DOCUMENT RESUME

ED 056 075 TM 000 878

AUTHOR Thorndike, Robert M.
TITLE Method of Factor Extraction and Simple Structure of Data from Diverse Scientific Areas.
INSTITUTION Western Washington State Coll., Bellingham.
PUB DATE Sep 71
NOTE 13p.; Paper presented at the Annual Meeting of the Western Psychological Association, San Francisco, California, April 1971

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Correlation; Criteria; *Factor Analysis; *Factor Structure; Mathematical Models; *Orthogonal Rotation; *Sciences; Transformations (Mathematics)

ABSTRACT

To study the applicability of simple structure logic for factorial data from scientific disciplines outside psychology, four correlation matrices from each of six scientific areas were factor analyzed by five factoring methods. Resulting factor matrices were compared on two objective criteria of simple structure before and after rotation. Factoring methods differed in their ability to provide a simple structure for a given set of data. Scientific disciplines differed in the extent to which they provided data which could be factor analyzed, but not in the simple structure of factors obtained. Varimax rotation did not systematically improve simple structure. (Author)

Method of Extraction and Simple Structure of

Data from Diverse Scientific Areas

Robert M. Thorndike

Western Washington State College

To study the applicability of simple structure logic for factorial data from scientific disciplines outside psychology, four correlation matrices from each of six scientific areas were factor analyzed by five factoring methods. Resulting factor matrices were compared on two objective criteria of simple structure before and after rotation. Factoring methods differed in their ability to provide a simple structure for a given set of data. Scientific disciplines differed in the extent to which they provided data which could be factor analyzed, but not in the simple structure of factors obtained. Varimax rotation did not systematically improve simple structure.

Method of Factor Extraction and Simple Structure

of Data from Diverse Scientific Areas

Robert M. Thorndike

Western Washington State College

The search for simple structure in sets of variables has assumed an important position psychological applications of factor analysis. It is reasonable to question whether the concept of simple structure deserves such a lofty place in factor analytic practice. This question has become increasingly appropriate with the growing use of factor analysis by researchers in disciplines outside psychology. Can the logical procedures which psychologists have developed for their own problems be applied to data from other areas of science? The present study examines several issues germain to the general use of factor analysis and the logic of simple structure.

## Method

To study the effects of changes in method of factor extraction and source of data on simple structure, five factoring models were applied to data from six scientific areas. The five methods and the type of reduction of the correlation matrix used by each are described in detail elsewhere (Thorndike, 1970). Briefly, the five methods were principal components with Kaiser criterion, minimum residuals, maximum likelihood, image covariance, and alpha. Data were obtained from the literature in the areas of medicine, economics, ability measurement, personality measurement, sociology, and taxonomy. Four examples of data from each area were analyzed by each factor model.

### Criteria of Simple Structure

Objectification of simple structure has been a problem for factor analysts since Thurstone proposed that factors should be rotated to a simple structure solution. The proportion of variables located in the hyperplane (the hyperplane count) has traditionally been used as a criterion. However, this method tends to locate, define, and evaluate factors by what they are not.

A revision of the simple structure model which is based on positive information and which provides an objective index of fit to the model has recently been proposed (Thorndike, 1971). Briefly, the revision states that maximum simplicity of structure is obtained when each variable loads on only one of the factors and all of its other loadings are zero. The index of goodness-of-fit to this model is given by

$$G = \sum_{j=1}^{m} \left[ 2 \sum_{i=1}^{m} \frac{(h_i - |a_{ij}|) + a_{ij}^2}{\sum_{i=1}^{n} h_i^2} - 1 \right] \qquad (1)$$

where $h_i^2$ = the communality of variable $i$

and $a_{ij}$ = the factor loading of variable $i$ on factor $j$

It has a range of 0 to 1.0 and gives at least ordinal information. The results from applying this index and the hyperplane count to the 24 sets of data for each of the analyses are given in Tables 1 through 3.

## Results

Tables 1 and 2 contain the values of the criteria for adequacy of approximation to simple structure which were described above. The values of G and proportion of hyperplane loadings in Table 1 are for the unrotated matrices while those in Table 2 are for the matrices after varimax rotation.

---

Insert Table 1 about here

---

The omissions in the tables indicate that a solution could not be obtained for that set of data by that method of extraction (for a discussion of these results see Thorndike, 1970). It is interesting that only the psychological areas yield data which are generally analyzable by all methods. We may also note that the principal components and minimum residual methods provide solutions for all cases.

The first aspect of the data which is apparent in both tables is that, with the exception of the results for some image matrices, the value of G always exceeds the proportion of variable loadings in the hyperplane. The reason that the hyperplane count exceeds G for some of the image matrices appears to be that one or more very small factors are retained by the decision rule used for the image analyses. These small fac[illegible] air all or nearly all hyperplane loadings, resulting in a substantially higher hyperplane count for the image matrices than for those of any other method. In every case at least half of the image loadings before rotation are in the hyperplane. The values for G are also unusually high, but they have not been raised as much as the hyperplane count by the inclusion of essentially null factors.

Close attention to Table 1 reveals some interesting relationships among the factor methods. In comparing principal components with minres on the value of G, it may be seen that minres is superior in 18 of the 24 cases, indicating that minres provides a better fit to the revised simple structure model than principal components does. The same conclusion is reached in regard to the traditional simple structure model and its criterion, the hyperplane count. Minres is superior in 20 of 23 cases and there is one tie. Expanding the comparison to include alpha, the three methods can be ordered on the sums of their ranks on the criteria for the 17 cases providing data. Again it will be seen that minres provides the best fit to either simple structure model. Alpha performs less well than minres, but better than principal ocmponents. When all five methods are compared for the 11 relevant cases, image is seen to give the best solution in almost every case. On the basis of the two criteria to fit to the simple structure models it is definitely superior to any other method. However, this finding is also an artifact of the tendency of the image method to retain null factors, as pointed out before. (It is worth noting that null factors do not appear to distort G as much as they do the hyperplane count.) A more interesting finding is that maximum likelihood ranks second, slightly ahead of minres. The two statistical methods, maximum likelihood and minres, provide the best fit to simple structure when the spuriously inflated values of image analysis are ignored. Alpha and prinicpal components retain their relative positions.

-------------------------

Insert Table 2 about here

-------------------------

From a comparison of Table 1 with Table 2 it may be seen that the various factor methods are differentially affected by varimax rotation. The superiority of minres over principal components is less marked for the rotated matrices than for the unrotated matrices on both criteria, with the G index showing a greater gain than the hyperplane count for the principal components method. On the second comparison there is little difference between minres and principal components on G, and alpha is decidedly third, while the hyperplane count orders the methods: minres, alpha, principal components. With all methods included, image analysis remains superior for the same reason cited above. According to G, principal componenets and maximum likelihood tie for second, minres is next, and alpha performs most poorly. The hyperplane count suggests a different ordering. Image remains first, followed by minres and maximum likelihood, tied. Alpha is next and principal components is least satisfactory. The primary differential effect of rotation is upon the fit of the principal components matrices to the revised simple structure model. Rotation improves principal components relative to the other methods. Why this should be so is not immediately clear. It is noteworthy that the effect is observed only for the revised simple structure model.

Aside from their relative magnitudes and fairly consistent ordering of the methods with regard to adequacy, there does not appear to be a consistent relationship between the two criteria. In only about 75% of cases did they agree on which initial solution was best from a simple structure viewpoint, and half of those agreements were for the image solution, which has a tendency to inflate both criteria because of the presence of null factors. Essentially the same rates of agreement were found for the rotated matrices and these agreements could also be attributed largely to the presence of the image solution. When the values for the image matrices are omitted from consideration, the rate of agreement drops below 50%. Thus, the two indices do not appear to agree at much above a chance level. It should be noted, however, that often the index values for different methods were not very different. Values of the two criteria may be differently affected by changes in the number of factors, possibly accounting for the lack of agreement.

Rotation

The effect of rotation on the values of the two indices is shown in Table 3. Entries are the amount of increase in each index due to rotation, negative values indicating a decrease in the index. It is obvious there is no consistent effect of rotation on either G or the hyperplane count.

Insert Table 3 about here

However, both indices generally agree on the effect of rotation for a single matrix and usually also agree for all matrices for a given set of data. There is no consistent effect of rotation for all of the data from any scientific area. Even in the psychological areas, where an analytic rotation is often taken as the final solution, there is no consistent tendency for varimax rotation to improve on the simple structure of the initial solution as it is evaluated by the criteria used in this study.

The only relatively consistent finding which appears in Table 3 is that variamax rotation, with one exception, makes the simple structure of factors obtained from image analysis worse. This is probably due to the fact that most of the original image matrices included several null factors which would yield inflated values for both criteria. A varimax rotation would tend to build up slightly these null factors, reducing the number of hyperplane loadings and causing a decrease in G because of the increasing number of non-zero and non-maximum loadings. The other methods, which do not generally retain null factors, do not show this consistent decrease in simple structure as a result of rotation.

There is a slight tendency for data from non-psychological areas to show better simple structure than do psychological data. This tendency is found for both criteria and it may therefore be inferred that the logic of simple structure is applicable to data from all areas of science included in this study. However, the most important observation which can be made is that the correlation matrices resulting from non-psychological areas of science, cannot, in general, be analyzed by all factoring methods. The researcher in non-psychological areas would probably be well advised to select a principal components solution, even though it may not result in an optimum simple structure.

## Discussion

There are few empirical grounds in this study on which to compare the two simple structure criteria. However, the logic by which they were derived does permit some evaluation of the information which may be conveyed. The hyperplane count was derived from Thurstone's criteria for rotation to simple structure. It yields information on the proportion of loadings which are within some specified range of zero. The assumption is that the remaining loadings will be large and meaningful. However, basing a criterion for adequacy of the obtained solution on negative information seems a questionable practice, just as rotating by finding hyperplanes with the maximum number of informationless variables is questionable.

By contrast, the proposed simple structure index, G, and its associated model have logical appeal because they treat as the simplest structure that structure which is most simple. The model is readily quantified in a manner which makes maximum use of the information, both positive and negative. Both the revised model and the index, G, give meaningful results for cases where a single factor is the appropriate solution. Also, and perhaps more importantly, values of G do not seem to be distorted q.ite so much by the image method's tendency to retain null factors.

Table I

Values of G-Index and Hyperplane Count before Rotation, by Factor Method

| | G | | | | | Hyperplane Count | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Matrix | Principal Components | Minimum Residual | Maximum Likelihood | Image | Alpha | Principal Components | Minimum Residual | Maximum Likelihood | Image | Alpha |
| CARD-1 | .330 | .388 | .452 | .442 | .288 | .248 | .311 | .393 | .506 | .238 |
| CARD-2 | .465 | .520 | .482 | .700 | .524 | .219 | .381 | .194 | .657 | .292 |
| CARD-3 | .471 | .521 | * | * | * | .335 | .435 | * | * | * |
| CARD-4 | .471 | .467 | * | * | * | .351 | .343 | * | * | * |
| ECON-1 | .496 | .477 | .585 | .707 | .451 | 0.000 | .179 | .143 | .515 | 0.000 |
| ECON-2 | .613 | .621 | * | * | .593 | .250 | .437 | * | * | .250 |
| ECON-3 | .693 | .628 | * | * | .670 | .196 | .326 | * | * | .174 |
| ECON-4 | .320 | .466 | * | * | .280 | .150 | .340 | * | * | .100 |
| ABIL-1 | .323 | .383 | .330 | .681 | .387 | .224 | .283 | .152 | .739 | .290 |
| ABIL-2 | .420 | .486 | .445 | .764 | .471 | .100 | .289 | .300 | .733 | .133 |
| ABIL-3 | .260 | .297 | .377 | .579 | .319 | .167 | .200 | .360 | .590 | .167 |
| ABIL-4 | .391 | .492 | * | * | * | .260 | .375 | * | * | * |
| PERS-1 | .310 | .347 | .432 | .584 | .355 | .231 | .231 | .298 | .584 | .231 |
| PERS-2 | .397 | .462 | .468 | .722 | .444 | .093 | .244 | .266 | .667 | .129 |
| PERS-3 | .394 | .463 | .517 | .646 | .418 | .286 | .399 | .413 | .654 | .386 |
| PERS-4 | .304 | 1.000 | 1.000 | .597 | .383 | .260 | .176 | .588 | .675 | .277 |
| SOC-1 | .468 | .512 | * | * | .462 | .029 | .294 | * | * | .029 |
| SOC-2 | .631 | .505 | .320 | .695 | .585 | .111 | .250 | .250 | .510 | .167 |
| SOC-3 | .315 | .375 | * | * | .336 | .167 | .067 | * | * | .167 |
| SOC-4 | .399 | .475 | * | * | .444 | .167 | .351 | * | * | .224 |
| TYPO-1 | .641 | .602 | * | * | * | .179 | .310 | * | * | * |
| TYPO-2 | 1.000 | .721 | * | * | * | 0.000 | .250 | * | * | * |
| TYPO-3 | .469 | .500 | * | * | * | .300 | .375 | * | * | * |
| TYPO-4 | .352 | .374 | * | * | * | .200 | .252 | * | * | * |

Table 2

Values of G-Index and Hyperplane Count after Rotation, by Factor Method

| | G | | | | | Hyperplane Count | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Matrix | Principal Components | Minimum Residual | Maximum Likelihood | Image | Alpha | Principal Components | Minimum Residual | Maximum Likelihood | Image | Alpha |
| CARD-1 | .613 | .649 | .654 | .650 | .601 | .475 | .570 | .592 | .643 | .505 |
| CARD-2 | .351 | .352 | .236 | .552 | .288 | .208 | .196 | .069 | .495 | .188 |
| CARD-3 | .600 | .666 | * | * | * | .461 | .558 | * | * | * |
| CARD-4 | .594 | .593 | * | * | * | .505 | .540 | * | * | * |
| ECON-1 | .668 | .604 | .676 | .700 | .685 | .146 | .250 | .143 | .457 | .214 |
| ECON-2 | .545 | .596 | * | * | .539 | .181 | .410 | * | * | .167 |
| ECON-3 | .568 | .471 | * | * | .469 | .065 | .152 | * | * | .043 |
| ECON-4 | .473 | .561 | * | * | .463 | .238 | .405 | * | * | .225 |
| ABIL-1 | .479 | .439 | .471 | .668 | .449 | .355 | .272 | .294 | .711 | .377 |
| ABIL-2 | .476 | .280 | .290 | .602 | .377 | .100 | .067 | .133 | .572 | 0.000 |
| ABIL-3 | .442 | .437 | .437 | .575 | .412 | .167 | .276 | .266 | .552 | .200 |
| ABIL-4 | .443 | .512 | * | * | * | .280 | .395 | * | * | * |
| PERS-1 | .465 | .596 | .582 | .625 | .468 | .308 | .442 | .413 | .530 | .338 |
| PERS-2 | .297 | .268 | .272 | .640 | .259 | .018 | .078 | .100 | .606 | .019 |
| PERS-3 | .545 | .616 | .602 | .669 | .541 | .417 | .573 | .538 | .697 | .474 |
| PERS-4 | .523 | 1.000 | 1.000 | .612 | .556 | .386 | .176 | .588 | .722 | .495 |
| SOC-1 | .430 | .447 | * | * | .425 | .029 | .235 | * | * | .029 |
| SOC-2 | .634 | .487 | .460 | .550 | .570 | .111 | .278 | .250 | .466 | .111 |
| SOC-3 | .455 | .557 | * | * | .431 | .234 | .167 | * | * | .266 |
| SOC-4 | .448 | .464 | * | * | .447 | .167 | .283 | * | * | .197 |
| TYPO-1 | .394 | .225 | * | * | * | .036 | .048 | * | * | * |
| TYPO-2 | 1.000 | .266 | * | * | * | 0.000 | 0.000 | * | * | * |
| TYPO-3 | .514 | .566 | * | * | * | .300 | .469 | * | * | * |
| TYPO-4 | .500 | .544 | * | * | * | .288 | .410 | * | * | * |

Table 3

Changes in G-Index and Hyperplane Count due to Rotation, by Factor Method

| Matrix | G: Principal Components | G: Minimum Residual | G: Maximum Likelihood | G: Image | G: Alpha | Hyperplane Count: Principal Components | Hyperplane Count: Minimum Residual | Hyperplane Count: Maximum Likelihood | Hyperplane Count: Image | Hyperplane Count: Alpha |
|---|---|---|---|---|---|---|---|---|---|---|
| CARD-1 | .28 | .26 | .20 | .21 | .31 | .23 | .26 | .20 | .14 | .26 |
| CARD-2 | -.15 | -.17 | -.25 | -.15 | -.23 | -.01 | -.18 | -.12 | -.16 | -.10 |
| CARD-3 | .13 | .05 | -- | -- | -- | .13 | .12 | -- | -- | -- |
| CARD-4 | .12 | .13 | -- | -- | -- | .15 | .13 | -- | -- | -- |
| ECON-1 | .17 | .12 | .09 | -.01 | .23 | .15 | .07 | .00 | -.06 | .21 |
| ECON-2 | -.13 | -.02 | -- | -- | -.05 | -.07 | -.03 | -- | -- | -.08 |
| ECON-3 | -.12 | -.16 | -- | -- | -.20 | [illegible] | -.17 | -- | -- | -.13 |
| ECON-4 | .15 | .09 | -- | -- | .18 | [illegible] | .07 | -- | -- | .12 |
| ABIL-1 | .16 | .06 | .14 | -.01 | .08 | .13 | -.01 | .14 | -.03 | .09 |
| ABIL-2 | .06 | -.21 | -.15 | -.16 | -.09 | .00 | -.22 | -.17 | -.16 | -.13 |
| ABIL-3 | .18 | .14 | .06 | .00 | .09 | .00 | .08 | -.09 | -.04 | .03 |
| ABIL-4 | .05 | .02 | -- | -- | -- | .02 | .02 | -- | -- | -- |
| PERS-1 | .16 | .25 | .15 | .04 | .11 | .18 | .21 | .11 | -.05 | .11 |
| PERS-2 | -.10 | -.19 | -.19 | -.08 | -.18 | -.07 | -.16 | -.17 | -.06 | -.11 |
| PERS-3 | .15 | .15 | .08 | .02 | .12 | .13 | .17 | .13 | .04 | .09 |
| PERS-4 | .22 | .00 | .00 | .01 | .28 | .13 | .00 | .00 | .04 | .22 |
| SOC-1 | -.04 | -.05 | -- | -- | -.04 | .00 | -.06 | -- | -- | .00 |
| SOC-2 | .00 | -.02 | .14 | -.14 | -.02 | .00 | -.03 | .00 | -.04 | -.05 |
| SOC-3 | .14 | .18 | -- | -- | .10 | .06 | .10 | -- | -- | .10 |
| SOC-4 | .05 | -.01 | -- | -- | .00 | .00 | -.07 | -- | -- | -.03 |
| TYPO-1 | -.25 | -.38 | -- | -- | -- | -.14 | -.26 | -- | -- | -- |
| TYPO-2 | .00 | -.45 | -- | -- | -- | .00 | -.25 | -- | -- | -- |
| TYPO-3 | .04 | .07 | -- | -- | -- | .00 | .09 | -- | -- | -- |
| TYPO-4 | .15 | .17 | -- | -- | -- | .08 | .16 | -- | -- | -- |

## References

Thorndike, R.M. Method of extraction, type of data, and adequacy of solutions in factor analysis. Dissertation Abstracts, 1970, p. 2970-B.

Thorndike, R.M. Simple structure: A revision and an objective criterion. Paper presented to the Psychometric Society, April 8, 1971. (mimeograph)