Factor Rotations in Factor Analyses.

Hervé Abdi[1]

The University of Texas at Dallas

INTRODUCTION

The different methods of factor analysis first extract a set a factors from a data set. These factors are almost always orthogonal and are ordered according to the proportion of the variance of the original data that these factors explain. In general, only a (small) subset of factors is kept for further consideration and the remaining factors are considered as either irrelevant or nonexistent (i.e., they are assumed to reflect measurement error or noise).

In order to make the interpretation of the factors that are considered relevant, the first selection step is generally followed by a rotation of the factors that were retained. Two main types of rotation are used: orthogonal when the new axes are also orthogonal to each other, and oblique when the new axes are not required to be orthogonal to each other. Because the rotations are always performed in a subspace (the so-called factor space), the new axes will always explain *less* variance than the original factors (which are computed to be optimal), but obviously the part of variance explained by the total subspace after rotation is the same as it was before rotation (only the partition of the variance has changed). Because the rotated axes are not defined according to a statistical criterion, their *raison d'être* is to facilitate the interpretation.

In this article, I illustrate the rotation procedures using the loadings of variables analyzed with principal component analysis (the so-called R-mode), but the methods described here are valid also for other types of analysis and when analyzing the subjects' scores (the so-called Q-mode).

Before proceeding further, it is important to stress that because the rotations always take place in a subspace (i.e., the space of the retained factors), the choice of this subspace strongly influences the result of the rotation. Therefore, in the practice of rotation in factor analysis, it is strongly recommended to try several sizes for the subspace of the retained factors in order to assess the robustness of the interpretation of the rotation.

NOTATIONS

To make the discussion more concrete, I will describe rotations within the principal component analysis (PCA) framework. PCA starts with a data matrix denoted $\mathbf{Y}$ with I rows and J columns, where each row represents a unit (in general subjects) described by J measurements which are almost always expressed as Z-scores. The data matrix is then decomposed into scores (for the subjects) or components and loadings for the variables (the loadings are, in general, correlations between the original variables and the components extracted by the analysis). Formally, PCA is equivalent to the singular value decomposition of the data matrix as:

$$\mathbf{Y} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^T \tag{1}$$

with the constraints that $\mathbf{Q}^T\mathbf{Q} = \mathbf{P}^T\mathbf{P} = \mathbf{I}$ (where $\mathbf{I}$ is the identity matrix), and $\boldsymbol{\Delta}$ being a diagonal matrix with the so-called singular values on the diagonal. The orthogonal matrix $\mathbf{Q}$ is the matrix of loadings (or projections of the original variables onto the components), where one row stands for one of the original variables and one column for one of the new factors. In general, the subjects' score matrix is obtained as $\mathbf{F} = \mathbf{P}\boldsymbol{\Delta}$ (some authors use $\mathbf{P}$ as the scores).

ROTATING: WHEN AND WHY?

Most of the rationale for rotating factors comes from Thurstone (1947) and Cattell (1978) who defended its use because this procedure simplifies the factor structure and therefore makes its interpretation easier and more reliable (i.e., easier to replicate with different data samples).

Thurstone suggested five criteria to identify a simple structure. According to these criteria, still often reported in the literature, a matrix of loadings (where the rows correspond to the original variables and the columns to the factors) is simple if:

1. each row contains at least one zero;

2. for each column, there are at least as many zeros as there are columns (i.e., number of factors kept);

3. for any pair of factors, there are some variables with zero loadings on one factor and large loadings on the other factor;

4. for any pair of factors, there is a sizable proportion of zero loadings;

5. for any pair of factors, there is only a small number of large loadings.

Rotations of factors can (and used to) be done graphically, but are mostly obtained analytically and necessitate to specify mathematically the notion of simple structure in order to implement it with a computer program.

ORTHOGONAL ROTATION

An orthogonal rotation is specified by a rotation matrix denoted $\mathbf{R}$, where the rows stand for the original factors and the columns for the new (rotated) factors. At the intersection of row m and column n we have the cosine of the

angle between the original axis and the new one: $r_{m,n} = \cos\theta_{m,n}$. For example the rotation illustrated in Figure 1 will be characterized by the following matrix:

$$\mathbf{R} = \begin{bmatrix} \cos\theta_{1,1} & \cos\theta_{1,2} \\ \cos\theta_{2,1} & \cos\theta_{2,2} \end{bmatrix} = \begin{bmatrix} \cos\theta_{1,1} & -\sin\theta_{1,1} \\ \sin\theta_{1,1} & \cos\theta_{1,1} \end{bmatrix} , \tag{2}$$

with a value of $\theta_{1,1} = 15$ degrees. A rotation matrix has the important property of being orthonormal because it corresponds to a matrix of direction cosines and therefore $\mathbf{R}^T\mathbf{R} = \mathbf{I}$.
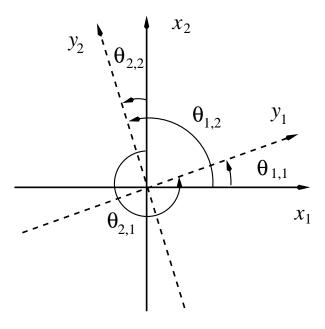

Figure 1: An orthogonal rotation in 2 dimensions. The angle of rotation between an old axis m and a new axis n is denoted by $\theta_{m,n}$.

Varimax

VARIMAX, which was developed by Kaiser (1958), is indubitably the most popular rotation method by far. For VARIMAX a simple solution means that each factor has a small number of large loadings and a large number of zero (or small) loadings. This simplifies the interpretation because, after a VARIMAX rotation, each original variable tends to be associated with one (or a small number) of factors, and each factor represents only a small number of variables. In addition, the factors can often be interpreted from the opposition of few variables with positive loadings to few variables with negative loadings.

Formally VARIMAX searches for a rotation (i.e., a linear combination) of the original factors such that the *variance* of the loadings is maximized, which amounts to maximizing

$$\mathcal{V} = \sum \left(q_{j,\ell}^2 - \bar{q}_{j,\ell}^2 \right)^2 , \tag{3}$$

3

Table 1: An (artificial) example for PCA and rotation. Five wines are described by seven variables.

	Hedonic	For meat	For dessert	Price	Sugar	Alcohol	Acidity
Wine 1	14	7	8	7	7	13	7
Wine 2	10	7	6	4	3	14	7
Wine 3	8	5	5	10	5	12	5
Wine 4	2	4	7	16	7	11	3
Wine 5	6	2	4	13	3	10	3

Table 2: Wine example: Original loadings of the seven variables on the first two components.

	Hedonic	For meat	For dessert	Price	Sugar	Alcohol	Acidity
Factor 1	-0.3965	-0.4454	-0.2646	0.4160	-0.0485	-0.4385	-0.4547
Factor 2	0.1149	-0.1090	-0.5854	-0.3111	-0.7245	0.0555	0.0865

with $q_{j,\ell}^2$ being the squared loading of the jth variable on the ℓ factor, and $\bar{q}_{j,\ell}^2$ being the mean of the squared loadings (for computational stability, each of the loadings matrix is generally scaled to length one prior to the minimization procedure).

Other orthogonal rotations

There are several other methods for orthogonal rotation such as the QUARTIMAX rotation, which minimizes the number of factors needed to explain each variable, and the EQUIMAX rotation which is a compromise between VARIMAX and QUARTIMAX. Other methods exist, but none approaches VARIMAX in popularity.

An example of orthogonal VARIMAX rotation

To illustrate the procedure for a VARIMAX rotation, suppose that we have 5 wines described by the average rating of a set of experts on their hedonic dimension, how much the wine goes with dessert, how much the wine goes with meat; each wine is also described by its price, its sugar and alcohol content, and its acidity. The data are given in Table 1.

A PCA of this table extracts four factors (with eigenvalues of 4.7627, 1.8101, 0.3527, and 0.0744, respectively), and a 2 factor solution (corresponding to the components with an eigenvalue larger than unity) explaining 94% of the variance is kept for rotation. The loadings are given in Table 2.

The VARIMAX rotation procedure applied to the table of loadings gives a clockwise rotation of 15 degrees (corresponding to a cosine of .97). This gives the new set of rotated factors shown in Table 3. The rotation procedure is illustrated in Figures 2 to 4. In this example, the improvement in the simplicity of the

Table 3: Wine example: Loadings, after VARIMAX rotation, of the seven variables on the first two components.

	Hedonic	For meat	For dessert	Price	Sugar	Alcohol	Acidity
Factor 1	−0.4125	−0.4057	−0.1147	0.4790	0.1286	−0.4389	−0.4620
Factor 2	0.0153	−0.2138	−0.6321	−0.2010	−0.7146	−0.0525	−0.0264

interpretation is somewhat marginal, maybe because the factorial structure of such a small data set is already very simple. The first dimension remains linked to price and the second dimension now appears more clearly as the dimension of sweetness.
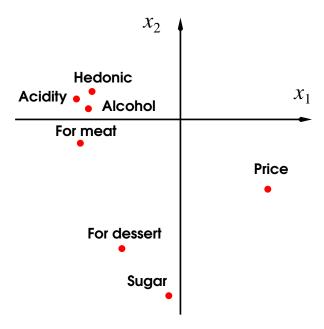

Figure 2: PCA: Original loadings of the seven variables for the wine example.

OBLIQUE ROTATIONS

In oblique rotations the new axes are free to take any position in the factor space, but the degree of correlation allowed among factors is, in general, small because two highly correlated factors are better interpreted as only one factor. Oblique rotations, therefore, relax the orthogonality constraint in order to gain simplicity in the interpretation. They were strongly recommended by Thurstone, but are used more rarely than their orthogonal counterparts.
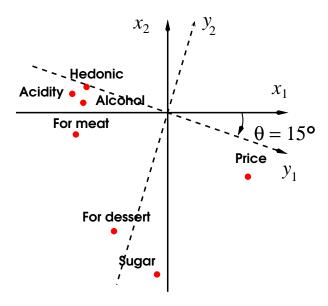
Figure 3: PCA: The loading of the seven variables for the wine example showing the original axes and the new (rotated) axes derived from VARIMAX.

Table 4: Wine example. PROMAX oblique rotation solution: correlation of the seven variables on the first two components.

	Hedonic	For meat	For dessert	Price	Sugar	Alcohol	Acidity
Factor 1	−0.8783	−0.9433	−0.4653	0.9562	0.0271	−0.9583	−0.9988
Factor 2	−0.1559	−0.4756	−0.9393	−0.0768	−0.9507	−0.2628	−0.2360

For oblique rotations, the PROMAX rotation has the advantage of being fast and conceptually simple. Its name derives from PROCRUSTEAN rotation because it tries to fit a *target* matrix which has a simple structure. It necessitates two steps. The first step defines the target matrix, almost always obtained as the result of a VARIMAX rotation whose entries are raised to some power (typically between 2 and 4) in order to "force" the structure of the loadings to become bipolar. The second step is obtained by computing a least square fit from the VARIMAX solution to the target matrix.

Even though, in principle, the results of oblique rotations could be presented graphically they are almost always interpreted by looking at the correlations between the rotated axis and the original variables. These correlations are interpreted as loadings.

For the wine example, using as target the VARIMAX solution raised to the 4th power, for a PROMAX rotation gives the loadings shown in Table 4. From these
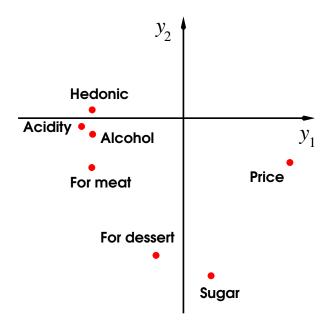
Figure 4: PCA: The loadings after VARIMAX rotation of the seven variables for the wine example.

loadings one can find that, the first dimension once again isolates the price (positively correlated to the factor) and opposes it to the variables: alcohol, acidity, "going with meat," and hedonic component. The second axis shows only a small correlation with the first one ($r = -.01$). It corresponds to the "sweetness/going with dessert" factor. Once again, the interpretation is quite similar to the original PCA because the small size of the data set favors a simple structure in the solution.

Oblique rotations are much less popular than their orthogonal counterparts, but this trend may change as new techniques, such as *independent component analysis*, are developed. This recent technique, originally created in the domain of signal processing and NEURAL NETWORKS, derives directly from the data matrix an oblique solution that maximizes statistical independence.

References

[1] Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: Wiley.
[2] Kim, J.O. & Mueller, C.W. (1978). *Factor analysis: statistical methods and practical issues*. Thousand Oaks (CA): Sage Publications.
[3] Kline, P. (1994). *An easy guide to factor analysis*. Thousand Oaks (CA): Sage Publications.

[4] Mulaik, S.A. (1972). *The foundations of factor analysis.* New York: McGraw-Hill.

[5] Nunnally, J.C. & Berstein, I.H. (1994). *Psychometric theory (3rd Edition).* New York: McGraw-Hill.

[6] Reyment, R.A. & Jöreskog, K.G. (1993). *Applied factor analysis in the natural sciences.* Cambridge: Cambridge University Press.

[7] Rummel, R.J. (1970). *Applied factor analysis.* Evanston: Northwestern University Press.