LI Analysis Training Series

Exploratory Factor Analysis with alpha method and varimax rotation.

(Last Revised: 01/05/00)

Melissa McDermeit, Rodney Funk, Mark Foss & Michael Dennis
Chestnut Health Systems
Bloomington IL 61701
309-827-6026
www.chestnut.org

Acknowledgement: This document was developed under grant No. TI11320 from the Center for Substance Abuse Treatment (CSAT) of the Substance Abuse and Mental Health Services Administration (SAMHSA). The opinions expressed here are solely those of the authors and do not reflect official positions of the U.S. Government.

Purpose: One of the uses of Factor Analysis is for the development and validation of scales within an inventory or test battery. Factor Analysis can be used to identify groups of similar items, create scales, and, thereby, reduce the number of variables used in further analysis. Factor Analysis can also be used to test the factor validity of test inventory or battery to make sure that items are more associated with the scales to which they are assigned that other scales or tests in the battery.

Background: Factor Analysis is used to model the relationships (correlations or covariances) between objects (in the examples discussed here the objects are items from an inventory). The Factor Analysis Model assumes that the interrelationships are due to latent variables called "common factors". The model further assumes that the variability within each item's responses can be divided (as additive components) into that due to the common factors and that which is unique to the item (i.e., unexplained by the model). Most Factor Analysis models also assume that the objects are fixed while the subjects (across which the correlations are obtained) are randomly sampled from a population. However, the *alpha method* of factor analysis was designed for situations where it is more appropriate to think of the objects as being randomly sampled from a population of items. When the objects being factored are items from an inventory that is assumed to have one or more scales, then the model for the alpha method is consistent with the classical true score model in test theory.

Once the appropriate number of factors (*r*) are selected the factor analysis produces a solution with that number of axes or dimensions. The axes however may have no intrinsic meaning. To uncover interpretable scales it is often necessary to rotate the axes (in the r-dimensional space). The goal of most rotational methods is to find what is called "simple structure." A simple structure is an idealized solution where each item is related to ("loads on") one and only one axis (scale). If it is reasonable to assume that the latent variables (factors, scales) are uncorrelated, then the Varimax rotation procedure often produces a result closest to simple structure.

However if one expects the scales that make up an inventory to be correlated (especially if the inter-scale correlations are moderate to high), then an oblique rotational method (e.g., Promax) should be used.

A word of caution about Principal Component Analysis (PCA), many textbooks often misrepresent PCA as a simplified version of Factor Analysis, it is not. PCA makes no assumption about latent variables nor the partitioning of the variance. Furthermore, although related, PCA maximizes the explanation of the variances in a set of items with the smallest set of dimensions while Factor Analysis tries to maximize the explanation of the covariances. Under certain condition the two methods can lead to similar results but one can not assume that is the case.

Data Requirements:

Factor analysis is conducted on the product-moments (correlations or covariances) between items. The items therefore should be interval scales (note that by definition dichotomous variables are interval scales). No assumption is made about the distribution of the variables during Factor Analysis and this only becomes an issue if various statistical test are then conducted on the factors.

The SPSS Factor command accepts input as either record level data or matrix data in the form of correlation, covariance or factor loading matrices. Factor can also write matrix materials to a matrix data file. There are no limitations to the number of analyses, variables, extractions or rotations, although the more of these you have, the longer it may take to perform the analysis. You can control the number of factors extracted, the number of iterations for extraction and rotation and other rotation parameters. Factor scores may be calculated and saved on the active file.

Factor analysis does not distinguish between independent and dependent variables. All variables are treated as a dependent set. The basis of the analysis is a correlation matrix built from the specified variables.

For our purposes, the Alpha extraction method with Varimax rotation is used. The following explanation of these choices may be used in a method section of an article:

‘The Alpha Method of Factor Analysis (Kaiser & Coffrey, 1965) was used because its factor model was specifically designed to be used in scale construction and testing. Varimax rotation was use because it typically produces an orthogonal set of interpretable dimensions.’

Sample Data

Data for this example come from the Drug Outcome Monitoring Study (Dennis, et. al., 1999). The study collected data from 8/96 to 12/97. The Global Appraisal of Individual Needs

(GAIN) was administered at intake, in person, by clinical staff. The total sample of 736 includes 271 adolescents from 10 adolescent treatment units and 465 adults from 11 adult treatment units.

Specifically, the data used for this example will be the mental health indices of the GAIN. These include the Somatic Symptom Index (SSI), the Anxiety Symptom Index (ASI), the Depressive Symptom Index (DSI), the In-Attention Index (IAI), the Hyperactive-Impulsivity Index (HII), and the Conduct Disorder Index (CDI).

Procedure: The procedures and equations used here can also be found on pages 323-334 in the *SPSS Base 10.0 User's Guide* manual (SPSS, 1999). To get the desired output, in SPSS, on the main menu, with an appropriate dataset open, choose:

Statistics
  Data Reduction
    Factor

This will open a dialog box. Select the series of variable you wish to include in the factor analysis. To highlight (select) several consecutive variables, select the first variable by pointing to it and clicking the mouse. Then, move the cursor to the last variable, hold the <SHIFT> key and click on the last variable in the list. The first, last and all intervening variables will be highlighted. To highlight several variables that are not consecutive, click the first variable. Hold the <CTRL> key and click any other variables you wish to include. In this example, click on work1, move to work9, hold the <SHIFT> key down and click on work9. When all variables have been highlighted, click the arrow next to the box labeled "Variables" to move the variables into the "Variable" box. (At this point, if you choose OK, Factor will perform a principal components analysis on the selected variables (no rotation) with listwise exclusion of missing values.)

To select alpha method, click the "Extraction" box. At the top of this dialog box, under Method, choose "Alpha factoring" which is the next to last entry. (Alternatively, with the Method box highlighted, type the letter 'A' and Alpha factoring will appear in the box. Click 'Scree Plot'. Then click 'Continue".

To select varimax rotation, click the "Rotation box, and select Varimax. To print the factor score coefficients click the "Scores" box, then check the last item, "Display factor score coefficient matrix". (This is the 'FSCORE' on the '/PRINT' line.) Click "Paste" to place the following into a syntax file.

```
FACTOR  /VARIABLES ssi dsi asi iai hii cdi
  /MISSING LISTWISE
  /ANALYSIS ssi dsi asi iai hii cdi
  /PRINT INITIAL EXTRACTION ROTATION FSCORE
  /PLOT EIGEN
```

```
 /CRITERIA MINEIGEN(1) ITERATE(25)
 /EXTRACTION ALPHA
 /CRITERIA ITERATE(25)
 /ROTATION VARIMAX .
```

If you wish to specify the number of factors, rather than having Factor select those with eigenvalues greater than 1 (or a number you specify). In the Extraction dialog box under Extract, choose "Number of Factors:" and enter the desired number in the box to the right. The pasted syntax will now look like this: (Note the difference in the first "/CRITERIA" line.)

```
FACTOR
 /VARIABLES ssi dsi asi iai hii cdi
 /MISSING LISTWISE
 /ANALYSIS ssi dsi asi iai hii cdi
 /PRINT INITIAL EXTRACTION ROTATION
 /CRITERIA FACTORS(2) ITERATE(25)
 /EXTRACTION ALPHA
 /CRITERIA ITERATE(25)
 /ROTATION VARIMAX .
```

To save the factor scores as variables in the active dataset, in the factor dialog box click the Scores button. Then choose "Save as variables" and "Display factor score correlation matrix". The pasted syntax will look like this: (Note the last line is new.) REG in the last line refers to Regression method used to calculate factor scores. Other available methods include Bartlett and Anderson-Rubin.

```
FACTOR
 /VARIABLES ssi dsi asi iai hii cdi
 /MISSING LISTWISE
 /ANALYSIS ssi dsi asi iai hii cdi
 /PRINT INITIAL EXTRACTION ROTATION
 /PLOT EIGEN
 /CRITERIA FACTORS(2) ITERATE(25)
 /EXTRACTION ALPHA
 /CRITERIA ITERATE(25)
 /ROTATION VARIMAX
 /SAVE REG(ALL).
```

In this example, factor will add 2 new variables to the end of the dataset called fac1_1 and fac1_2. In these variable names, 'fac' identifies that the variable is the result of a factor analysis. The first digit (to the left of the underscore character) identifies the number of the analysis run, while the number to the right of the underscore character identifies the factor level. Thus, fac1_1 is the first factor score of the first factor analysis, while fac1_2 is the second factor score of the first factor analysis. If you were to run this command again, two additional variables called fac2_1 and fac2_2. To specify your own root name (rather than the default 'fac') for the factor scores (available only in syntax), include the name in the /SAVE subcommand. For example, the subcommand /SAVE REG(ALL, MHFac) would result in factors named MHFac1 and MHFac2.

The output for the factor matrices can also be formatted. When in the Factor Analysis dialogue box, click on the 'Options' button. Under 'Coefficient Display Format', click on 'Sorted by size' to have the coefficients sorted by size. The 'Suppress absolute values less than:' will leave blank the coefficients that are less than the value specified, in the factor matrices. The syntax line, which would follow the /Print subcommand, would be:

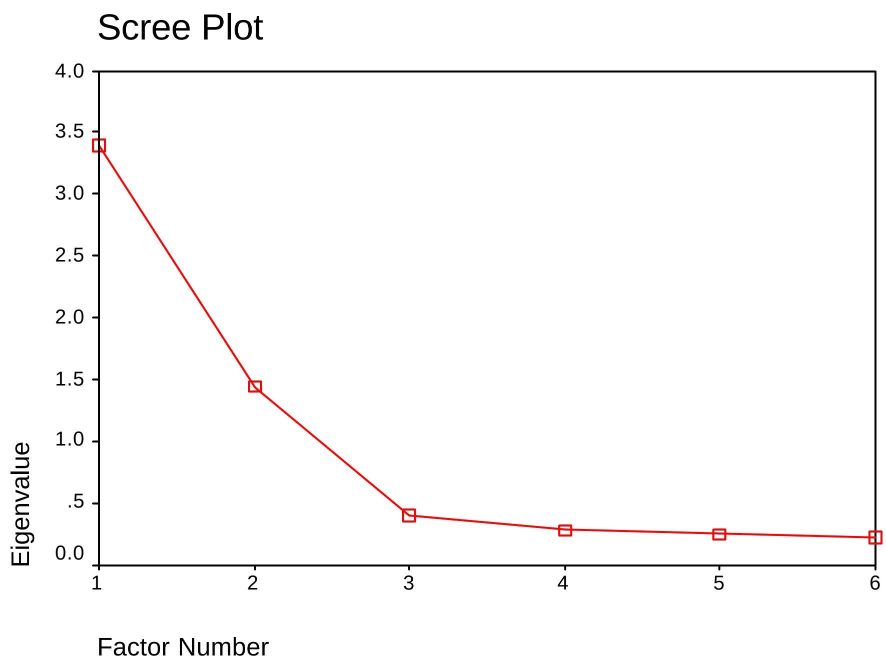
```
/FORMAT SORT BLANK(.3).
```

This will sort the coefficients by size on the first factor and any loading less than the absolute value of .3 will be left blank.

Output: The primary output used from Factor includes the Scree Plot. It is used to determine the optimal number of factors to retain. The scree plot is a graphical representation of the eigenvalues of the initial factors. The eigenvalue is the ratio of the between-groups sum of squares to the within-groups sum of squares or the amount (not the percent) of variance in the observed variables accounted for by each component or factor. A general rule of thumb is not to accept eigenvalues less than 1. For Alpha Factor Analysis this not just a rule of thumb, since the eigenvalues have special meaning in this model. An eigenvalue represents the generalizability coefficient (α) for the dimension, such that an eigenvalue value of one is an α of zero, that is nothing in common among the items. Another general rule is to accept factors where the slope is steep. For a good factor analysis, the scree plot will look roughly like the intersection of two lines. You want to select factors that are to the left of this intersection (not including the factor AT the intersection). Figure 1 shows the scree plot for this example.

The second key output is the "Rotated Factor Matrix". This matrix displays how each variable 'loads' on each factor. A loading is the partial correlation between the item and the factor. Loadings higher than .4 indicate that the variable is highly correlated with the factor. It is preferable that each variable included in the analysis load high on one of the retained factors. There should be three items with a loading of at least .4 on each (rotated) factor and the presence of every item in at least one dimension. By examining the pattern of variables that load high on a given factor, you can begin to interpret the results. Table 1 shows the Rotated Factor Matrix for this example. By default, Factor Analysis displays all factors with eigenvalues greater than 1. If you decide to keep more or fewer than the number selected by default, you must re-run the analysis using a specified number of factors as described above.

Figure 1.



As you can see from the above scree plot, there were 2 factors with eigenvalues greater than 1. Examining the rotated factor matrix in Table 1 shows three items with factor loadings of 0.4 or greater for each of the two factors. For factor 1, the SSI, DSI and ASI had high loadings. The scales IAI, HII and CDI were loaded high on factor 2.

Table 1: Rotated Factor Matrix

Rotated Factor Matrix		
	Factor	
	1	2
SSI Somatic Symptom Index	**0.8049**	0.2016
DSI Depressive Symptom Index	**0.8618**	0.1766
ASI Anxiety Symptom Index	**0.8501**	0.2189
IAI In-attention Index	0.1993	**0.8405**
HII Hyperactivity-Impusivity Index	0.2708	**0.8123**
CDI Conduct Disorder Index	0.1198	**0.7305**
Extraction Method: Alpha Factoring. Rotation Method: Varimax with Kaiser Normalization.		
a	Rotation converged in 3 iterations.	

This solution meets the above mentioned criteria for a selecting the number of factors. We have two factors with eigenvalues greater than 1, there are three items with loadings of .4 or higher on each of the factors, and each scale is present on at least one of the factors (loading > .4). The next step is to label the factors. Here we could label factor 1, with the somatic, anxiety and depressive scales as an 'Internal Distress' dimension. Factor 2, with in-attention, hyper-activity and conduct disorder could be labeled as an 'External Distress' dimension.

Comments. Not all factor solutions will come out this clean. There will be times when you will have many factors with eigenvalues greater than 1. By examining the scree plot, when the line is starting to level off (elbow), the number of factors can be cut off there. Also, some factors may have less than three items with loadings greater than .4, then force a solution with fewer factors and recheck the loadings. Some items may not load high on the rotated factors, even after forcing fewer factors to be selected. These items may need to be dropped from the analysis.

Describing These Procedures. These procedures would normally be described:

The number of factors was selected based on multiple criteria including (a) visual inspection of the scree; (b) eigenvalues greater than 1, (c) 3 item loadings of at least .4 on each (rotated) factor, and (d) the presence of every scale in at least 1 dimension. There were only two factors with an eigenvalue greater than 1 and the solution met our remaining criteria. The two factors were labeled as an Internal Distress dimension and an External Distress dimension. These dimensions were then calculated for the intake GAIN indices, using the rotated factor loadings. Equations 1 and 2 show the respective factor loadings:

(1) ID = .80(SSI) + .86(DSI) + .85(ASI)

(2) ED = .84(IAI) + .81(HII) + .73(CDI).

Annotated Bibliography

Dennis, M. Scott, C.K, Godley, M.D., & Funk, R. (1999). Comparisons of Adolescents and Adults by ASAM profile using GAIN data from the Drug Outcome Monitoring Study (DOMS). Bloomington, IL: Chestnut Health Systems (http://www.chestnut.org/li/posters).

This is the data source for the examples used in this paper. Information is useful any study using the GAIN.

Statistical Program for the Social Sciences (SPSS 1999). SPSS Base 10.0 User's Guide. Chicago, IL: Author (www.spss.com).

This is the most recent SPSS manual describing the procedures for running Factor Analysis. For the most part, the syntax has not changed from earlier versions.

Kaiser, H. F. & Coffrey, J. (1965). Alpha factor analysis, *Psychometrika, 30*, 1-14.

This is the primary citation. It introduced the analytic method for conducting Alpha Factor Analysis.

Kim, J. & Mueller, C. W. (1978). *Introduction to Factor Analysis: What it is and how it works.* Beverly Hills, CA: Sage.

Kim, J. & Mueller, C.W. (1978). *Factor Analysis: Statistical methods and practical issues.* Beverly Hills, CA: Sage.

These two little Sage books give a nice overview of Factor Analysis. They assume a general statistical background (1st year graduate course) and a basic understanding of matrix algebra (or at least notation). However, for a little more money you might want to try the paperback version of the following.

Kline, P. (1993). *An easy guide to Factor Analysis.* New York: Routledge.

A very good overview of Factor Analysis, complete with examples and practical applications. Also assumes a general statistical background (1st year graduate course) and a basic understanding of matrix algebra.

Harmon, H. H. (1976). *Modern factor Analysis (3rd ed., revised).* Chicago: The University of Chicago Press.

A classic, this book is still used widely as a textbook for a first course on Factor Analysis in graduate schools. A background with multivariate statistics and working familiarity with matrix algebra is recommended. However, for a more advanced textbook it is still quick and readable.