PSYCHOMETRIKA—VOL. 9, NO. 4
DECEMBER, 1944

A SIMPLE METHOD OF FACTOR ANALYSIS

KARL J. HOLZINGER
THE UNIVERSITY OF CHICAGO

A simple method for extracting correlated factors simultaneously is described. The method is based on the idea that the centroid pattern coefficients for the sections of unit rank of the complete matrix may be interpreted as structure values for the entire matrix. Only the routine centroid average process is required.

The simple method here presented is applicable to the factoring of a correlation matrix in case the latter can be sectioned into portions of approximate rank unity. Such sectioning can often be accomplished by inspection of the whole matrix, by the use of B-coefficients,* and by the nature of the variables involved.

After the matrix is sectioned, centroid coefficients for the variables in each section are computed. These coefficients may be based upon communalities or other values in the diagonals. For any *one section* the first centroid coefficients a_{js} may be interpreted as pattern values† or as *structure* values, since they show correlation between a variable z_j and the centroid C_s of a particular section. The first centroids of the various sections will be correlated, however, and the coefficients a_{js} extended to all variables may then be interpreted as structure values of s_{js} for the entire correlation matrix. It is this simple idea that is the basis of the present method.

The adequacy of the solution may be tested in each section, regarding the a_{js} of that section as pattern values which will yield pattern correlations for a given section. A complete pattern may also be found if desired in order to test for overlapping of factors in the remainder of the correlation matrix.

The reproduced correlations in a section are

$$r'_{jk} = a_{js}\, a_{ks}\,, \tag{1}$$

where the factor pattern of a section has the form,

* Karl J. Holzinger and Harry H. Harman, Factor analysis, p. 24. Chicago: University of Chicago Press, 1941.

† *Ibid.*, p. 16. A "structure" S_{js} is a matrix of correlations between tests and factors.

$$\left.\begin{array}{l} z_1 = a_{1s}C_s \\ z_2 = a_{2s}C_s \\ \\ \cdot\quad\cdot\quad\cdot\quad\cdot \\ \\ z_p = a_{ps}C_s \end{array}\right\} \begin{array}{l} j = 1, 2, 3, \cdots, p. \\ p = \text{number of variables in a section.} \\ s = 1, 2, 3, m. \\ m = \text{number of centroids.} \\ \\ \text{Specific factors not shown.} \end{array} \tag{2}$$

A simple discussion of the centroid method as applied here will next be given. Assume that the variables z_1, z_2, and z_3 are in a given section and that z_u is not. The theoretical pattern would then have the form,

$$\begin{array}{l} z_1 = a_1 C \\ z_2 = a_2 C \\ z_3 = a_3 C \\ \hline z_u = a_u C \end{array} \tag{3}$$

wherein the second subscript has been dropped for simplicity.

The correlation matrix for pattern (3) may be written

$$\begin{array}{l} R = \\ \text{Sum} \end{array} \left\| \begin{array}{ccc|c|c} a_1a_1 & a_1a_2 & a_1a_3 & & a_1a_u \\ a_2a_1 & a_2a_2 & a_2a_3 & & a_2a_u \\ a_3a_1 & a_3a_2 & a_3a_3 & & a_3a_u \\ \hline T_1 & T_2 & T_3 & T = T_1 + T_2 + T_3 & T_u \end{array} \right\| .$$

The centroid coefficient for z_1 may then be written in the form,

$$a_1 = \frac{T_1}{\sqrt{T}} = \frac{a_1(a_1 + a_2 + a_3)}{\sqrt{a_1(a_1 + a_2 + a_3) + a_2(a_1 + a_2 + a_3) + a_3(a_1 + a_2 + a_3)}}$$

$$= \frac{a_1(a_1 + a_2 + a_3)}{\sqrt{(a_1 + a_2 + a_3)^2}} = a_1 = \frac{r_{11} + r_{12} + r_{13}}{\sqrt{\text{Sum of } 9r\text{'s}}}. \tag{4}$$

The coefficient for z_u is taken as

$$a_u = \frac{T_u}{\sqrt{T}}, \tag{5}$$

employing the T for z_1, z_2, z_3 as before. It is therefore apparent that within the group a_1, a_2, a_3 are both pattern and structure values, whereas the values a_u are structure values for the whole matrix. For this reason the values from formulas (4) and (5) are denoted as s_{js} in Table 1. The structure is required in case estimates of the factors are desired. The calculations may be made by the Doolittle* or similar method.

* *Ibid.*, p. 390.

A complete outline for the calculation is given on the work sheet, using the data of Table 7.1 for eight physical variables* with bi-factor communalities from Table 8.4.† The matrix has been sectioned (1,2,3,4), (5,6,7,8), as shown on the work sheet.

After the values s_{js} have been computed, they may be arranged as in the following table:

TABLE 1
Common Structure S_{js}

Variable	s_{j1}	s_{j2}
1	.919	.484
2	.942	.434
3	.907	.399
4	.893	.454
5	.455	.932
6	.375	.813
7	.312	.739
8	.412	.724

These values agree with those of Table 11.2 within rounding error of .001.‡

The fit of the correlations for the sections yielding the factors may be obtained from the reproduced correlations at the bottom of the work sheet. The residuals in general are small.

In case the complete pattern is required, it is necessary to obtain the correlation between factors. This correlation may be found from the intercorrelations of the variables in the inter-group sections. In the present example the intercorrelations among variables in the groups (1,2,3,4) and (5,6,7,8) would be employed. The average of these sixteen values, using the subtotals from the work sheet, is

$$\frac{1.554 + 1.394 + 1.281 + 1.457}{16} = .3554.$$

This last average must be corrected, however, in order to obtain the values for the correlations in the common-factor space.||

The values $s_{11} = .919$, $s_{21} = .942$, $s_{31} = .907$, and $s_{41} = .893$ may be interpreted as the lengths of the vectors $\overline{z_1}$, $\overline{z_2}$, $\overline{z_3}$, and $\overline{z_4}$ projected on the $0C_1$ axis. Their average length is

* *Ibid.*, p. 169.

† *Ibid.*, p. 192.

‡ *Ibid.*, p. 245.

|| *Ibid.*, p. 61. See formula 3.50, which is here applied to averages of variables.

$$\frac{.919 + .942 + .907 + .893}{4} = .9152.$$

Similarly the average length of the vectors z_5, z_6, z_7, and z_8 projected on the $0C_2$ axis is

$$\frac{.932 + .813 + .739 + .724}{4} = .8020.$$

The correlation between factors C_1 and C_2 in the common-factor space is then

$$r_{C_1 C_2} = \frac{.3554}{.9152 \times .8020} = .4842.$$

From the structure S_{js} and the correlation between factors ϕ_{ss} the pattern may be found by the method of Appendix G.3.*

TABLE 2
Common Pattern A_{js}

Variable	a_{j1}	a_{j2}
1	.894	.051
2	.956	—.029
3	.932	—.052
4	.879	.029
5	.005	.930
6	—.024	.824
7	—.060	.768
8	.080	.685

It is now possible to check the fit of the whole observed correlation matrix by obtaining the complete matrix of reproduced correlations. Denoting this matrix as R^+, then

$$R^+_{jj} = A_{js}\phi_{ss}A'_{js}, \qquad (6)\dagger$$

but since

$$S_{js} = A_{js}\phi_{ss}, \qquad (7)\ddagger$$

then

$$S'_{js} = \phi_{ss}A'_{js} \qquad (8)$$

and

$$R^+ = A_{js}S'_{js} = S_{js}A'_{js}. \qquad (9)$$

* *Ibid.*, p. 386.
† *Ibid.*, p. 19.
‡ *Ibid.*, p. 327. (Here, T instead of S denotes structure.)

This last equation is very convenient for obtaining the reproduced correlations inasmuch as the correlation ϕ_{ss} is not explicitly required. From Tables 1 and 2 the product $S_{js}A'_{js}$ is found to be

$$\left\| \begin{matrix} .846 & .865 & .831 & .822 & .455 & .377 & .317 & .405 \\ .864 & .888 & .855 & .841 & .408 & .335 & .277 & .373 \\ .831 & .856 & .825 & .809 & .376 & .307 & .252 & .346 \\ .821 & .841 & .809 & .798 & .427 & .353 & .295 & .382 \\ .454 & .408 & .376 & .427 & .869 & .757 & .688 & .675 \\ .377 & .335 & .307 & .353 & .758 & .661 & .602 & .587 \\ .317 & .277 & .252 & .296 & .689 & .601 & .549 & .531 \\ .405 & .373 & .346 & .383 & .675 & .587 & .531 & .529 \end{matrix} \right\|$$

Upon comparing the original correlation matrix given on the work sheet with the above reproduced correlations, it is apparent that all the residuals are negligible. The fit of the pattern is therefore an excellent one.

It will be observed that the present method obviates the usual procedure of first obtaining an orthogonal centroid solution for the entire correlation matrix, and then rotating to oblique axes satisfying what Thurstone would call a special case of "simple structure." Such "simple structure" is tested here analytically by checking the rank of the submatrixes as shown at the bottom of the work sheet, and by testing the goodness of fit of the entire correlation matrix from equation (9). If large residuals occur in either case, the variables may be rearranged and the correlation matrix resectioned to secure a better fit. There is no guarantee, of course, that any correlation matrix can be factored in the above manner, but if either a bi-factor pattern or this type of "simple structure" exists, then the above method is applicable and is much more simple and direct than those now in use.

WORK SHEET FOR SIMPLE METHOD

	1	2	3	4		5	6	7	8
1	.854	.846	.805	.859		.473	.398	.301	.382
2	.846	.897	.881	.826		.376	.326	.277	.415
3	.805	.881	.833	.801		.380	.319	.237	.345
4	.859	.826	.801	.783		.436	.329	.327	.365
Sum	3.364	3.450	3.320	3.269	$T = 13.403$	1.665	1.372	1.142	1.507
a_{1j}	.919	.942	.907	.893	$\sqrt{T} = 3.661$	.455	.375	.312	.412
					$\frac{1}{\sqrt{T}} = .2731$				

5	.473	.376	.380	.436		.870	.762	.730	.629	
6	.398	.326	.319	.329		.762	.687	.583	.577	
7	.301	.277	.237	.327		.730	.583	.521	.539	
8	.382	.415	.345	.365		.629	.577	.539	.579	$T = 10.297$
Sum	1.554	1.394	1.281	1.457		2.991	2.609	2.373	2.324	$\sqrt{T} = 3.209$
	.484	.434	.399	.454	a_{2j}	.932	.813	.739	.724	$\frac{1}{\sqrt{T}} = .3116$

	a_{j1}	$\times$	a'_{j1}				$=$		R'			
1	.919	$\times$	.919	.942	.907	.893	$=$	1	.845	.866	.834	.821
2	.942							2	.866	.887	.854	.841
3	.907							3	.834	.854	.823	.810
4	.893							4	.821	.841	.810	.797

	a_{j2}	$\times$	a'_{j2}				$=$		R'			
5	.932	$\times$	.932	.813	.739	.724	$=$	5	.869	.758	.689	.675
6	.813							6	.758	.661	.601	.589
7	.739							7	.689	.601	.546	.535
8	.724							8	.675	.589	.535	.524