DOCUMENT RESUME

ED 111 828 TM 004 299

AUTHOR Mayeske, George W.; Beaton, Albert E., Jr.
TITLE Reverse Commonality and Some Illustrative Applications.
INSTITUTION Office of Education (DHEW), Washington, D.C. Office of Planning, Budgeting, and Evaluation.
PUB DATE Aug 74
NOTE 45p.

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS Algorithms; Computer Programs; *Correlation; Mathematical Applications; *Matrices; Measurement Techniques; *Multiple Regression Analysis; Predictor Variables; Statistical Analysis; *Transformations (Mathematics)
IDENTIFIERS *Commonality; Reverse Commonality

ABSTRACT

The results of an algorithm which is designed to take a set of commonality coefficients, either real or manipulated, and, if possible, produce one or more sets of regressor correlations that are consistent with them are examined. A number of different ways of resolving the higher order commonality values into their lower orders were tried and the number and nature of solutions generated from them were examined for their meaningfulness and variability. In general it was found that this could be a meaningful exercise because it allows an analyst to test his assumptions about the nature of the confounding to see if a solution can be obtained. If one cannot be obtained then his assumptions must be revised. But if one can be obtained, then he can examine them to see which kinds of variables have the greatest sensitivity to the assumptions. The resulting output also allows one to gauge the variability of regression coefficients that will satisfy the same set of commonality values and the effectiveness of the regression system. This technique called "Reverse Commonality" is best suited for an interactive computing arrangement so that an analyst can use it rapidly in a sequential manner. Since, for a large number of variables the sheer volume of commonality values become unmanageable and, since a number of different ways are available for resolving the confounding, an algorithm is needed which uniquely resolves the higher order values according to one's assumptions. (Author/DEP)

TM

REVERSE COMMONALITY AND SOME ILLUSTRATIVE APPLICATIONS

by

George W. Mayeske
Albert E. Beaton, Jr.

Office of Planning, Budgeting and Evaluation

U.S. Office of Education

August 1974

The results of a "Commonality Analysis" yield measures of the extent to which each variable contributes to the explanation of variance in the dependent variable independently of the other variables in the analysis. Such an analysis also yields measures of the extent to which two or more of the variables share some of their explanatory power. This shared explanatory power may arise from the functional interplay of the variables over time and/or merely from their joint occurrence. For example, one might observe that children from more affluent families tend to attend schools that are better staffed and tend to do better on standardized achievement tests than their less affluent counterparts. Such observations would produce a degree of correlation among these three classes of variables so that the effects of affluence and school type on achievement would be difficult to disentangle. However, for heuristic reasons an analyst might want to attempt to disentangle their "possible effects". This paper includes one such technique and some illustrative applications.

We have called this technique "Reverse Commonality" because it takes as given a set of commonality coefficients, from them reproduces all possible squared multiple correlations (RSQ's) and then attempts to produce one or more sets of regressor (independent) variable inter-correlations that are consistent with these RSQ's. When the commonality coefficients have been obtained from real data, one admissible set of regressor intercorrelations will be obtained which corresponds, within rounding error, to the actual or real data intercorrelations. When manipulated coefficients are read in to the program, the computer will inform the analyst as to the number of admissible solutions. If any exist the algorithm will print out the regressor intercorrelations, their standardized regression weights and the variance explained by the regression.

Before proceeding with the details of the technique we shall develop the discussion around a three variable commonality analysis. Later, greater numbers of variables will be dealt with.

Beaton (1973) has synthesized the work of other investigators on commonality and has extended and clarified its properties. Consequently, we shall deal only briefly here with its nature for the three variable case. Given three regressor variables X_1, X_2 and X_3, their relationship with a dependent variable Y can be expressed in terms of all of their possible squared simple and multiple correlations with Y as follows: RSQ (X_1); RSQ (X_2); RSQ (X_3); RSQ (X_1X_2); RSQ (X_1X_3); RSQ (X_2X_3); RSQ $(X_1X_2X_3)$. For m variables there are $2^m - 1$ correlations that can be obtained. Commonality analysis merely forms sums or differences of these RSQ's so that the Y variance explained by any X_i can be expressed in terms of the proportion it has in common with the other X's plus what it brings uniquely to the analysis.* These latter unique terms can be obtained as follows:

*Beaton (1973) has a matrix formulation of this computational rationale for the general case.

(1) $U(X_1) = RSQ\ (X_1X_2X_3) - RSQ\ (X_2X_3)$

(2) $U(X_2) = RSQ\ (X_1X_2X_3) - RSQ\ (X_1X_3)$

(3) $U(X_3) = RSQ\ (X_1X_2X_3) - RSQ\ (X_1X_2)$

The second order commonality coefficients are obtained by:

(4) $C(X_1X_2) = RSQ\ (X_1X_2X_3) - RSQ\ (X_3) - U(X_1) - U(X_2)$

(5) $C(X_1X_3) = RSQ\ (X_1X_2X_3) - RSQ\ (X_2) - U(X_1) - U(X_3)$

(6) $C(X_2X_3) = RSQ\ (X_1X_2X_3) - RSQ\ (X_1) - U(X_2) - U(X_3)$

while the third order coefficient is obtained by:

(7) $C(X_1X_2X_3) = RSQ\ (X_1X_2X_3) - C(X_1X_2) - C(X_1X_3) - C(X_2X_3)$
$- U(X_1) - U(X_2) - U(X_3)$

By virtue of this foregoing, the RSQ's for X_1, X_2 and X_3 can each be expressed as a sum of their commonality coefficients as follows:

(8) $RSQ(X_1) = C(X_1X_2X_3) + C(X_1X_2) + C(X_1X_3) + U(X_1)$

(9) $RSQ(X_2) = C(X_1X_2X_3) + C(X_1X_2) + C(X_2X_3) + U(X_2)$

(10) $RSQ(X_3) = C(X_1X_2X_3) + C(X_1X_3) + C(X_2X_3) + U(X_3)$

and the total RSQ is given by:

(11) $RSQ\ (X_1X_2X_3) = C(X_1X_2X_3) + C(X_1X_2) + C(X_1X_3) + C(X_2X_3)$
$+ U(X_1) + U(X_2) + U(X_3)$

Arranging these in tabular form we have

$$\begin{array}{l} \\ U(X_i) \\ C(X_1X_2) \\ C(X_1X_3) \\ C(X_2X_3) \\ C(X_1X_2X_3) \end{array} \left[\begin{array}{ccc} \underline{1} & \underline{2} & \underline{3} \\ a & b & c \\ d & d & - \\ e & - & e \\ - & f & f \\ g & g & g \end{array}\right]$$

where the alphabetic entries represent empirically observed values and the values in each column sum to the RSQ for the variable number heading the column. For example, RSQ $(X_1) = a + d + e + g$ (note that the empirically observed value for a coefficient is repeated under its respective column as appropriate). These empirically observed values can be manipulated in a number of different ways and then submitted to the algorithm developed by Beaton and described in Appendix A, to see if they yield a consistent set of correlations. The remainder of this paper deals with the results of the systematic application of this Reverse Commonality procedure.

1. Three Variable Reverse Commonality for a Fixed Level of Explained Variation

The results of a three variable commonality analysis are given in Table 1.1. The variables used here are composites taken from Mayeske, et. al. (1974). The first variable is a weighted* sum of a student's values on his Socio-Economic Status and Family Structure indices as

* The weights are taken from the first Principal Component of the variables under consideration.

well as a value for his Ethnic Group* and Boy-girl group** membership. The second variable is a weighted sum of his scores on four attitudinal and motivational indices*** while the third variable is the average achievement of students in the grade level that one attends school with. These three variables are called respectively: Social Background (SB); Motivational (MTVTN); and School (SCH) factors. The dependent variable, Y, is each student's score on an Achievement composite.†

* Whites were scored highest, Oriental-Americans intermediate, while all others were scored low.

**Girls were scored highest.

***These are: Expectations for Excellence; Attitude Toward Life; Educational Plans and Desires; and, Study Habits. For details see Mayeske, et. al., 1972.

†See Mayeske, et. al., 1972 for more details on the nature of this composite.

Table 1.1. - Commonality Analysis of Achievement With Social Background, Motivational and School Factors*

	1 SB	2 MTVTN	3 SCH
$U(X_i)$	.0236	.0307	.1149
$C(X_1X_2)$	.0654	.0654	
$C(X_1X_3)$	.1025		.1025
$C(X_2X_3)$		-.0071	-.0071
$C(X_1X_2X_3)$	.1117	.1117	.1117
$RSQ(X_i)$	.3032	.2008	.3220
$RSQ(X_1X_2X_3)$		.4417	

* There are 123,305 students and their 2370 schools included in these analyses.

Given this table of commonality coefficients we may ask: "How do the regressor intercorrelations produced from these coefficients compare with those actually observed"? The observed (0) values are given in the first row of Table 1.2 with their reproduced values (R) directly below them.

Table 1.2. - Observed (O) and Reproduced (R) Regressor Intercorrelations for SB, MTVTN and SCH and their Correlations With Achievement

		SB	MTVTN	SCH	ACHIEVEMENT
Social Background (SB)	O	1	.5879	.5219	.5506
	R	1	.5879	.5219	.5506
Motivation (MTVTN)	O		1	.2623	.4481
	R		1	.2623	.4480
School Factors (SCH)	O			1	.5674
	R			1	.5675

Inspection of the values shows them to be identical to the third decimal digit in every case and to the fourth decimal digit in almost every case with the errors occurring in the regressor-regressand correlations (i.e., Column 4). Clearly, then for the case of these three regressors, the algorithun reproduces the correlations to a satisfactory degree of accuracy.

Next we may ask: "If the coefficients in Table 1.1 are manipulated in some simple manner, might they yield a consistent set of regressor correlations?" The simplest change that suggests itself is to get rid of the negative value. To do this we shall assume that it arose from the variable called MTVTN and its relationship with the other variables.

By adding the value of -.0071 to the second order coefficient for SB, MTVTN (viz., $C(X_1X_2)$) thereby reducing it to .0583, and setting the $C(X_2X_3)$ value to zero, we obtain a simplified set of coefficients. The application of the algorithm showed that a single consistent set of intercorrelations could be produced which differed very little from the observed values. They are compared in Table 1.3.

Table 1.3. Observed (O) Correlations Compared With Those Reproduced (R) From Manipulated Commonalities

Social Background (SB)	O	1	.5879	.5219	.5506
	R	1	.5556	.5235	.5442
Motivation (MTVTN)	O		1	.2623	.4481
	R		1	.2820	.4480
School Factors (SCH)	O			1	.5674
	R			1	.5737

The application of this algorithm raises two main concerns: (1) how to reduce the sheer volume of input and output that we have to assimilate when there is more than one admissible solution (after all we don't want to have to resort to a new table comparing the O's and the R's each time); and, (2) how to manipulate the commonality coefficients in a systematic way. Let us deal with each of these in turn.

To make the volume of input more comprehensible and to render the volume of output more manageable we shall arrange the values for our commonality coefficients in columnar form as follows:

OBSERVED VALUES TRANSFORMED

	Observed	T_1	T_2	T_3	....	T_n
$U(X_1)$						
$U(X_2)$						
$U(X_3)$						
$C(X_1X_2)$						
$C(X_1X_3)$						
$C(X_2X_3)$						
$C(X_1X_2X_3)$						
$RSQ(X_1X_2X_3)$						
Number of Admissible Solutions						

The first column will contain the observed values while the succeeding columns will contain various transformations or manipulations of these values. There are two ways to perform these transformations. The way that we have selected might be called "sequential resolution". By this we mean that the higher order commonality values are split up and pushed into the next lower order each time. For each such transformation the algorithm indicates the number of admissible solutions*.

*The number of admissible solutions is affected by the tolerance limits one sets (see Appendix A). Generally, the more lax this criterion is, the greater are the number of admissible solutions. Almost without exception, we have used a limit of .005 or less (see Appendix A).

A second way which might be called "bypass resolution", splits up the higher order values and relegates them directly to the uniques (U(Xi)'s) without passing them through the intermediate orders. For large numbers of variables (five or more) these two approaches may not yield the same resolutions (viz., once all the values have been relegated to the uniqueness, the unique values may differ for the two techniques). We have chosen the former rather than the latter because we have often found the results of each step in a sequence to be informative.*

What kinds of theory then might we use as a guide to our sequential resolutions. The one that most readily comes to mind is a "state of ignorance" theory. Such a theory says that we don't know how best to allocate the higher order values and therefore that we shall split them up so that each variable gets an equal share. In Table 1.4 **the transformations under the columns numbered 2 and 3 represent these resolutions (we have already examined results for the observed commonalities and those for the first transformation). Our procedure for these was as follows:

---

* We do not know which one of these techniques is optimal.

** In order to facilitate the paper's readability the reader is advised to unstaple the pages and pull out the tables that are being referred to.

"State of Ignorance" Transformations

Take the simplified values from the first transformation (we shall use the simplified commonalities for these and all succeeding transformations):

T2 - Split $C(X_1X_2X_3)$ into 1/3's and put 1/3 on each of $C(X_1X_2)$, $C(X_1X_3)$ and $C(X_2X_3)$

T3 - Split each of the $C(X_iX_j)$'s (as modified in T2) in half and put 1/2 in each of its relevant uniques (e.g., $\frac{1}{2}$ $C(X_1X_2)$ is added to $U(X_1)$ and the other half to $U(X_2)$, perform similar operations for $C(X_1X_3)$ with respect to $U(X_1)$ and $U(X_3)$; and for $C(X_2X_3)$ with respect to $U(X_2)$ and $U(X_3)$).

We can see from inspection of Table 1.4 that T2 did not yield any admissible solutions (probably because of the large $C(X_2X_3)$ value produced) whereas T3 yielded five. The correlations obtained from these five solutions are given in Table 1.5 along with the intercorrelations for the first transformation and those obtained from the observed values.* For the first solution, (3A), the regressor intercorrelations are all zero. Each regressor correlation takes on a very high value in turn for the next three solutions while the others remain low whereas, for the last solution, all the values are high. Examination of the regression coefficients given in Table 1.6 show that they vary in magnitude enormously over the five solutions. For example, variable 1 (Social Background) takes on a high value of 2.03 and a low of - 11.25.

* We may recall that the regressor-regressand correlations are fixed for any single resolution (or transformation).

Variable 2 (Motivation) ranges from 4.76 down to -1.68 while Variable 3 (School factors) remains positive all the time but ranges from a low of .45 to a high of 7.06. Clearly then for regressor correlations that yield solutions consistent with the fixed regressor-regressand relationships given in Table 1.5 (those are r14, r24 & r34), their regression coefficients may differ over an enormously, for a fixed RSQ $(X_1X_2X_3)$ of .4417. Our "state of ignorance" theory then allows a great deal of variation in the admissible solutions.

Theory Guided Transformations.

Let us consider next how a few content oriented theories might help us resolve these values. They are:

T4 - All the variance that is accounted for by SB (Social Background) is also caused by it. Therefore, take every higher order coefficient that has a one in its subscript and add it's value to the unique value for SB (viz. add them to $U(X_1)$).

T5 - All the variance that is accounted for by MTVTN (Motivational factors) is also caused by it. Therefore, take every higher order coefficient that has a 2 in its subscript and add its value to the unique value for MTVTN (viz. add them to $U(X_2)$).

T6 - All the variance that is accounted for by SCH (School factors) is also caused by it. Therefore, take every higher order coefficient that has a 3 in its subscript and add its value to the unique value for SCH (viz.add them to $U(X_3)$).

Inspection of Table 1.4 shows that these transformations have 5, 2 and 2 admissible solutions respectively. Let us examine their correlations in Table 1.5 and their regression coefficients in Table 1.6.

Examination of Table 1.5 shows that the solutions for T4 (our everything associated with Social Background belongs to Social Background theory) vary from zero through .58 for r12, and to about .9 for r13 and r23. Clearly for these regressor-regressand relationships the correlations between SCH (School factors) and other factors (either SB or MTVTN) can vary to a greater extent than can those between SB and MTVTN. For T5 (our everything associated with MTVTN belongs to MTVTN theory) the correlations between SB and MTVTN and between MTVTN and SCH stay null while that between SB and SCH varies from a low of .47 to a high of about .9. Not only are these solutions fewer in number but they are also more compact (i.e. less variable). For T6 (our everything that is associated with SCH belongs to SCH theory) the two solutions are also compact with the SCH related correlations staying null while those for SB and MTVTN range from about .52 almost up to 1.0.

In examining the behavior of the regression coefficients for these solutions, given in Table 1.6, we can note for T4 that they range from: a low of about .5 to a high of about 1.5 for SB; a high of about .3 for MTVTN to a low of about -.3; and, from a low of -1.3 for SCH to a high of about .6. Clearly then the variability is greater over this set of solutions for SCH than for any of the others. For T5 the variability is least for MTVTN and greatest for SB. For T6 the variability is null for SCH but very large for SB and MTVTN. What these results along with those in Tables 1.4 and 1.5 suggest is that the variable or variables which have

Table 1.4 Transformed Commonality Values and Their Number of Admissible Solutions

Commonality Coefficients*	Observed	TRANSFORMATIONS 1	2	3	4	5	6
$U(X_1)$	.0236	.0236	.0236	.1412	.2961	.0236	.0236
$U(X_2)$	.0307	.0307	.0307	.0971	.0307	.2007	.0307
$U(X_3)$	.1149	.1149	.1149	.2034	.1149	.1149	.3291
$C(X_1X_2)$	.0654	.0583	.0955	0	0	0	.0583
$C(X_1X_3)$	.1025	.1025	.1397	0	0	.1025	0
$C(X_2X_3)$	-.0071	0	.0372	0	0	0	0
$C(X_1X_2X_3)$	.1117	.1117	0	0	0	0	0
$RSQ(X_1X_2X_3)$	.4417	.4417	.4416	.4417	.4417	.4417	.4417
NUMBER OF SOLUTIONS	1	1	NONE	5	5	2	2

Commonality Coefficients*	TRANSFORMATIONS 7	8	9
$U(X_1)$	.2111	.0236	.0731
$U(X_2)$	.1157	.0307	.0546
$U(X_3)$	.1149	.1149	.1498
$C(X_1X_2)$	0	.0769	.0385
$C(X_1X_3)$	0	.1211	.0606
$C(X_2X_3)$	0	.0186	.0093
$C(X_1X_2X_3)$	0	.0558	.0558
$RSQ(X_1X_2X_3)$	.4417	.4416	.4417
NUMBER OF SOLUTIONS	3	NONE	NONE

*VARIABLES
1 - Social Background (SB)
2 - Motivation (MTVTN)
3 - School (SCH)

Table 1.5 Regressor and Regressand Correlations Reproduced from Commonalities

TRANSFORMATIONS

r*	Observed	1	3A	3B	3C	3D	3E
12	.5879	.5556	0	.9827	0	0	.9827
13	.5219	.5235	0	0	.9836	0	.9836
14	.5506	.5442	.3758	.3758	.3758	.3758	.3758
23	.2623	.2820	0	0	0	.9353	.9353
24	.4480	.4480	.3116	.3116	.3116	.3116	.3116
34	.5675	.5737	.4510	.4510	.4510	.4510	.4510

r	4A	4B	4C	4D	4E	5A	5B
12	0	.5835	0	0	.5835	0	0
13	0	0	.8976	0	.8976	.4709	.9031
14	.5442	.5442	.5442	.5442	.5442	.3551	.3551
23	0	0	0	.8158	.8158	0	0
24	.1752	.1752	.1752	.1752	.1752	.4480	.4480
34	.3390	.3390	.3390	.3390	.3390	.4663	.4663

r	6A	6B	7A	7B	7C
12	.5192	.9973	0	.9564	0
13	0	0	0	0	.9555
14	.2862	.2862	.4595	.4595	.4595
23	0	0	0	0	0
24	.2983	.2983	.3402	.3402	.3402
34	.5737	.5737	.3389	.3389	.3389

*VARIABLES

1 – Social Background (SB)
2 – Motivation (MTVTN)
3 – School (SCH)

Table 1.6 Regression Coefficients Computed from Reproduced Regressor-Regressand Correlations

TRANSFORMATIONS

β*	Observed	1	3A	3B	3C	3D	3E
1	.2152	.2146	.3758	2.0305	-2.0818	.3758	-11.2492
2	.2170	.2158	.3116	-1.6838	.3116	-.8809	4.7600
3	.3982	.4005	.4510	.4510	2.4986	1.2749	7.0633

β	4A	4B	4C	4D	4E	5A	5B
1	.5442	.6700	1.234	.5442	1.5198	.1741	-.3577
2	.1752	-.2158	.1752	-.3030	.3731	.4480	.4480
3	.3390	.3390	-.7688	.5862	-1.3295	.3842	.7893

β	6A	6B	7A	7B	7C
1	.1797	-2.082	.4595	1.5739	1.5569
2	.2050	2.375	.3402	-1.1652	.3402
3	.5737	.5737	.3389	.3389	-1.1487

*VARIABLES

1 - Social Background (SB)
2 - Motivation (MTVTN)
3 - School (SCH)

the largest unique values after resolution tend to show less variability in their regression coefficients than those with lower unique values--a not surprising result.

Mixed Transformations

It may also be meaningful to perform mixed transformations - that is - transformations that are guided by a theory but at points where the theory does not provide guidance a random rationale can be adopted. Some of these are:

T7 - The SCH factors should have only the value of their uniqueness ($U(X_3)$) but we do not know how the values for the higher order coefficients should be spread among SB & MTVTN. Therefore, split $C(X_1X_2X_3)$ into 1/3 and add this value to each of the second order coefficients (viz. $C(X_iX_j)$). Then add to $U(X_1)$ half of the new $C(X_1X_2)$ and all of the new $C(X_1X_3)$. Similarly add to $U(X_2)$ half of the new $C(X_1X_2)$ and all of the new $C(X_2X_3)$.

T8 - This approach postulates that only half of the variance we observe in the third order can be resolved into the second orders, whereas the rest should remain where it is. Therefore, take half of $C(X_1X_2X_3)$, split this value into 1/3's (i.e. a value of 1/6 of $C(X_1X_2X_3)$) and add this much to each of the $C(X_iX_j)$'s.

T9 - This approach takes one half of the newly created second order values from T8, splits them in one-half (i.e. 1/4 of $C(X_1X_2)$, 1/4 of $C(X_1X_3)$etc.) and adds each of these to its appropriate uniqueness. That is, $U(X_1)$ gets 1/4 of $C(X_1X_2)$ and 1/4 of $C(X_1X_3)$; $U(X_2)$ gets 1/4 of $C(X_1X_3)$ and 1/4 of $C(X_2X_3)$, etc.

Inspection of Table 1.4 shows that T7 yielded three admissible solutions whereas T8 and T9 did not yield any. These results along with those for T2 suggest that theories about only part of the shared variance are not as likely to yield admissible solutions.

Let us examine the nature of the solutions to T7 in Tables 1.5 and 1.6. The correlations in Table 1.5 show that the values for SB and MTVTN and SB and SCH range from zero to about .95 but never simutaneously, while the correlation between MTVTN and SCH stays at zero. The regression coefficients for the T7 solutions, in Table 1.6, vary from a low of about .45 to a high of about 1.6 for SB, whereas those for MTVTN vary downward from .34 to -1.16 and those for SCH vary from about .34 down to -1.15. However, the coefficients for MTVTN take on a slightly greater range than that of SCH while both of those take on a greater range than those of SB.

In this section we have examined a number of different ways that might be used to resolve higher order commonality values into their lower orders for a fixed level of variance explained (viz. a fixed value of the squared multiple correlation). In the next section we examine results obtained from varying the level of explained variance.

2. Three Variable Reverse Commonality for Varying Levels of Explained Variation

The algorithm's results may show that for a fixed level of variance explained (i.e. in our case a fixed value of RSQ $(X_1X_2X_3)$) an analysts' particular resolution does not yield any admissible solutions. As a consequence, the analyst may want to explore neighboring levels to see if they yield solutions. In this section we examine results in which the level of explained variance is systematically varied.

These computations were performed by taking the simplified commonalities (called T1 in Table 1.4), dividing them by their RSQ ($X_1X_2X_3$) value of .4417 so that they sum to 1.00, and then scaling these values by multiplying them by: .1; .2; .3; .4; .5; .6; .7; .8; .9; 1.0 respectively; these values are given in Table 2.1. We can observe that each of these yielded one and only one admissible solution. Examination of the regressor intercorrelations showed that they stayed the same for each scaling condition (these values are given in Table 1.5 under column T1). What did change were the regressor-regressand correlations for the different scaling conditions. These values, given in Table 2.2, show a progressive increase as the scaling factor approached 1.0. The regression coefficients, given in Table 2.3, also increased progressively in magnitude as the scaling factor increased. Clearly then for the commonality proportions fixed on here, a wide range of RSQ ($X_1X_2X_3$)'s will yield admissible solutions.

Table 2.1 Fixed Commonality Proportions and Their Number of Admissible Solutions for Increasing Levels of Variance Explained

COMMONALITY COEFFICIENTS*	OBSERVED SIMPLIFIED	TRANSFORMATIONS .1	.2	.3	.4	.5	.6
$U(X_1)$	.0236	.0053	.0107	.0160	.0214	.0267	.0321
$U(X_2)$	.0307	.0069	.0139	.0208	.0278	.0348	.0417
$U(X_3)$	.1149	.0260	.0520	.0780	.1040	.1301	.1561
$C(X_1X_2)$	.0583	.0132	.0264	.0396	.0528	.0660	.0792
$C(X_1X_3)$	.1025	.0232	.0464	.0696	.0928	.1161	.1393
$C(X_2X_3)$	0	0	0	0	0	0	0
$C(X_1X_2X_3)$	.1117	.0253	.0506	.0759	.1011	.1265	.1517
$RSQ(X_1X_2X_3)$	.4417	.0999	.2000	.2999	.3999	.5000	.6001
NUMBER OF SOLUTIONS	1	1	1	1	1	1	1

COMMONALITY COEFFICIENTS*	.7	.8	.9	1.0
$U(X_1)$	.0374	.0427	.0481	.0534
$U(X_2)$	.0486	.0556	.0626	.0695
$U(X_3)$	.1821	.2081	.2341	.2601
$C(X_1X_2)$	.0924	.1056	.1188	.1320
$C(X_1X_2)$	.1625	.1857	.2089	.2321
$C(X_2X_3)$	0	0	0	0
$C(X_1X_2X_3)$	.1770	.2023	.2276	.2529
$RSQ(X_1X_2X_3)$	.7000	.8000	.9001	1.0000
NUMBER OF SOLUTIONS	1	1	1	1

*VARIABLES

1 - Social Background (SB)
2 - Motivation (MTVTN)
3 - School (SCH)

Table 2.2 Regressor-Regressand Correlations Reproduced from Fixed Commonality Proportions and Varying Levels of Variance Explained

		TRANSFORMATIONS						
r*	OBSERVED SIMPLIFIED	.1	.2	.3	.4	.5	.6	.7
14	.5442	.2588	.3662	.4484	.5178	.5791	.6343	.6851
24	.4480	.2131	.3015	.3692	.4263	.4768	.5221	.5639
34	.5737	.2729	.3860	.4728	.5458	.6105	.6687	.7222

r*	.8	.9	1.0
14	.7323	.7768	.8188
24	.6029	.6395	.6741
34	.7721	.8189	.8632

*VARIABLES

1 - Social Background (SB)
2 - Motivation (MTVTN)
3 - School (SCH)

Table 2.3 Regression Coefficients Computed from Admissible Solutions for Fixed Commonality Proportions and Varying Levels of Variance Explained

		TRANSFORMATIONS						
β*	OBSERVED SIMPLIFIED	.1	.2	.3	.4	.5	.6	.7
1	.2146	.1019	.1445	.1768	.2043	.2283	.2502	.2702
2	.2158	.1025	.1452	.1778	.2054	.2298	.2515	.2716
3	.4005	.1906	.2694	.3300	.3810	.4262	.4668	.5042

β*	.8	.9	1.0
1	.2887	.3063	.3229
2	.2905	.3082	.3248
3	.5390	.5716	.6026

*VARIABLES

1 - Social Background (SB)
2 - Motivation (MTVTN)
3 - School (SCH)

We also took the same set of commonality proportions as are given in column 1.0 of Table 2.3, relegated them to the uniquenesses according to our theory called T7* in the previous section, and then scaled these proportions by increments of .1, ranging from .1 to 1.0. Each one of these yielded 3 admissible solutions which, upon examination, were found to have the same regressor values across the different $RSQ(X_1X_2X_3)$'s. One solution was the identity matrix, another had a .96 in the r 12 location and zero's elsewhere while the third solution had a .96 in the r 13 position and zero's elsewhere.

From these two sets of results it would appear that if a set of solutions can be found for fixed commonality proportions and one value of an $RSQ(X_1X_2X_3)$, then they can also be found for neighboring and even quite distant $RSQ(X_1X_2X_3)$ values. There is nothing astounding about these results for they follow directly from the algebra of the algorithm as set forth in Appendix A. However, these latter statements are so only for commonality proportions that are all positive. It seems obvious that if some of the proportions are negative while others are positive, and, if both increase in their absolute magnitudes, then a point (or value of $RSQ(X_1X_2X_3)$) will be reached beyond which admissible solutions cannot be obtained.

3. Three Variable Reverse Commonality for Varying Levels of Unique Variation

In the preceding sections we examined results when: the level of explained variance was held constant and the relative proportions of the variables were manipulated; and, when the relative proportions were kept fixed and the level of explained variance was systematically varied. An analyst may also

*This is the theory that SCH factors should have only their unique value but, that we don't know how to split up the higher order commonalities and, therefore, a random rationale is adopted.

want to examine how large or small a unique value a variable may take on and still yield an admissible solution. For example, in the equation

$$(12) \qquad RSQ(X_1) = C(X_1X_2X_3) + C(X_1X_2) + C(X_1X_3) + U(X_1)$$

the empirical value for $U(X_1)$ can be made to take on progressively larger or smaller values. Such a manipulation will alter not only the RSQ for that variable but also the RSQ for all three variables combined (as can be seen from equation 11 in the introductory section). In a sense this kind of analysis tests how large or how small a relationship one variable might have with a given dependent variable and still be consistently related to other regressor variables. For this example we have chosen to reduce the unique value of one variable to see to what extent this reduction affects the overall results.

In Table 3 we have systematically reduced the size of the unique value for variable X_3, which represents the set of School factors (SCH). The first set of columns in the uppermost part of Table 3 gives the set of transformations, as follows:

T1 - this is a simplification of the observed commonalities and is included for purposes of comparison with the suceeding transformations. In previous sections it has been called T1.

T20 - the unique value for X_3 in T1 is reduced by .05 while all other values are kept constant.

T21 - the unique value for X_3 in T1 is reduced by .10 while all other values are kept constant.

T22 - the unique value for X_3 in T1 is set equal to zero while all other values are kept constant.

For each of these transformations the algorithm indicated that there was only one admissible solution. In the center of the table we can observe the trend in these solutions. The correlations between variables: 1 and 2, 1 and 4; and, 2 and 4 remain the same across the different transformations and this is to be expected since the coefficients for these variables were not manipulated. The correlation between variables 1 and 3 increased substantially and this increase can be attributed to the fixed second order coefficient ($C(X_1X_3)$) and the fixed third order coefficient ($C(X_1X_2X_3)$). In a sense, as the unique variability of a variable is reduced but its common variability kept fixed, its correlation with the other variables must of necessity, increase. This trend also occurs for the correlation between variables 2 and 3, but to a lesser extent. The correlation between variables 3 and 4 decline progressively, which is to be expected since this is the relationship we manipulated when we reduced $U(X_3)$.

It is perhaps of even greater interest to examine the trend in the regression coefficients for these solutions. The coefficient for variable 1 shows a progressive increase, that for variable 2 stays about the same while that for variable 3 declines appreciably but does not become zero. Clearly then School factors derives a portion of its explanatory power from its relationship with the other regressor variables.*

* We also conducted a series of analyses in which the second order coefficient for variables 2 and 3 (viz. $C(X_2X_3)$) was made increasingly negative while the other values were kept constant. These analyses did not yield any admissible solutions. However, when the negative coefficient for $C(X_2X_3)$ was kept at its observed value (see Table 1.1) and the coefficients were unitized and then scaled by increments of .1, from .1 to 1.0, a single admissible solution was obtained for each such transformation.

Table 3 Reduced Unique Values and Their Effects

COMMONALITY COEFFICIENTS*	OBSERVED SIMPLIFIED (1)	TRANSFORMATIONS 20	21	22
$U(X_1)$	.0236	.0236	.0236	.0236
$U(X_2)$	.0307	.0307	.0307	.0307
$U(X_3)$	.1149	.0649	.0149	0
$C(X_1X_2)$	.0583	.0583	.0583	.0583
$C(X_1X_3)$	.1025	.1025	.1025	.1025
$C(X_2X_3)$	0	0	0	0
$C(X_1X_2X_3)$	.1117	.1117	.1117	.1117
$RSQ(X_1X_2X_3)$	.4417	.3917	.3417	.3268
NUMBER OF SOLUTIONS	1	1	1	1

REGRESSOR-REGRESSAND CORRELATIONS

r*	OBSERVED SIMPLIFIED (1)	TRANSFORMATIONS 20	21	22
12	.5556	.5556	.5556	.5556
13	.5235	.5944	.7252	.8505
14	.5442	.5442	.5442	.5442
23	.2820	.3114	.3528	.3688
24	.4480	.4480	.4480	.4480
34	.5737	.5283	.4786	.4628

REGRESSION COEFFICIENTS

B*	OBSERVED SIMPLIFIED (1)	TRANSFORMATIONS 20	21	22
1	.2146	.2301	.2732	.3489
2	.2158	.2195	.2253	.2233
3	.4005	.3232	.2011	.0837

*VARIABLES

1 - Social Background (SB)
2 - Motivation (MTVTN)
3 - School (SCH)

In the next section we examine results of the algorithm for larger numbers of variables.

4. Seven Variable Reverse Commonality

As the number of regressor variables increase so too do the potential number of admissible solutions and their corresponding volume of output. For example, in an analysis with 4 regressor variables one resolution yielded 16 admissible solutions; for 5 regressors one resolution yielded 52 admissible solutions; and, for 6 regressors a single resolution yielded 193 admissible solutions. Needless to say, we have no desire to report on that many results. However, a certain number of difficulties do arise when working with a large number of regressors and since the number of admissible solutions were fairly low for our 7 regressor case we have chosen to deal with some of these results in this section (seven regressor variables happens also to be the maximum number that the algorithm can currently accomodate).

The variables used in these analyses are composites or indices as developed in Mayeske, et.al., 1972. They are:

1. Socio-Economic Status
2. Family Structure and Stability
3. Ethnic Group Membership
4. Expectations for Excellence
5. Attitude Toward Life
6. Study Habits
7. School Factors (the student body's Achievement Level)
8. Individual Student Achievement

The first 7 variables are used as regressors while the eighth is the regressand (or dependent variable).

First we should recall that with 7 variables one obtains 127 different commonality coefficients (2^m-1 where m is the number of variables). This is usually too large a number to work with even if the volume of output were small. However, we decided to submit our <u>observed</u> results of the 7 variable commonality analysis to the Reverse Commonality algorithm to see if it would yield a single admissible solution. Actually, it yielded 3 admissible solutions all very similar to one another and to the observed correlations, save for a few variations in the magnitude of the correlation between variables 4 and 5 and, between variables 4 and 6, as can be seen from Table 4.1. The regression coefficients obtained from

Table 4.1 Regressor and Regressand Correlations Reproduced from Seven Variable Commonalities

		REPRODUCED FROM OBSERVED COMMONALITIES			REPRODUCED FROM SIMPLIFIED COMMONALITIES			
r*	OBSERVED	1	2	3	1	2	3	4
12	.3567	.3574	.3574	.3574	.3703	.3703	.3703	.3703
13	.3714	.3719	.3719	.3719	.3748	.3748	.3748	.3748
14	.2787	.2793	.2793	.2793	.2740	.2740	.2740	.2740
15	.3736	.3739	.3739	.3739	.3795	.3795	.3795	.3795
16	.3968	.3972	.3972	.3972	.4120	.4120	.4120	.4120
17	.4346	.4345	.4345	.4345	.4472	.4472	.4472	.4472
18	.4973	.4974	.4974	.4974	.4850	.4850	.4850	.4850
23	.2999	.3007	.3007	.3007	.3019	.3019	.3019	.3019
24	.3725	.3734	.3734	.3734	.3839	.3839	.3839	.3839
25	.4707	.4707	.4707	.4707	.4854	.4854	.4854	.4854
26	.4723	.4715	.4715	.4715	.4967	.4967	.4967	.4967
27	.2908	.2894	.2894	.2894	.3032	.3032	.3032	.3032
28	.3259	.3260	.3260	.3260	.3028	.3028	.3028	.3028
34	.1684	.1678	.1678	.1678	.1550	.1550	.1550	.1550
35	.2253	.2253	.2253	.2253	.2146	.2146	.2146	.2146
36	.2219	.2217	.2217	.2217	.2101	.2101	.2101	.2101
37	.6165	.6173	.6173	.6173	.6420	.6420	.6420	.6420
38	.4908	.4909	.4909	.4909	.4784	.4784	.4784	.4784
45	.5172	.5143	.7782	.5143	.5699	.7148	.5699	.7148
46	.4894	.4876	.4876	.8264	.5451	.5451	.7639	.7639
47	.1558	.1553	.1553	.1553	.1562	.1562	.1562	.1562
48	.2552	.2546	.2546	.2546	.2356	.2356	.2356	.2356
56	.6623	.6603	.6603	.6603	.7112	.7112	.7112	.7112
57	.2321	.2312	.2312	.2312	.2306	.2306	.2306	.2306
58	.3820	.3818	.3818	.3818	.3635	.3635	.3635	.3635
67	.2331	.2327	.2327	.2327	.2344	.2344	.2344	.2344
68	.3674	.3670	.3670	.3670	.3523	.3523	.3523	.3523
78	.5674	.5673	.5673	.5673	.5598	.5598	.5598	.5598

* The variables are: 1-Socio-Economic Status; 2-Family Structure and Stability; 3-Ethnic Group Membership; 4-Expectations for Excellence; 5-Attitude Toward Life; 6-Study Habits; 7-School Factors; 8-Individual Student Achievement

Table 4.2 Standardized Regression Weights from Observed and Reproduced Commonalities for Seven Variables

		REPRODUCED FROM OBSERVED COMMONALITIES			REPRODUCED FROM SIMPLIFIED COMMONALITIES			
β*	OBSERVED	1	2	3	1	2	3	4
1	.2116	.2116	.2120	.2091	.2084	.2079	.2057	.1998
2	.0050	.0056	.0086	.0068	-.0203	-.0189	-.0205	-.0256
3	.1626	.1622	.1632	.1636	.1527	.1529	.1535	.1542
4	.0104	.0103	-.0738	-.0625	-.0071	-.0496	-.0490	-.1140
5	.1328	.1332	.1957	.1335	.1377	.1677	.1382	.1777
6	.0766	.0762	.0747	.1333	.0737	.0753	.1082	.1350
7	.3235	.3235	.3208	.3213	.3268	.3258	.3259	.3244
RSQ	.4517	.4516	.4537	.4528	.4253	.4264	.4262	.4298

* The variables are: 1-Socio-Economic Status; 2-Family Structure and Stability; 3-Ethnic Group Membership; 4- Expectations for Excellence; 5-Attitude Toward Life; 6-Study Habits; 7-School Factors; 8-Individual Student Achievement

these solutions, given in Table 4.2, show that they tend to be very similar to those obtained from the observed correlations save for variable 4 (Expectations for Excellence) which takes on small negative values and variables 5 (Attitude Toward Life) and 6 (Study Habits) which take on slightly larger values for the second and third solutions, respectively.

Next, in order to reduce the sheer number of coefficients, we set equal to zero any empirical value that had zero's for its first three decimal digits.* The algorithm indicated that there were four admissible solutions for this simplified set of coefficients. The correlations and regression coefficients obtained from these solutions are given in the last four columns of Tables 4.1 and 4.2 respectively. Inspection of these values shows them to be virtually identical to one another except for the correlations between variables 4 and 5 and variables 4 and 6. These latter alternate in taking on higher or lower values. Although these correlations are not very different from one another they do differ enough from the observed correlations to result in somewhat different values for some of the regression coefficients, as can be seen from Table 4.2. This is especially so for variables 2 (Family Structure) and, 6 (Study Habits). It appears to be these attitudinal and motivational variables then that show the greatest sensitivity to simplification of the commonalities.

In an attempt to further simplify the number of empirical values one had to work with we set equal to zero any of these commonalities that had zero's in the first two decimal digits.** For these simplifications the algorithm indicated that there were not any admissible solutions.+ We then took these

*e.g. .0004 would be set to zero but.0040 would not.
**e.g. .0040 would be set to zero but .04000 would not.
+However, when the criterion was relaxed to a value of 1.0, the algorithm indicated that there were 20 admissible solutions (see Appendix A). Whether or not it is desirable to relax this criterion will depend upon the analysts' objectives.

further simplified values, split them up and allocated them to the unique values in two distinctly different ways. For each of these the algorithm indicated that there were 15 admissible solutions. We shall not dwell on the nature of these results other than to note that the regressor correlations often varied over a considerable range. Some of them assumed values in the range from zero to as high as .8 or .9.

5. Summary

In this paper we have examined the results of an algorithm which is designed to take a set of commonality coefficients, either real or manipulated, and, if possible, produce one or more sets of regressor correlations that are consistent with them. A number of different ways of resolving the higher order commonality values into their lower orders were tried and the number and nature of solutions generated from them were examined for their meaningfulness and variability. In general it was found that this could be a meaningful exercise because it allows an analyst to test his assumptions about the nature of the confounding to see if a solution can be obtained. If one cannot be obtained then his assumptions must be revised. But if one can be obtained, then he can examine them to see which kinds of variables have the greatest sensitivity to the assumptions. The resulting output also allows one to gauge the variability of regression coefficients that will satisfy the same set of commonality values and the effectiveness of the regression system.

This technique called "Reverse Commonality" is best suited for an interactive computing arrangement so that an analyst can use it rapidly in a sequential manner. Since, for a large number of variables the sheer volume of commonality values becomes unmanageable and, since a number of different ways are available for resolving the confounding, an algorithm is needed which uniquely resolves the higher order values according to one's assumptions.

References Cited

Beaton, A. E., Jr.1973 "Commonality" Unpublished Manuscript, Princeton, New Jersey

Maysske, et. al., 1972, A Study of Our Nation's Schools. Washington, D.C.: U.S. Government Printing Office.

Mayeske, G. W., and Beaton, A. E., Jr. 1974, Special Studies of Our Nation's Students, Washington, D.C.: U.S. Government Printing Office (In Press).

A.E. Beaton
6/10/74

Reverse Commonality (REVCOM)

1. Discussion

Let us assume that there is a variable y which has been used in a regression analysis in which variables $x_j (j=1,2,...,m)$ have been used as predictors. Let us further assume that the statistics available to us from the regression analysis is either the commonalities or squared multiple correlations of y with each possible combination of predictors. The signs of the correlations of individual x_j with y are known. Our problem is to compute the intercorrelations among the predictors from the known simple and multiple correlations.

This problem must have at least one valid solution if the known coefficients were computed from real data. There need not be a unique solution, however, and in the two predictor case there will ordinarily be two solutions. Early experience with test problems indicates that there are not very many solutions for multi-predictor systems. If artificial commonalities or squared multiple correlations are supplied then there may be no solutions indicating that the correlations are not consistent with any possible set of real data.

This paper presents an algorithm for computing all possible interpredictor correlation matrices from a given vector of commonalities and the signs of the simple correlations between y and the x_j. Commonalities can be simply converted to squared simple and multiple correlations. We will use only the more familiar multiple correlation coefficients for discussion here.

Discussion

The information is a vector of all possible squared multiple correlations of a dependent variable y with the independent variables $x_j (j=1,2,\ldots m)$ which is created from the commonalities. There are 2^m-1 such correlations which will be indicated by subscripts $R^2_{y1}, R^2_{y2}, \ldots R^2_{ym}, R^2_{y12}, R^2_{y123} \ldots R^2_{y123\ldots m}$.

The signs of the simple correlations $r_{yj} = \sqrt{R^2}$ are also known. We wish to compute the simple correlations $r_{ij} (i=1,2,\ldots m-1; j=i+1, i+2,\ldots m)$ among the dependent variables. There are $m(m-1)/2$ such correlations.

The general strategy is to compute the possible values of r_{ij} that are consistent with r^2_{yi}, r^2_{yj}, and R^2_{yij}, then to check these for consistency with each other known correlation, then each other pair of correlations, and so forth until they are demonstrated to be consistent with all other variables collectively. If a possible r_{ij} does not pass every such test it is eliminated from further consideration.

The first step is to compute the initial set of possible values for the squared multiple correlation for predicting y from x_i and x_j may be written

$$R^2_{yij} = \frac{r^2_{yi} + r^2_{yj} - 2r_{yi}r_{yj}r_{ij}}{1-r^2_{ij}}$$

which may be solved for r_{ij} by

$$r_{ij} = \frac{2r_{yi}r_{yj} \pm \sqrt{r^2_{yi}r^2_{yj} - R^2_{yij}(R^2_{yi}+r^2_{yj}) - R^4_{yij}}}{R^2_{yij}}$$

Depending on the value of the value inside the radical, r_{ij} may have either no solution, one solution or two solutions. If the value of the radicand is negative there is no possible r_{ij} that could combine with r_{yi} and r_{yj} to create such a R^2_{yij} and thus the given values could not come from real data. In the rare case that the radicand is zero, r_{ij} is unique. If the radicand is positive then there are exactly two solutions which we will refer to as $r_{ij}^{S_{ij}}$ where S_{ij} = + or - depending on the sign of the radical.

By computing $r_{ij}^{S_{ij}}$ for all i and j we have computed all r_{ij} that are possible in two variable prediction. Since there are up to 2 possible values for each r_{ij} there are $2^{m(m-1)/2}$ possible combinations of r_{ij} thus the same number of intercorrelation matrices which must be considered. Many of these may be eliminated through evaluation of higher order relationships. To investigate higher order relationships we have two checks:

(1) Are the intercorrelations internally consistent? and

(2) Does this combination of intercorrelations produce the correct higher order multiple correlation?

These two questions can be answered by standard multiple regression techniques. The SWP operator (Beaton (1964))* is a simple method of implementing these computations. Basically, the process begins with 3 predictor regressions in which case there are three off-diagonal co-relations and up to 8 possible combinations of intercorrelations. Each is examined in order by forming a 4x4 correlation matrix (including the 3 independent variables and y) and sweeping out the independent variables. If the SWP operator encounters a negative pivot then this combination is internally inconsistent. If the multiple correlation of these X_j and y is not the known value within a tolerance limit, then this combination does not fit the problem and is rejected. After computing the (up to) 8 combinations only the $r_{ij}^{S_{ij}}$ that are in at least one acceptable combination are retained. Our experience indicates that many of the potential correlations are lost at this point. If no combination is acceptable then there can be no solution to these problems and thus the given multiple correlations cannot come from real data.

After trying all 3 variable combinations, all 4 variable, 5 variable, and higher order combinations are tried until finally the m variable regression is performed. At this point the problem is solved since the remaining matrices have been demonstrated to have a positive determinant and fit all given multiple correlations.

If the algorithm worked through all possible combinations the computer costs might be prohibitive for large m. The algorithm saved considerable time by approaching the problem by successive elimination. The 3 variable phase usually reduces the potential $r_{ij}^{S_{ij}}$ substantially so that there are usually not too many 4 variable combinations worth trying; and the 4 variable combination does the same for the 5, and so forth.

*Beaton, A. E. The Use of Special Matrix Operators in Statistical Calculus, Research Bulletin 64-51, Educational Testing Service, October, 1964.

2. REVCOM PROGRAM FLOW

(1) Read in problem control card (if blank card to to (8)).
NI = # of independent varialbes.
NV = NI+1 (program adds on the dependent variable).

(2) Read in signs to be used for the independent vs dependent correlations.

(3) Call subroutine 'CRSQ'.
(a) Read in inputted commonality table.
(b) Compute and print all possible R squares ($2^{NI}-1$).
(c) Return to main program.

(4) Develop the two-valued NVxNV correlation matrix using the R square table. Report any discrepancies.

(5) If any correlation cannot be developed go to (1).

(6) Call subroutine 'ELIM'.
(a) Set NVB = 3.
(b) Select one combination of NI variables taken NVB at a time. If all combinations computed go to (f).
(c) For these selected NVB variables set up all possible correlation matrices from the two-valued matrix computed in step (4). Compute the R square for each of these matrices.
(d) If one of the two values for any of the NVBx(NVB+1)/2 possible correlations never gave an R square solution that checked against the table computed in step (3.b), eliminate it. If both values are eliminated stop the program.
(e) If NVB=NI print each matrix that checks against the R square table along with its inverse and label it a 'GOOD' case.
(f) Go to (b).
(g) Print out the two valued correlation matrix showing which of the NIx(NI+1)/2 x 2 values have been elimated by the NVB order elimination.

(h) NVB=NVB+1 if NVB$\leq$NI go to (b).

(i) Return to main program.

(7) Go to (1).

(8) END.

3. REVCOM Program Instructions

Purpose - Given a commonality table derive the original correlation matrix.

Input

1) Control Card

Cols.	Var.	Explanation
5	N	$\leq$ 7 Number of Independent variables in the commonality table.
6-60	Title	Any heading information.

2) Sign control card and R square tolerance

Cols.	Var.	Explanation
1		'+' or '-' (This is the sign of the correlation between the first independent var. and the dependent var.
2		'+', or '-' (sign of the correlation between the second independent var. and the dependent var.
3		'+', or '-' (3rd sign)
4		'+', or '-' (4th sign)
⋮		⋮
N		'+', or '-' (Nth sign)

Cols.	Var.	Explanation
[optional] 11-20	TOL	This is the largest absolute deviation allowed when checking the derived R-squares against the R-square computed from the commonality table. If columns 11-20 are left blank the program will use a default value of .005.

3) Commonality Table.

The commonality table consists of 2^n-1 input cards. There is one card input for each uniqueness and each commonality combination. The card contains the commonality variable numbers identifying the combination as well as the commonality value. The first order commonmition are read in first, followed by the possible second order, third, etc.

EX. Four Variable Problem $-(2^n-1 = 15)$

1	xxx
2	xxx
3	xxx
4	xxx
12	xxx
13	xxx
14	xxx
23	xxx
24	xxx
34	xxx
123	xxx
124	xxx
134	xxx
234	xxx
1234	xxx

Cols.	Var.	Explanation
1-7		Commonality Level
11-20		Commonality Value (F10.4)

If another problem is to be run repeat steps 1-3. If this is the last problem insert a blank card after the input to bring the problem to a normal termination.