Constraints: Restricted (EXP=16; PERC = 500)

Budget Alterna-tives	01 CERT	02 CLAS	03 BENE	04 SUPL	05 INST	06 CONT	07 TRAV	08 CAPI	09 PERC	10 COMP	SELECTION TALLY	BUDGET AMOUNT
01	X	X	X	X	X	X	X	X	X	X	10	87.5
02	X	X	X	X	X	X	X	X	X	X	10	44.5
03	X	X		X			X	X			5	34.5
04	X	X	X	X	X	X	X	X	X	X	10	71.5
05	X	X	X	X	X	X	X	X	X	X	10	70.5
06											--	32.5
07	X	X	X	X	X	X	X	X	X	X	10	51.5
08											--	1.5
09	X	X		X			X	X			5	43.0
10											--	4.0
11	X	X	X	X	X	X	X	X	X	X	10	54.0
12											--	1.0
13											--	5.5
14											--	4.0
15	X	X	X	X	X	X	X	X	X	X	10	116.0
16			X		X	X			X	X	5	23.0
17	X	X	X	X	X	X	X	X	X	X	10	107.0
18											--	13.0
19											--	2.0
20											--	1.0
21											--	16.0
22											--	10.5
23			X		X	X			X	X	5	55.0
24											--	4.5
25											--	2.5
26											--	19.0
27											--	1.0
28											--	1.0
29											--	2.0
30											--	12.0
31											--	2.5
	10	10	10	10	10	10	10	10	10	10		
O.F. Value:	340.7	274.5	217.9	433.9	330.0	362.1	50.0	534.6	496.2	680.5		
Iteration at Optimality:	36	69	76	115	228	27	114	51	5000+	369		
Time (secs):	.266	.298	.288	.325	.384	.264	.383	.274	4.498	.850		
Roll-Back Savings:	680.0	680.0	680.5	680.0	680.5	680.5	680.0	680.0	680.5	680.5		
(- Cut)	(-213.5)	(-213.5)	(-213.0)	(-213.5)	(-213.0)	(-213.0)	(-213.5)	(-213.5)	(-213.0)	(-213.0)		

Note: Total Initial Budget = 893.5 ($1000's)

Figure 10.

Effect Upon Budget Deallocation Decisions Based Upon the Variable Forms of a Cyclic Objective Function, and the Interaction of a "Minimized, Restricted" Constraint Interative Problem.

Objective = Minimization Constraints = Restricted (8xP=16; PERC=500)

Note: Total Initial Budget = 893.5 ($1000's)

Figure 10.

Effect Upon Budget Deallocation Decisions Based Upon the Variable Forms of a Cyclic Objective Function, and the Interaction of a "Minimized, Restricted" Constraint Interative Problem.

Objective = Minimization Constraints = Restricted (8xP=16; PERC=500)

Budget Alterna-tives	01 CERT	02 CLAS	03 BENE	04 SUPL	05 INST	06 CONT	07 TRAV	08 CAPI	09 PERC	10 COMP	SELECTION TALLY	BUDGET AMOUNT
01		X	X	X	X	X		X	X		7	87.5
02		X	X	X	X	X		X	X		7	44.5
03			X		X	X			X		4	34.5
04		X	X	X	X	X		X	X		7	71.5
05		X	X	X	X	X		X	X		7	70.5
06											--	32.5
07		X	X	X	X	X		X	X		7	51.5
08											--	1.5
09			X		X	X		X	X		4	43.0
10											--	4.0
11		X	X	X	X	X		X	X		7	54.0
12											--	1.0
13											--	5.5
14											--	4.0
15		X	X	X	X	X		X	X		7	116.0
16		X		X				X			3	23.0
17		X	X	X	X	X		X	X		7	107.0
18											--	13.0
19											--	2.0
20											--	1.0
21											--	16.0
22											--	10.5
23		X		X				X			3	55.0
24											--	4.5
25											--	2.5
26											--	19.0
27											--	1.0
28											--	1.0
29											--	2.0
30											--	12.0
31											--	2.5
	10	10	10	10	10	10	10	10	10	10		
O.F. Value:	--	234.5	197.0	366.8	314.8	313.0	--	482.6	489.0	--		
Iteration at Optimality:	--	5000+	686	200	85	902	--	203	53	--		
Time (sec):	--	4.581	.933	.563	.304	1.193	--	.407	.256	--		
Roll-Back Savings:	--	680.5	680.0	680.5	680.0	680.0	--	680.5	680.0	--		
(-Cut)	--	(-213.0)	(-213.5)	(-213.0)	(-213.5)	(-213.5)	--	(-213.0)	(-213.5)	--		

Note: Total Initial Budget = 893.5 ($1000's)

Figure 11.

‹Figure 12› and ‹Figure 13› display those solution sets resulting from the optimizations within a relaxed environment. A total of 17 distinct solution set vectors are formed; and thus the selection tally matrix demonstrates greater variability than existent within the restricted orientation.

Discriminant functions were computed for the relaxed modeling setting first, requiring a separate discriminant execution for each of the distinct solution vectors resulting from the MAM analysis. As noted in an earlier section to this paper, criterion strength was evaluated utilizing the three composites vectors:

DISCRIMINANT CRITERION INCLUSION VECTOR

DISCRIMINANT CRITERION ENTRY VECTOR

DISCRIMINANT WEIGHTING SUMMARY VECTOR.

The first vector is composed of binary (1,0) entries signifying whether a specific criterion was entered into the discriminant analysis for explaining the variance within the solution set. The second vector contains entries of 1,2,3,... , such that the order-of-entry for the discriminant criteria is represented. Finally, the third vector contains a factor-weight entry for each of the 'entered' vectors, to measure the relative importance of each of the discriminating criterion references.

The notion of decisioning reliability was evaluated utilizing two techniques:

CANONICAL CORRELATION

RE-CLASSIFICATION ANALYSIS.

‹Figure 14› contains the discriminant results for solutions accountable to maximization within a relaxed region. The first ten columns contain the information from the discriminant analyses for each of the ten simulated solution sets. The ordinal numerals represent order-of-entry, while the bracketed entries

Effect Upon Budget Deallocation Decisions Based Upon the Variable Flows of a Cyclic Objective Function, and the Interaction of a "Maximized, Relaxed" Constraint Iterative Problem.

Objective = Maximization Constraints: Relaxed (EXP=10; PERC = 600)

Budget Alternatives	01 CERT	02 CLAS	03 BENE	04 SUPL	05 INST	06 CONT	07 TRAV	08 CAPI	09 PERC	10 COMP	SELECTION TALLY	BUDGET AMOUNT
01	X	X	X	X	X	X	X	X	X	X	10	87.5
02	X	X	X	X	X	X		X			7	44.5
03	X	X	X				X	X			5	34.5
04	X	X	X	X	X	X	X	X	X	X	10	71.5
05	X		X	X	X	X	X		X	X	8	70.5
06	X			X	X		X	X	X		6	32.5
07	X	X		X	X	X	X	X	X	X	9	51.5
08											--	1.5
09	X	X	X	X	X	X	X	X	X	X	10	43.0
10						X					1	4.0
11		X	X	X		X		X		X	6	54.0
12											--	1.0
13					X		X				2	5.5
14									X		1	4.0
15	X	X	X	X	X	X	X	X	X	X	10	116.0
16						X		X	X	X	4	23.0
17	X	X	X	X	X	X	X	X	X	X	10	107.0
18				X					X		2	13.0
19		X									1	2.0
20											--	1.0
21	X		X								2	16.0
22								X			1	10.5
23		X	X		X		X		X	X	6	55.0
24											--	4.5
25											--	2.5
26	X	X					X			X	4	19.0
27				X							1	1.0
28				X					X		2	1.0
29										X	1	2.0
30						X					1	12.0
31											--	2.5
	12	12	11	13	11	12	12	12	13	12		
O.F. Value	485.4	425.5	316.1	615.9	476.6	477.7	100.0	659.04	600.0	700.0		
Starting at Optimality:	20	60	202	16	43	52	163	65	5000+	457		
Time (sec):	.246	.359	.416	.227	.297	.337	.589	.310	6.022	1.166		
Roll-Back Savings:	685.5	685.5	699.5	693.0	684.5	684.5	693.5	675.5	675.5	700.0		
	(-208.0)	(-208.0)	(-194.0)	(-200.5)	(-209.0)	(-209.0)	(-200.0)	(-218.0)	(-218.0)	(-193.5)		

Note: Total Initial Budget = 893.5 ($1000's)

Figure 12.

Effect Upon Budget Deallocation Decisions Based Upon the Variable Flows of a Cyclic Objective. Function, and the Interaction of a "Minimized, Relaxed" Constraint Iterative Problem.

Objective = Minimization   Constraints: Relaxed (EXP=10; PERC = 600)

Budget Alterna-tives	01. CERT	02 CLAS	03 BENE	04 SUPL	05 INST	06 CONT	07 TRAV	08 CAPI	09 PERC	10 COMP	SELECTION TALLY	BUDGET AMOUNT
01	X		X	X	X	X	X		X		7	87.5
02	X			X	X	X			X		5	44.5
03				X	X	X					3	34.5
04	X		X	X	X	X	X		X		7	71.5
05			X	X	X		X		X		5	70.5
06			X			X	X		X		4	32.5
07	X		X	X		X	X		X		6	51.5
08											--	1.5
09	X		X		X	X	X		X		5	43.0
10											--	4.0
11	X		X		X		X		X		5	54.0
12											--	1.0
13			X								1	5.5
14											--	4.0
15	X		X	X	X	X	X		X		7	116.0
16	X		X	X	X		X				5	23.0
17	X		X	X	X	X	X		X		7	107.0
18	X		X		X	X					4	13.0
19											--	2.0
20											--	1.0
21							X				1	16.0
22	X										1	10.5
23	X			X		X					3	55.0
24											--	4.5
25											--	2.5
26				X	X	X	X				4	19.0
27											--	1.0
28			X								1	1.0
29											--	2.0
30											--	12.0
31				X	X						2	2.5

	01 CERT	02 CLAS	03 BENE	04 SUPL	05 INST	06 CONT	07 TRAV	08 CAPI	09 PERC	10 COMP
	12	--	13	12	13	12	12	--	10	--
Optimal Value	238.8	--	110.0	304.7	265.9	198.3	50.0	--	482.3	--
Iteration at Optimality:	869	--	5000+	197	625	1030	5000+	--	34	--
Time (sec):	1.665	--	4.143	.548	.731	1.823	4.010	--	.249	--
Roll-Back Savings:	676.5	--	676.0	682.5	686.0	675.0	691.5	--	678.0	--
(- Cut)	(-217.0)	--	(217.5)	(-211.0)	(-207.5)	(-218.5)	(-202.0)	--	(-215.5)	--

Note: Total Initial Budget = 893.5 ($1000's)

Figure 13.

x.xx] contain the factor-weights computed from dividing each of the standardized anonical discriminant coefficients by the smallest such coefficient for each iscriminant analysis. For example in the first column signifying the results f discriminating the solution computed from maximizing 'certificated salaries', criteria were required to explain available variance within the solution set. he criterion 'budgetary composites' was entered first, and represents a factor f 2.51 in its importance to the remaining 4 criterion discriminants. The riterion 'certificated salaries' was entered secondly, and represents a factor f 3.17 in its relative importance for discriminating the solution set being nalyzed; and so forth. The selection tally vector is similarly analyzed via iscriminant functions.

For understanding the dimension of decisioning reliability, computed canonical correlation coefficients existed as follows, for maximized-relaxed solutions:

Objective Function	Canonical Coefficient	Percent Variance Explained	Relative Rank
CERT	.9056	82.0	3
CLAS	.8633	74.5	6
BENE	.8729	76.2	4
SUPL	.9077	82.4	2
INST	.9339	87.2	1
CONT	.8679	75.3	5
TRAV	.8614	74.2	7
CAPI	.8419	70.9	8
PERC	.7870	61.9	9
COMP	.7281	53.0	10.

Thus it would seem, that a formalized objective of "maximizing" the expenditures associated with instructional materials in determining which programs to refund during a period of scant resourses, produced the highest correlation between the criterion matrix of 10 vectors and the proposed solution set vector constructed from the MAM analysis execution. Likewise, the maximization of

Summary of Criterion Vector Order-of-Entry, in Discriminating the Solution Set Vector for Each Cyclic MAXIMIZATION within a RELAXED Region. (Note: Source of Discriminant Criterion Inclusion Vector; Discriminant Criterion Entry Vector; and Discriminant Weighting Summary Vector)

‹ VALUE OF OBJECTIVE FUNCTION DURING CYCLIC-OPTIMIZATION EVALUATIONS ›

Criterion Vector	CERT	CLAS	BENE	SUPL	INST	CONT	TRAV	CAPI	PERC	COMP	Selection Tally Vector	Discriminant Function #
Certificated Salaries	2 [3.17]	4 [1.00]	--	--	--	--	--	--	--	--	--	--
Classified Salaries	--	2 [3.08]	--	4 [1.57]	--	--	--	--	3 [2.04]	2 [1.00]	5	5
Employee Benefits	--	3 [1.48]	2 [2.74]	--	--	5 [1.00]	--	--	--	--	--	--
Supplies & Materials	5 [1.00]	--	--	1 [3.56]	--	3 [1.21]	--	2 [1.46]	--	--	--	--
Instructional Materials	--	--	5 [1.00]	--	1 [3.13]	--	2 [1.78]	3 [1.00]	4 [2.03]	--	2	2
Contractual Services	4 [2.16]	--	3 [1.15]	5 [1.00]	--	2 [1.87]	3 [2.19]	--	--	--	--	--
Travel Expenditures	--	--	--	--	4 [1.00]	--	4 [1.17]	--	5 [1.00]	--	--	--
Capital Outlay	3 [2.40]	--	4 [1.29]	--	--	4 [1.02]	5 [1.00]	1 [3.65]	--	--	3	3
Administrative Perception	--	--	--	3 [1.68]	3 [1.25]	--	--	--	2 [2.59]	--	4	4
Budgetary Composites	1 [2.51]	1 [2.29]	1 [2.91]	2 [3.46]	2 [3.03]	1 [1.71]	1 [2.63]	--	1 [3.04]	1 [3.67]	1	1
Number of Mis-inclusions	--	2	2	1	--	--	1	2	2	2	4	Number of Over-Estimates
Number of Mis-Exclusions	--	--	--	--	--	--	1	--	2	3	5	Number of Under-Estimates
Re-Prediction Accuracy (%)	100.0	93.6	93.6	96.8	100.0	100.0	93.6	93.6	87.1	83.9	71.0	Re-Prediction Accuracy (%)

* (No integer-feasible solution possible; Optimality not achieved)

'budgetary composites' produced the lowest correlation, explaining only 53.0 percent of independent variance within the MAM solution vector.

The second 'phase' of measuring decisioning reliability exists in the accuracy of re-predicting solution set membership based upon the classification function coefficients generated via the discriminant analysis. The bottom portion of igure 14 portrays these results for each of the 10 solution vectors formed by the varying criterion focus of the objective function. The results of e-classification for the selection tally vector are also displayed.

Figure 15> illustrates the similar results from applying discriminant function nalyses to the solution vectors formed by minimization within a relaxed setting he three vectors for denoting criterion strength are easily distinguishable rom the 7 successful (columns) optimizations. The re-classification portion of easuring decisioning reliability is also shown.

he computed canonical correlation coefficients for minimized-relaxed solutions:

Objective Function	Canonical Coefficient	Percent Variance Explained	Relative Rank
CERT	.7721	59.6	6
CLAS	--	--	--
BENE	.7902	62.4	5
SUPL	.8194	67.1	2
INST	.7675	58.9	7
CONT	.8000	64.0	3
TRAV	.7928	62.9	4
CAPI	--	--	--
PERC	.9343	87.3	1
COMP	--	--	--

nstrated, that solution set formulated by minimizing the 'administrative eption' entries in determining a solution, to be the best fit with the

Summary of Criterion Vector Order-of-Entry, in Discriminating the Solution Set Vector for Each Cyclic MINIMIZATION within a RELAXED Region. (Note: Source of Discriminant Criterion Inclusion Vector; Discriminant Criterion Entry Vector; and Discriminant Weighting Summary Vector)

❬ VALUE OF OBJECTIVE FUNCTION DURING CYCLIC-OPTIMIZATION EVALUATIONS ❭

Criterion Vector	CERT	CLAS	BENE	SUPL	INST	CONT	TRAV	CAPI	PERC	COMP	Selection Tally Vector	Discriminant Function #
Certificated Salaries	2 [2.37]	*	—	3 [1.41]	—	3 [1.22]	3 [1.01]	*	—	*	—	—
Classified Salaries	—	*	2 [2.16]	—	5 [1.00]	4 [1.52]	—	*	—	*	—	—
Employee Benefits	—	*	—	—	—	5 [1.00]	4 [1.00]	*	—	*	2	6
Supplies & Materials	5 [1.00]	*	5 [1.00]	2 [1.97]	—	—	—	*	2 [1.37]	*	6	4
Instructional Materials	4 [1.38]	*	—	—	—	—	—	*	4 [1.00]	*	—	—
Contractual Services	—	*	—	5 [1.00]	4 [1.27]	2 [1.33]	—	*	—	*	3	5
Travel Expenditures	—	*	4 [1.45]	—	—	—	—	*	—	*	4	3
Capital Outlay	3 [1.99]	*	—	—	3 [1.90]	—	2 [1.63]	*	3 [1.17]	*	5	2
Administrative Perception	—	*	3 [1.34]	4 [1.58]	2 [2.74]	—	—	*	—	*	—	—
Budgetary Composites	1 [1.68]	*	1 [2.95]	1 [2.30]	1 [2.08]	1 [2.28]	1 [2.24]	*	1 [1.07]	*	1	1
Number of Mis-inclusions	2	*	1	1	1	1	2	*	—	*	3	Number of Over-Estimates
Number of Mis-Exclusions	1	*	3	1	2	1	1	*	—	*	2	Number of Under-Estimates
Re-Prediction Accuracy (%)	90.3	*	87.1	93.6	90.3	93.6	90.3	*	100.0	*	83.9	Re-Prediction Accuracy (%)

* (No integer-feasible solution possible; Optimality not achieved)

Figure 15.

overall criterion matrix; and the solution from minimizing 'instructional materials', the least 'best' fit.

Regarding the results of optimizing (both maximally and minimally) within the restricted environment, ‹Figure 16› illustrates the discriminant function analysis framework. Similarly, the canonical coefficients were computed as:

Solution Vector	Canonical Coefficient	Percent Variance Explained	Relative Rank
#1	.8947	80.0	1
#2	.8628	74.4	2

Figure 15.

Summary of Criterion Vector Order-of-Entry, in Discriminating the Two Distinct Solution Set Vectors Resulting from the Cyclic MAXIMIZATION and MINIMIZATION within a RESTRICTED Region. (Note: Source of Discriminant Criterion Inclusion Vector; Discriminant Criterion Entry Vector; and Discriminant Weighting Summary Vector)

Criterion Vector	Solution Set #1	Solution Set #2
Certificated Salaries	4 [1.05]	—
Classified Salaries	5 [1.00]	—
Employee Benefits	—	—
Supplies & Materials	2 [1.70]	—
Instructional Materials	—	3 [1.00]
Contractual Services	—	2 [1.65]
Travel Expenditures	—	—
Capital Outlay	3 [1.25]	—
Administrative Perception	—	—
Budgetary Composites	1 [2.11]	1 [3.24]
Number of Mis-inclusions	1	—
Number of Mis-exclusions	—	1
Re-Prediction Accuracy (%)	96.8	96.8

Figure 16.

SUMMARY OF FINDINGS

The use of discriminant functions in providing a useful post-hoc evaluation strategy for multiple alternatives decision-making has been studied within two separate real-world settings: the closure of schools; and the deallocation of program unit budgetary items. Two generalized issues of content and process were the main foci: content, in as much as there is a need to relate criteria used to the decisions made; and process, in order to verify the reliability of the decisioning procedures based upon the criteria utilized.

The author maintains, that two related "abilities" are necessary for prudent and trustworthy decision-making. The first ability refers to that knowledge which clarifies (1) which criteria 'effected' the decisions, and to what extent; and (2) to what degree did this 'effect' vary across the results of the cyclical optimizations. The second ability relates the need to study (1) the relationship between the 'optimizing vector' (objective function) and the results of a discriminant analysis; and (2) the relationship between the extent of feasibility region constraint (relaxed v. restricted) and the results of a discriminant analysis. To accomplish these ends, the multiple linear regression technique, discriminant functions analysis, is utilized to measure the topics of criterion strength and decisioning reliability.

The results of these discriminant analyses illustrate the superior efficacy found in relating multiple correlational strategies to discovering relationships between solution vectors and the criterion vectors (matrice) supporting those decisions. Three measures of criterion strength and two measures of decisioning reliability are illustrated for the reader -- all measures normally products of discriminant function(s) formulation.

It is a fundamental by-product of this study though all to important not to note, that the formation of "classification coefficients" within the discriminant process provides an excellent way of projecting expected impact from a newly collected set of data variables. By utilizing the linear combinations of this new data, 'expected correlative' decisions can be computed which maintain the same variance relationship as the decisions utilized originally in the initial discriminant analyses.

In summary, the use of discriminant functions in addressing the issues of criterion strength and decisioning reliability has been illustrated to hold great promise for the decision-maker, evaluator and otherwise problem-solver. Increased accountability, visibility and responsibility are the maximized ends.

Anderson, T.W.
with linea

Bentler, P.M.
the analys
(3), 588-6

Cooley, W.W.
1978, 7, 9

Cooley, W.W. a
Irvington

Jöreskog, K.G.
Psychometr

Jöreskog, K.G.
Psychometr

Jöreskog, K.G
analysis.
1968, 21,

Jöreskog, K.G
equation
national

Jöreskog, K.G
estimatin
national

Lawley, D.N.
likelihoo
64-82.

Lawley, D.N.
York: Am

Maruyama, G.
maximum-l

itial

criterion
romise
d

References

Anderson, T.W. Asymptotically efficient estimation of covariance matrices with linear structure. Annals of Statistics, 1973, 1, 135-141.

Bentler, P.M. and Bonnett, D.G. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin ,1980, 88 (3), 588-606.

Cooley, W.W. Explanatory observational studies. Educational Researcher, 1978, 7, 9-15.

Cooley, W.W. and Lohnes, P. Evaluation research in education. New York: Irvington Publishers, Inc., 1976.

Jöreskog, K.G. A general approach to confirmatory maximum-likelihood analysis. Psychometrika, 1969, 34, 183-202.

Jöreskog, K.G. Structural analysis of covariance and correlation matrices. Psychometrika, 1978, 43, 443-447.

Jöreskog, K.G. and Lawley, D.N. New methods in maximum likelihood factor analysis. The British Journal of Mathematical and Statistical Psychology, 1968, 21, 85-96.

Jöreskog, K.G. and Sörbom, D. LISREL III-Estimation of linear structural equation systems by maximum likelihood methods. Chicago, Il.: International Educational Services, 1976.

Jöreskog, K.G. and Sörbom, D. LISREL IV - A general computer program for estimating a linear structural equation system. Chicago, Il.: International Educational Resources, 1978.

Lawley, D.N. The estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society of Edinburgh, 1940, 60, 64-82.

Lawley, D.N. and Maxwell, A.E. Factor analysis as a statistical method. New York: American Elsevier, 1971.

Maruyama, G. and McGarvey, B. Evaluating causal models: An application of maximum-likelihood. Psychological Bulletin, 1980, 87, 502-512.

Appendix I

Synthetic "True" Covariance Structure for Misspecification Categories I and III: The Covariance Structure used as the Input Matrix for the Simulation of Data Sets whith a Multivariate Normal Distribution.

	Y_1	Y_2	Y_3	X_1	X_2	X_3	X_4
Y_1	1.338						
Y_2	.781	1.1175					
Y_3	1.9525	2.54375	6.453975				
X_1	.84	.62	1.55	1.1			
X_2	.42	.31	.775	.5	.35		
X_3	.78	.55	1.375	.8	.4	1.1	
X_4	.234	.165	.4125	.24	.12	.3	.19

Appendix II

Synthetic "True" Covariance Structure for Misspecification Category II: The Covariance Structure used as the Input Matrix for the Simulation of Data Sets with a Multivariate Normal Distribution.

	Y_1	Y_2	Y_3	X_1	X_2	X_3	X_4
Y_1	1.538						
Y_2	1.18	1.5175					
Y_3	2.9675	3.0439	7.9094				
X_1	.84	.62	1.55	1.1			
X_2	.42	.31	.775	.53	.35		
X_3	.78	.55	1.375	.81	.41	1.1	
X_4	.234	.165	.4125	.13	.13	.32	.19

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 11, NUMBER 3 SUMMER 1982

USING LINEAR MODELS TO SIMULTANEOUSLY ANALYZE A SOLOMON FOUR GROUP DESIGN

John D. Williams
University Of North Dakota

Isadore Newman
University Of Akron

Solomon (1949) devised a design to control threats to design validity (Campbell and Stanley, 1966). Using the notation of Campbell and Stanley, the four groups can be diagrammed as

R	0_1	X	0_2
R	0_3		0_4
R		X	0_5
R			0_6

The four groups are respectively Group One: An experimental groups that has been pretested and posttested; Group Two: A control group that has been pretested and posttested; Group Three: An experimental group that has been posttested only; and Group Four: A control group that has been posttested only.

Campbell and Stanley state, "There is no singular statistical procedure which makes use of all six sets of observations simultaneously." (p. 24)

The Solomon Four Group Design, while very simple conceptually, can be very misleading depending upon the statistical analysis. Campbell and Stanley (1966) have a preferred approach, in which they set up a 2x2 factorial design.

---

resented at AERA 1982, MLR Special Interest Group
ot refereed by editorial staff

		Treatment: Yes	Treatment: No
Pre-Test	Yes	Gr1	Gr2
	No	Gr3	Gr4

In this design only the posttest scores are statistically analyzed. This procedure does not allow one to control for the pretest scores in groups 1 and 2, thereby losing some power; it does however estimate the effects of treatments that are independent of individuals having a pretest and treatment pretest interaction. It also tests for the effects of pretesting, independent of treatment and pretest-treatment interaction, on posttest scores. Finally, the approach estimates the effects of pretest-treatment interaction, on posttest scores.

One of the advantages of writing specific regression models which reflect research questions is that one is less likely to have a statistical answer that is unrelated to the researcher's question of interest. The following are a variety of regression models which will reflect potential research questions that can be ascertained from the Solomon Four Group Design. It should be remembered that there is not one correct answer.

Recently, Newman, Benz, and Williams (1980) devised a way to analyze data that, by extension, might be applied to Solomon type designs. A unique property of this technique is that, the statement by Campbell and Stanley not withstanding, a single statistical procedure can be employed which makes use of all six sets of observations simultaneously. On the other hand, the solution(s) may prove to be no more satisfactory than existing possibilities that split the data into two sets. In the end, the Solomon Four Group design may prove to be one of those recalcitrant research situations that leave the would be analysts foundered on the shoal of a simple design whose simplicity is only a deception.

Consider the following research situation. Five people in each group have scores such that one experimental group has been pretested and posttested and one experimental group has been posttested only. Two similarly tested control groups are also included. Data for such a situation are given in Table 1.

Table 1

Data for a Solomon Four Group Design

Experimental: Group One		Control: Group Two		Experimental: Group Three	Control: Group Four
Pretest	Posttest	Pretest	Posttest	Posttest	Posttest
5	15	5	8	13	9
7	12	4	7	10	8
5	10	4	8	12	6
12	17	6	6	11	3
6	11	6	6	14	4

Several different approaches might be tried. One approach would be to divide the data into two sets: Groups One and Two (those who were both pretested and posttested) as one set, and the posttested only groups (Groups Three and Four) as the second set. The latter set can be simply tested by the use of the t test:

$t = 4.24$ ($p<.05$).

The former data set (Groups One and Two) can be conceived either as a repeated measures design or as a problem that can be approached through the analysis of covariance (or related techniques such as residual gain analysis).

To approach the problem first as an analysis of covariance, the following variables can be defined:

Y = the criterion, or posttest score;

X_1 = the pretest score;

X_2 = 1 if the score is from the experimental group, 0 if the score is from the control group;

X_3 = 1 if the score is from the control group, 0 if the score is from the control group.

Then either of two full models can be used:

$$Y = b_1X_1+b_2X_2+b_3X_3+e_1, \qquad (1)$$

or

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

Equation 2 utilizes the unit vector in the process of generating a constant whereas equation 1 does not. Either model will yield the same R^2 value.

The restricted model (with equation 2 as the full model) is of the form:

$$Y = b_0+b_1X_1+e_2. \qquad (3)$$

For this data set R_2^2 = .79379, R_3^2 = .42334, $F = \frac{(.79379 - .42334)/1}{(1 - .79379)/7} = 12.58$, $p<.05$.

<u>Using a Repeated Measures Approach</u>

If the problem is visualized as a repeated measures design wherein the pretest is the first measure and the posttest is the second measure, then the design is like the Type I design shown in Lindquist (1953) and can be achieved through a regression approach (Williams, 1974). For a regression formulation, see Table 2.

Table 2
Design Matrix for a Repeated Measures Problem

Y	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	X_1	X_2	X_3	X_4	X_5
5	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1
7	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1
5	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1
12	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1
6	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1
15	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0
12	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0
10	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0
7	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0
11	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0
5	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0
6	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0
6	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0
8	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0
7	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0
8	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0
6	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0
6	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0

Here, Y = the criterion test score;

P_1 thru P_{10} are binary coded person vectors (1 if the person, 0 otherwise);

X_1 = 1 if the score comes from a person in the experimental group, 0 otherwise;

X_2 = 1 if score comes from a person in the control group, 0 otherwise;

X_3 = 1 if the score occurs with a pretest situation, 0 otherwise;

X_4 = 1 if the score occurs with a posttest situation, 0 otherwise; and

$X_5 = X_1 \cdot X_3$.

Several models can be used to generate an analysis. The use of the following is instructive:

$$Y = b_0+b_1P_1+b_2P_2+\ldots+b_9P_9+e_3; \quad (4)$$

(or alternatively, $Y = b_1P_1+b_2P_2+\ldots+b_{10}P_{10}+e_3$)

$$Y = b_0+b_1X_1+e_4; \quad (5)$$

$$Y = b_0+b_3X_3+e_5; \quad (6)$$

$$Y = b_0+b_1X_1+b_3X_3+e_6; \quad (7)$$

$$Y = b_0+b_1X_1+b_3X_3+b_5X_5+e_7; \qquad (8)$$

and

$$Y = b_0+b_1P_1+b_2P_2+\ldots+b_9P_9+b_{10}X_3+b_{11}X_5+e_8. \qquad (9)$$

For the preceding, $R_4^2 = .54297$;

$R_5^2 = .31250$;

$R_6^2 = .31250$;

$R_7^2 = .62500$;

$R_8^2 = .70312$ and $R_9^2 = .93359$.

What might have occured if a model of the following form were used?

$$Y = b_0+b_1P_1+b_2P_2+\ldots+b_9P_9+b_{10}X_1+b_{11}X_3+b_{12}X_5+e_9.$$

It would not sensibly yield $R^2 = .54297 + .70312 = 1.24609$. Such a model would fail because the effect for experimental-control is "nested" in the subject (or person) effect. To test for the experimental-control effect,

$$F = \frac{R_5^2/1}{(R_4^2-R_5^2)/(P-1-1)} = \frac{.31250/1}{(.54297-.31250)/(10-2)} = 10.85,$$

$p<.05$.

To test for the test-retest effect,

$$F = \frac{R_6^2/1}{(1-R_9^2)/(N-P-1-1)}$$

$$= \frac{.31250}{.06641/8} = 37.65,\ p<.01.$$

The interaction is tested by

$$F = \frac{(R_8^2-R_7^2)/1}{(1-R_9^2)/(N-P-1-1)} = \frac{(.70312-.62500)/1}{.06641/8} = 9.41,\ p<.05.$$

Note that the interaction effect can be conceptualized as actually being additional evidence for the experimental effect. The higher increases in the experimental group will show up in part as interaction for a repeated measures design.

The usual summary table for the repeated measures design can be constructed. The summary table is shown in Table 3.

Table 3

Summary Table for Repeated Measures Design

	df	SS	MS	F
Subjects	9	139.00		
Experimental-Control	1	80.00	80.00	10.85
Error (a)	8	59.00	7.375	
Within Subjects	10	117.00		
test-retest	1	80.00	80.00	37.65
interaction	1	20.00	20.00	9.41
error (b)	8	17.00	2.125	
Total	19	256.00		

Using All Six Groups Simultaneously

As the Solomon design is approached, several conceptual issues ensue. Is this to be seen as a six group design with attendant solutions? If the researcher opts for a six group design, person vector information needs to be excluded. Indeed, this was also true in the previous section. At no time were the four groups and person vectors used simultaneously; if it were, the R^2 was theoretically to be 1.24609, obviously an impossibility. If a six group design is to be used, what dimensions would be appropriate? This could be considered to be a one-way lay-out, a two-way lay-out, or a three-way lay-out (but with two missing cells) only the one-way and three-way layouts are discussed here. First hypotheses with a one-way lay-out as addressed.

Consider the following variables:

Y = the criterion score;

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

X_2 = 1 if the score is a posttest score from a member of the experimental group that has been pretested, 0 otherwise;

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

X_4 = 1 if the score is a posttest score from a member of the control group that was pretested, 0 otherwise;

X_5 = 1 if the score is from a member of the experimental group that was <u>not</u> pretested, 0 otherwise; and

X_6 = 1 if the score is from a member of the control group that was <u>not</u> pretested, 0 otherwise.

For the six group situation, the full model is:

$$Y = b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_9. \quad (10)$$

At least two different sets of restrictions might make sense in addressing the Solomon design. One such set would be $b_2-b_1=b_4-b_3$, which addresses the hypothesis $\overline{Y}_2-\overline{Y}_1=\overline{Y}_4-\overline{Y}_3$, as the hypothesis that the gains in the twice tested experimental and control groups are equal; also, the second restriction is $b_5=b_6$ as the once tested experimental groups have equal means: $\overline{Y}_5=\overline{Y}_6$.

The first restriction can be rewritten as $b_2=b_4-b_3+b_1$: Placing these two restrictions on the Full Model:

$$Y = b_1X_1+(b_4-b_3+b_1)X_2+b_3X_3+b_4X_4+b_5X_5+b_5X_6+e_{10} \quad (11)$$

$$Y = b_1(X_1+X_2)+b_3(X_3-X_2)+b_4(X_4+X_2)+b_5(X_5+X_6)+e_{10}. \quad (12)$$

Letting $D_1 = X_1+X_2$;

$D_2 = X_3-X_2$;

$D_3 = X_4+X_2$; and

$D_4 = X_5+X_6$, the restricted model is:

$$Y = b_1D_1+b_3D_2+b_4D_3+b_5D_4+e_{10}. \quad (13)$$

Here, $R^2_{10} = .71183$; $R^2_{13} = .42882$.

$$F = \frac{(R^2_{10}-R^2_{13})/2}{(1-R^2_{10})/(N-6)} = \frac{.28301/2}{(1-.71183)/24} = 11.79,\ p<.01.$$

This F test tests simultaneously $\overline{Y}_2-\overline{Y}_1=\overline{Y}_4-\overline{Y}_3$ and $\overline{Y}_5=\overline{Y}_6$; placing both sets of restrictions allows the rejection of the null hypotheses. If these restrictions are equivalent to hypotheses the researcher had in mind, then there is no further problem. Translating the meaning of these two hypotheses into English may leave the researcher somewhat uneasy; however, one attempt at a translation into English is: It is not simultaneously true that there is no differences in the means of the non-pretested group and that there is no differences in the gains of the pre-tested groups.

One approach would be to test each of these hypotheses separately and using Dunn's (1961) test for multiple comparisons. Imposing the first restriction separately ($b_2-b_1=b_4-b_3$) yields $Y = b_1X_1+(b_4-b_3+b_1)X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_{11}$;

$Y = b_1(X_1+X_2)+b_3(X_3-X_2)+b_4(X_4+X_2)+b_5X_5+b_6X_6+e_{11}$. (14)

Then using D_1, D_2 and D_3 as previously defined, $Y = b_1D_1+b_3D_2+b_4D_3+b_5X_5+b_6X_6+e_{11}$. (15)

R^2_{15} = .66038 and

$$F = \frac{(R^2_{10}-R^2_{15})/1}{(1-R^2_{10})/(N-6)} = \frac{.71183-.66038}{(1-.71183)/24} = \frac{.05145}{(1-.71183)/24} = 4.29.$$

$t = \sqrt{F} = 2.07$. Since two contrasts are planned, a value of 2.39 is necessary for significance of the .05 level, hence the hypothesis $\mu_2-\mu_1=\mu_4-\mu_3$, corresponding to $\overline{Y}_2-\overline{Y}_1=\overline{Y}_4-\overline{Y}_3$ cannot be rejected. The imposition of the second restriction ($b_5=b_6$) yields:

$Y = b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_{12}$;

$Y = b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5(X_5+X_6)+e_{12}$. (16)

Using D_4,

$Y = b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5D_4+e_{12}$. (17)

$$R^2_{17} = .48028 \text{ and } F = \frac{(R^2_{10}-R^2_{17})/1}{(1-R^2_{10})/(N-6)} = \frac{.71183-.48028}{(1-.71183)/24} = \frac{.23155}{(1-.71883)/24} = 19.28,$$

$t = \sqrt{F} = 4.39$, $t > 3.09$ from Dunn's table, so that $p < .01$. Note also that from the numerator of these two tests that .05145 + .23155 = .28300, within rounding error of the numerator when both restrictions were applied; this is because these contrasts are independent. From these calculations, it can be seen that the greatest portion of the rejection of the hypotheses tested by the restrictions in equation 13 is due to the differences in the groups that were posttested only rather than due to differential increases.

A second set of restrictions (actually, a single restriction) is given as $(b_2-b_1)-(b_4-b_3)=b_5-b_6$. This restriction tests the hypothesis related to $(\bar{Y}_2-\bar{Y}_1)-(\bar{Y}_4-\bar{Y}_3)=\bar{Y}_6-\bar{Y}_5$; that is, the difference between the mean of the gain scores is equal to the difference in posttest measures of the non-pretested group. The restriction can be stated as $b_2=b_5-b_6+b_1+b_4-b_3$. Imposing this restriction yields:

$$Y = b_1X_1+(b_5-b_6+b_1+b_4-b_3)X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_{13}; \qquad (18)$$

$$Y = b_1(X_1+X_2)+b_3(X_3-X_2)+b_4(X_4+X_2)+b_5(X_5+X_2)+b_6(X_6-X_2)+e_{13} \qquad (19)$$

Using D_1, D_2, D_3 and defining $D_5 = X_5+X_2$ and $D_6 = X_6-X_2$, equation 19 can be rewritten as $Y = b_1D_1+b_3D_2+b_4D_3+b_5D_5+b_6D_6+e_{13}$. (20)

$R^2_{20}=.70326$.

Then $F = \dfrac{(.71183-.70326)/1}{(1-.71183)/24} = \dfrac{.01857}{(1-.71183)/24} = 1.55$,

which is non-significant. Thus, while we have previously showed that the differences between the posttested groups is significant ($p<.01$) and the differences in gains in the pretested groups are non-significant ($p>.05$), there are no significant differences between the gain of the mean scores and the posttested only groups differences. This is not to say the outcomes for the Solomon design are uninterpretable; it does say that the interpretations are tricky.

Viewing the Solomon as a Three-Way Design

It is possible to view the Solomon design as a 2x2x2 design with two missing cells. The missing cells are planned, as was the case in missing cell design described by Williams and Wali (1979). In diagramatic form, the three dimensional case can be seen as:

	Pretested: Pre	Pretested: Post	Non-Pretested: Pre	Non-Pretested: Post
Experimental	Group 1	Group 2	X	Group 5
Control	Group 3	Group 4	X	Group 6

To test for the experimental-control main effect (A effect), the following restriction can be imposed:

$$b_1+b_2+b_5=b_3+b_4+b_6$$

which yields

$$Y = b_2(X_2-X_1)+b_3(X_3+X_1)+b_4(X_4+X_1)+b_5(X_5-X_1)+b_6(X_6+X_1)+e_{14}. \quad (21)$$

Defining $D_7 = X_2-X_1$;

$D_8 = X_3+X_1$;

$D_9 = X_4+X_1$;

$D_{10} = X_5-X_1$; and

$D_{11} = X_6+X_1$

$$Y = b_2D_7+b_3D_8+b_4D_9+b_5D_{10}+b_6D_{11}+e_{14}. \quad (22)$$

$R^2_{22} = .29159$;

$$F = \frac{(.71183-.29159)/1}{(1-.71183)/24} = \frac{.42024}{.28817/24} = 35.00,\ p<.01.$$

To test the effect of pretesting (the B effect), several rival hypotheses might be used to serve as the main effect.

One such hypothesis is $b_1+b_2+b_3+b_4=b_5+b_6$. This hypothesis does not test the more appropriate hypothesis of interest, since the pretested scores are being compared to the scores which have been posttested only. More interesting is $b_2+b_4=b_5+b_6$ or $b_2=b_5+b_6-b_4$.

Then,

$$Y = b_1X_1+b_3X_3+b_4(X_4-X_2)+b_5(X_5+X_2)+b_6(X_6+X_2)+e_{15}. \qquad (23)$$

Defining $D_{12} = X_4-X_2$;

$D_{13} = X_5+X_2$;

$D_{14} = X_6+X_2$,

$$Y = b_1X_1+b_3X_3+b_4D_{12}+b_5D_{13}+b_6D_{14}+e_{15}. \qquad (24)$$

$R^2_{24} = .69897$;

$$F = \frac{(.71183-.69897)/1}{(1-.71183)/24} = \frac{.01286}{(1-.71183)/24} = 1.07,\ p>.05.$$

The outcome of this test would suggest that the effect of pretesting per se is minimal for this data set.

To test for pre-post differences (the C main effect), the restriction $b_1+b_3=b_2+b_4$ or $b_1=b_2+b_4-b_3$ can be imposed. Then

$$Y = (b_2+b_4-b_3)X_1+b_2X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_{16}, \text{ or}$$

$$Y = b_2(X_2+X_1)+b_3(X_3-X_1)+b_4(X_4+X_1)+b_5X_5+b_6X_6+e_{16}. \qquad (25)$$

Letting $D_{15} = X_3-X_1$, equation 25 can be rewritten

$$Y = b_2D_1+b_3D_{15}+b_4D_9+b_5X_5+b_6X_6+e_{16}. \qquad (26)$$

$R^2_{26}=.50600$;

$$F = \frac{(.71183-.50600)/1}{(1-.71183)/24} = \frac{.20583}{(1-.71183)/24} = 17.14,\ p<.01,$$

indicating a pre-test increase in scores.

Interactions in the Three-Way Design

First of all, the two missing cells will cause the non-existence of two interactions. The three way interaction will not exist, since it is impossible to have non-pretested groups who were pretested. For the same reason, the BC interaction will fail to exist. To test for the AB interaction, that is, the interaction between the experimental-control condition

(A) and the effect of pretesting (B), the restriction on the full model would be:

$b_2 - b_5 = b_4 - b_6$ or $b_2 = b_4 - b_6 + b_5$.

Then $Y = b_1X_1 + (b_4 - b_6 + b_5)X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + e_{17}$, or

$$Y = b_1X_1 + b_3X_3 + b_4(X_4 + X_2) + b_5(X_5 + X_2) + b_6(X_6 - X_2) + e_{17}. \quad (27)$$

Using previously defined transformations, $Y = b_1X_1 + b_3X_3 + b_4D_3 + b_5D_{13} + b_6D_6 + e_{17}$. (28)

$R^2_{17} = .71183$; R^2_{17} is identical to the R^2 for the full model. This is circumstantially so because $\overline{Y}_2 - \overline{Y}_5 = \overline{Y}_4 - \overline{Y}_6 = 13-12 = 7-6$. Thus, the AB interaction is equal to zero.

To test the AC interaction, that is, the experimental-control condition (A) with pre-post differences (C), the restriction $b_2 - b_1 = b_4 - b_3$ would be imposed on the full model. This in fact was already done in equation 15, yielding $R^2_{15} = .66038$, $F = 4.29$, $p > .05$. The results from the three-way analysis can be placed into a summary table; see Table 4.

Table 4

Summary Table for a Three-Way Solution to the Solomon Design

Effect	Restriction	R^2	df	SS	MS	F
Full Model						
A (experimental-control)	$b_1+b_2+b_5=b_3+b_4+b_6$	.29159	1	163.33	163.33	35.00
B (pretesting)	$b_2+b_4=b_5+b_6$	.69897	1	5.00	5.00	1.07
C (pre-post differences)	$b_1+b_3=b_2+b_4$	.50600	1	80.00	80.00	17.14
AB	$b_2-b_5=b_4-b_6$	.71183	1	0	0	0
AC	$b_2-b_1=b_4-b_3$	.66038	1	20.00	20.00	4.29
Deviation from Full Model		.28817	24	112.00	4.67	

Finding the sum of squares in Table 4 is facilitated by knowing $SS_T = 388.67$. Also, the C effect and the AC effect are identical to the same effects as shown in Table 3.

References

Campbell, D.T. and Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand-McNally, 1966.

Dunn, O.J. Multiple comparisons among means. Journal of the American Statistical Association, 1961, 56, 52-64.

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

Newman, I., Benz, C.R. and Williams, J.D. Repeated measures: Suggested approach for simultaneously independent and dependent measures. Paper presented at the American Psychological Association Annual Meeting, Montreal, Canada, September 1980.

Solomon, R.L. An extension of control group design. Psychological Bulletin, 1949, 46, 137-150.

Williams, J.D. Regression analysis in educational research. New York: MSS Publishing Company, 1974.

Williams, J.D. and Wali, M.K. Missing cells and a curious case of degrees of freedom. Multiple Linear Regression Viewpoints, 1979, No. 5, 75-87.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 11, NUMBER 3 SUMMER 1982

TEACHING APPLIED RESEARCHERS TO CREATE THEIR OWN STATISTICAL MODELS

Joe H. Ward, Jr.
Brooks Air Force Base, Texas

Earl Jennings
The University of Texas at Austin

The purpose of the following remarks is to give you something of the flavor of a novel approach to the teaching of statistical model building and manipulation. Historically, it evolved out of an applied environment in which many of the classical models appeared to be inadequate or at least deficient in one or more respects. Students in applied areas who have been exposed to the approach respond enthusiastically to it, and, in general, the more "traditional" work they have had, the greater their enthusiasm. The response of teachers has been mixed. Many of the critics make remarks similar to those criticisms that are directed at the "new math." It is certainly accurate to state that students of this approach get very little practice in arithmetic for even the most elementary models. In fact, the primary text [6] is almost totally devoid of computing formulae.

With respect to mathematical and statistical foundations, we rely very heavily on the theory of the classical fixed-x linear model, and the text bears some superficial resemblance to a typical text on linear models. However, a great deal of the material covered in a typical linear models text will be

Presented at AERA 1982, MLR Special Interest Group
Not refereed by editorial staff

found in ours only indirectly, if at all. Conversely, the concepts we identify and the skills we try to develop are only indirectly inferable from the typical text.

In general, our approach has the following characteristics:

1. A technical vocabulary of minimal length.
2. Very few special symbols and computational formulae. In those places where a new special symbol or formula would ordinarily be introduced, we make every effort to identify the concept as a special case of a more general concept and the formula as a special case of a more general formula. The cumulative effect of this is, we believe, a hierarchical structuring of the content that enhances learning. See Appendix A for an example of the way we summarize the models of one-way analysis of variance, a test for non-linearity, and simple regression analysis, and Appendix B for a summary of a two-factor problem. Students are assumed to have access to a computer, so very little arithmetic is required.
3. An emphasis on the idea that a model is a way of formalizing an argument.
4. Practice in translating natural language into models with unambiguously specified properties. The kind of skill required to do this is similar to the skill required to translate elementary algebra "word problems" into algebraic equations.
5. Extensive practice in the algebraic manipulation of models. This skill is frequently necessary to create an assumed model with specified properties and almost always required to produce a restricted model that can be used in tests of hypotheses about the parameters of the

assumed model. Although the amount of algebra required is burdensome for some models, the level of skill required is minimal.

Some of the features of the approach can best be understood by an example. Suppose we were interested in evaluating the differential effects of two different methods of teaching reading in the second grade. Students are randomly assigned to the two conditions. A measure of reading achievement is obtained before instruction begins, and another measure is obtained at the end of instruction. Because girls tend to read better at this age than boys, we can probably increase the precision of our estimations and the power of our tests by considering sex in the model. Moreover, there is a possibility that sex might interact with teaching method, initial performance, or both.

Ultimately, we are going to argue that if we can reject the hypotheses

$$E(1, \text{boy}, x) = E(2, \text{boy}, x)$$

$$E(1, \text{girl}, x) = E(2, \text{girl}, x)$$

we are in a position to conclude that the methods are not equally effective. Stated in prose, the hypothesis is that the expected posttest performance for a Method 1 boy with initial performance x is the same as the expected posttest performance for a Method 2 boy with the same initial performance, x. A similar statement is made for girls, and x takes on all possible values of initial performance. Suppose the potential range of x is 20 to 80. We seek a model that will produce 2 (methods) X 2 (sexes) X 61 (values of x) = 244 estimates of expected values. If we are not willing to make any simplifying assumptions about the relationships among the expected values, we need a model with 244 parameters, which we refer to as the mutually exclusive categorical model. Fortunately, in this problem, it seems reasonable to assume that the expected difference in posttest performance per unit difference in initial performance is constant (sometimes called the linearity assumption), although perhaps a different constant for each of the four groups. If this assumption is true, then the 244 expected values are expressible as a function of only eight parameters. In the text, we discuss ways of investigating the tenability of this assumption. Although there are an infinite number of ways of parameterizing a model to estimate the eight parameters, one with intuitive appeal is

$$Y = a_1 B^{(1)} + a_2 B^{(2)} + a_3 G^{(1)} + a_4 G^{(2)} + c_1(X * B^{(1)}) + c_2(X * B^{(2)}) + c_3(X * G^{(1)}) + c_4(X * G^{(2)}) + E^{(1)}$$

where

$\underset{\sim}{Y}$ is a column vector of dimension n containing the observed posttest scores.

$\underset{\sim}{B}^{(i)}$ is a column vector of dimension n containing a one if the corresponding value in Y was observed on a boy in method i; zero otherwise. (i = 1,2)

$G^{(i)}$ is defined for girls similar to $B^{(i)}$ for boys.

X is a column vector of dimension n containing pretest scores arranged in the same order as Y.

The a's and c's are unknown scalars, and $E^{(1)}$ is an unknown column vector. A least squares solution to Model 1 might produce values that could be represented as in Figure 1.

The a's are the intercepts and the c's the slopes of the four separate straight lines. They are also estimates of the eight parameters which are assumed to yield the expected values. We could proceed to investigate our

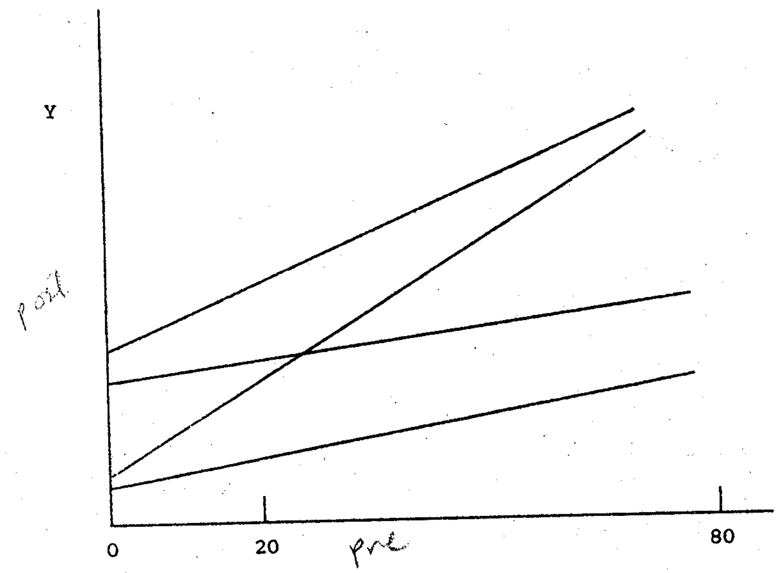


Figure 1. Possible results for Model 1.

ultimate hypothesis using Model 1 as an assumed model. However, such a test based on the F distribution would involve four degrees of freedom in the numerator and would not produce an unqualified recommendation with respect to method.

This kind of problem is frequently approached in standard methods by a factorial analysis of covariance in which the assumed model is a subspace of Model 1 incorporating the assumption that each c is an estimate of the same parameter. This assumption is frequently referred to as the homogeneity of regression assumption. If this assumption is true, then the 244 expected values are expressible in terms of only five parameters. A model to estimate these parameters is

$$Y = a_1B^{(1)} + a_2B^{(2)} + a_3G^{(1)} + a_4G^{(2)} + cX + E^{(2)}$$

A least squares solution to Model 2 might be represented as in Figure 2.

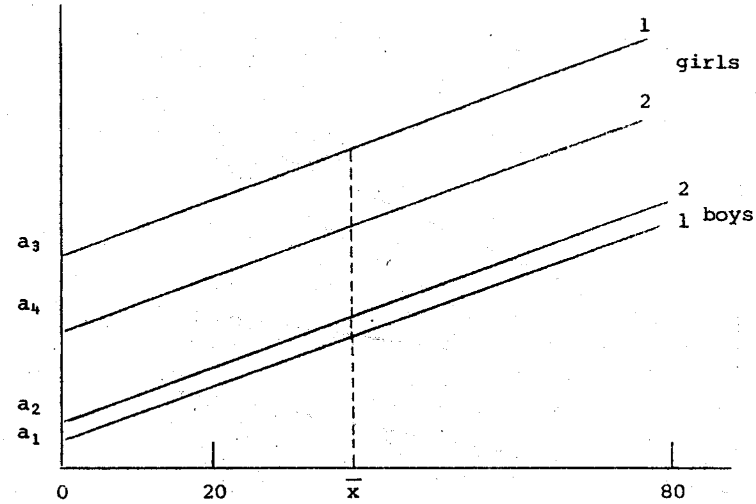


Figure 2. Possible results for Model 2.

In Model 2, the a's are the intercepts of the four lines in Figure 2, and c is the common slope. The test for "treatment effect" involves a comparison of what are called the "adjusted means," namely

$$\frac{(a_1 + c\bar{x}) + (a_3 + c\bar{x})}{2} = \frac{(a_2 + c\bar{x}) + (a_4 + c\bar{x})}{2}$$

which simplifies to $a_1 + a_3 = a_2 + a_4$.

A sufficiently large non-zero difference leads to a relatively large F, a rejection of the hypothesis, and the conclusion that the methods differ. Such a conclusion seems defensible, but we are still not in a position to make an unqualified recommendation with respect to method. In Figure 2, $a_1 + a_3$ is greater than $a_2 + a_4$, yet the available data seem to suggest that Method 1 is better for girls and Method 2 is better for boys.

A number of possibilities exist to reduce this ambiguity. The standard covariance sex by method interaction test is relevant information, but it does not directly address the issue. We could conduct pair-wise investigations

($a_1 = a_2$ and $a_3 = a_4$) and suffer the problems of an increased experimentwise Type I error rate or adopt some post hoc test and suffer the consequent loss of power.

An alternative is to consider an assumed model that avoids the ambiguity altogether. For example, if we are willing to assume the following relationships among the expected values

$$E(1, \text{boys}, x) - E(2, \text{boys}, x) =$$
$$E(1, \text{girls}, x) - E(2\ \text{girls}, x)$$

and

$$E(1, \text{boys}, x_1) - E(2, \text{boys}, x_1) =$$
$$E(1, \text{boys}, x_2) - E(2, \text{boys}, x_2)$$

and

$$E(1, \text{girls}, x_1) - E(2, \text{girls}, x_1) =$$
$$E(1, \text{girls}, x_2) - E(2, \text{girls}, x_2)$$

where

$$x, x_1, x_2 = 20, 21, \ldots 80 \qquad x_1 \neq x_2$$

the 244 expected values are expressible as a function of only five parameters as in Model 2, but because we are making different assumptions, the model we create will have different properties than Model 2. The skills required to create a model that incorporates the desired assumptions are identical to the skills required to test the assumptions. Involved is a simple substitution for the expected values above, their estimates in symbolic form from Model 1, and an algebraic simplification that results in three implied restrictions. Substituting the symbolic estimates from Model 1 for the expected values above,

$$a_1 + c_1x - a_2 - c_2x = a_3 + c_3x - a_4 - c_4x \qquad (1)$$

$$a_1 + c_1x_1 - a_2 - c_2x_1 = a_1 + c_1x_2 - a_2 - c_2x_2 \tag{2}$$

$$a_3 + c_3x_1 - a_4 - c_4x_1 = a_3 + c_3x_2 - a_4 - c_4x_2 \tag{3}$$

Equation (2) can be simplified to

$$(c_1 - c_2)(x_1 - x_2) = 0$$

Since $x_1 \neq x_2$, c_1 must equal c_2, and they can be given a common name.

$$c_1 = c_2 = b, \text{ a common value} \tag{4}$$

Similarly, Equation (3) can be simplified to

$$(c_3 - c_4)(x_1 - x_2) = 0$$

implying

$$c_3 = c_4 = g, \text{ a common value} \tag{5}$$

Substituting (4) and (5) into (1), we achieve

$$a_1 + bx - a_2 - bx = a_3 + gx - a_4 - gx$$

which can be written

$$a_1 - a_2 - a_3 + a_4 = 0 \tag{6}$$

a_1 through a_4 can be renamed so that they satisfy (6) as follows:

$$\begin{aligned} a_1 &= d_1 \\ a_2 &= d_1 + d_3 \\ a_3 &= d_2 \\ a_4 &= d_2 + d_3 \end{aligned} \tag{7}$$

In effect, we have renamed the eight parameter estimates in Model 1 in terms of only five names: d_1, d_2, d_3, b, and g.

If the new names are substituted in Model 1, we get

$$Y = d_1B^{(1)} + (d_1 + d_3)B^{(2)} + d_2G^{(1)} + (d_2 + d_3)G^{(2)} +$$
$$b(X * B^{(1)}) + b(X * B^{(2)}) + g(X * G^{(1)}) +$$
$$g(X * G^{(2)}) + E^{(3)}$$

Expanding and simplifying yields

$$Y = d_1(B^{(1)} + B^{(2)}) + d_2(G^{(1)} + G^{(2)}) + d_3(B^{(2)} + G^{(2)}) +$$

$$b(X * B^{(1)} + X * B^{(2)}) + g(X * G^{(1)} + X * G^{(2)}) + E^{(3)}$$

A least squares solution to Model 3 might appear as in Figure 3.

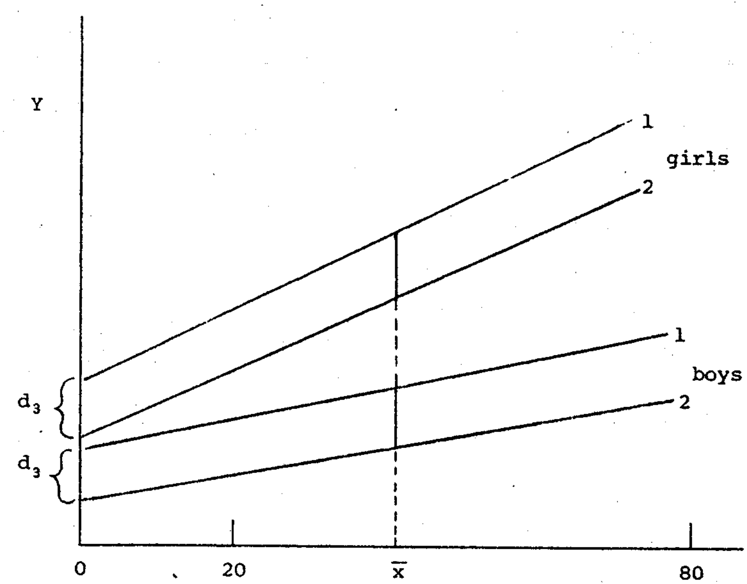


Figure 3. Possible results for Model 3.

The essential property of Model 3 for our purpose is that the expected difference between any pair of persons having the same sex and initial performance, differing only in the method of instruction, is estimated by the same constant, namely d_3. When the properties of a model are not immediately obvious by inspection, we encourage the practice of verifying that the model has the claimed properties. This involves writing the symbolic expressions that estimate the expected values and verifying that the symbolic expressions are related as the expected values are assumed to be, as shown in Table 1.