DOCUMENT RESUME

ED 328 589 TM 016 103

AUTHOR Rowell, R. Kevin
TITLE Partitioning Predicted Variance into Constituent Parts: How To Conduct Commonality Analysis.
PUB DATE 24 Jan 91
NOTE 22p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 25-27, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; Life Satisfaction; Mathematical Models; Nursing Homes; Older Adults; *Predictive Measurement; *Regression (Statistics)
IDENTIFIERS *Commonality Analysis; *Variance (Statistical)

ABSTRACT

This paper explains how commonality analysis (CA) can be conducted using a specific Statistical Analysis System (SAS) procedure and some simple computations. CA is used in educational and social science research to partition the variance of a dependent variable into its constituent predicted parts. CA determines the proportion of explained variance that is unique to a predictor variable and the proportion that is common to two or more predictors. Whereas the ordering of the predictors using stepwise regression may lead to faulty data interpretations, CA is a method by which all possible predictor combinations are tested to determine the model that best explains predicted variance. Data from a study of life satisfaction (LS) among 198 elderly residents in 17 Texas nursing homes illustrate procedures for conducting CA with regression results. The subjects completed a LS questionnaire to determine if their self-reports of LS differed from those of the elderly living outside of nursing homes. Eight subscale components and the number of years in the nursing home were analyzed by regression to determine which variable best predicted nursing home satisfaction. Meaning was the dominant factor in predicting nursing home satisfaction and accounted for about 80% of all explained variance in the sample. In addition, a SAS computer program for obtaining all possible R-squared values is discussed as an efficient method of implementing the required analyses. CA offers a fairly straightforward method of analysis when no more than four independent variables are of interest. Three tables of data are presented, and the R-squares of LS scales are included. (SLD)

krsera.wp1 1/24/91

Partitioning Predicted Variance into Constituent Parts: How to Conduct Commonality Analysis

R. Kevin Rowell

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX, January 25, 1991.

Abstract

Commonality analysis is used in educational and social science research to partition the variance of a dependent variable into its constituent predicted parts. Commonality analysis determines the proportion of explained variance that is unique to a predictor variable and the proportion that is common to two or more predictors. Whereas the ordering of the predictors using stepwise regression may lead to faulty interpretation of the data, commonality analysis is a method by which all possible predictor combinations are tested to determine the model that best explains predicted variance. Data from a study of life satisfaction among nursing home residents are used to illustrate the procedures for conducting commonality analysis with regression results. In addition, a SAS computer program procedure for obtaining all possible R^2 values is discussed as an efficient method of implementing the required analyses. Four tables of data are presented.

Partitioning Predicted Variance into Constituent Parts: How to Conduct Commonality Analysis

Multiple regression analysis continues to be used more frequently by educational and social researchers as a means of describing the relationship among a given set of independent variables and in predicting the impact of these same variables on a dependent variable (Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985; Willson, 1982). It is often difficult to determine the "true" effects of the independent variables upon the dependent variables. More often than not, these independent variables are correlated, even substantially, and this increases the complexity of sifting through the data for accurate explanations of obtained results (Pedhazur, 1982).

To better understand the relative contribution of each independent variable, researchers may choose among a number of variance partitioning methods by which the squared multiple correlation (R^2) can be reduced into constituent portions that can be attributed to the independent variables. Among these methods, commonality analysis offers a useful method of partitioning variance because it does not depend upon a priori knowledge of the influence of the predictors. According to Cooley and Lohnes (1976), "such neutrality allows the information inherent in the data about the value of organizing observations in a certain framework (that of the domains of predictors) to emerge" (p. 219). Because commonality analysis views all possible orders of entry of the predictors into the model, there is no distortion of the results that may occur with stepwise regression analysis (Snyder, 1991), and essentially the predictors fall where they may. In support of this stance, Thompson, Smith, Miller, and Thomson (1991) contend that conventional stepwise regression analysis can lead to erroneous interpretations due to inflated Type I errors, the variables selected after k steps may not include all or even any of the variables in the best predictor set of size k, and the order of entry provides limited information regarding variable importance.

Seibold and McPhee (1979) explain that commonality analysis decomposes the squared multiple correlation into the proportion of the explained variance of the dependent variable associated with each independent variable and with the common effects of each. They also state that this decomposition of R^2 into its unique and common components is rarely conducted and argue that:

> Advancement of theory and the useful application of research findings depend not only on establishing that a relationship exists among predictors and the criterion, but also upon determining the extent to which those independent variables, singly and in all possible combinations, share variance with the dependent variable. Only then can we fully know the relative importance of independent variables with regard to the dependent variable in question. (p. 355)

The purpose of the present paper is not to argue the utility of commonality analysis as against other methods of analysis, since others have already presented these arguments (Cooley & Lohnes, 1976; Creager, 1971; Daniel, 1989; Mood, 1969; Seibold & McPhee, 1979; Thompson, 1985; Wisler, 1972). Instead, the paper explains how commonality analysis can be conducted using a specific SAS procedure and some simple computations. To make this discussion concrete, data involving life satisfaction among elderly nursing home residents are used for heuristic purposes.

Defining the Components of Commonality Analysis

The unique contribution of an independent variable can be specifically defined as the squared semipartial correlation between the dependent variable and the selected independent variable after all other independent variable components have been partialed out (Wisler, 1972). As an example, suppose that in a model with two independent variables, we are given U(1) and U(2) as the unique contribution of variables 1 and 2 respectively, $R^2_{Y.12}$ as the squared multiple correlation of Y with variables 1 and 2, $R^2_{Y.1}$ as the squared correlation of Y with variable 1, and $R^2_{Y.2}$ as the squared correlation of Y with variable 2. The unique contributions of variables 1 and 2 are:

$$U(1) = R^2_{Y.12} - R^2_{Y.2}$$

$$U(2) = R^2_{Y.12} - R^2_{Y.1}$$

The commonality of variable 1 and 2, i.e., the proportion of variance in Y predictable using either variable 1 or variable 2, can be written:

$$C(12) = R^2 - U(1) - U(2)$$

As a result, three variance components can be derived from the R^2 of the model, namely U(1), U(2), and C(12).

In general, the number of possible combinations of unique and commonality components is determined by $2^P - 1$ where P is the number of independent variables examined in the model. Since P independent factors are considered, the number of unique components equals P as well. The number of commonality components can then be derived as the difference between the total number of components and the number of unique components, or $(2^P - 1) - P$.

Since the number of possible unique and commonality components is exponentially determined, five or more independent variables of interest will render the analysis extremely burdensome. For example, with five variables, the total possible components are $2^5 - 1 = 31$, with 5 being unique components and 26 being commonality components. Thus, the *number* of components or variance partitions increases very rapidly as additional predictors are considered.

The rules for calculating the unique and commonality components are fairly straightforward algebraic product expansions of the independent variables; however, as the number of independent variables increases, the *complexity* of the respective component calculations also increases. Table 1 presents the necessary formulas for 2-, 3-, and 4-variable models. As illustrated in this table, deriving all required unique and commonality component values is somewhat tedious since all possible R^2 combinations are necessary for these calculations. For a more detailed explanation of the required calculations and their derivations, the reader can consult Mood (1969), Pedhazur (1982), or Wisler (1972). In addition, Seibold and McPhee (1979) offer the formulas for a 5-variable model.

Insert Table 1 about here.

When five or more independent variables are involved in the model, some have recommended alternative grouping of these variables into meaningful subsets through such methods as cluster analysis, factor analysis, or theoretical constructs (Mood, 1969; Seibold & McPhee, 1979; Wisler, 1972). One inherent problem however is that the primary reason for conducting commonality analysis is to make some sense of intercorrelated variables and to maintain neutrality in determining the most meaningful predictors. High intercorrelations may render grouping of these data into meaningful subsets impossible. An alternate solution, which was used in this study, is to limit the number of independent variables to four by initially selecting the best predictors through a series of preliminary analyses.

Commonality analysis requires every possible R^2 value for all variable combinations. SAS provides a useful program (PROC RSQUARE) that will print out in ascending order the R^2 values of all possible combinations of the independent variables in the model. This SAS routine makes commonality analysis much simpler, since the calculation of the required R^2 values is fully automated. Appendix A presents the SAS file used to execute the analysis for the present example.

The R^2's obtained from PROC RSQUARE are then subjected to the appropriate arithmetic computations suggested in Table 1. These can be rapidly done using a microcomputer spreadsheet program.

Next, once the variance components have been determined, the results can be placed in tabular summary that is easy to interpret and allows for a quick check of arithmetic (Pedhazur, 1982). Row entries are the specific unique and commonality effects of each independent variable. The column totals of each independent variable will equal to the R^2 of the regression model in which that independent variable is the only variable entered into the model. Another check is that the sum of all unique and commonality values should equal the R^2 value of the regression model when all the independent variables are entered into the model.

An Application of Commmonality Analysis

Data from a previous study involving the life satisfaction of nursing home residents can be employed to illustrate the steps in the process. In this study, 198 elderly nursing home residents in 17 Texas nursing homes completed a life satisfaction questionnaire to determine if their self-report of life satisfaction differed from that of the elderly living outside of nursing homes. In addition, eight subscale components and the number of years of stay in the nursing home were analyzed by regression to determine which of the variables best predicted nursing home satisfaction.

For purposes of discussion only, and not as part of the commonality analysis, a stepwise regression analysis was computed. Using a .15 level of significance for entry into the model, a forward stepwise regression analysis was conducted. These results are presented in Table 2. Only three of the nine independent variables were entered into the model used to predict satisfaction--past and present life having meaning (beta = .3407), need for social contact (beta = -.0389), and years in the nursing home (beta = .0250).

Insert Table 2 about here.

One's feeling that his or her life has been and continues to be meaningful was the best predictor of nursing home satisfaction. Additionally, the results suggest that the less one needs social contact and the longer one stays in the nursing home, the better is adjustment. The problem however is that this implies that these factors cannot be influenced by mental health interventions. Because of the high degree of correlation between the predictor variables, commonality analysis can be used to determine the unique and common components of these variables so that a more accurate explanation of satisfaction derivatives can be obtained.

The first step was to determine from the SAS printout of R^2 values which four variables best accounted for variance in the dependent measure of satisfaction. Inspection of these values revealed that the variable, planning for new goals, adds more to the model than do the remaining variables (R^2 increased .0032 while all others added only .0010 or less). This gives four independent variables of interest--meaning, need for social contact, years of stay in the nursing home, and planning for new goals. In this particular sample, meaning, social, and years were also the predictors selected by the stepwise procedure; however, this may not necessarily be the case for other studies. That is, stepwise results can be more anamolous than they were in the present study.

With these variables now having been selected, the second step is to obtain the 15 equations necessary for computing the unique and commonality components of a 4-variable model. These are obtained from Table 1.

The third step is to then extract all R^2 values from the SAS printout (Appendix A) and substitute these accordingly into the 15 equations. The computations can then be conducted using a standard calculator or computer routine. For example:

$$\begin{aligned} U1 \text{ (meaning)} &= -R^2(234) + R^2(1234) \\ &= -.13534 + .54451 \\ &= .40917 \end{aligned}$$

Therefore the unique contribution of the variable, meaning, to the proportion of total dependent variable variance explained is .40917, or approximately 41%. Also, as an example, the commonality between years (1) and meaning (2) is calculated as:

$$\begin{aligned} C14 &= -R^2(23) + R^2(123) + R^2(234) - R^2(1234) \\ &= -.11929 + .13534 + .54022 - .54451 \\ &= .01176 \end{aligned}$$

Therefore, the common variance of the model shared by meaning and years is .01176, or approximately 1.2%.

The fourth step is to arrange these obtained values into a commonality analysis table, like the one presented in Table 3. Once in tabular form, the previously mentioned arithmetic checks can be performed. For instance, summing down column 4 (meaning) results in an value of .52985, which is the R^2 value for the regression model when only the variable, meaning, is entered. Additionally, the sum of all 15 unique and commonality components equals .54451, which is the R^2 value for the regression model when all four independent variables have been entered into the model.

Insert Table 3 about here.

Discussion

The commonality summary table presented in Table 3 indicates that the unique predicted variance contribution of the predictor, meaning, is approximately 41% (.40917) and its total commonality variance with one or more of the other predictors is approximately 12% (.12068). In this particular example, the variable, meaning, is the dominant factor in predicting nursing home satisfaction and alone accounts for about 80% of all explained variance in the sample. The other variables offer little unique contribution to the variance. In fact, the other variables (goals, social, and years) have greater commonality components than their respective unique components. Because meaning, social, and years are factors which are based on life experience, and are not "mutable" conditions of the nursing home environment, providing new meaningful goals for nursing home residents would not significantly enhance their life satisfaction, according to the results of this commonality analysis.

It should be noted that some instances negative commonalities may occur, as in this particular example with C12 and C134, as reported in Table 3. As Thompson (1985) explains, this result can be "counterintuitive since the result could be taken to mean that ... predictor variables have in common the ability to explain less than 0% of the variance" (p. 54). But the presence of negative commonalities is typically attributable to suppressor effects and is more likely to occur with higher order partitions, as in this case (Beaton, 1973; Creager, 1971; DeVito, 1976).

Commonality analysis is but one method of partitioning variance in regression analysis of educational and social models. It offers a fairly straightforward method of analysis when no more than four independent variables are of interest, and with the assistance of the SAS PROC RSQUARE routine, the most difficult aspect of commonality analysis can be greatly simplified. As such, commonality analysis can be readily employed in research. This analysis can be very useful as a supplement to conventional regression analysis.

References

Beaton, A. E. (1973). Commonality. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction No. ED 111 829)

Cooley, W. W., & Lohnes, P. R. (1976). Evaluation research in Education. New York: Irvington.

Creager, J. A. (1971). Orthogonal and nonorthogonal methods of partitioning regression variance. American Educational Research Journal, 8, 671-676.

Daniel, L. G. (1989). Commonality analysis with multivariate data sets. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction No. ED 314 483)

DeVito, P. J. (1976). The use of commonality analysis in educational research. Paper presented at the annual meeting of the New England Educational Research Association, Provincetown, MA. (ERIC Document Reproduction No. ED 146 218)

Elmore, P., & Woehlke, P. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17, 19-20.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14, 5-11.

Mood, A. R. (1969). Macro-analysis of the American educational system. Operations Research, 17, 770-784.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart, and Winston.

Seibold, D. R., & McPhee, R. D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analyses. Human Communication Research, 5, 355-365.

Snyder, P. (1991). Three reasons why stepwise regression methods should not be used by researchers. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 99-106). Greenwich, CT: JAI Press.

Thompson, B. (1985). Alternate methods for analyzing data from education experiments. Journal of Experimental Education, 54, 50-55.

Thompson, B., Smith, Q. W. Smith, Miller, L. M., & Thomson, W. A. (1991, January). Methods lead to bad interpretations: Better alternatives. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.

Willson, V. (1982, January). Misuses of regression approaches to ANOVA and ANCOVA. Paper presented at the annual meeting of the Southwestern Educational Research Association, Austin, TX.

Wisler, C. E. (1969). Partitioning the variance explained in a regression analysis. In G. Mayeske, C. Wisler, A. Beaton, Jr., F. Weinfield, W. Cohen, T. Okada, J. Proshek, & K. Tabler, A Study of Our Nations Schools. Washington, DC: U.S. Government Printing Office.

Table 1
Formulas for Unique and Commonality Components of Variance

Two Independent Variables

$$U1 = -R^2(2) + R^2(12)$$
$$U2 = -R^2(1) + R^2(12)$$
$$C12 = R^2(1) + R^2(2) - R^2(12)$$

Three Independent Variables

$$U1 = -R^2(23) + R^2(123)$$
$$U2 = -R^2(13) + R^2(123)$$
$$U3 = -R^2(12) + R^2(123)$$
$$C12 = -R^2(3) + R^2(13) + R^2(23) - R^2(123)$$
$$C13 = -R^2(2) + R^2(12) + R^2(23) - R^2(123)$$
$$C23 = -R^2(1) + R^2(12) + R^2(13) - R^2(123)$$
$$C123 = R^2(1) + R^2(2) + R^2(3) - R^2(12) - R^2(13) - R^2(23) + R^2(123)$$

Four Independent Variables

$$U1 = -R^2(234) + R^2(1234)$$
$$U2 = -R^2(134) + R^2(1234)$$
$$U3 = -R^2(124) + R^2(1234)$$
$$U4 = -R^2(123) + R^2(1234)$$
$$C12 = -R^2(34) + R^2(134) + R^2(234) - R^2(1234)$$
$$C13 = -R^2(24) + R^2(124) + R^2(234) - R^2(1234)$$
$$C14 = -R^2(23) + R^2(123) + R^2(234) - R^2(1234)$$
$$C23 = -R^2(14) + R^2(124) + R^2(134) - R^2(1234)$$
$$C24 = -R^2(13) + R^2(123) + R^2(134) - R^2(1234)$$
$$C34 = -R^2(12) + R^2(123) + R^2(124) - R^2(1234)$$

$$C123 = -R^2(4) + R^2(14) + R^2(24) + R^2(34) - R^2(124) - R^2(134) - R^2(234) + R^2(1234)$$

$$C124 = -R^2(3) + R^2(13) + R^2(23) + R^2(34) - R^2(123) - R^2(134) - R^2(234) + R^2(1234)$$

$$C134 = -R^2(2) + R^2(12) + R^2(23) + R^2(24) - R^2(123) - R^2(124) - R^2(234) + R^2(1234)$$

$$C234 = -R^2(1) + R^2(12) + R^2(13) + R^2(14) - R^2(123) - R^2(124) - R^2(134) + R^2(1234)$$

$$C1234 = R^2(1) + R^2(2) + R^2(3) + R^2(4) - R^2(12) - R^2(13) - R^2(14) - R^2(23) - R^2(24) - R^2(34) + R^2(123) + R^2(124) + R^2(134) + R^2(234) - R^2(1234)$$

Table 2
STEPWISE REG OF LIFE SAT SUBSCALES TO NH_SAT
LEVEL OF SIGNIFICANCE = .15 FOR ENTRY

STEPWISE REGRESSION PROCEDURE FOR DEPENDENT VARIABLE NH_SAT

STEP 1 VARIABLE MEANING ENTERED R SQUARE = 0.52985330 C(P) = 5.20979602

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	241.21705263	241.21705263	220.89	0.0001
ERROR	196	214.03547262	1.09201772		
TOTAL	197	455.25252525			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-0.46113731				
MEANING	0.32772707	0.02205074	241.21705263	220.89	0.0001

BOUNDS ON CONDITION NUMBER: 1, 1

STEP 2 VARIABLE SOCIAL ENTERED R SQUARE = 0.53612074 C(P) = 4.55416484

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	244.07032233	122.03516116	112.68	0.0001
ERROR	195	211.18220292	1.08298566		
TOTAL	197	455.25252525			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-0.03327342				
MEANING	0.34250707	0.02377241	224.80978765	207.58	0.0001
SOCIAL	-0.03775607	0.02326091	2.85326970	2.63	0.1062

BOUNDS ON CONDITION NUMBER: 1.171945, 4.68778

(Table 2 cont.)

STEP 3 VARIABLE YEARS ENTERED R SQUARE = 0.54129874 C(P) = 4.36015322

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	246.42761762	82.14253921	76.31	0.0001
ERROR	194	208.82490763	1.07641705		
TOTAL	197	455.25252525			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-0.08562938				
MEANING	0.34075652	0.02372971	221.96469121	206.21	0.0001
SOCIAL	-0.03891501	0.02320348	3.02766996	2.81	0.0951
YEARS	0.02498929	0.01688641	2.35729529	2.19	0.1405

BOUNDS ON CONDITION NUMBER: 1.174865, 10.06175

Table 3
Commonality Analysis Summary Table

Component	1 Years	2 Social	3 Goals	4 Meaning
U1	.00429			
U2		.00587		
U3			.00321	
U4				.40917
C12	-.00041	-.00041		
C13	.00089		.00089	
C14	.01176			.01176
C23		.00078	.00078	
C24		.00759		.00759
C34			.07840	.07840
C123	.00003	.00003	.00003	
C124	.00274	.00274		.00274
C134	-.00552		-.00552	-.00552
C234		.02546	.02546	.02546
C1234	.00025	.00025	.00025	.00025
Total	.01403	.04231	.10350	.52985
U	.00429	.00587	.00321	.40917
C	.00974	.03644	.10029	.12068

APPENDIX A

RSQUARES OF LIFE SAT SCALES TO NH_SAT

N=198 REGRESSION MODELS FOR DEPENDENT VARIABLE: NH_SAT
MODEL: MODEL1

NUMBER IN MODEL	R-SQUARE	VARIABLES IN MODEL
1	0.01403466	YEARS
1	0.04230736	SOCIAL
1	0.10350445	GOALS
1	0.52985330	MEANING
2	0.05373485	YEARS SOCIAL
2	0.11929491	GOALS SOCIAL
2	0.12188332	GOALS YEARS
2	0.53464821	MEANING YEARS
2	0.53475748	MEANING GOALS
2	0.53612074	MEANING SOCIAL
3	0.13533964	GOALS YEARS SOCIAL
3	0.53864125	MEANING GOALS YEARS
3	0.54021848	MEANING GOALS SOCIAL
3	0.54129874	MEANING YEARS SOCIAL
4	0.54450878	MEANING GOALS YEARS SOCIAL