



## An SPSS R-Menu for Ordinal Factor Analysis

**Mário Basto**  
Polytechnic Institute of  
Cávado and Ave

**José Manuel Pereira**  
Polytechnic Institute of  
Cávado and Ave

---

### Abstract

Exploratory factor analysis is a widely used statistical technique in the social sciences. It attempts to identify underlying factors that explain the pattern of correlations within a set of observed variables. A statistical software package is needed to perform the calculations. However, there are some limitations with popular statistical software packages, like SPSS. The R programming language is a free software package for statistical and graphical computing. It offers many packages written by contributors from all over the world and programming resources that allow it to overcome the dialog limitations of SPSS. This paper offers an SPSS dialog written in the R programming language with the help of some packages, so that researchers with little or no knowledge in programming, or those who are accustomed to making their calculations based on statistical dialogs, have more options when applying factor analysis to their data and hence can adopt a better approach when dealing with ordinal, Likert-type data.

*Keywords:* polychoric correlations, principal component analysis, factor analysis, internal reliability.

---

## 1. Introduction

In SPSS (IBM Corporation 2010a), the only correlation matrix available to perform exploratory factor analysis is the Pearson's correlation, few rotations are available, parallel analysis and Velicer's minimum average partial criteria (MAP) to determine the number of factors to retain are not available, and internal reliability is based mainly on Cronbach's alpha. Some implications of software limitations, like those referred, can be seen in many research done by social scientists. Thereby, for extraction and rotation of factors, principal component analysis and varimax rotation are frequently used. Also commonly used, are the Kaiser criterion and/or the scree test to decide the correct number of factors to retain. The analysis is almost always performed with Pearson's correlations even when the data is ordinal.

These procedures are usually the default option in statistical software packages like SPSS, but it will not always yield the best results. This paper offers a SPSS dialog to overcome some of the SPSS dialog limitations and also offers some other options that may be or become useful for someone's work. The SPSS dialog is written in the R programming language (R Development Core Team 2011), and requires SPSS 19 (IBM Corporation 2010a), the R plug-in 2.10 (IBM Corporation 2010b) and the following R packages: **polycor** (Fox 2009), **psych** (Revelle 2011), **GPArotation** (Bernaards and Jennrich 2005), **nFactors** (Raiche and Magis 2011), **corp-cor** (Schaefer, Opgen-Rhein, Zuber, Duarte Silva, and Strimmer 2011) and **ICS** (Nordhausen, Oja, and Tyler 2008).

## 2. Principal component and factor analysis

Principal component analysis (PCA) is the default method of extraction in many statistical software packages, including SPSS. Principal components analysis is a technique for forming new variables (called principal components) which are linear composites of the original variables and are uncorrelated among themselves. It is computed using the variances of the manifest variables, in such a way that the the first principal component accounts for as much as possible of the variance, the second principal component accounts for as much as possible of the remaining variance, and successively for the rest of the principal components. Principal components analysis is often confused with factor analysis, a related but a conceptually distinct technique. Both techniques can be used to reduce a larger number of variables to a smaller number of components or factors. They are interdependent procedures which means that they do not assume the existence of a dependent variable. However, the aim of factor analysis is to uncover the latent structure of a set of variables, i.e., to reveal any latent variables that explain the correlations among the variables, called dimensions. Hence, factor analysis is based on the assumption that all variables are correlated to some degree. Therefore, those variables that share similar underlying dimensions should be highly correlated, and those variables that measure different dimensions should yield low correlations (Sharma 2007; Ho 2006). During factor extraction the shared variance of a variable is partitioned into unique variance and error variance but only shared variance appears in the solution. Usually to test the adequacy of the correlation matrix to perform a factor analysis, one tests if the correlation matrix is the identity matrix. SPSS provides the  $\chi^2$  test of Bartlett (1951). There are two more  $\chi^2$  tests, available in the package **psych**, the test of Jennrich (1970) and the test of Steiger (1980). Besides the Bartlett test, these two tests are also available in the present menu. Nevertheless,  $\chi^2$  tests tend to reject the null hypothesis for big samples, so they have questionable interest for practical applications.

### 2.1. Rotations

The goal of rotation is to simplify and clarify the data structure. A rotation can help to choose the correct number of factors to retain and can also help the interpretation of the solution. One can do a factor analysis with different values of the number of factors extracted, performing rotations, and, eventually, choosing the solution that has the more logical interpretation. Interpretation of factors is subjective, and usually can be made less subjective after a rotation. Rotation consists of multiplying the solution by a nonsingular linear application called the rotation matrix. The objective is to obtain a matrix of loadings whose entries, in absolute

value, are close to zero or to one. There are some choices available in popular software packages. In SPSS, one has varimax, quartimax, and equamax as orthogonal methods of rotation and direct oblimin and promax as oblique methods of rotation. Orthogonal rotations produce factors that are uncorrelated while oblique methods allow the factors to be correlated. In the social sciences where one expects some correlation among the factors, oblique rotations should theoretically drive a more accurate solution. If the best factorial solution involves factors uncorrelated, orthogonal and oblique rotations produce nearly identical results (Costello and Osborne 2005; Fabrigar, Wegener, MacCallum, and Strahan 1999).

Among rotations, varimax is the most popular one. The purpose of this rotation is to achieve a solution where each factor has a small number of large loadings and a large number of small loadings, simplifying interpretation, since each variable tends to have high loadings with only one or with only few factors. The quartimax rotation minimizes the number of factors needed to explain each variable, and the equamax rotation is a compromise between varimax and quartimax rotations (IBM Corporation 2010a). The options concerning oblique rotations in SPSS are the oblimin family and promax rotation, which performs faster than oblimin and tries to fit a target matrix. Promax consists of two steps. The first step defines the target matrix, a solution obtained after a orthogonal rotation (almost always a varimax rotation) whose entries are raised to some power ( $\kappa$ , that typically is some value between 2 and 4). The second step is obtained by computing a least square fit from the solution to the target matrix (Hendrickson and White 1964). The present menu gives the choice of the orthogonal rotation performed in first step (including the no rotation option, which may have little or no interest, but it is available to someone who wants to try it). Oblimin is a family of methods for oblique rotations. The parameter delta allows the solution to be more or less oblique. For delta equal to zero, the solution is the most oblique, and the rotation is called quartimin. For  $\gamma = 1/2$  one has the the bi-quartimin criterion and is called covarimin for  $\gamma = 1$ .

There are several other methods that are less popular and less used and known, and are briefly in the next paragraph. Some may not have been tested or applied in factor analysis practice, perhaps because these options are not available in computer programs and hence require programming for their implementation (Bernaards and Jennrich 2005). Some of these rotations are available in the present SPSS menu. They can be used as alternatives to achieve an interpretable and meaningful solution. Several orthogonal and oblique rotations are described in Browne (2001) and Bernaards and Jennrich (2005).

Almost all rotations are based on choosing a criterion to minimize. Promax is an exception. The family for the Crawford-Ferguson method is a set of rotations parametrized by  $\kappa$  (Crawford and Ferguson 1970). For different values of  $\kappa$  there are different names for the rotations. Being the loading matrix a  $p$  by  $m$  matrix ( $p$  variables and  $m$  factors), for  $\kappa = 0$  the rotation is named quartimax, for  $\kappa = 1/p$  varimax, for  $\kappa = m/(2p)$  equamax, for  $\kappa = (m - 1)/(p + m - 2)$  parsimax, and for  $\kappa = 1$  factor parsimony (Bernaards and Jennrich 2005). Infomax is a rotation that gives good results for both orthogonal and oblique rotations according to Browne (2001). The simplest entropy criterion works well for orthogonal rotation but is not appropriate for oblique rotation (Bernaards and Jennrich 2005). Similarly, the McCammon minimum entropy works for orthogonal rotation, but it is unsatisfactory in oblique rotation (Browne 2001). Oblimax criterion is another oblique rotation available. The algorithms from the package **GPArotation** (Bernaards and Jennrich 2005) implement both the oblique and orthogonal cases for Bentler criteria. The tandem criteria are two rotation methods, Principles I and II, to be used sequentially. Principle I is used to determine the

number of factors, and Principle II is used for final rotation (Bernaards and Jennrich 2005). Geomin criteria is available for both orthogonal and oblique rotations but may be not optimal for orthogonal rotation (Browne 2001). Simplimax is an oblique rotation method proposed by Kiers (1994). It was defined to rotate so that a given number of small loadings are as close to zero as possible, hence, to perform this rotation an extra argument  $k$  is necessary. This is the number of loadings close to zero (Bernaards and Jennrich 2005). Generally, the greater the number of small loadings, the simpler the loading matrix.

## 2.2. Polychoric correlations

Given that factor analysis is based on correlations between measured variables, a correlation or covariance matrix for the variables must be computed. The variables should be measured at least at the ordinal level, although two-category nominal variables can also be used. Frequently, when dealing with Likert-type data, the computed correlation matrix is the Pearson's correlation matrix, due in part to the fact that in almost all the popular statistical packages, only Pearson's correlations are available to perform a principal component or factor analysis, although literature suggests that it is incorrect to treat nominal and ordinal data as interval or ratio (Bernstein and Teng 1989; Gilley and Uhlig 1993; Stevens 1946). Applying traditional factor analysis procedures to item-level data almost always produces misleading results. Several authors explain the scale problem and suggest alternative procedures when attempting to establish the validity of a Likert scale since ordinal variables do not have a metric scale and tend to be attenuated due to the restriction on range (Marcus-Roberts and Roberts 1987; Mislevy 1986; Muthèn 1984; Muthèn and Kaplan 1985). Also, traditional factor analysis procedures produce meaningful results only if the data is continuous and is also multivariate normal. Item-level data almost never meets these requirements (O'Connor 2010; Bernstein and Teng 1989).

The correlation between two items is affected by their substantive similarity and by the similarities of their statistical distributions (Bernstein, Garbin, and Teng 1988). Items with similar distributions tend to correlate more strongly with one another than do with items with dissimilar distributions (Bernstein *et al.* 1988; Nunnally and Bernstein 1994). Factor analyses using Pearson's correlations, when dealing with Likert-type data may produce factors that are based solely on item distribution similarity. The items may appear multidimensional when in fact they are not (Bernstein *et al.* 1988).

An ordinal variable can be thought of as a crude representation of an unobserved continuous variable. The estimates of the correlations between the unobserved variables are called polychoric correlations. Hence, they are used when both variables are dichotomous or ordinal but both are assumed to reflect underlying continuous variables. These type of correlations extrapolate what the categorical variables distributions would be if they were continuous, adding tails to the distribution. As such it is an estimate strongly based on the assumption of an underlying continuous bivariate normal distribution. Polychoric correlations coefficients are maximum likelihood estimates of the Pearson's correlations for those underlying normally distributed variables. When both variables are dichotomous the polychoric correlations may be called tetrachoric correlations. Factor analysis for ordinal data should be conducted on the raw-data matrix of polychoric correlations and not on Pearson's correlations. Studies suggest that polychoric correlations should used when dealing with ordinal data, or in the presence of strong skewness or kurtosis (Muthèn and Kaplan 1985; Gilley and Uhlig 1993), as is often the case of Likert items.

### 2.3. Number of factors to retain

When doing a factor analysis and after factor extraction, one must decide how many factors to retain. The correct number is fundamental. One disadvantage that can occur is when the correct number of non-trivial principal components is not retained for subsequent analysis. In this case either relevant information is lost (underestimation) or noise is included (overestimation), provoking an inaccurate description of the underlying patterns of the correlations among the variables.

Extracting too few factors (underextraction) results in a loss of relevant information, and can overlook potentially relevant factors, in an inaccurate unifying of two or more factors, and an increase in the error of the loadings. Extracting too many factors (overextraction) increases results with noise included, factor splitting, factors with few high loadings, and gives too much substantive importance to trivial factors (O'Connor 2000; Wood, Tataryn, and Gorsuch 1996; Zwick and Velicer 1986; Peres-Neto, Jackson, and Somers 2005). Wood *et al.* (1996) have studied the effects of under and overextraction in principal factor analysis with varimax rotation. They found that overextraction generally leads to less error than does underextraction.

The default in most statistical popular software packages, such as SPSS, when dealing with a correlation matrix, is the Kaiser criterion, to retain all factors with eigenvalues greater than one and the scree test. The Kaiser criterion may overestimate or underestimate the true number of factors, but it usually overestimates the true number of factors (Costello and Osborne 2005; Lance, Butts, and Michels 2006; Zwick and Velicer 1986). The scree test involves examining the graph of the eigenvalues and looking for the bend in the data where the curve flattens out.

Two less well-known procedures, parallel analysis and Velicer's minimum average partial (MAP) criteria, usually yield optimal solutions for the number of factors to retain (Wood *et al.* 1996; Zwick and Velicer 1986).

Parallel analysis is often recommended as the best method to assess the true number of factors (Velicer, Eaton, and Fava 2000; Lance *et al.* 2006; Zwick and Velicer 1986). It retains the components that account for more variance than the components derived from generated random data. Normally distributed random data are generated, and the eigenvalues for the correlation matrices of random data representing the same number of cases and variables are computed. Then, the mean or a particular percentile, usually the 95th percentile, of the distribution of random eigenvalues computed are compared and plotted together with the actual eigenvalues (Cota, Longman, Holden, and Fekken 1993; Glorfeld 1995; Turner 1998). The intersection of the two lines determines the number of factors to be extracted. The SPSS dialog presented in this paper gives the choice of choosing the data generated. Hence, one can choose between normally distributed random data or permuted data, to perform parallel analysis. This last procedure has the advantage of having the possibility of being conducted with different correlation matrices beyond Pearson's correlation, and has the advantage of following the original data distribution. Also, the SPSS dialog presented has the option to choose between the eigenvalues obtained from a principal component analysis or from a factor analysis.

Velicer's MAP criteria was developed by Velicer, and is similar to parallel analysis in the results achieved (Velicer 1976; Velicer and Jackson 1990). With this criteria, components are retained as long as the variance in the correlation matrix represents systematic variance. The number of

components is determined by the step number  $k$  ( $k$  varies from zero to the number of variables less one), that results in the lowest average squared partial correlation between variables after removing the effect of the first  $k$  principal components. For  $k = 0$ , Velicer's MAP corresponds to the average values for the original correlations. If the minimum is achieved for  $k = 0$  no component should be retained. Zwick and Velicer (1986) found that Velicer's MAP test was very accurate for components with large eigenvalues, or where there was an average of eight or more variables per component. Zwick and Velicer (1986), through simulation's studies, observed that the performance of the Velicer's MAP method was quite accurate across several situations, although, specifically when there are few variables loading in a particular component or when there are low variables loadings, MAP criteria can underestimate the real number of factors (Zwick and Velicer 1986; Ledesma and Valero-Mora 2007). Peres-Neto *et al.* (2005) applied Velicer's MAP after removing the step  $k = 0$ . Velicer's MAP test has been revised by Velicer *et al.* (2000), with the partial correlations raised to the 4th power (rather than squared). These results are available in the SPSS dialog created.

Although accurate and easy to use, these two methods are not available in popular statistical software packages, like SPSS. O'Connor (2000) wrote programs in SPSS and SAS syntax to allow computation of parallel analyses and Velicer's MAP in SPSS and SAS. Parallel analyses on both principal components and common factor analysis (also called principal factor analysis or principal axis factoring) can be conducted. Instead of working with the original correlation matrix as in principal component analysis, common factor analysis works with a modified correlation matrix, on which the diagonal elements are replaced by estimates of the communalities. Hence, while principal components looks for the factors that explain the total variance of a set of variables, common factor analysis only looks for the factors that can account for the shared correlation between the variables. When one wants to conduct a common factor analysis, there is no agreement if one should use principal component eigenvalues or common factor eigenvalues to determine the number of factors to retain. The procedure usually used in common factor analysis is the extraction and examination of principal component eigenvalues, but some authors argue that if one wants to conduct a common factor analysis, the common factor eigenvalues should be used to determine the number of factors to retain (O'Connor 2000, 2010). The present SPSS dialog allows both options.

Besides those criteria, other procedures not available in SPSS are available in the present menu. VSS criterion proposed by Revelle and Rocklin (1979), and two non graphical solutions to the scree test proposed by Raiche, Riopel, and Blais (2006), the optimal coordinates and an acceleration factor, are available.

To determine the optimal number of factors to extract, Revelle and Rocklin (1979) proposed using the very simple structure criterion (VSS). The complexity of the items, indicated with  $c$ , determines the structure of the simplified matrix. For each item, all loadings of a factor matrix, except the biggest  $c$ , are set to zero. Then, the VSS criterion compares the correlation matrix reproduced by the simplified version of the factor loadings to the original correlation matrix. VSS criterion will tend to have its great value at the most interpretable number of factors according to Revelle and Rocklin (1979) and Revelle (2011). Revelle (2011) states that this criterion will not work very well if the data has a factor structure very complex, and simulations suggest that it will work well if the complexities of some of the items are no more than two.

For  $n$  eigenvalues, the optimal coordinates for factor  $i$  are, as described in the **nFactors** manual, the extrapolated eigenvalue  $i$ , made by a line passing through the eigenvalue  $(i + 1)$

and the last eigenvalue  $n$ . There are  $n - 2$  lines like this. The factors are retained as long as the observed eigenvalues are over the extrapolated ones. Simultaneously, the parallel analysis criterion must be satisfied.

The scree plot (Cattell 1966) is a graphical representation of the eigenvalues, where the eigenvalues are ordered by magnitude, and the eigenvalues are plotted against the factors number. The choice of how many factors to retain is visual heuristic, looking for an elbow of the curve. The number of factors to keep are the ones above the elbow in the plot. The reason is that when the factors are important, the slope must be steep, while when the factors correspond to error variance, the slope must be flat. According to the **nFactors** manual and Raiche *et al.* (2006), the acceleration factor corresponds to a numerical solution of the elbow of the scree plot. It corresponds to the maximum value of the numerical solution of the approximation of the second derivative by finite differences,  $f''(i) \approx f(i+1) - 2f(i) + f(i-1)$ . The number of factors to retain corresponds to the solution found minus one. Simultaneously, as with optimal coordinates, the parallel analysis criterion must be satisfied. This criterion is a tentative of substituting the subjective visual heuristic of the elbow of the scree plot. However, it seems that this criterion will tend to underestimate the number of factors, since it can identify the elbow when the slope is steep.

## 2.4. Quality of adjustment

Assessing the quality of a particular factorial model can be made, in a heuristic way, by computing the differences between the correlations observed in the sample data with the ones estimated by the factorial model, called residuals. If those residuals are high, the model poorly reproduces the data. The corresponding matrix is called the residual matrix and, in an empirical way, it is considered that a high percentage of residuals less than 0.05 is an indicator of good adjustment. Another way to assess the quality of the factorial model is by using goodness-of-fit statistics, commonly used in structural equation analysis, like GFI (goodness-of-fit index), AGFI (adjusted goodness-of-fit index) and RMSR (root mean square residual, off-diagonal). GFI can be interpreted as the fraction of the correlations of the observed variables that are explained by the model. Usually, values above 0.90 indicate a good fit and above 0.95 indicate a very good fit. The AGFI is the GFI statistic adjusted for the degrees of freedom, since GFI tends to overestimate the true value of the adjustment. The RMSR employs the residual matrix and is also used to measure the quality of the adjustment. Values less than 0.05 are considered very good while values below 0.1 are good. Higher values are an indicator of bad adjustment.

The root mean square partial correlations controlling factors, named RMSP in the present menu, is similar to RMSR, but calculated in the matrix of the partial correlations between variables after the effects of all factors have been removed. These partial correlations reveal how much variance each pair of variables share that is not explained by the factors extracted. This index appears, according to our knowledge, only in the output of SAS (SAS Institute Inc. 2010) and a model will be much better adjusted the lower its value.

## 2.5. Reliability

A measurement is said to be reliable or consistent if the measurement can produce similar results if used again in similar circumstances. Hence, the reliability of a scale is the correlation of that scale with the hypothetical one which truly measures what it is supposed to. Lack of

reliability may result from negligence, guessing, differential perception, recording errors. Internal consistency reliability (e.g., psychological tests) refers to the extent to which a measure is consistent with itself.

Cronbach's alpha is the most common internal reliability coefficient as it is the standard approach for summated scales built from ordinal or continuous items. It requires multinormal linear relations and assumes unidimensionality. A variation (Kruider-Richardson's formula) can be used with dichotomous items. In classical test theory, the observed score of an item (variable measurement) can be decomposed into two components, the true score and the error score (which in turn is decomposed into systematic error and random error). The true score compared to the error score reflects the reliability of a measurement, being the reliability higher when the proportion between those two scores components is higher. Cronbach's alpha equals zero when the true score is not measured at all and there is only an error component. Cronbach's alpha equals one when all items measure only the true score and there is no error component. Cronbach's alpha ( $\alpha$ ) is given by:

$$\alpha = \left( \frac{2 \sum_{j=i+1}^m \sum_{i=1}^{m-1} cov(x_i, x_j)}{m-1} \right) / \left( \frac{\sum_{i=1}^m V(x_i)}{m} + \frac{2 \sum_{j=i+1}^m \sum_{i=1}^{m-1} cov(x_i, x_j)}{m} \right)$$

where  $m$  is the number of components,  $V(x_i)$  is the variance of  $x_i$  and  $cov(x_i, x_j)$  is the covariance between  $x_i$  and  $x_j$ .

Standardized alpha ( $\alpha_{st}$ ) is Cronbach's alpha applied to standardized items, which means that it is Cronbach's alpha measured for items with equal variances:

$$\begin{aligned} \alpha_{st} &= \left( \frac{2 \sum_{j=i+1}^m \sum_{i=1}^{m-1} co(x_i, x_j)}{m-1} \right) / \left( 1 + \frac{2 \sum_{j=i+1}^m \sum_{i=1}^{m-1} co(x_i, x_j)}{m} \right) = \\ &= \left( G \frac{2 \sum_{j=i+1}^m \sum_{i=1}^{m-1} co(x_i, x_j)}{m-1} \right) / \left( G + G \frac{2 \sum_{j=i+1}^m \sum_{i=1}^{m-1} co(x_i, x_j)}{m} \right) \\ &= \left( \frac{2 \sum_{j=i+1}^m \sum_{i=1}^{m-1} G \cdot co(x_i, x_j)}{m-1} \right) / \left( \frac{\sum_{i=1}^m G}{m} + \frac{2 \sum_{j=i+1}^m \sum_{i=1}^{m-1} G \cdot co(x_i, x_j)}{m} \right) = \alpha \end{aligned}$$

where  $co(x_i, x_j)$  is the correlation between  $x_i$  and  $x_j$  and  $G = V(x_i)$ ,  $i = 1, \dots, m$ , since the variance is constant.

When the factor loadings of each variable on the common factor are not equal, Cronbach's alpha is a lower bound of the reliability, giving a negatively biased estimate of the theoretical reliability (Zumbo, Gadermann, and Zeisser 2007). Simulations conducted by Zumbo *et al.* (2007) show that the negative bias is even more evident under the condition of negative skewness of the variables.

Armor's reliability theta ( $\theta$ ) is a similar measure of reliability developed by Armor (1974). Reliability theta is a coefficient that can be interpreted similar to other reliability coefficients. It measures the internal consistency of the items in the first factor scale derived from a principal component analysis. It is calculated using the equation:

$$\theta = \frac{p}{p-1} \left( 1 - \frac{1}{\lambda_1} \right)$$

where  $p$  is the number of items in the scale and  $\lambda_1$  is the first and largest eigenvalue from the principal component analysis of the correlation matrix of the items comprehending the scale.



Zumbo *et al.* (2007) proposed a polychoric correlation matrix to calculate Cronbach's alpha coefficient and Armor's reliability theta coefficient, which they named ordinal reliability alpha and ordinal reliability theta, respectively. They concluded that the ordinal reliability alpha and theta provide consistently suitable estimates of the theoretical reliability regardless of the magnitude of the theoretical reliability, the number of scale points, and the skewness of the scale point distributions. These coefficients show to be appropriate to measure reliability in the presence of skewed data, whereas alpha coefficient is affected by it, giving a negatively biased estimate of the reliability. Hence, ordinal reliability alpha and ordinal theta will normally be higher than the corresponding Cronbach's alpha (Zumbo *et al.* 2007).

Besides Cronbach's alpha, the others coefficients are not computed by popular statistical software packages, like SPSS, but the SPSS dialog presented in this paper can estimate them.

### 3. Examples

The objective of the following examples is to show and explain some of the features available on the menu that can help researchers work.

#### 3.1. Example 1

The data set used in this example is in the file `Example1.sav`. This is simulated data and has 14 ordinal variables and 390 cases. The objective is to exhibit some features of the menu. The first table of the output reports the number of valid cases (if `listwise` checked) or the number of valid cases for each pair of variables (if `pairwise` checked). To determine the number of factors that explain the correlations among the set of variables, polychoric correlations were estimated by a two-step method (Figure 1).

The correlation matrix used for simulations of parallel analysis was also the polychoric one estimated by a quick two-step method (Figure 2). To perform a parallel analysis it is necessary to simulate random correlation matrices with the same number of variables and cases of the actual data. The simulated data is then subject to principal component analysis or common factor analysis and the mean or a particular quantile of their eigenvalues is computed and compared to the eigenvalues produced by the actual data. The criterion for factor extraction is where the eigenvalues generated by random data exceed the eigenvalues produced by the actual data. In this example, parallel analysis was performed with the original data randomized (permutation data), based on components model and on the 95th percentile, which means that the data generated has been performed on a polychoric correlation matrix following the original data distribution and a principal component analysis was performed to each permutation with the 95th percentile for each set of eigenvalues computed and compared to the eigenvalues produced by the actual data (Figures 2 and 3). The number of factors to retain according to different rules are displayed in Figures 3, 4, 5, and 6. Based on the previous results, a factor analysis was performed and four factors were retained. The extraction was conducted by a principal component analysis from the polychoric correlation matrix estimated by a two-step method (Figure 7). An orthogonal rotation of the factors, varimax rotation, was also performed (Figure 8). One of the goals of factor analysis is to balance the percentage of variation explained with limitation of the number of factors to extract. In this example, the first four factors account for 89.726% of the variance in all the variables (Figure 9). For factor analysis to be appropriate, the correlation matrix must reveal a substantial number

**Number of Components/Factors estimated for the following Correlation Matrix**

Correlation Matrix	Missing values
Polychoric (Two Step)	listwise

Figure 1: Correlation matrix.

**Parameters for Parallel Analysis**

Number of Samples	Model	Data	Measure	Correlation matrix	Matrix for simulations
100	components	permutation	quantile	Polychoric (Two Step)	Polychoric (Two Step)

Figure 2: Parameters for parallel analysis.

**Parallel analysis on empirical data by columns permutation**

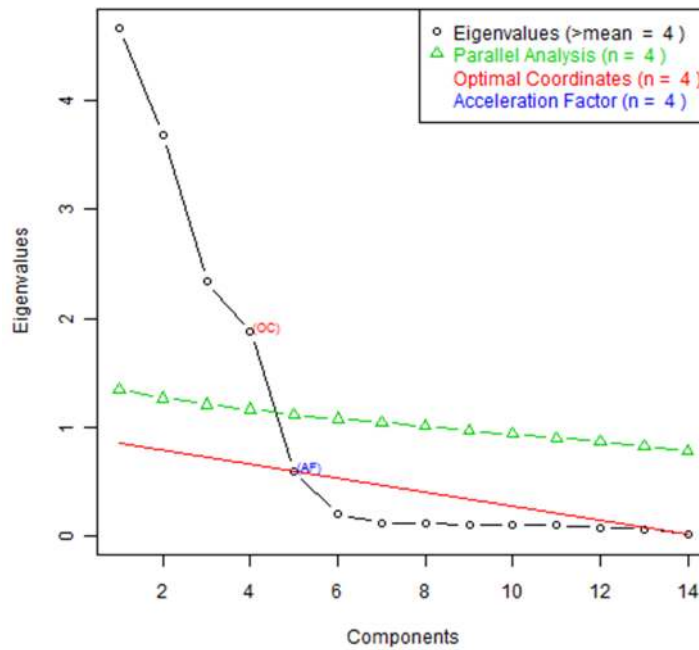


Figure 3: Scree plot and other criteria.

**Number of components/factors to retain according to different rules**

Optimal coordinates	Acceleration factor	Parallel analysis	Kaiser rule
4	4	4	4

Figure 4: Number of factors for different rules.

**Velicer's Minimum Average Partial Test**

	Velicer's Minimum	
	Minimum	Components to retain
Squared MAP	.050	4
4th power MAP	.015	3

Figure 5: Velicer's MAP.

**Very Simple Structure (Number of factors)**

	Very Simple Structure (VSS)					
	Rotation	Factoring method	Max VSS complexity 1	N factors complexity 1	Max VSS complexity 2	N factors complexity 2
Values	varimax	pc	.918	4	.988	6

Figure 6: Very simple structure.

**Factor Analysis**

Correlation Matrix	Factor Extraction	Missing values
Polychoric (Two Step)	pca	listwise

Figure 7: Factor analysis.

**Rotation**

method	orthogonal	convergence
varimax	TRUE	TRUE

Figure 8: Factor rotation.

**Variance Explained (extracted)**

	Sums of Squared Loadings	% of Variance	Cumulative %
1	4.664	33.312	33.312
2	3.689	26.347	59.659
3	2.339	16.705	76.364
4	1.871	13.362	89.726

(a) Variance explained.

**Rotated Variance Explained (extracted)**

	Sums of Squared Loadings	% of Variance	Cumulative %
1	4.004	28.599	28.599
2	4.059	28.996	57.595
3	2.598	18.558	76.153
4	1.900	13.573	89.726

(b) Variance explained after rotation.

Figure 9: Variance explained by the four factors.

Tests of whether a correlation matrix is an identity matrix				KMO - Kaiser-Meyer-Olkin measure of sampling adequacy		MSA - Measures of sampling adequacy	
	Chisquare tests			KMO	MSA		
	Chisquare	Degrees of freedom	p-value				
Bartlett's Test	11357.859	91	.000	.659	V1	.841	
Steiger Test	9032.080	91	.000		V2	.541	
Jennrich Test	2418.152	91	.000	V3	.625		
				V4	.496		
				V5	.811		
				V6	.826		
				V7	.681		
				V8	.443		
				V9	.504		
				V10	.840		
				V11	.685		
				V12	.922		
				V13	.487		
				V14	.676		

(a) Tests.

(b) KMO.

(c) MSA.

Figure 10: Measures and tests of sampling adequacy.

**Fit of the model to the correlation matrix**

	Residual fit values	
	Residuals > 0.05	% residuals > 0.05
Values	5	5.495

Figure 11: Residuals.

GFI simplified (Goodness of Fit Index)	AGFI simplified	RMSR (Root Mean Square Residual)	Root mean square partial correlations controlling factors
GFI ----- .994	AGFI ----- .984	RMSR ----- .024	RMSP ----- .224
(a) GFI.	(b) AGFI.	(c) RMSR.	(d) RMSP.

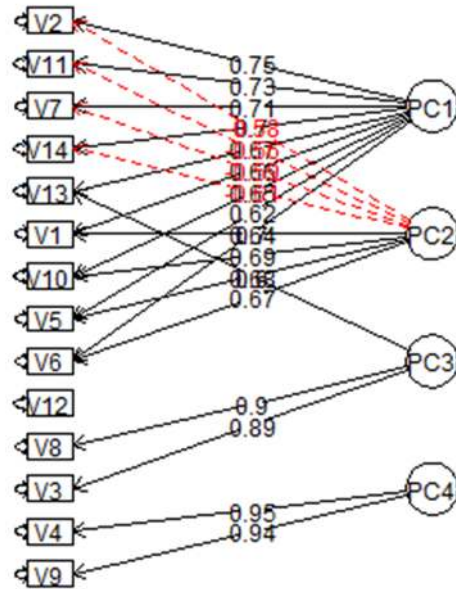
Figure 12: Goodness-of-fit statistics.

**Sorted Rotated Factor Loadings**

	PC1	PC2	PC3	PC4
V2	.975	.069	.072	.023
V14	.966	.020	.056	.021
V11	.961	.070	.043	.013
V7	.957	.036	.069	.024
V1	.035	.960	.055	-.009
V10	.047	.957	.028	.001
V6	.046	.931	.053	-.013
V5	.020	.923	.039	-.021
V12	.054	.693	.023	-.036
V8	-.023	.039	.976	-.004
V3	-.002	.077	.964	-.022
V13	.520	.054	.832	-.008
V4	.020	-.030	-.003	.974
V9	.039	-.030	-.024	.973

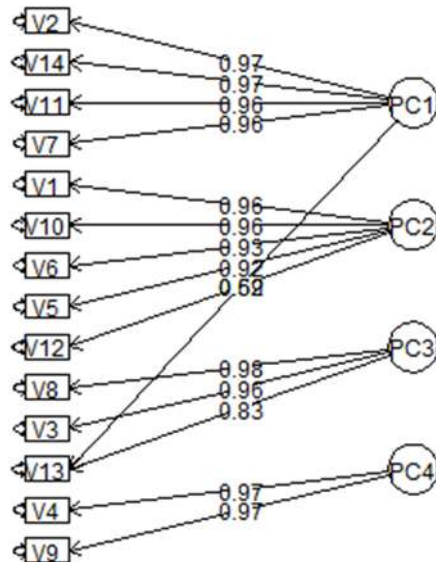
Figure 13: Loadings after rotation.

**Factor Diagram (initial solution)**



(a) Factor diagram before rotation.

**Factor Diagram (rotated solution)**



(b) Factor diagram after rotation.

Figure 14: Factor diagrams.

**Ordinal coefficient alpha**

S 1	S 2	S 3	S 4
.978	.938	.935	.947

(a) Ordinal alpha.

**Armor's reliability theta**

S 1	S 2	S 3	S 4
.975	.935	.920	.964

(b) Armor teta.

**Ordinal coefficient theta**

S 1	S 2	S 3	S 4
.978	.942	.935	.947

(c) Ordinal teta.

Figure 15: Reliability coefficients.

of significant correlations and most of the partial correlations must be low. By *Bartlett's*, *Jennrich* and *Steiger* tests,  $p$ -values less than 0.001 are obtained, indicating that the correlation matrix is not an identity matrix. The measures given by KMO (KMO overall statistic) and MSA (KMO statistic for each individual variable) show that all variables are adequate to factor analysis. Although  $KMO = 0.659$  reveals a mediocre value, all values of MSA are over 0.40 (Figure 10). Hence, one can proceed with factor analysis without dropping any variable. The residuals, the difference between the correlations estimated by the model and the observed ones, are displayed in Figure 11, showing only 5.495% of residuals greater than 0.05. Although not usual in exploratory factor analysis, some goodness-of-fit statistics are displayed (Figure 12). The loadings obtained after rotation are displayed in Figure 13. These loadings are helpful in resolving the structure underlying the variables. To have a better idea which variables load on each factor, a factor diagram is displayed. Loadings lower than 0.5 were omitted. For comparisons purposes, the factor diagram before rotation is also displayed (Figure 14). One can also display some internal reliability coefficients regarding the output of factor analysis. Automatically, each item is assigned to the factor with the higher loading. In Figure 15, the three coefficients displayed show good reliability of the four factors. Also, the scores of the factors can be recorded in a new database for further analysis.

### 3.2. Example 2

The data for this example was randomly generated in a linear subspace of dimension 7 from a 129 dimensional space. Three sets of data, with increasing noise added to the linear subspace, were generated. The three data sets have 129 variables and 196 observations. The objective is to compare different rules for the number of factors to retain. This data has different levels of noise added. The files are named `Example21.sav` (few noise), `Example22.sav` (more noise) and `Example23.sav` (even more noise). The parameters used for parallel analysis are displayed in Figure 16. The number of factors to retain according to Kaiser rule, parallel analysis and nongraphical scree tests are displayed in Figure 17.

Number of Samples	Model	Data	Measure	Correlation matrix	Matrix for simulations
100	components	permutation	mean	Pearson	Pearson

Figure 16: Example 2. Parameters for parallel analysis.

Number of components/factors to retain according to different rules

Optimal coordinates	Acceleration factor	Parallel analysis	Kaiser rule
7	1	7	7

(a) With little noise. Number of factors to retain.

Number of components/factors to retain according to different rules

Optimal coordinates	Acceleration factor	Parallel analysis	Kaiser rule
5	1	5	18

(b) With more noise. Number of factors to retain.

Number of components/factors to retain according to different rules

Optimal coordinates	Acceleration factor	Parallel analysis	Kaiser rule
3	1	3	33

(c) With even more noise. Number of factors to retain.

Figure 17: Example 2. Number of factors to retain according to different rules.

Velicer's Minimum Average Partial Test

	Velicer's Minimum	
	Minimum	Components to retain
Squared MAP	.006	7
4th power MAP	.000	7

Figure 18: Example 2. Velicer's map criterion for the first three cases (few noise, more noise and even more noise).

The Velicer's MAP criteria points to the correct number of factors in the three cases (Figure 18). For increasing values of noise, the Kaiser rule begins to overestimate the correct number of factors, while the opposite happens with parallel analysis. In this example, Kaiser rule shows a great difficulty to deal with noise.

VSS criterion, as items were not of complexity one or two, underestimates the correct number of factors (one factor for complexity one and two factors for complexity two).

Figure 19 shows the scree plot and some other criteria. Observe that the acceleration factor points to only one factor. That is due to the big gap between the first two eigenvalues.



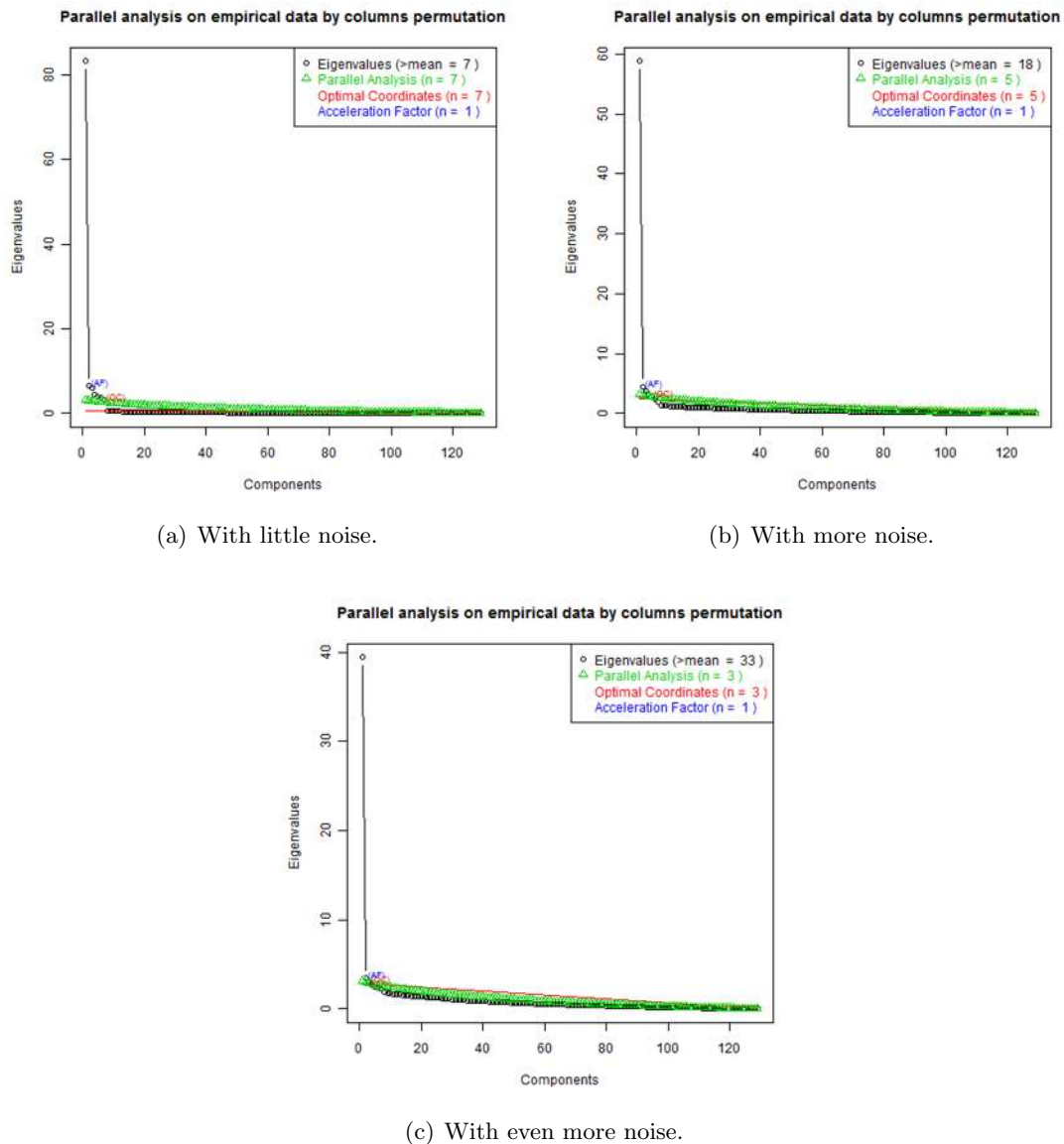


Figure 19: Example 2. Scree plot and other criteria.

Also, a fourth set of data, `Example24.sav`, derived from the previous dataset `Example23.sav`, and transformed into ordinal data with four levels, was analyzed. Again, Velicer's MAP criteria points to the correct number of factors, although the revised version with the partial correlations raised to the 4th power do not. The Kaiser rule behaves even worse, with the increasing number of factors to retain. The results are shown in Figure 20.

### 3.3. Example 3

The data source for this example was drawn from SASS (Schools and Staffing and Teacher Follow-up Surveys). The file named `Example3.sav` corresponds to the first 300 cases of 88 ordinal variables from the file `SASS_99_00_S2a_v1_0.sav` of the dataset

**Number of components/factors to retain according to different rules**

Optimal coordinates	Acceleration factor	Parallel analysis	Kaiser rule
3	1	3	39

(a) Parallel analysis and Kaiser rule.

**Velicer's Minimum Average Partial Test**

	Velicer's Minimum	
	Minimum	Components to retain
Squared MAP	.006	7
4th power MAP	.000	4

(b) Velicer's map.

Figure 20: Example 2. Number of factors according to different rules for the last data set.

**Number of components/factors to retain according to different rules**

Optimal coordinates	Acceleration factor	Parallel analysis	Kaiser rule
11	1	11	23

(a) Example 3. Number of factors (pearson correlations).

**Number of components/factors to retain according to different rules**

Optimal coordinates	Acceleration factor	Parallel analysis	Kaiser rule
12	1	12	22

(b) Example 3. Number of factors (polychoric correlations).

Figure 21: Example 3. Number of factors according to different rules.

<b>Velicer's Minimum Average Partial Test</b>	<b>Velicer's Minimum Average Partial Test</b>																						
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th rowspan="2"></th> <th colspan="2">Velicer's Minimum</th> </tr> <tr> <th>Minimum</th> <th>Components to retain</th> </tr> </thead> <tbody> <tr> <td>Squared MAP</td> <td style="text-align: center;">.009</td> <td style="text-align: center;">12</td> </tr> <tr> <td>4th power MAP</td> <td style="text-align: center;">.000</td> <td style="text-align: center;">14</td> </tr> </tbody> </table>		Velicer's Minimum		Minimum	Components to retain	Squared MAP	.009	12	4th power MAP	.000	14	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th rowspan="2"></th> <th colspan="2">Velicer's Minimum</th> </tr> <tr> <th>Minimum</th> <th>Components to retain</th> </tr> </thead> <tbody> <tr> <td>Squared MAP</td> <td style="text-align: center;">.016</td> <td style="text-align: center;">12</td> </tr> <tr> <td>4th power MAP</td> <td style="text-align: center;">.001</td> <td style="text-align: center;">13</td> </tr> </tbody> </table>		Velicer's Minimum		Minimum	Components to retain	Squared MAP	.016	12	4th power MAP	.001	13
		Velicer's Minimum																					
	Minimum	Components to retain																					
Squared MAP	.009	12																					
4th power MAP	.000	14																					
	Velicer's Minimum																						
	Minimum	Components to retain																					
Squared MAP	.016	12																					
4th power MAP	.001	13																					

(a) Example 3. Velicer's map (pearson correlations).

(b) Example 3. Velicer's map (polychoric correlations).

Figure 22: Example 3. Velicer's map.

SASS\_1999-00\_TFS\_2000-01\_v1\_0\_SPSS\_Datasets.zip. The dataset is available online at <http://nces.ed.gov/EDAT/>. The results were obtained using Pearson and polychoric correlations. For parallel analysis, the actual eigenvalues were compared to the mean of the Pearson's correlations evaluated from 100 normally distributed random variables. As in the previous example, the results show that the Kaiser rule overestimates the number of factors to retain. The results are shown in Figures 21 and 22. Based only on these results, one would point to 12 factors to retain. The results obtained by using polychoric correlations show the approach to this number, especially for the parallel analysis and Velicer's MAP.

## 4. Final remarks and conclusions

Exploratory factor analysis is a widely used applied statistical technique in the social sciences that often deals with ordinal data. Some of the limitations of popular statistical software packages can be overcome by this SPSS dialog. Also, this dialog offers some options that a researcher may find useful. This is especially helpful for those researchers with little or no knowledge in programming, or those who are accustomed to making their analysis with statistical dialogs. The availability of the polychoric correlation matrix, besides Pearson's one, is useful. More rotations are available to apply to data, although the benefits of some of these rotations are still not fully investigated. Velicer's MAP and parallel analysis are also available to help determining the correct number of factors to retain. In the above examples, Velicer's MAP showed the best performance and Kaiser rule, the default criteria in SPSS, showed its bias to overestimate the true number of factors. Measures of quality of adjustment and of internal consistency are also available.

Future updates to the software will be hosted on SourceForge at <http://SourceForge.net/projects/spssmenu/>.

## References

- Armor DJ (1974). "Theta Reliability and Factor Scaling." In HL Costner (ed.), *Sociological Methodology*, pp. 17–50. Jossey-Bass, San Francisco.
- Bartlett MS (1951). "The Effect of Standardization on a Chi Square Approximation in Factor Analysis." *Biometrika*, **38**, 337–344.
- Bernaards CA, Jennrich RI (2005). "Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis." *Educational and Psychological Measurement*, **65**, 676–696. URL <http://www.stat.ucla.edu/research/gpa/>.
- Bernstein IH, Garbin C, Teng G (1988). *Applied Multivariate Analysis*. Springer-Verlag, New York.
- Bernstein IH, Teng G (1989). "Factoring Items and Factoring Scales are Different: Spurious Evidence for Multidimensionality Due to Item Categorization." *Psychological Bulletin*, **105**, 467–477.
- Browne MW (2001). "An Overview of Analytic Rotation in Exploratory Factor Analysis." *Multivariate Behavioral Research*, **36**(1), 111–150.

- Cattell RB (1966). "The Scree Test For The Number Of Factors." *Multivariate Behavioral Research*, **1**, 245–276.
- Costello AB, Osborne JW (2005). "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis." *Practical Assessment, Research & Evaluation*, **10**(7).
- Cota AA, Longman RS, Holden RR, Fekken GC (1993). "Comparing Different Methods for Implementing Parallel Analysis: A Practical Index of Accuracy." *Educational & Psychological Measurement*, **53**, 865–875.
- Crawford CB, Ferguson GA (1970). "A General Rotation Criterion and its Use in Orthogonal Rotation." *Psychometrika*, **35**, 321–332.
- Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999). "Evaluating the Use of Exploratory Factor Analysis in Psychological Research." *Psychological Methods*, **4**(3), 272–299.
- Fox J (2009). *polycor: Polychoric and Polyserial Correlations*. R package version 0.7-7, URL <http://CRAN.R-project.org/package=polycor>.
- Gilley WF, Uhlig GE (1993). "Factor Analysis and Ordinal Data." *Education*, **114**(2), 258–264.
- Glorfeld LW (1995). "An Improvement on Horn's Parallel Analysis Methodology for Selecting the Correct Number of Factors to Retain." *Educational & Psychological Measurement*, **55**, 377–393.
- Hendrickson AE, White PO (1964). "Promax: A Quick Method for Rotation to Oblique Simple Structure." *British Journal of Statistical Psychology*, **17**(1), 65–70.
- Ho R (2006). *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*. Chapman & Hall/CRC, New York.
- IBM Corporation (2010a). *IBM SPSS Statistics 19*. IBM Corporation, Armonk, NY. URL <http://www-01.ibm.com/software/analytics/spss/>.
- IBM Corporation (2010b). *SPSS R Plug-in 2.10*. URL <http://www.spss.com/devcentral/>.
- Jennrich RI (1970). "An Asymptotic Chi-Square Test for the Equality of Two Correlation Matrices." *Journal of the American Statistical Association*, **65**, 904–912.
- Kiers HAL (1994). "SIMPLIMAX: Oblique Rotation to an Optimal Target with Simple Structure." *Psychometrika*, **59**, 567–579.
- Lance CE, Butts MM, Michels LC (2006). "The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say?" *Organizational Research Methods*, **9**(2), 202–220.
- Ledesma RD, Valero-Mora P (2007). "Determining the Number of Factors to Retain in EFA: An Easy-to-Use Computer Program for Carrying Out Parallel Analysis." *Practical Assessment, Research & Evaluation*, **12**(2), 1–11.

- Marcus-Roberts HM, Roberts FS (1987). “Meaningless Statistics.” *Journal of Educational Statistics*, **12**, 383–394.
- Mislevy RJ (1986). “Recent Developments in the Factor Analysis of Categorical Variables.” *Journal of Educational Statistics*, **11**, 3–31.
- Muthèn BO (1984). “A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators.” *Psychometrika*, **49**, 115–132.
- Muthèn BO, Kaplan D (1985). “A Comparison of some Methodologies for the Factor Analysis of Non-Normal Likert Variables.” *British Journal of Mathematical and Statistical Psychology*, **38**, 171–189.
- Nordhausen K, Oja H, Tyler DE (2008). “Tools for Exploring Multivariate Data: The Package **ICS**.” *Journal of Statistical Software*, **28**(6), 1–31. URL <http://www.jstatsoft.org/v28/i06/>.
- Nunnally J, Bernstein I (1994). *Psychometric Theory*. McGraw-Hill, New York.
- O’Connor BP (2000). “SPSS and SAS Programs for Determining the Number of Components Using Parallel Analysis and Velicer’s MAP Test.” *Behavior Research Methods, Instrumentation, and Computers*, **32**, 396–402.
- O’Connor BP (2010). “Cautions Regarding Item-Level Factor Analyses.” Access 2010-02-22, URL <https://people.ok.ubc.ca/briocomm/nfactors/nfactors.html>.
- Peres-Neto PR, Jackson DA, Somers KM (2005). “How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited.” *Computational Statistics & Data Analysis*, **49**, 974–997.
- Raiche G, Magis D (2011). *nFactors: Parallel Analysis and Non Graphical Solutions to the Cattell Scree Test*. R package version 2.3.3, URL <http://CRAN.R-project.org/package=nFactors>.
- Raiche G, Riopel M, Blais JG (2006). “Non Graphical Solutions for the Cattell’s Scree Test.” Paper presented at the International Annual meeting of the Psychometric Society, Montreal, URL <http://www.er.uqam.ca/nobel/r17165/RECHERCHE/COMMUNICATIONS/>.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Revelle W (2011). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois. R package version 1.1-2, URL <http://personality-project.org/r/psych.manual.pdf>.
- Revelle W, Rocklin T (1979). “Very Simple Structure – Alternative Procedure for Estimating the Optimal Number of Interpretable Factors.” *Multivariate Behavioral Research*, **14**(4), 403–414.
- SAS Institute Inc (2010). *The SAS System, Version 9.2*. SAS Institute Inc., Cary, NC. URL <http://www.sas.com/>.

- Schaefer J, Opgen-Rhein R, Zuber V, Duarte Silva AP, Strimmer K (2011). *corpcor: Efficient Estimation of Covariance and (Partial) Correlation*. R package version 1.6.0, URL <http://CRAN.R-project.org/package=corpcor>.
- Schumacher RE, Lomax RG (1996). *A Beginner's Guide to Structural Equation Modeling*. Lawrence Erlbaum, Mahwah, NJ.
- Sharma S (2007). *Applied Multivariate Techniques*. John Wiley & Sons, New York.
- Steiger JH (1980). "Testing Pattern Hypotheses on Correlation Matrices: Alternative Statistics and some Empirical Results." *Multivariate Behavioral Research*, **15**, 335–352.
- Stevens SS (1946). "On the Theory of Scales of Measurement." *Science*, **103**, 677–680.
- Turner NE (1998). "The Effect of Common Variance and Structure on Random Data Eigenvalues: Implications for the Accuracy of Parallel Analysis." *Educational & Psychological Measurement*, **58**, 541–568.
- Velicer WF (1976). "Determining the Number of Components from the Matrix of Partial Correlations." *Psychometrika*, **41**, 321–327.
- Velicer WF, Eaton CA, Fava JL (2000). "Construct Explication Through Factor or Component Analysis: A Review and Evaluation of Alternative Procedures for Determining the Number of Factors or Components." In RD Goffin, E Helmes (eds.), *Problems and Solutions in Human Assessment*, pp. 41–71. Springer-Verlag.
- Velicer WF, Jackson DN (1990). "Component Analysis Versus Common Factor-Analysis – Some Further Observations." *Multivariate Behavioral Research*, **25**(1), 97–114.
- Wood JM, Tataryn DJ, Gorsuch RL (1996). "Effects of Under- and Overextraction on Principal Axis Factor Analysis with Varimax Rotation." *Psychological Methods*, **1**, 354–365.
- Zumbo BD, Gadermann AM, Zeisser C (2007). "Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales." *Journal of Modern Applied Statistical Methods*, **6**, 21–29.
- Zwick WR, Velicer WF (1986). "Comparison of Five Rules for Determining the Number of Components to Retain." *Psychological Bulletin*, **99**, 432–442.

## A. The menu

Obtain and install the appropriate version of R (R 2.10.1 for IBM SPSS Statistics 19) from the R website before installing the R essentials and plugin. Download the R essentials for IBM SPSS Statistics version 19 from <http://sourceforge.net/projects/ibmspssstat/files/VersionsforStatistics19/>. If one has Windows Vista or Windows 7, before installing the R essentials, disable `user account control` and download and install the appropriate version of the R essentials. Then install the file named `R-Factor.spd`. The file is installed following the commands: `Utilities` → `Custom Dialogs` → `Install Custom Dialog...`, and then point to the file `R-Factor.spd` and the dialog will be available in `Analyse` → `Dimension Reduction` → `R Factor...`. Eventually, for Windows Vista or Windows 7, users should enable again the `user account control`.

Within the program, each line should not exceed 251 bytes. This can be a problem if one has many input variables and/or their names are too long. To overcome this difficulty, one must push the button `Past` to open the syntax window, and then, manually, break the line `mdata` `<- spssdata.GetDataFromSPSS(variables = c("..."))`.

### A.1. Main dialog

The options available to deal with missing data are `pairwise` or `listwise`. The output shows the number of valid cases for the option choosed.

One can perform several analyses:

- `Correlations`. Tests multivariate normality and displays correlation matrices of different kinds.
- `PCA`. Performs a principal component analysis.
- `FA`. Performs a factorial analysis.
- `FA options`. Options for FA.
- `Number factors`. Criteria to help choosing the number of factors to retain.
- `Score items`. Scores items to a new database and calculates internal consistency coefficients.

### A.2. Correlations

The following correlation matrices can be analysed by checking in “correlations” the appropriate type:

- `Pearson`. Function `corr.test` from package `psych`.
- `Polychoric (two step)`. Function `hetcor` from package `polycor`, computing quick two-step estimates.
- `Polychoric (max.lik.)`. Function `hetcor` from package `polycor`, computing maximum-likelihood estimates.
- `Spearman`. Function `corr.test` from package `psych`.

Tests of multivariate normality can be performed if checked. The tests for multivariate normality are based on kurtosis (function `mvnorm.kur.test` from package **ICS**) and on skewness (`mvnorm.skew.test` from package **ICS**).

Correlations differences. The differences on the values of different correlations types can be performed.

### A.3. PCA (Principal component analysis)

Check box PCA to perform a principal component analysis (function `princomp`).

One can choose the Correlation Matrix type where the analysis is going to be performed: Pearson, Polychoric (two step), Polychoric (max.lik.) or Spearman or the Covariance matrix.

Checking the box sort loadings by size sort loadings (function `ICLUST.sort` from package **psych**).

### A.4. FA (Factor analysis)

Check box FA to perform a factor analysis.

Extraction of factors make use of packages **psych** and **nFactors**. Allows one to specify the method of factor extraction.

Available methods from package **psych** are:

- Principal components (function `principal`).
- Maximum likelihood (function `fa`).
- Principal axis factor (function `fa`). According to the author of the package **psych**, test cases comparing the output to SPSS suggest that the algorithm matches what SPSS calls unweighted least squares.
- Minimum residual (function `fa`).
- Weighted least squares (function `fa`).
- Generalized least squares (function `fa`).

Available methods from packages **nFactors** are (four different choices of initial communalities are given: Components, Maximum correlation, Generalized inverse and Multiple correlation):

- Principal axis (function `iterativePrincipalAxis`).
- Principal axis factor (function `principalAxis`).

For each Rotation, T denotes orthogonal rotation and Q denotes oblique rotation. All the rotations are from package **GPARotation**. Available are the following rotations: Varimax-T, Quartimax-T, Oblimin, Quartimin-Q, Equamax-T, Parsimax-T, Factor parsimony-T, Oblimin Biquartimin-Q, Oblimin Covarimin-Q, Oblimax-Q, Entropy-T, Simplimax-Q,



BentlerT-T, BentlerQ-Q, TandemI-T, TandemII-T, GeominT-T, GeominQ-Q, InfomaxT-T, InfomaxQ-Q, Mccammon-T.

Checking box **Kaiser normalization**, Kaiser normalization is done before rotation and then undone after rotation, for rotations from **GPArotation**.

By selecting **Do promax rotation**, one can perform a promax oblique rotation (a two-stage procedure where, commonly, the first stage consists of an orthogonal rotation such as varimax rotation). A promax rotation is only performed after an orthogonal rotation or no rotation. The code is derived from the function **Promax** from package **psych**. It performs a promax rotation after an orthogonal rotation specified in *Rotation* (includes the case where no previously rotation is selected). **Kappa for promax rotation** is the value of the power (typically between 2 and 4) to which the loadings are raised to obtain the target matrix.

If **simplimax rotation** is selected, enter the number of loadings one wants close to zero in the box **k for simplimax rotation**. If no number is written, the program assumes it equal to  $p \cdot (m - 1)$  for  $p$  variables and  $m$  factors.

Write how many factors to extract in the box **Number of factors to extract**. If this number is excessive it is automatically adjusted to the maximum allowed.

Checking box **Sort loadings by size**, sort loadings (function **fa.sort** from **psych**).

One can choose the **Correlation Matrix** type where the analysis is going to be performed.

## A.5. FA options

The following menus only work if **FA** (factor analysis) is selected.

In **Display** one can choose to display some useful coefficients and statistics:

- **Scree Plot**. Plot of the Cattell's scree test (function **plotuScree** from package **nFactors**).
- **Residual correlations**.
- **Partial correlations controlling variables** (partial correlations calculated by function **cor2pcor** from package **corpcor**).
- **Partial correlations controlling factors**.
- **KMO**. Kaiser-Meyer-Olkin measure of sampling adequacy.
- **MSA**. Measures of sampling adequacy for each of the variables.
- **GFI (simplified)**. Goodness-of-fit index simplified (Schumacher and Lomax 1996).
- **AGFI (simplified)**. Adjusted goodness-of-fit index (Schumacher and Lomax 1996).
- **RMSR**. Root mean square residuals.
- **Root mean square partials**. **RMSP**, root mean square partial correlations controlling factors.
- **Determinant**.

- **Bartlett's test, Steiger test and Jennrich test.** Test whether or not the correlation matrix is an identity matrix (functions `cortest.bartlett`, `cortest` and `cortest.jennrich` from package **psych**).

One can display the **Unrotated Factor Plot** or the **Rotated Factor Plot**. The last one is only displayed if a rotation is selected in FA (function `factor.plot` from package **psych**). Plot factor loadings from two factors and assign items to clusters by different colours (the red is for items not assigned to any cluster).

- **Plot factor loadings and Assign items to clusters if absolute values of loadings are greater than some cut point** choosed (the default is zero).
- One can choose which **factors to plot** and to assign clusters.

One can display the **Unrotated Diagram** or the **Rotated Diagram**. The last one is only displayed if a rotation is selected in FA (function `fa.diagram` from package **psych**). Graphical representation of factor loadings where all loadings with an absolute value greater than some cut point are represented as an edge (path).

- Check box **Sort loadings by size** to sort loadings.
- One can check **Only the biggest loading per item is shown** to have each item assigned to only one factor.
- One can define the cut point in **Show loadings with absolute value greater than** (default is 0.5).

## A.6. Number factors

One can choose **Parallel Analysis**, **Velicer's MAP** or **Very Simple Structure (VSS)** to help finding the number of factors to retain.

*Correlation matrix to analyse.* This is the correlation matrix used to determine the components or the factors of the actual data to perform the analysis.

**Velicer's MAP.** Check if one wants to calculate the Velicer's MAP. Velicer's MAP with the partial correlations squared and raised to the 4th power are shown in the output. Code is based on [O'Connor \(2000\)](#).

**Very simple structure (VSS).** Indicates the number of factors to retain by comparing the original correlation matrix to that reproduced for different complexities (function `VSS` from package **psych**). The output shows the VSS maximum for complexity 1 and 2. Options available:

- **Fit the diagonal.** Check if one wants to fit the diagonal.
- **Number of factors to extract.**
- **Extraction.** Specify the method of factor extraction.
- **Rotation.** Options are none, varimax, oblimin or promax.

Parallel Analysis and Non Graphical Scree Test. Options:

- **Number of samples.** Write the number of samples to be drawn.
- **Simulations based on:** **Mean** or **Quantile**, which means that the mean or a particular quantile from each set of eigenvalues computed from simulated data is compared to the actual data eigenvalues. In case **Quantile** is checked, the quantile of the eigenvalues must be mentioned (default is 0.95, percentile 95).
- **Model** indicates if parallel analysis is based on **Components** or on **Factors**, i.e., if simulated data is subjected to principal component analysis or factor analysis.
- **Samples from.** Choose if samples are generated by random uncorrelated standardized normal variables (check **Normal distribution**, where Pearson's correlation matrix is used), or by permutations of actual data (check **Permutation**). If one select **Permutation**, one must choose if it is done with replacement or without, and the correlation matrix to be calculated for each permutation (default is **Pearson**). Part of the code is adapted from functions **parallel** and **eigenBootParallel** from package **nFactors**.

The output contains several criteria of components/factors to retain:

Parallel analysis, the Kaiser rule, the acceleration factor (AF), the optimal coordinates (OC) (functions **nScree** and **plotnScree** from package **nFactors**).

## A.7. Score items

Check **Scores** to save the scores to a database according to the scales defined in **Items** and **Clusters**. The function **score.items** from package **psych** is used. It will adjust scores for reverse scored items.

- **Save scores to database named:.** Give a name to the database.
- **Total or average scores?.** Choose to write total scores or average scores.
- **Missing values.** Check **Replace missing values with the item median or mean** or **Do not score cases with missing values**.
- **If replacing missing values.** If one check to replace missing values, choose between **median** or **mean**.

**Reliability of Scales.** Check **Internal consistency** and choose which coefficients are going to be calculated. The scales to calculate the internal consistency are defined in **Items** and **Clusters**. It will adjust scores for reverse scored items.

- **Cronbach's alpha.** Calculates raw and standardized coefficients.
- **Ordinal coefficient alpha.** The same as Cronbach's alpha coefficient but uses a polychoric correlation matrix instead.

- **Armor's reliability theta.** It is given by the formula  $\theta = \frac{p}{1-p} \left(1 - \frac{1}{\lambda}\right)$ , where  $p$  is the number of items in the scale and  $\lambda$  is the largest eigenvalue from the principal component analysis of the correlation matrix of the items of the scale.
- **Ordinal coefficient theta.** The same as Armor's reliability theta coefficient but uses a polychoric correlation matrix instead.

One can choose the method to estimate the polychoric correlations for ordinal coefficients in `Estimation of polychoric correlations (if needed)`.

`Items and clusters`. `Scores or Internal consistency` must be checked. If not, no output is shown.

- **Scales.**
  - `Factors to clusters (FA required)`. Factor analysis must be checked. From the matrix of loadings of a factor analysis, each item, according to the minimum absolute loading indicated in `Extract items with loadings absolute values greater than (factors to clusters must be checked)`, is assigned to the cluster corresponding to the largest factor loading for that item. Function `factor2cluster` from package **psych** is used.
  - `All items equal (one cluster)`.
  - `Enter keys manually`. Function `make.keys` from package **psych** is used to create a keys matrix of  $-1$ ,  $0$  or  $1$ , to score each item of composite scales. The scales are defined in `Enter keys manually (enter keys manually must be checked)`. It will adjust scores for reverse scored items ( $-1$ ).
- `Extract items with loadings absolute values greater than (factors to clusters must be checked)`. To score items from a factor analysis.
- `Enter keys manually (enter keys manually must be checked)`. To score items assigned manually to scales.

Requesting to save the scores to a database or reliability of scales, provides some output performed by the function `score.items` from package **psych**:

- Number of items for each scale.
- The average correlation within scales.
- The intercorrelation of all the scales.
- The unattenuated correlations of scales with coefficients raw alpha on the diagonal.
- Correlation of items with scales not corrected for item overlap.
- Correlation of items with scales corrected for item overlap.

**Affiliation:**

Mário Basto  
Department of Science  
Polytechnic Institute of Cávado and Ave  
4750-810 Vila Frescainha S. Martinho BCL, Portugal  
E-mail: [mbasto@ipca.pt](mailto:mbasto@ipca.pt)

José Manuel Pereira  
Department of Accounting and Taxation  
Polytechnic Institute of Cávado and Ave  
4750-810 Vila Frescainha S. Martinho BCL, Portugal  
E-mail: [jpereira@ipca.pt](mailto:jpereira@ipca.pt)