

## Statistics:

### 3.3 Factor Analysis

Rosie Cornish. 2007.

## 1 Introduction

This handout is designed to provide only a brief introduction to factor analysis and how it is done. Books giving further details are listed at the end.

As for principal components analysis, factor analysis is a multivariate method used for data reduction purposes. Again, the basic idea is to represent a set of variables by a smaller number of variables. In this case they are called **factors**. These factors can be thought of as underlying constructs that cannot be measured by a single variable (e.g. happiness).

## 2 Assumptions

Factor analysis is designed for interval data, although it can also be used for ordinal data (e.g. scores assigned to Likert scales). The variables used in factor analysis should be linearly related to each other. This can be checked by looking at scatterplots of pairs of variables. Obviously the variables must also be at least moderately correlated to each other, otherwise the number of factors will be almost the same as the number of original variables, which means that carrying out a factor analysis would be pointless.

## 3 The steps in factor analysis

The factor analysis model can be written algebraically as follows. If you have  $p$  variables  $X_1, X_2, \dots, X_p$  measured on a sample of  $n$  subjects, then variable  $i$  can be written as a linear combination of  $m$  factors  $F_1, F_2, \dots, F_m$  where, as explained above  $m < p$ . Thus,

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + e_i$$

where the  $a_{is}$ s are the factor loadings (or scores) for variable  $i$  and  $e_i$  is the part of variable  $X_i$  that cannot be 'explained' by the factors.

There are three main steps in a factor analysis:

1. Calculate initial factor loadings.

This can be done in a number of different ways; the two most common methods are described very briefly below:

- **Principal component method**

As the name suggests, this method uses the method used to carry out a principal

components analysis. However, the factors obtained will not actually be the principal components (although the loadings for the  $k^{th}$  factor will be proportional to the coefficients of the  $k^{th}$  principal component).

- **Principal axis factoring**

This is a method which tries to find the lowest number of factors which can account for the variability in the original variables that is associated with these factors (this is in contrast to the principal components method which looks for a set of factors which can account for the total variability in the original variables).

These two methods will tend to give similar results if the variables are quite highly correlated and/or the number of original variables is quite high. Whichever method is used, the resulting factors at this stage will be uncorrelated.

## 2. Factor rotation

Once the initial factor loadings have been calculated, the factors are rotated. This is done to find factors that are easier to interpret. If there are 'clusters' (groups) of variables — i.e. subgroups of variables that are strongly inter-related — then the rotation is done to try to make variables within a subgroup score as highly (positively or negatively) as possible on one particular factor while, at the same time, ensuring that the loadings for these variables on the remaining factors are as low as possible. In other words, the object of the rotation is to try to ensure that all variables have high loadings only on one factor.

There are two types of rotation method, **orthogonal** and **oblique** rotation. In orthogonal rotation the rotated factors will remain uncorrelated whereas in oblique rotation the resulting factors will be correlated. There are a number of different methods of rotation of each type. The most common orthogonal method is called **varimax** rotation; this is the method that many books will recommend.

## 3. Calculation of factor scores

When calculating the final factor scores (the values of the  $m$  factors,  $F_1, F_2, \dots, F_m$ , for each observation), a decision needs to be made as to how many factors to include. This is usually done using one of the following methods:

- Choose  $m$  such that the factors account for a particular percentage (e.g. 75%) of the total variability in the original variables.
- Choose  $m$  to be equal to the number of eigenvalues over 1 (if using the correlation matrix). [A different criteria must be used if using the covariance matrix.]
- Use the scree plot of the eigenvalues. This will indicate whether there is an obvious cut-off between large and small eigenvalues.

In some statistical packages (e.g. SPSS) this choice is actually made at the outset. The second method, choosing eigenvalues over 1, is probably the most common one.

The final factor scores are usually calculated using a regression-based approach.

## 4 Carrying out factor analysis in SPSS

- **Analyze**
- **Data Reduction**
- **Factor**
- Select the variables you want the factor analysis to be based on and move them into the **Variable(s)** box.
- In the **Descriptives** window, you should select **KMO and Bartlett's test of sphericity**. KMO is a statistic which tells whether you have sufficient items for each factor. It should be over 0.7. Bartlett's test is used to check that the original variables are sufficiently correlated. This test should come out significant ( $p < 0.05$ ) — if not, factor analysis will not be appropriate. Click on **Continue**.
- In the **Extraction** window, you can select the extraction method you want to use (e.g. principal components, etc.). Under **Analyze** ensure that **Correlation Matrix** is selected (this is the default). The default is also to extract eigenvalues over 1 but if you want to extract a specific number of factors you can specify this. Click on **Continue**.
- In the **Rotation** window you can select your rotation method (as mentioned above, **Varimax** is the most common). You can also ask SPSS to display the rotated solution. Once you have selected this click on **Continue**.
- In the **Scores** window you can specify whether you want SPSS to save factor scores for each observation (this will save them as new variables in the data set). Under **Method** choose **Regression**. You can also ask SPSS to display the factor score coefficients (the  $a_i$ s). Click on **Continue**.
- **OK**

## 5 References

- Manly, B.F.J. (2005), **Multivariate Statistical Methods: A primer**, Third edition, Chapman and Hall.
- Rencher, A.C. (2002), **Methods of Multivariate Analysis**, Second edition, Wiley.