PSYCHOMETRIKA—VOL. 4, NO. 2
JUNE, 1939

THE CONTRIBUTION OF AN ORTHOGONAL MULTIPLE FACTOR SOLUTION TO MULTIPLE CORRELATION

P. S. DWYER
The University of Michigan

A method is indicated by which multiple factor analysis may be used in determining a number, r, and then in selecting r "predicting" variables out of n variables so that each of the remaining n-r variables may be predicted almost as well from the r variables as it could be predicted from all the n—1 variables.

One of the unsolved questions of multiple factor analysis deals with the relation of multiple factor analysis to multiple correlation. Roff (**1**) has previously given some theorems dealing with this subject, but his approach emphasized the application of multiple correlation results in determining communality. The present paper emphasizes the contribution of the factor analysis results to multiple correlation.

Suppose that the intercorrelations of variables 1, 2, 3, $\cdots$, j, $\cdots$, n are subjected to a multiple factor analysis which results in r common factors. The resulting weights are indicated by a_{ji} where the first subscript indicates the variable and the second subscript indicates the factor. Let the communality of test j be indicated by h^2_j, and the uniqueness by u^2_j (**2**, p. 63).

We let the r common factors be represented by r orthogonal unit reference vectors and prove the theorem, given previously by Roff (**1**, p. 2) with somewhat different emphasis, "The multiple correlation of variable j with the r unit orthogonal reference vectors is equal to the square root of the communality of variable j."

We first construct the matrix of the correlations involving variable j and the r reference vectors. We note first that the correlation of reference vector i with itself is unity and that the correlation of reference vector i with any other reference vector is 0. Furthermore, the correlation of variable j with reference vector i is the correlation bebetween

$$x_j = a_{j1}\,y_1 + a_{j2}\,y_2 + \cdots + a_{ji}\,y_i + \cdots + a_{jr}\,y_r + u_j\,z_j$$

and y_i, and is

$$r_{x_j\,y_i} = \frac{a_{ji}}{\sqrt{a^2_{j1} + a^2_{j2} + \cdots + a^2_{jr} + u^2_j}} = a_{ji}\,.$$

This result was previously indicated by Roff (**1**, p. 3). In case $r = 3$, the determinant of the correlations becomes

$$\Delta = \begin{vmatrix} 1 & a_{j1} & a_{j2} & a_{j3} \\ a_{j1} & 1 & 0 & 0 \\ a_{j2} & 0 & 1 & 0 \\ a_{j3} & 0 & 0 & 1 \end{vmatrix}.$$

If now the elements of the second column are multiplied by $-a_{j1}$, the elements of the third column are multiplied by $-a_{j2}$, the elements of the fourth column by $-a_{j3}$, and the results added to the first column, we get

$$\Delta = \begin{vmatrix} u^2_j & a_{j1} & a_{j2} & a_{j3} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = u^2_j,$$

where $u^2_j = 1 - a^2_{j1} - a^2_{j2} - a^2_{j3}$.

The proof for the general case of r factors follows in an identical manner so that we have

$$\Delta = u^2_j, \text{ where } u^2_j = 1 - \sum_{i=1}^{r} a^2_{ji} = 1 - h^2_j.$$

It is at once apparent that the determinant obtained by deleting the first row and the first column is 1 so that $\Delta_{11} = 1$. The usual determinantal formula for multiple correlation (**3**, p. 301) then gives

$$r_{j\cdot 12\cdots r} = \sqrt{1 - \frac{\Delta}{\Delta_{11}}} = \sqrt{1 - u^2_j} = h_j.$$

Another method of establishing this theorem consists in correlating

$$x_j = a_{j1} y_1 + a_{j2} y_2 + \cdots + a_{ji} y_i + \cdots + a_{jr} y_r + u_j z_j$$

with the "common" part of x_j,

$$x'_j = a_{j1} y_1 + a_{j2} y_2 + \cdots + a_{ji} y_i + \cdots + a_{jr} y_r,$$

and gives

$$r_{jj'} = \frac{a^2_{j1} + a^2_{j2} + \cdots + a^2_{jr}}{[(a^2_{j1} + a^2_{j2} + \cdots + a^2_{jr} + u^2_j)(a^2_{j1} + a^2_{j2} + \cdots + a^2_{jr})]^{\frac{1}{2}}}$$

$$= \sqrt{a^2_{j1} + a^2_{j2} + \cdots + a^2_{jr}} = \sqrt{h^2_j} = h_j .$$

A second theorem is "The multiple correlation of variable j with the $n - 1$ remaining variables and the r reference vectors is equal to the square root of the communality of variable j." This theorem was also given by Roff (**1**, p. 2) (with somewhat different emphasis) who implied that it followed from the first theorem with the statement, "Since the correlation between any two tests j and k is 0 after common factors have been removed, a second theorem follows immediately."

A more formal proof is outlined below.

The matrix of the n variables and the r reference vectors is composed of four different types of matrices if the n variables precede the r factors both horizontally and vertically. In the upper right hand corner there is the factorial matrix, in the lower left hand corner there is the transpose of the factorial matrix, while the lower right hand corner contains the identity matrix. For example, when $n = 3$ and $r = 2$, we have

$$\begin{bmatrix} 1 & r_{12} & r_{13} & a_{11} & a_{12} \\ r_{21} & 1 & r_{23} & a_{21} & a_{22} \\ r_{31} & r_{32} & 1 & a_{31} & a_{32} \\ a_{11} & a_{21} & a_{31} & 1 & 0 \\ a_{12} & a_{22} & a_{32} & 0 & 1 \end{bmatrix} .$$

The process used in establishing Theorem II is similar to that used in establishing Theorem I. The elements of column 4 are multiplied by $-a_{11}$, the elements of column 5 are multiplied by $-a_{12}$, and the results added to the elements of column 1. Similarly the elements of column 4 are multiplied by $-a_{21}$, the elements of column 5 are multiplied by $-a_{22}$, and the results are added to column 2. Also the elements of column 4 are multiplied by $-a_{31}$, the elements of column 5 are multiplied by $-a_{32}$, and the results added to column 3. The results give a matrix, whose determinant, Δ, equals the determinant of the original matrix, of the form

$$\begin{bmatrix} u^2_1 & 0 & 0 & a_{11} & a_{12} \\ 0 & u^2_2 & 0 & a_{21} & a_{22} \\ 0 & 0 & u^2_3 & a_{31} & a_{32} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{where } u^2_j = 1 - a^2_{j1} - a^2_{j2} ,$$

and it is immediately apparent that the determinant, Δ, of the matrix is

$$\Delta = u^2_1 u^2_2 u^2_3 .$$

By similar reasoning it can be shown that the determinant of the intercorrelation coefficients of $n + r$ rows and columns can be reduced to a determinant which is composed of four parts. The upper left entries constitute a diagonal matrix with elements u^2_j, the upper right entries constitute the factorial matrix, the lower left entries constitute the 0 matrix, while the lower right entries constitute an identity matrix. Hence

$$\Delta = u^2_1 u^2_2 \cdots u^2_r .$$

This argument can be reduced to matrix notation. If the fundamental matrix equation is $AA' = R - U^2$ and if we let $\Delta(M)$ denote the determinant of the matrix, M, then

$$\Delta\begin{bmatrix} R & A \\ A' & I \end{bmatrix} = \Delta\begin{bmatrix} R - AA' & A \\ 0 & I \end{bmatrix} = \Delta\begin{bmatrix} U^2 & A \\ 0 & I \end{bmatrix} = \Delta(U^2).$$

If Δ_{jj} is the determinant obtained by deleting variable j from the variables, then

$$\Delta_{jj} = u^2_1 u^2_2 \cdots u^2_{j-1} u^2_{j+1} \cdots u^2_r .$$

Δ_{jj} is also the cofactor of r_{jj} and $\dfrac{\Delta}{\Delta_{jj}} = u^2_j$.

If we let R_j be the multiple correlation of variable j with the n—1 remaining variables and the r factors, then

$$R_j = \sqrt{1 - \frac{\Delta}{\Delta_{jj}}} = \sqrt{1 - u^2_j} = h_j .$$

An important corollary follows at once from Theorem II. "The multiple correlation of variable j with the n—1 remaining variables is equal to or less than the square root of the communality of variable j." This, too, was given essentially by Roff (**1**, p. 4) who emphasized the fact that the multiple correlation might be used to provide an estimate of the communality. The emphasis in the present paper is on the fact that the square root of the communality, or even an approximation to it as indicated by the Thurstone method (**2**, p. 89), may be used as an upper bound for the multiple correlation. (In later papers it will be shown how multiple and partial correlations can be found from multiple factor results). In the present paper the attempt is made to answer the question, "Is it possible to find a small number, r, of variables, from which it is possible to predict each of the n—r remaining variables almost as well as though each variable were predicted from all the n—1 other variables?" In answering this question,

the square root of the communality is used as an upper limit against which one may measure the approximate loss of predictive power.

A simple problem involving only four variables has been chosen to illustrate the method, as it permits the examination of all possible multiple correlation coefficients. The method has also been applied with reasonable success to the prediction of the scores on 15 variables from the scores of four other variables (**4**).

The simple illustration is one previously used by Hotelling (**5**). The variables are, in order:

(1) memory for words
(2) memory for numbers
(3) memory for meaningful symbols
(4) memory for meaningless symbols.

In Table I are presented the intercorrelation coefficients which Hotelling obtained from T. L. Kelley, rounded to two decimal places.

TABLE I. *Intercorrelations of Four Memory Tests*

Test	1	2	3	4
1	1.00	.96	.77	.54
2	.96	1.00	.86	.70
3	.77	.86	1.00	.82
4	.54	.70	.82	1.00

With such a small number of variables it is feasible to secure all possible multiple correlation coefficients. These are given in Table II. The predicting tests are indicated in the first column.

TABLE II. *All Possible Multiple Correlations of the Four Tests*

Predicting Test		Predicted Test 1	2	3	4	Average multiple correlation, not including predicting tests
One test	1	1.00	.96	.77	.54	.76
	2	.96	1.00	.86	.70	.84
	3	.77	.86	1.00	.82	.82
	4	.54	.70	.82	1.00	.69
Two tests	1,2	1.00	1.00	.88	.84	.86
	1,3	1.00	.98	1.00	.83	.905
	1,4	1.00	.98	.90	1.00	.94
	2,3	.97	1.00	1.00	.82	.895
	2,4	.98	1.00	.91	1.00	.945
	3,4	.79	.86	1.00	1.00	.825
Three tests	1,2,3	1.00	1.00	1.00	.88	.88
	1,2,4	1.00	1.00	.91	1.00	.91
	1,3,4	1.00	.98	1.00	1.00	.98
	2,3,4	.99	1.00	1.00	1.00	.99

It appears from the above table that there is a definite increase in predictive power as one goes from one-test prediction to two-test prediction. There is also considerable variation in the predictive power of the different combinations of two tests. The best predicting combination seems to be the tests 2,4 with the combination 1,4 a close second. In the case of three-test prediction, there is a slight improvement in predictive power. Specifically, the tests 2,4 predict all the tests 1,2,3,4 almost as well as do the tests 1,2,4. That is, the inclusion of test 1 does not greatly improve the predictive power, except that of test 1 itself. The inclusion of test 3 does make some improvement.

It is possible to determine by exhaustive methods, as in the above example, just how many tests and what tests should be chosen. In the general case involving a very large number of variables, however, such exhaustive methods are out of the question.

A technique for exhaustive methods has been worked out by Frisch (**6**). As a matter of fact his Confluence Analysis should be studied closely by all who work with multiple correlations in which the predicting variables are correlated. However, his technique is hardly feasible when the number of variables is larger than 12. According to his estimate (**6**, p. 97), his "tilling," which takes only a little over four hours for a five-variable problem, would take over 565 hours for a twelve-variable problem. In such a case it is preferable, if the results can be used in obtaining the essential information, to utilize a centroid solution, particularly if, as is customary, the number of factors is appreciably less than the number of variables.

The correlation coefficients of Table I were subjected to a centroid solution. The approximation scheme, which Thurstone recommends, for placing in the diagonals at each step the absolute value of the largest entry in the appropriate row, was followed. The values of the communalities resulting are first approximations to the true communalities. These approximations are not so good with n and r small as with n and r large.

In a few minutes it was possible to show that, aside from small residuals, the intercorrelations could be explained on the basis of two factors. The factor loadings, with the corresponding h^2_j and h_j are given in Table III.

TABLE III. *The Factor Loadings — Thurstone Approximation Method*

Test	I	II	h^2_j	h_j
1	.901	.402	.973	.986
2	.968	.233	.991	.995
3	.924	—.193	.891	.944
4	.801	—.418	.816	.903

Roff suggested that the multiple correlations might be placed in the main diagonal and communalities computed. These are available from Table II. The resulting factor weights do not differ much from those of Table III.

In multiple factor theory it is important to isolate primary traits and to work out the regression equations giving the primary traits in terms of the original variables. The values of specified variables can then be obtained by providing regression equations giving the values of these variables in terms of the primary traits. It is our purpose to attain this end, prediction of actual values, more directly by obtaining regression equations giving the values of the $n - r$ variables in terms of the values of r of the original variables, hereafter called "fundamental" variables. We would be willing to sacrifice some of the predictive power if such direct methods might be made available. It is evident that this problem, as stated, demands a least squares solution and need not be translated into multiple factor analysis.

If, however, each of the primary traits of a multiple factor analysis coincided with one of the original variables, it would be possible to use the values of these fundamental variables as independent variables and to compute the values of other variables from them without the necessity of considering primary traits as such. Even though no variables are coincident with the primary traits, it may be possible to find a variable in the vicinity of the primary trait which may serve as a fundamental variable. The selection of such a fundamental variable usually results in some small loss of predictive power. An approximate estimate of the maximum loss of predictive power may be obtained by comparing the multiple correlation resulting from these fundamental variables with the square root of the communality of the predicted variable.

If simple structure is present and if at least one of the variables is near each of the primary traits, the method is quite certain to give satisfactory results. However, the demonstration of simple structure and the isolation of primary traits take considerable time.

It is sometimes possible to make a good selection of the fundamental variables from the centroid weightings themselves. Plot the coordinates of the factor loadings and determine from the graph approximately which points may serve as vertices so that the hyperplanes will come as nearly as possible to passing through the points representing the other variables. With such an informal method it is possible that different combinations of the original variables might have approximately the same predictive power. The justification is

not in a formal method of selection, but in the fact that in the results the fundamental variables account for the major portion of the possible multiple correlation.

It should be emphasized that the fundamental variables so selected are not to be considered as primary in a unique psychological sense. They are simply variables which may be substituted for the complete set of variables with little loss of predictive power. If unique meaning is desired, one should follow the method as outlined by Thurstone (**2**).

In the illustration above the factor loadings can be used as coordinates, the points representing the variables (1),(2),(3),(4).

The question then arises: Which two-point combination of the four points should be chosen? We wish to choose two points so that the line passing through these points will come as nearly as possible to passing through the other two points. Inspection reveals that the combinations 1,2 and 3,4 are unsatisfactory. These combinations were the least satisfactory combinations as revealed by Table II. Inspection also shows that the combinations 1,4 and 2,4 are among the best. This result also agrees with Table II.

As a matter of fact there is little to choose between the 1,4 and the 2,4 combinations. Actually the 2,4 combination is slightly better, as judged by the standard deviation of errors. Computing the errors, we get

$$e_{1.24} = .011 \qquad e_{2.14} = .008$$

$$e_{3.24} = .004 \qquad e_{3.14} = .009$$

so that the line through 2,4 fits the four points slightly better than the line through 1,4. This agrees with the result of Table II.

With the tests selected the problem becomes a straight least squares problem. The regression coefficients may be computed from the correlation coefficients or from the actual original scores, if available. Methods of solving the groups of regression equations which are demanded here have been outlined (**7**), (**8**).

The multiple correlation coefficients are then found by the formula

$$r_{j\cdot 12\cdots r} = \sqrt{\beta_1 r_{j1} + \beta_2 r_{j2} + \cdots + \beta_r r_{jr}}\,.$$

If there are three factors, the graphing may be done on a sphere with the use of augmented coordinates. If there are four or more factors, the first three may be plotted in three-dimensional space and the values of the loadings on the smaller factors considered in the selection of the fundamental variables. If a formal test of best selection is

desired, the distance from each point (representing a variable) to the hyperplane may be computed through the tentative fundamental vertices, and the fundamental vertices may be so chosen that the standard deviation of these distances is a minimum. In many cases, however, this mathematical work is not demanded, since the justification of the selection of the tentative variables is shown by the fact that the multiple correlation coefficient is nearly as large as the square root of the communality. In the above illustration, for example,

$$r_{1.24} = .978 \quad \text{while} \quad h_1 = .986 \ ;$$
$$r_{3.24} = .913 \quad \text{while} \quad h_3 = .944 \ .$$

SUMMARY

A method is indicated by which a centroid solution may be used in determining

1. an approximation to the greatest possible multiple correlation;
2. an indication of the number of "fundamental" variables;
3. rough methods for indicating tentative selections.

The limitations of the method are apparent. It (1) is not very formal, (2) leaves considerable to the ingenuity of the worker, and (3) provides no guarantee that the results are the best that can be obtained. Its virtue is in the fact that in many cases and with a relatively small amount of effort, it provides a reasonably satisfactory answer to the problem of reducing the number of independent variables. It permits the approximate solution of some problems which, because of the enormous amount of computation which is demanded by a more formal method, would not be attempted.

REFERENCES

1. Roff, M. "Some Properties of the Communality in Multiple Factor Theory," *Psychometrika*, June, 1935, **1**, 1-6.
2. Thurstone, L. L. *The Vectors of Mind.* Chicago: University of Chicago Press, 1935.
3. Kelley, T. L. *Statistical Method.* New York: Macmillan, 1923.
4. Dwyer, P. S. "An Analysis of 19 Occupational Scores, etc.," *Journal of Applied Psychology*, 1938, **22**, 8-16.
5. Hotelling, H. "Simplified Calculation of Principal Components," *Psychometrika*, 1936, **1**, 27-35.
6. Frisch, R. *Statistical Confluence Analysis by Means of Complete Regression Systems.* Oslo, 1934.
7. Dwyer. P. S. "The Simultaneous Computation of Groups of Regression Equations and Associated Multiple Correlation Coefficients," *Annals of Math. Stat.*, 1937, **8**, 224-231.
8. Tucker, L. R. "A Method for Finding the Inverse of a Matrix." *Psychometrika*, 1938, **3**, 189-197.