DOCUMENT RESUME

ED 334 239 TM 016 771

AUTHOR	Rowell, R. Kevin
TITLE	Double Cross-Validation in Multiple Regression: A Method of Estimating the Stability of Results.
PUB DATE	21 Mar 91
NOTE	25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Equations (Mathematics); *Estimation (Mathematics); Heuristics; *Mathematical Models; *Multiple Regression Analysis
IDENTIFIERS	*Double Cross Validation; *Invariance; Research Replication

ABSTRACT

In multiple regression analysis, where resulting predictive equation effectiveness is subject to shrinkage, it is especially important to evaluate result replicability. Double cross-validation is an empirical method by which an estimate of invariance or stability can be obtained from research data. A procedure for double cross-validation is discussed, using heuristic and actual research data to illustrate a non-generalizable outcome and a generalizable outcome. The procedure involves the use of two samples or subsamples to produce two pairs of regression equations from which respective shrinkages can be determined. The more closely the shrinkage estimates approach zero, the greater the degree of stability across the subsamples and the more confidence the researcher can vest in the replicability of the results. The first example uses a heuristic data set of 20 (2 subsets of 11 and 9 subjects, respectively) with 5 predictors to illustrate that weights must be compared empirically rather than subjectively. In the second example, data are drawn from a study of the life satisfaction of 200 nursing home residents. Seven tables present study data. An appendix contains the Statistical Analysis System program used to analyze the data. (SLD)

doublecro.ws 3/21/91

Double Cross-Validation in Multiple Regression: A Method of Estimating the Stability of Results

R. Kevin Rowell

Texas A&M University 77843-4425

Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 5, 1991.

Abstract

Tests of statistical significance are widely used in educational and psychological research to facilitate interpretation of findings. But such tests do not reflect the degree of stability of findings across samples, and in multiple regression where resulting predictive equation effectiveness is subject to "shrinkage", it is especially important to evaluate result replicability. Indeed, since all parametric analytic methods are special cases of regression (just as all univariate and multivariate methods are special cases of canonical correlation analysis), evaluating result replicability is important in <u>all</u> sorts of studies. Double cross-validation is an empirical method by which an estimate of invariance or stability can be obtained from research data in hand. This paper discusses the procedure for double cross-validation using both a heuristic data set and an actual research data set to illustrate both a nongeneralizable outcome and a generalizable outcome.

Double Cross-Validation in Multiple Regression:

A Method for Estimating the Stability of Results

A growing trend in educational and psychological research has been the recognition that sole reliance upon statistical significance testing presents insufficient evidence to determine the importance of research findings (Carver, 1978; Craig, Eison, & Metze, 1976; Thompson, 1989). The most compelling argument against statistical significance testing addresses the issue of sample size effects on the outcome of the null hypothesis test.

More specifically, the larger the sample size, the greater is the likelihood that a null hypothesis will be rejected (Carver, 1978). As a result, if one's sample size is sufficient, then even the most trivial research findings will become statistically significant and thereby seem "important. On the other hand, it is likely that results based or small sample sizes may not be statistically significant, but may nevertheless reveal noteworthy results. Since the null hypothesis in social sciences research is seldom, if ever, exactly true, a sufficiently large sample size will almost always yield a statistically significant result (Fish, 1986; Sandler, 1987).

An even more compelling problem with statistical significance testing is that it offers the researcher no indication of the replicability of the results, that is, the likelihood that such results would be reproduced in the future (Thompson, 1989). In essence, relying only on statistical significance in determining the merits of research findings "represents a corrupt form of the scientific method" (Carver, 1978, p. 378).

Given the limitations of statistical significance testing in determining the stability of research results, certainly the best predictor of result generalizability, or stability across samples, would be to conduct replications on as many samples as possible, thereby empirically validating result stability. But given that in educational and psychological research such a solution is often impractical, the stability of research results across samples can be estimated by using invariance techniques (Fish, 1986). As described by Englehard (1989), "invariance can also be viewed more broadly as the quest for generality in science" (p. 32). Borrowing from Englehard's discussion on the history of invariance, the concept of invariance within social sciences research is best described by Stevens (1951):

> The scientist is usually looking for invariance whether he knows it or not . . . The quest for invariant relations is essentially the aspiration toward generality, and in psychology, as in physics, the principles that have wide applications are those we prize. (p. 20)

Because methods of invariance investigation depend to a large extent upon the analytic method used, the number of invariance procedures is quite large. The scope of the present papers focuses upon invariance procedures used with multiple regression analysis, and more specifically, double cross-validation. However, given that all statistical analytic methods are interrelated, the logic illustrated in the present paper can certainly be generalized to other analytic methods.

Result Stability in Multiple Regression

The Problem of Shrinkage

In multiple regression analysis, the researcher seeks to find those independent variables (and their respective weights) that correlate most highly with the dependent variable. The analysis produces weights that can be applied to the predictor variables to yield a predicted score, YHAT, for each subject. When the variables are all in z-score form these weights are called beta weights, and when the variables are unstandardized the weights are called b weights.

The weights are developed subject to the restriction that the YHAT scores must come as close as possible to the Y scores in the sample, for the sample as a whole. The deviation of a given subject's YHAT from the subject Y is the subject's e score. Thus, the weights are derived to minimize the e scores, or, more specifically to minimize the sum of the squared e scores, also called SOS error or SOS within.

The multiple correlation, R, is the correlation between the predicted scores, i.e., the YHAT scores, and the observed criterion scores, i.e., the Y scores. R^2 represents the proportion of variance of the dependent variable that is shared with the independent variables as a set, the set being represented by the YHAT scores (Pedhazur, 1982).

It should be noted, however, that R^2 is the maximum mathematical value for the given sample due to "overfitting," or the capitalization on sampling error in the derivation of the "optimal" weights for the sample data (Mitchell & Klimoski, 1986). Mosier (1951) describes this chance factor as involving the idiosyncratic characteristics of the sample. Pedhazur (1982) explains that such overfitting is due to the treatment of zero-order correlations as being error-free, which is never true. Pedhazur further contends that the degree of overestimating R is affected by the ratio of the number of independent variables to sample size and that as the number of independent variables approaches the sample size, the likelihood of overestimating R (and R^2) gets larger.

These difficulties pose problems in applying the regression equation to other samples. If the derived sample weights were applied to the predictor scores of another sample, the resulting multiple correlation between the predicted scores and the criterion scores of the second sample would almost always be less than the original multiple correlation (Pedhazur, 1982). In other words, shrinkage of R is certain to occur when the regression equation is applied across samples, and in fact, predictor variables that prove to be statistically significant in the derivation sample may "shrink" to nonsignificant values in the second sample. In this context, if invariance represents stability across samples, then it can be viewed as inversely related to the degree of shrinkage. The smaller the degree of shrinkage, the greater is the invariance, and thus the more generalizable is the regression equation.

The Double Cross-Validation Procedure

"Double cross-validation" is an empirical invariance procedure used in multiple regression that essentially involves the use of two samples or subsamples to produce two pairs of regression equations from which respective shrinkages can be determined. Double cross-validation offers a greater level of confidence in generalizability when applied to two separate samples than when applied to two subsamples creating by splitting a single sample; however, if the sample size is sufficiently large, then two randomly assigned subsamples can provide a fair estimate of result reproducibility. And using some estimate of result stability or invariance is almost always better than failing to conduct any empirical evaluation of result replicability.

Because educational and psychological researchers often encounter problems in obtaining data from more than one sample, the use of two random subsamples derived by splitting a sample may be a useful procedure. For this reason, the procedure for double cross-validation described in the present paper is offered in the context of comparing two subsamples rather than two separate samples.

If a single sample is used, the first step in double cross-validation requires that the sample be divided randomly into two subsamples (e.g., 50% + 50%; 51% + 49%; 75% + 25%). Fish (1986) contends that the subsamples should be unequal, for if the results of a disproportionately smaller subsample (e.g., 25%) prove to be replicable, one might be willing to vest even more confidence in the generalizability of the results. However, such "acid" tests may be counterproductive or overconservative. Some researchers will want to use subsamples of more nearly equal size, to provide greater likelihood that invariance will be found.

After creating the two subsamples, the variables within the two sets are converted to z-scores. The z-scores are standardized with the means and SDs of subsample 1 for the subsample 1 data, and with the means and SDs of subsample 2 for the subsample 2 data. Separate regression analyses are also performed with each subsamples' data. As a result, separate beta weights are derived for each subsample's set of z-scores. The resulting betas are then used to compute predicted Y values, YHAT, for the cases in each subsample, such that:

$$YHAT_{11} = beta_{11} z_{11} + beta_{12} z_{12} + beta_{13} z_{13} + \ldots beta_{1j} z_{1j}$$

$$YHAT_{22} = beta_{21} z_{21} + beta_{22} z_{22} + beta_{23} z_{23} + \ldots beta_{2j} z_{2j}$$

In this notation, the first subscript for the YHATs indicate which sample's z-scores were used to calculate the YHATs, while the second subscript for YHAT indicates which sample's beta weights were employed. The first subscript for the beta weights and for z-scores indicate which sample yielded the weights or the z-scores, while the second subscript indicates the sequence number of the predictor, ranging from 1 through the j^{th} predictor variable.

With two composite scores, $YHAT_{11}$ and $YHAT_{22}$, thus determined, the next procedure is to "cross" beta weights and compute two new sets of predicted YHAT values, namely $YHAT_{12}$ and $YHAT_{21}$, by the same methods. In computing $YHAT_{12}$, the betas from the regression of subsample 2 are applied to the z-scores of Subsample 1. Conversely, the betas of subsample 1 are applied to the z-scores of Subsample 2 in order to calculate $YHAT_{21}$. Thus, these estimates take the form:

$$YHAT_{12} = beta_{21} z_{11} + beta_{22} z_{12} + beta_{23} z_{13} +\ldots beta_{2j} z_{1j}$$

$$YHAT_{21} = beta_{11} z_{21} + beta_{12} z_{22} + beta_{13} z_{23} +\ldots beta_{1j} z_{2j}$$

Upon completing the computation of the four sets of YHAT scores, two for subjects in each of the two subsample groups, various combinations of the scores can be correlated. The correlation of $YHAT_{11}$ with the Y scores of subjects in subsample 1 will yield the R for that subsample. The correlation of $YHAT_{22}$ with the Y scores of subjects in subsample 2 will yield the R for that subsample.

The invariance of the results can be evaluated in either of two ways. First, the shrinkage can be evaluated for each group as:

$$INV_1 = R^2_{11} - R^2_{12}$$

$$INV_2 = R^2_{22} - R^2_{21}$$

The more closely these two shrinkage estimates approach zero, the greater is the degree of stability across the subsamples, and hence the more confidence the researcher can vest in the replicability of the results. It should be noted that R^2_{11} and R^2_{22} will almost certainly be greater than R^2_{12} and R^2_{21}, respectively, since the first two R^2's are the mathematical optimums for their respective subsamples.

One aspect of this method of evaluating shrinkage, however, is that the result has no set metric. For example, shrinkage from an R^2 of 70% to one of 60% is not the same as shrinkage from an R^2 of 10% to one of 0%, since the former result is still quite noteworthy, while the latter is not. These difficulties can be overcome by comparing the r^2 of Y with $YHAT_{11}$ (i.e., the actual R^2 of subsample 1) against the r^2 of Y with $YHAT_{12}$, and by comparing the r^2 of Y with $YHAT_{22}$ (i.e., the actual R^2 of subsample 2) against the r^2 of Y with $YHAT_{21}$. These two correlation coefficients can be called invariance coefficients. The more closely these invariance r's approach one, the greater is the degree of confidence obtained in the stability of the regression equation across different configurations of subjects.

Numerical Examples of Double Cross-Validation

In order to illustrate the double cross-validation procedure, two numerical examples will be presented. The first example uses a smaller heuristic data set (n = 20) of 5 predictors to illustrate an invariance situation in which the weights appear to be different across subsamples, but in fact yield reasonably equivalent results across at least one subsample. This example is utilized to drive home the point that weights must be compared empirically, rather than subjectively (Thompson, 1989).

In the second example, the data are drawn from an actual study of life satisfaction in elderly nursing home residents. In this study, 200 nursing home residents were administered a life satisfaction inventory which consisted of 8 subscales. These 8 subscales were used as independent variables to predict overall nursing home satisfaction as measured on a self-report Likert scale.

Example 1. The first step in the cross-validation procedure requires that the data be randomly sorted into two relatively equal subsamples. In this example, Subsample 1 contains 11 subjects, and Subsample 2 has 9 subjects. The hypothetical data are presented in Table 1. The SAS program used to analyze the data is presented in Appendix A.

Insert Table 1 about here

In the second step, the scores within each subsample were converted to z-scores by first computing means and variances of the independent variables, X1-X5, in an initial computer run. These results were then used in the second computer run to obtain z-scores, Z1_X1-Z1_X5 and Z2_X1-Z2_X5, for both samples. In addition, separate regression analyses were conducted within each subsample in order to generate each subsample's respective beta weights. These results are presented in Tables 2 and 3.

Insert Tables 2 and 3 about here

At this point in the analysis z-scores and betas have been computed for subsample 1 and subsample 2. From these values, estimates of Y can be predicted for each of the subsamples:

Let YHAT_11 = predicted Y scores for Subsample 1

YHAT_22 = predicted Y scores for Subsample 2

therefore:

YHAT_11 = (0.99311 x Z1_X1) + (-0.17191 x Z1_X2)

$$
\begin{aligned}
&\quad + (0.25976 \times \text{Z1\_X3}) + (-0.30172 \times \text{Z1\_X4}) \\
&\quad + (0.07974 \times \text{Z1\_X5}) \\
\text{YHAT\_22} &= (-0.47831 \times \text{Z2\_X1}) + (0.39201 \times \text{Z2\_X2}) \\
&\quad + (0.74884 \times \text{Z2\_X3}) + (0.68759 \times \text{Z2\_X4}) \\
&\quad + (0.60224 \times \text{Z2\_X5})
\end{aligned}
$$

The third step is to "cross" the betas in subsample 2 with the z-scores in subsample 1 in order to compute the invariance coefficient for the subsample 2 weights. Likewise, the betas in subsample 1 are crossed with the z-scores in subsample 2 in order to calculate the invariance for the subsample 1 weights, as follows:

Let YHAT_12 = invariance composite scores for Subsample 1

YHAT_21 = invariance composite scores for Subsample 2

therefore:

$$
\begin{aligned}
\text{YHAT\_12} &= (-0.47831 \times \text{Z1\_X1}) + (0.39201 \times \text{Z1\_X2}) \\
&\quad + (0.74884 \times \text{Z1\_X3}) + (0.68759 \times \text{Z1\_X4}) \\
&\quad + (0.60224 \times \text{Z1\_X5}) \\
\text{YHAT\_21} &= (0.99311 \times \text{Z2\_X1}) + (-0.17191 \times \text{Z2\_X2}) \\
&\quad + (0.25976 \times \text{Z2\_X3}) + (-0.30172 \times \text{Z2\_X4}) \\
&\quad + (0.07974 \times \text{Z2\_X5})
\end{aligned}
$$

As a result of these computations, each subsample has two sets of predicted YHAT scores, namely YHAT_11 and YHAT_12 for subsample and YHAT_21 and YHAT_22 for subsample 2. These values are then correlated to yield the invariance coefficients. The results for the example are presented in Table 4.

Insert Table 4 about here

The resulting correlation of YHAT_11 with YHAT_12 revealed an invariance estimate of .95, indicating that the weights from the two subsamples yield very similar estimates of YHAT. At first pale this result may seem surprising, since the beta weights ("STANDARDIZED ESTIMATEs") presented in Tables 2 and 3 appear to be very different, e.g., +.079 for ZX5 in subsample 1 versus +.602 for ZX5 in subsample 2.

However, the invariance correlation of YHAT_22 with YHAT_21 was .54, indicating that the sets of weights were not equally effective when they were both to the data for subsample 2. This finding illustrates the utility of "doubly" cross-validating, both ways. The discrepancy between these two invariance estimates would contraindicate stability of predictor weights across samples.

Example 2. Regarding the nursing home satisfaction example, the nursing home data were randomly sorted into two relatively equal subsamples. Subsample 1 contained 101 subjects, and Subsample 2, 99 subjects. Subsequently, within each subsample, z-scores and beta weights were obtained. The results for the two subsamples are presented in Tables 5 and 6.

Insert Tables 5 and 6 about here

From these values, nursing home satisfaction (NH_SAT) can be predicted for each of the subsamples:

Let NH_SAT11 = predicted NH_SAT for Subsample 1

NH_SAT22 = predicted NH_SAT for Subsample 2

therefore:

$$\text{NH\_SAT11} = (.80567 \times \text{Z1\_MEAN}) + (-.09939 \times \text{Z1\_GOAL}) + (-.01853 \times \text{Z1\_SOCL}) + (.10754 \times \text{Z1\_YEAR})$$

$$\text{NH\_SAT22} = (.74817 \times \text{Z2\_MEAN}) + (-.03272 \times \text{Z2\_GOAL}) + (-.13812 \times \text{Z2\_SOCL}) + (.04420 \times \text{Z2\_YEAR})$$

The cross-validation YHATs were computed as:

Let NH_SAT12 = invariance composite scores for Subsample 1

NH_SAT21 = invariance composite scores for Subsample 2

therefore:

$$\text{NH\_SAT12} = (.74817 \times \text{Z1\_MEAN}) + (-.03272 \times \text{Z1\_GOAL}) + (-.13912 \times \text{Z1\_SOCL}) + (.04420 \times \text{Z1\_YEAR})$$

$$\text{NH\_SAT21} = (.80567 \times \text{Z2\_MEAN}) + (-.09939 \times \text{Z2\_GOAL}) + (-.01853 \times \text{Z2\_SOCL}) + (.10754 \times \text{Z2\_YEAR})$$

As a result of these computations, each subsample has two sets of predicted NH_SAT scores, namely NH_SAT11 and NH_SAT12 for subsample 1 and NH_SAT22 and NH_SAT21 for subsample 2.

The invariance coefficients for this analysis are presented as a part of Table 7. The correlation of NH_SAT11 with NH_SAT12 yielded an invariance estimate of .98, indicating that the weights from the two subsamples were very stable. Similarly, the correlation of NH_SAT22 with NH_SAT21 yielded an invariance estimate of .98, thus indicating that weights in the second cross-validation performed very well also. Such a high degree of stability yields a large degree of confidence that the original regression equation for the full sample is an accurate predictor of nursing home satisfaction in this sample and that the equation is fairly stable across samples.

Insert Table 7 about here

And it is always the equation based on the full sample that is ultimately the basis for interpretation. The subsample analyses are conducted to get a feel for the stability of the full sample results, and not to provide a basis for direct interpretation.

Conclusion

Double cross-validation is a method by which investigators using multiple regression analyses can simultaneously conduct two estimates of invariance either across two separate samples or two subsamples drawn from one sufficiently large sample. The advantage of using double cross-validation is that it provides a second "replication" of the results which is useful in comparing to the first set of results.

In educational and psychological research, the importance of a study is typically determined by some test of statistical significance. Whereas these statistical significance tests are widely accepted as measures of importance, they are not very dependable indicators of result reproducibility. A much more accurate estimation of generalizability would be to empirically test the findings across samples and to determine the degree of stability across these samples rather than relying solely upon tests of significance to indicate reproducibility.

References

Carver, R. C. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Craig, J. R., Eison, C. L., & Metze, L. P. (1976). Significance tests and their interpretation: An example utilizing published research and omega-squared. Bulletin of the Psychonomic Society, 7, 280-282.

Englehard, Jr., G. (1989, March). Historical views of the concept of invariance and measurement theory in the behavioral sciences. Paper presented at the Fifth International Objective Measurement Workshop, University of California, Berkeley, CA. (ERIC Document Reproduction Service No. ED 313 388)

Fish, L. (1986, November). The importance of invariance procedures as against tests of statistical significance. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis, TN. (ERIC Document Reproduction Service No. ED 278 707)

Mitchell, T. W., & Klimoski, R. J. (1986). Estimating the validity of cross-validity estimation. Journal of Applied Psychology, 71, 311-317.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). Ft. Worth: Holt, Rinehart, & Winston.

Sandler, A. B. (1987, January). The use of invariance and bootstrap procedures as a method to establish the reliability of research results. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX. (ERIC Document Reproduction Service No. ED 285 902)

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), Handbook of experimental psychology (pp. 1-49). New York: Wiley.

Thompson, B. (1989). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. Measurement and Evaluation in Counseling and Development, 22, 2-6.

Table 1
Hypothetical Data for Example 1 (n=20)

ID	GROUP	Y	X1	X2	X3	X4	X5
		YHAT_11	YHAT_12	YHAT_21	YHAT_22		
1	1	1.4494	2.6535	1.6129	1.4791	0.8249	1.3615
		1.6706	2.8344	.	.		
2	1	0.1235	0.9332	1.0165	-0.4782	0.3678	0.0549
		0.0358	-0.0791	.	.		
3	1	-0.1411	0.5796	-0.9072	0.3909	0.9591	-0.5203
		0.2184	0.2745	.	.		
4	1	-0.6855	-0.7859	0.4845	-0.9226	-1.7435	-0.7351
		-0.5731	-2.0287	.	.		
5	1	1.1238	1.2541	0.7067	0.4088	0.4282	0.6427
		0.6294	1.0881	.	.		
6	1	0.3943	-0.5134	-0.4392	0.7841	-1.1065	0.6111
		0.2953	0.6620	.	.		
7	1	-2.2715	-1.7856	-1.5240	-0.3698	-0.6803	-0.6170
		-0.9373	-1.1105	.	.		
8	1	-1.7230	-1.0883	-0.6883	-1.5188	-1.1091	0.7939
		-0.8318	-1.5652	.	.		
9	1	-1.5258	-0.8672	-1.2947	-0.1193	0.6647	-1.9800
		-0.8131	-1.1258	.	.		
10	1	1.0084	0.9381	-1.1084	0.5831	1.7002	0.7223
		0.4423	1.5776	.	.		
11	1	-0.5916	-0.4931	-1.3715	-0.3252	-0.4884	0.4942
		-0.1362	-0.5275	.	.		
12	2	-1.8914	-0.5908	-0.0792	-0.7090	-1.6626	0.2773
		.	.	-0.2533	-1.1022		
13	2	-0.5829	0.2041	0.1175	0.0811	0.8644	-0.6684
		.	.	0.1969	0.0285		
14	2	1.3889	0.3041	-0.2253	-0.4971	0.3586	2.0554
		.	.	0.6474	0.5248		
15	2	-1.4594	-1.3000	0.0148	-1.1782	-0.4230	-0.5962
		.	.	-1.7938	-0.4300		
16	2	2.8002	1.3113	2.3249	1.5848	0.9300	0.0506
		.	.	1.5619	1.6242		
17	2	-0.6989	-0.5218	-0.4050	1.2660	-1.3866	-1.3967
		.	.	0.0817	-0.9846		
18	2	0.8433	-0.2492	-0.5100	0.9812	0.3150	-0.1616
		.	.	0.0437	0.3601		
19	2	0.5183	0.8607	0.1861	-0.0563	0.3345	1.1504
		.	.	1.3755	0.1125		
20	2	-0.0161	-0.7832	0.5174	-2.1399	1.2817	-1.0705
		.	.	-1.8595	-0.1332		

Note. Y is the dependent variable. X1 to X5 are the predictor variables. GROUP is the hypothetical variable randomly created to divide the sample into two subsamples. The YHAT values for each case are also presented.

Table 2
SAS Results for Subsample 1 in Example 1

CORRELATION FOR GROUP 1

	Y	X1	X2	X3	X4	X5
Y	1.00000	0.89693	0.61959	0.75530	0.51278	0.61643
X1	0.89693	1.00000	0.70185	0.70028	0.66596	0.54554
X2	0.61959	0.70185	1.00000	0.28096	0.03819	0.41411
X3	0.75530	0.70028	0.28096	1.00000	0.59227	0.33971
X4	0.51278	0.66596	0.03819	0.59227	1.00000	0.07339
X5	0.61643	0.54554	0.41411	0.33971	0.07339	1.00000

REGRESSION OF GROUP 1
DEP VARIABLE: Y
ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	5	13.22347310	2.64469462	6.992	0.0262
ERROR	5	1.89135420	0.37827084		
C TOTAL	10	15.11482730			

ROOT MSE	0.6150373	R-SQUARE	0.8749
DEP MEAN	-0.2581	ADJ R-SQ	0.7497
C.V.	-238.294		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARDIZED ESTIMATE
INTERCEP	1	-0.40105606	0
X1	1	0.93953203	0.99311254
X2	1	-0.19418815	-0.17190653
X3	1	0.37886145	0.25975559
X4	1	-0.34675263	-0.30172221
X5	1	0.10282402	0.07973646

Table 3
SAS Results for Subsample 2 in Example 1

CORRELATION FOR GROUP 2

	Y	X1	X2	X3	X4	X5
Y	1.00000	0.78854	0.60199	0.47657	0.63885	0.40768
X1	0.78854	1.00000	0.58245	0.56626	0.44290	0.51467
X2	0.60199	0.58245	1.00000	0.22532	0.47329	-0.00927
X3	0.47657	0.56626	0.22532	1.00000	-0.11255	-0.02053
X4	0.63885	0.44290	0.47329	-0.11255	1.00000	0.10272
X5	0.40768	0.51467	-0.00927	-0.02053	0.10272	1.00000

REGRESSION OF GROUP 2
FOR CROSS-VALIDATION
DEP VARIABLE: Y

ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	5	15.48388211	3.09677642	5.428	0.0972
ERROR	3	1.71140563	0.57046854		
C TOTAL	8	17.19528774			

ROOT MSE	0.7552937	R-SQUARE	0.9005
DEP MEAN	0.1002222	ADJ R-SQ	0.7346
C.V.	753.619		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARDIZED ESTIMATE
INTERCEP	1	-0.08466587	0
X1	1	-0.84449003	-0.47831374
X2	1	0.67590502	0.39200759
X3	1	0.90571977	0.74884287
X4	1	0.98266984	0.68758602
X5	1	0.80860873	0.60223747

Table 4
Invariance Results for Example 1

	Y	YHAT_11	YHAT_12	YHAT_21	YHAT_22
Y	1.00000	0.93137	0.87360	0.63737	0.93282
YHAT_11	0.93137	1.00000	0.94721		
YHAT_12	0.87360	0.94721	1.00000		
YHAT_21	0.63737			1.00000	0.53884
YHAT_22	0.93282			0.53884	1.00000

Note. The r between Y and YHAT_11 is the R for subsample 1; the r between Y and YHAT_22 is the R for subsample 2. The r's between YHAT_11 and YHAT_12 and between YHAT_22 and YHAT_21 are the invariance coefficients.

Table 5
Regression Results for Subsample 1 in Example 2

CORRELATION FOR GROUP 1

	NH_SAT	MEANING	GOALS	SOCIAL	YEARS
NH_SAT	1.00000	0.75018	0.34277	0.25527	0.17272
MEANING	0.75018	1.00000	0.56846	0.34327	0.06857
GOALS	0.34277	0.56846	1.00000	0.15091	-0.12116
SOCIAL	0.25527	0.34327	0.15091	1.00000	0.11374
YEARS	0.17272	0.06857	-0.12116	0.11374	1.00000

REGRESSION FOR GROUP 1
BETAS USED IN ESTIMATING NH_SAT
FOR CROSS-VALIDATION

DEP VARIABLE: NH_SAT
ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	4	150.39325	37.59831128	33.717	0.0001
ERROR	96	107.05230	1.11512812		
C TOTAL	100	257.44554			

ROOT MSE	1.055996	R-SQUARE	0.5842
DEP MEAN	5.366337	ADJ R-SQ	0.5668
C.V.	19.67816		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARDIZED ESTIMATE
INTERCEP	1	-0.97756288	0
MEANING	1	0.41552367	0.80566633
GOALS	1	-0.06019197	-0.09939076
SOCIAL	1	-0.009329782	-0.01852470
YEARS	1	0.03907011	0.10754426

Table 6
Regression Results for Subsample 2 in Example 2

CORRELATION FOR GROUP 2

	NH_SAT	MEANING	GOALS	SOCIAL	YEARS
NH_SAT	1.00000	0.68295	0.26892	0.13953	0.10225
MEANING	0.68295	1.00000	0.46981	0.38664	0.08048
GOALS	0.26892	0.46981	1.00000	0.36867	0.02391
SOCIAL	0.13953	0.38664	0.36867	1.00000	0.00997
YEARS	0.10225	0.08048	0.02391	0.00997	1.00000

REGRESSION FOR GROUP 2
BETAS USED IN ESTIMATING NH_SAT
FOR CROSS-VALIDATION

DEP VARIABLE: NH_SAT

ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	4	109.26900	27.31724916	22.346	0.0001
ERROR	94	114.91282	1.22247682		
C TOTAL	98	224.18182			

ROOT MSE	1.105657	R-SQUARE	0.4874
DEP MEAN	5.242424	ADJ R-SQ	0.4656
C.V.	21.09056		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARDIZED ESTIMATE
INTERCEP	1	1.28855291	0
MEANING	1	0.30692720	0.74817030
GOALS	1	-0.02182330	-0.03272233
SOCIAL	1	-0.05702813	-0.13811796
YEARS	1	0.01538654	0.04419563

Table 7
Invariance Results for Example 2

	NH_SAT	NH_SAT11	NH_SAT12	NH_SAT21	NH_SAT22
NH_SAT	1.00000	0.76367	0.74721	0.68422	0.69785
NH_SAT11	0.76367	1.00000	0.97975		
NH_SAT12	0.74721	0.97975	1.00000		
NH_SAT21	0.68422			1.00000	0.98171
NH_SAT22	0.69785			0.98171	1.00000

Appendix A:
SAS Program to Analyze Table 1 Data

```
INFILE ABC;
INPUT ID GROUP Y X1 X2 X3 X4 X5;

   PROC REG;
      MODEL Y = X1 X2 X3 X4 X5 / STB;
      TITLE  'REGRESSION FOR ALL DATA';
   PROC CORR;
      VAR Y X1 X2 X3 X4 X5;
      TITLE 'CORRELATION FOR ALL DATA';

DATA GROUP1;
   SET STATS;
IF GROUP=1;
   Z1_X1 = (X1 - 0.0750) / 1.6888;
   Z1_X2 = (X2 + 0.3193) / 1.1845;
   Z1_X3 = (X3 + 0.0080) / 0.7105;
   Z1_X4 = (X4 + 0.0166) / 1.1444;
   Z1_X5 = (X5 - 0.0753) / 0.9089;

   YHAT_11 = (.99311*Z1_X1) - (.17191*Z1_X2) + (.25976*Z1_X3)
           - (.30172*Z1_X4) + (.07974*Z1_X5);
   YHAT_12 =(-.47831*Z1_X1) + (.39201*Z1_X2) + (.74884*Z1_X3)
           + (.68759*Z1_X4) + (.60224*Z1_X5);
PROC UNIVARIATE;
   VAR Y X1 X2 X3 X4 X5;
   TITLE  'UNIVARIATE STATISTICS FOR GROUP 1';
PROC REG;
   MODEL Y = X1 X2 X3 X4 X5 / STB;
   TITLE1 'REGRESSION OF GROUP 1';
   TITLE2 'FOR CROSS-VALIDATION';
PROC CORR;
   VAR Y X1 X2 X3 X4 X5;
   TITLE 'CORRELATION FOR GROUP 1';

DATA GROUP2;
   SET STATS;
IF GROUP=2;
   Z2_X1 = (X1 + 0.0850) / 0.6895;
   Z2_X2 = (X2 - 0.2157) / 0.7230;
   Z2_X3 = (X3 + 0.0742) / 1.4693;
   Z2_X4 = (X4 - 0.0680) / 1.0524;
   Z2_X5 = (X5 + 0.0400) / 1.1923;

   YHAT_21 = (.99311*Z2_X1) - (.17191*Z2_X2) + (.25976*Z2_X3)
           - (.30172*Z2_X4) + (.07974*Z2_X5);
   YHAT_22 =(-.47831*Z2_X1) + (.39021*Z2_X2) + (.74884*Z2_X3)
           + (.68759*Z2_X4) + (.60224*Z2_X5);
PROC UNIVARIATE;
   VAR Y X1 X2 X3 X4 X5;
   TITLE  'UNIVARIATE STATISTICS FOR GROUP 2';
PROC REG;
   MODEL Y = X1 X2 X3 X4 X5 / STB;
```

```
   TITLE1 'PEGRESSION OF GROUP 2';
   TITLE2 'FOR CROSS-VALIDATION';
PROC CORR;
   VAR Y X1 X2 X3 X4 X5;
   TITLE 'CORRELATION FOR GROUP 2';
DATA REGALL;
   SET GROUP1 GROUP2;
   PROC CORR;
      VAR Y YHAT_11 YHAT_12 YHAT_21 YHAT_22;
      TITLE 'INVARIANCE RESULTS';
   PROC PRINT;
     VAR ID GROUP Y X1 X2 X3 X4 X5 YHAT_11 YHAT_12 YHAT_21 YHAT_22;
```