DOCUMENT RESUME

ED 284 873 TM 870 253

AUTHOR McMurray, Mary Anne
TITLE The Concepts Underlying Structure Coefficients, Communality, and Factor Scores in the Exploratory Factor Analytic Case.
PUB DATE 30 Jan 87
NOTE 17p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 29-31, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Correlation; *Factor Analysis; *Factor Structure; *Goodness of Fit; Higher Education; Mathematical Models; *Matrices; Multiple Regression Analysis; Research Methodology; Simulation; Statistical Studies; *Weighted Scores
IDENTIFIERS *Communality; Z Scores

ABSTRACT

This paper illustrates the transformation of a raw data matrix into a matrix of associations, and then into a factor matrix. Factor analysis attempts to distill the most important relationships among a set of variables, thereby permitting some theoretical simplification. In this heuristic data, a correlation matrix was derived to display associations among the variables. Multiple regression, both with and without Z-score transformation, was used to generate structure coefficients and the factor matrix. This matrix shows the correlation of each variable with the factors extracted from the association pattern. Two factors were extracted in this case. For each, the scores constitute a new variable, a weighted combination of the scores on each of the original variables. Each factor is treated as a hypothetical construct, to be interpreted from the pattern of factor scores. Communality is the proportion of variance of each variable that is reproduced in the extracted factors. If multiple regression is the factoring method, then the sum of communalities is equivalent to the multiple correlation coefficient. (LPG)

THE CONCEPTS UNDERLYING STRUCTURE COEFFICIENTS, COMMUNALITY, AND FACTOR SCORES IN THE EXPLORATORY FACTOR ANALYTIC CASE

Mary Anne McMurray

University of New Orleans 70148

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, January 30, 1987.

Abstract

This paper illustrates the transformation of a raw data matrix into a matrix of associations, and then into a factor matrix. In the process of extracting two factors from heuristic data, the nature of structure coefficients is explored, the derivation and utility of factor scores are discussed, and the concept of communality is related to the multiple correlation coefficient.

Although factor analysis has several possible uses, an important application is the service the procedures provide to summarizing scientific data, thus enabling empirical relationships to be more easily conceptualized (Gorsuch, 1974, p. 2). The analysis evaluates a data matrix defined by two dimensions and can be visualized as a rectangular table with the columns defining one set of entities (such as variables, people, or occasions of measurement) and the rows defining a group of replicates over which those entities are measured.

According to Horst (1965, p. 17), "a primary concern of factor analysis with such a table of numbers is to determine whether the table may be simplified in some way". The original data matrix may contain random or unreliable information in addition to relevant or systematic information (p. 469). A simplification of the matrix, it is expected, will help to disentangle complex interrelationships so that they may be more easily understood as they exist in real life (Rummel, 1970, pp. 3-4). Factor analysis "distills" the information in the variables such that the data are more meaningful.

Several values derived through factor analytic procedures allow the researcher to examine different aspects of the relationships among the variables of a study. The present paper discusses the concepts underlying these indices. Specifically, this paper focuses upon structure coefficients, factor scores, and communality coefficients.

Hypothetical Data

The raw data matrix presented in Table 1 is analyzed in the present paper for illustrative purposes. The variables are ten instructional strategies rated by a sample of 53 college students on a scale of 1 to 10 with respect to their helpfulness to learning. The strategies analyzed consisted of the following:

1. The teacher provides some kind of outline in advance to indicate what is to be studied.

2. The teacher directs the manipulation of materials leading to discovery of concepts.

3. The teacher asks the class questions that can be answer "yes" or "no."

4. The teacher directs the writing of term papers or work on projects to be completed outside of class time.

5. The teacher provides opportunities for discussion of concepts with classmates.

6. The teacher writes notes on board for students to copy.

7. The teacher provides opportunities for work/discussion in small groups.

8. The teacher assigns reading from sources other than textbooks.

9. the teacher directs the manipulation of materials that illustrate concepts presented.

10. The teacher encourages students to study with classmates.

Included in Table 1 is the data utilized in the study.

Insert Table 1 about here.

An initial step in factor analysis is the computation of a matrix of association coefficients from the raw data matrix. In the present study a correlation matrix was derived. Perusal of that matrix, shown in Table 2, reveals that some pairs of variables have higher correlations with one another than do others. A pattern may exist among the correlation coefficients.

Insert Table 2 about here.

Determination of that pattern is the next logical step in factor analysis. A new matrix is derived, the elements of which can be termed "structure coefficients." A structure coefficient indicates the correlation between a variable and a variable composite derived by a process of weighting and aggregating (Thompson, [illegible]). In the factor analytic case the derived entity is a factor and can be a composite of portions of some or all of the original variables. In the new matrix, the factor matrix, the original variables form the rows and factors form the columns. For the illustrative data in Table 1, two factors were initially extracted using the principal components procedure, and the results were subjected to varimax rotation to clarify their meaning. The two factors seem to distinguish between strategies which promoted communication with classmates and dependence upon the instructor (Factor I) and strategies which encouraged self-involvement of the student with learning materials (Factor II). The structure coefficients of the 10 variables for each of the two factors are presented in Table 3.

Insert Table 3 about here.

Factor Interpretation

Factors are hypothetical constructs. The description of the factors in terms of the observed variables is one of the problems with which factor analysis is concerned (Harman, 1967, p. 345). The computation of factor scores serves as a means to continue investigation of the constructs. A factor score is a new variable, a weighted combination of the scores on each of the variables (Kachigan, 1982, p. 244). A variety of methods exist for deriving factor scores, and several are described by Harman (1967, pp. 346-374) and by Comrey (1973, pp. 232-238). In the present study the regression method was used to estimate two different sets of factor scores for each individual on the two factors. One procedure derived the factor scores based on Z-scores and used the matrix formula:

$$Z\ R^{-1}\ S = F$$

where Z is the Z-score matrix, R^{-1} is the inverse of the correlation matrix, S is the structure coefficient matrix, and F is the factor score matrix. These scores are presented in Table 1.

When the 10 variables are correlated with the factor scores derived from the variables' Z-scores, the nature of structure coefficients as the proportion of a variable that is reproduced within a factor becomes apparent. To this end, Table 4 presents the correlation of each of the variables with each of the factor scores.

Insert Table 4 about here.

However, factor scores need not be based upon Z-scores. Thompson (1983) describes a non-conventional procedure in which the raw data are transformed only by the division of each raw score by the standard deviation of the variable, with the result that each variable has a standard deviation of one, but a non-zero mean. Thus this procedure is the same except that a Z matrix is not employed. Both types of factor scores are presented in Table 5. The means of non-conventional factor scores can be calculated and compared across the factors. Thus the conclusion might be drawn that strategies which promote communication with classmates and dependence upon the instructor seemed to be perceived as slightly more helpful to learning than those which encouraged self-involvement of the student with learning materials.

Insert Table 5 about here.

Not all of the variance of the original ten variables is represented in the two factors extracted. The proportion of the variance of each variable that is reproduced in the extracted factors is called "communality" (Gorsuch, 1974, p. 26). Table 6 indicates the communality of the ten original variables. The sum of these communalities is 5.52. Divided by the number of variables (10), the result (0.552 or 55.2%) represents the proportion of the variance that is contained in the two factors. When multiple regression procedures were used to analyze the correlation of the pair of factor scores with each variable, the square of the multiple correlation ($\underline{R}^2$) was seen to be exactly equal to the communality. Hence a measure of communality may be considered a multiple correlation coefficient in the factor analytic case.

Insert Table 6 about here.

Discussion

Through factor analysis the complex web of interrelationships among variables can be disentangled. Research can focus upon relevant information, and patterns can be made evident. This paper illustrated the transformation of a raw data matrix into a matrix of associations, and then into a factor matrix. In the process of extracting two factors, the nature of structure coefficients has been explored, the derivation and utility of factor scores has been discussed, and the concept of communality has been related to the multiple correlation coefficient.

REFERENCES

Comrey, A. L. (1973). A First Course in Factor Analysis. New York: Academic Press.

Gorsuch, R. L. (1974). Factor Analysis. Philadelphia: W. B. Saunders Company.

Harman, H. H. (1967). Modern Factor Analysis. Chicago: The University of Chicago Press.

Horst, P. (1965). Factor Analysis of Data Matrices. New York: Holt, Rinehart and Winston, Inc.

Kachigan, S. K. (1982). Multivariate Statistical Analysis. New York: Radius Press.

Rummel, R. J. (1970). Applied Factor Analysis. Evanston: Northwestern University Press.

Thompson, B. (1983). The Calculation of Factor Scores: An Alternative. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston.

Thompson, B., & Borello, G. M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Table 1

Data Set

Case	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	FS1	FS2
	Original Variables										Factor Scores	
1	10	10	6	8	3	10	6	6	10	8	-0.25775	.51156
2	10	6	7	10	9	9	8	9	7	6	.43408	.27908
3	9	4	2	10	8	8	6	9	10	3	-0.90742	.67611
4	9	7	6	6	8	8	8	7	7	6	.22568	-0.33626
5	9	10	1	7	6	1	6	7	10	4	-1.39963	1.04917
6	10	10	6	10	10	8	10	8	8	9	1.02903	.90456
7	10	10	2	2	10	10	10	10	10	10	1.03996	.64755
8	8	10	1	8	9	7	8	9	9	6	-0.23241	.97578
9	10	8	8	6	8	9	8	8	8	8	.71512	.06139
10	9	10	8	9	10	9	10	9	8	10	1.31422	.75254
11	10	7	7	5	10	10	9	8	8	6	.86942	-0.13003
12	10	9	4	10	6	3	7	10	10	1	-1.24802	1.58537
13	6	6	2	7	5	2	3	4	6	2	-2.07572	-0.88726
14	10	10	4	10	10	8	10	9	10	7	.60116	1.46195
15	9	7	5	2	8	5	8	6	8	8	.19082	-0.56436
16	8	8	4	2	4	6	5	2	8	4	-1.05590	-1.18935
17	9	9	5	4	10	8	8	5	5	9	.86680	-0.92194
18	2	5	3	8	5	9	6	8	7	5	-1.20917	-0.73877
19	10	9	2	8	10	6	10	7	10	10	.62187	1.01546
20	10	9	7	2	9	8	9	5	6	7	.90388	-0.93932
21	10	10	10	10	10	10	10	10	10	10	1.52936	1.34170
22	10	10	10	10	10	10	10	2	10	10	1.61685	.30659
23	9	10	2	5	9	5	9	7	8	10	.4642	.30198
24	10	9	4	8	7	1	7	9	10	1	-1.21434	1.33681
25	10	10	1	8	8	8	5	7	10	5	-0.61065	.91836
26	10	10	10	10	10	9	9	10	10	10	1.34029	1.36663
27	10	10	5	5	5	8	8	7	10	5	-0.24543	.49058
28	10	6	7	8	8	10	9	8	9	7	.59585	.25993
29	10	10	10	1	8	10	8	2	2	5	1.05031	-2.29448
30	3	6	3	4	8	10	9	6	5	6	-0.05807	-1.58545
31	10	10	6	10	8	10	10	2	8	10	1.08772	.00632
32	9	9	5	9	9	10	8	9	9	10	.72610	.80795
33	6	7	4	4	9	3	9	5	7	8	.00940	-0.70838
34	10	10	1	4	7	5	1	4	8	1	-1.64117	-0.26365
35	10	8	5	4	8	9	8	6	7	9	.65757	-0.53181
36	7	5	5	4	9	10	6	3	5	7	.24091	-1.90536
37	6	8	10	8	8	10	8	8	8	8	.65899	-0.08915
38	8	9	8	9	9	8	8	7	9	6	.39217	.48967
39	8	9	8	5	9	9	8	3	8	8	.81992	-0.71412
40	7	6	9	6	7	9	5	3	7	6	-0.03674	-1.30606
41	5	8	1	2	5	10	2	2	4	1	-1.60939	-2.30603
42	10	6	10	1	9	10	10	5	5	10	1.67414	-1.71527
43	10	10	5	7	8	6	8	6	9	8	.27700	.53945

(table continues)

Case	Original Variables										Factor Scores	
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	FS1	FS2
44	5	4	2	8	7	2	6	8	9	3	-1.64350	.11601
45	9	9	3	10	6	9	4	8	8	9	-0.43643	.57203
46	7	10	5	8	9	3	8	7	10	7	-0.16371	.91484
47	5	10	4	9	3	9	4	4	10	4	-1.43726	.07280
48	8	6	1	6	10	1	9	7	5	5	-0.46887	-0.38375
49	6	8	3	6	7	7	7	6	6	5	-0.55892	-0.66482
50	1	8	5	9	5	6	2	8	8	6	-1.65453	-0.17991
51	10	10	2	9	6	4	5	9	10	4	-1.2150	1.41772
52	10	9	2	10	6	7	6	9	6	2	-1.01871	.55128
53	8	7	4	3	8	8	9	3	6	8	.44593	1.37567

Table 2

Correlation Matrix

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
V1	1.00000									
V2	.44008	1.00000								
V3	.21506	.07076	1.00000							
V4	.0543	.18742	.02269	1.00000						
V5	.37875	.05358	.31884	-0.01260	1.00000					
V6	.09428	.04883	.52394	-0.06303	.16872	1.00000				
V7	.42105	.12693	.43832	.00028	.74207	.24176	1.00000			
V8	.16704	.07501	-0.10493	.53452	.20687	-0.13630	.21239	1.00000		
V9	.27054	.38923	-0.06837	.59362	-0.03472	-0.16951	.07509	.50366	1.00000	
V10	.24897	.22979	.49499	.00252	.58438	.48772	.67759	.04256	.08867	1.00000

Table 3

Structure Coefficients

	Factor 1	Factor 2
V1	.47882	.41166
V2	.21368	.48407
V3	.70389	-0.11386
V4	-0.09875	.75442
V5	.75509	.13769
V6	.59279	-0.24966
V7	.83191	.19110
V8	.00680	.75011
V9	-0.05879	.85799
V10	.84452	.08172

Table 4

Correlation Between Original Variables and Factor Scores

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
FS1	.4788	.2137	.7039	-0.0987	.7551	.5928	.8319	.0068	-0.0588	.8445
	(53)	(53)	(53)	(53)	(53)	(53)	(53)	(53)	(53)	(53)
	P= .000	P= .062	P= .000	P= .241	P= .000	P= .000	P= .000	P= .481	P= .338	P= .000
FS2	.4117	.4841	-0.1139	.7544	.1377	-0.2497	.1911	.7501	.8580	.0817
	(53)	(53)	(53)	(53)	(53)	(53)	(53)	(53)	(53)	(53)
	P= .001	P= .000	P= .208	P= .000	P= .163	P= .036	P= .085	P= .000	P= .000	P= .280

(COEFFICIENT / (CASES) 1-TAILED SIG)

Table 5

Two Sets of Factor Scores

Case	Conventional Scores		Non-centered Scores	
	Factor 1	Factor 2	Factor 1	Factor 2
1	-0.25775	.51156	4.04	4.90
2	.43408	.27908	4.57	4.67
3	-0.90742	.67611	3.39	5.07
4	.22568	-0.33626	4.36	4.05
5	-1.39963	1.04917	2.89	5.44
6	1.02903	.90456	5.21	5.30
7	1.03996	.64755	5.33	5.04
8	-0.23241	.97578	4.01	5.37
9	.71512	.06139	4.90	4.45
10	1.31422	.75254	5.50	5.14
[illegible]	.86942	-0.13003	5.06	4.26
12	-1.24802	1.58537	3.05	5.98
13	-2.07572	-0.88726	2.00	3.50
14	.60116	1.46195	4.90	5.85
15	.19082	-0.56436	4.38	3.83
16	-1.05590	-1.18935	3.13	3.20
17	.86680	-0.92194	4.89	3.47
18	-1.20917	-0.73877	2.92	3.65
19	.62187	1.01546	4.92	5.41
20	.90388	-0.93932	4.98	3.45
21	1.52936	1.34170	5.82	5.73
22	1.61685	.30659	5.91	4.70
23	.46425	.30198	4.65	4.69
24	-1.21434	1.33681	3.08	5.73
25	-0.61065	.91836	3.68	5.31
26	1.34029	1.36663	5.63	5.76
27	-0.24543	.49058	4.05	4.88
28	.59585	.25993	4.84	4.65
29	1.05031	-2.29448	4.91	2.10
30	-0.05807	-1.58545	3.97	2.81
31	1.08772	.00632	5.27	4.40
32	.72610	.80795	4.97	5.20
33	.00940	-0.70838	4.14	3.68
34	-1.64117	-0.26365	2.54	4.13
35	.65757	-0.53181	4.79	3.86
36	.24091	-1.90536	4.26	2.49
37	.65899	-0.08915	4.84	4.30
38	.39217	.48967	4.63	4.88
39	.81992	-0.71412	5.01	3.68
40	-0.03674	-1.30606	4.09	3.08
41	-1.60939	-2.30603	2.36	2.08
42	1,67414	-1.71527	5.70	2.68
43	.27700	.53945	4.52	4.93

(table continues)

Case	Conventional Scores		Non-centered Scores	
	Factor 1	Factor 2	Factor 1	Factor 2
44	-1.64350	.11601	2.60	4.51
45	-0.43643	.57203	3.75	4.96
46	-0.16371	.91484	4.13	5.31
47	-1.43726	.07280	2.86	4.46
48	-0.46887	-0.38375	3.55	4.01
49	-0.55892	-0.66482	3.52	3.73
50	-1.65453	-0.17991	2.53	4.21
51	-1.21508	1.41772	3.08	5.81
52	-1.01871	.55128	3.06	4.94
53	.44593	-1.37567	4.52	3.01
Mean	.000	.000	4.18	4.39
SD	1.000	1.000	.999	1.000

Table 6

Communality Coefficients

V1	.39874
V2	.27998
V3	.50842
V4	.57890
V5	.58912
V6	.41373
V7	.72860
V8	.56271
V9	.73960
V10	.71990