

# Overview of Factor Analysis

Jamie DeCoster

Department of Psychology  
University of Alabama  
348 Gordon Palmer Hall  
Box 870348  
Tuscaloosa, AL 35487-0348

Phone: (205) 348-4431  
Fax: (205) 348-8648

August 1, 1998

If you wish to cite the contents of this document, the APA reference for them would be

DeCoster, J. (1998). *Overview of Factor Analysis*. Retrieved <month, day, and year you downloaded this file> from <http://www.stat-help.com/notes.html>

For future versions of these notes or help with data analysis visit  
<http://www.stat-help.com>

ALL RIGHTS TO THIS DOCUMENT ARE RESERVED.

# Contents

1	Theoretical Introduction	1
2	Exploratory Factor Analysis	2
3	Confirmatory Factor Analysis	5
4	Combining Exploratory and Confirmatory Factor Analyses	7

# Chapter 1

## Theoretical Introduction

- Factor analysis is a collection of methods used to examine how underlying constructs influence the responses on a number of measured variables.
- There are basically two types of factor analysis: exploratory and confirmatory.
  - Exploratory factor analysis (EFA) attempts to discover the nature of the constructs influencing a set of responses.
  - Confirmatory factor analysis (CFA) tests whether a specified set of constructs is influencing responses in a predicted way.
- Both types of factor analyses are based on the *Common Factor Model*, illustrated in figure 1.1. This model proposes that each observed response (measure 1 through measure 5) is influenced partially by underlying common factors (factor 1 and factor 2) and partially by underlying unique factors (E1 through E5). The strength of the link between each factor and each measure varies, such that a given factor influences some measures more than others. This is the same basic model as is used for LISREL analyses.

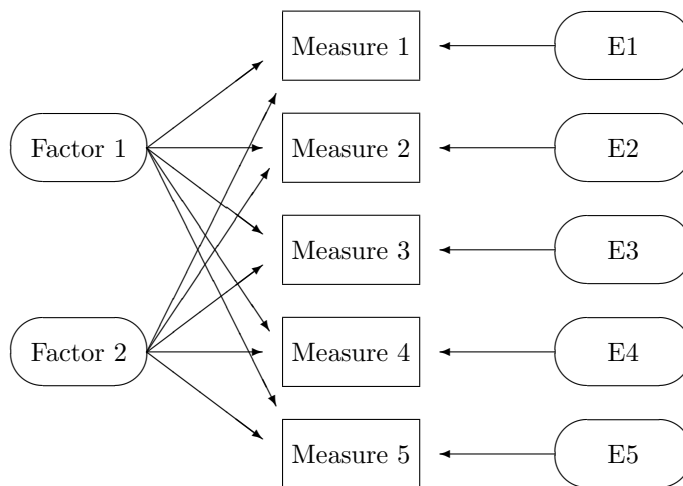


Figure 1.1: The Common Factor Model.

- Factor analyses are performed by examining the pattern of correlations (or covariances) between the observed measures. Measures that are highly correlated (either positively or negatively) are likely influenced by the same factors, while those that are relatively uncorrelated are likely influenced by different factors.

## Chapter 2

# Exploratory Factor Analysis

### 2.1 Objectives

- The primary objectives of an EFA are to determine
  1. The number of common factors influencing a set of measures.
  2. The strength of the relationship between each factor and each observed measure.
- Some common uses of EFA are to
  - Identify the nature of the constructs underlying responses in a specific content area.
  - Determine what sets of items “hang together” in a questionnaire.
  - Demonstrate the dimensionality of a measurement scale. Researchers often wish to develop scales that respond to a single characteristic.
  - Determine what features are most important when classifying a group of items.
  - Generate “factor scores” representing values of the underlying constructs for use in other analyses.

### 2.2 Performing EFA

There are seven basic steps to performing an EFA:

1. **Collect measurements.** You need to measure your variables on the same (or matched) experimental units.
2. **Obtain the correlation matrix.** You need to obtain the correlations (or covariances) between each of your variables.
3. **Select the number of factors for inclusion.** Sometimes you have a specific hypothesis that will determine the number factors you will include, while other times you simply want your final model to account for as much of the covariance in your data with as few factors as possible. If you have  $k$  measures, then you can at most extract  $k$  factors. There are a number of methods to determine the “optimal” number of factors by examining your data. The *Kaiser criterion* states that you should use a number of factors equal to the number of the eigenvalues of the correlation matrix that are greater than one. The “Scree test” states that you should plot the eigenvalues of the correlation matrix in descending order, and then use a number of factors equal to the number of eigenvalues that occur prior to the last major drop in eigenvalue magnitude.
4. **Extract your initial set of factors.** You must submit your correlations or covariances into a computer program to extract your factors. This step is too complex to reasonably be done by hand. There are a number of different extraction methods, including maximum likelihood, principal component, and principal axis extraction. The best method is generally maximum likelihood extraction, unless you seriously lack multivariate normality in your measures.

5. **Rotate your factors to a final solution.** For any given set of correlations and number of factors there are actually an infinite number of ways that you can define your factors and still account for the same amount of covariance in your measures. Some of these definitions, however, are easier to interpret theoretically than others. By rotating your factors you attempt to find a factor solution that is equal to that obtained in the initial extraction but which has the simplest interpretation.

There are many different types of rotation, but they all try make your factors each highly responsive to a small subset of your items (as opposed to being moderately responsive to a broad set). There are two major categories of rotations, *orthogonal rotations*, which produce uncorrelated factors, and *oblique rotations*, which produce correlated factors. The best orthogonal rotation is widely believed to be Varimax. Oblique rotations are less distinguishable, with the three most commonly used being Direct Quartimin, Promax, and Harris-Kaiser Orthoblique.

6. **Interpret your factor structure.** Each of your measures will be linearly related to each of your factors. The strength of this relationship is contained in the respective *factor loading*, produced by your rotation. This loading can be interpreted as a standardized regression coefficient, regressing the factor on the measures.

You define a factor by considering the possible theoretical constructs that could be responsible for the observed pattern of positive and negative loadings. To ease interpretation you have the option of multiplying all of the loadings for a given factor by -1. This essentially reverses the scale of the factor, allowing you, for example, to turn an “unfriendliness” factor into a “friendliness” factor.

7. **Construct factor scores for further analysis.** If you wish to perform additional analyses using the factors as variables you will need to construct factor scores. The score for a given factor is a linear combination of all of the measures, weighted by the corresponding factor loading. Sometimes factor scores are idealized, assigning a value of 1 to strongly positive loadings, a value of -1 to strongly negative loadings, and a value of 0 to intermediate loadings. These factor scores can then be used in analyses just like any other variable, although you should remember that they will be strongly collinear with the measures used to generate them.

## 2.3 Factor Analysis vs. Principal Component Analysis

- Exploratory factor analysis is often confused with principal component analysis (PCA), a similar statistical procedure. However, there are significant differences between the two: EFA and PCA will provide somewhat different results when applied to the same data.
- The purpose of PCA is to derive a relatively small number of components that can account for the variability found in a relatively large number of measures. This procedure, called *data reduction*, is typically performed when a researcher does not want to include all of the original measures in analyses but still wants to work with the information that they contain.
- Differences between EFA and PCA arise from the fact that the two are based on different models. An illustration of the PCA model is provided in figure 2.1. The first difference is that the direction of influence is reversed: EFA assumes that the measured responses are based on the underlying factors while in PCA the principal components are based on the measured responses. The second difference is that EFA assumes that the variance in the measured variables can be decomposed into that accounted for by common factors and that accounted for by unique factors. The principal components are defined simply as linear combinations of the measurements, and so will contain both common and unique variance.
- In summary, you should use EFA when you are interested in making statements about the factors that are responsible for a set of observed responses, and you should use PCA when you are simply interested in performing data reduction.

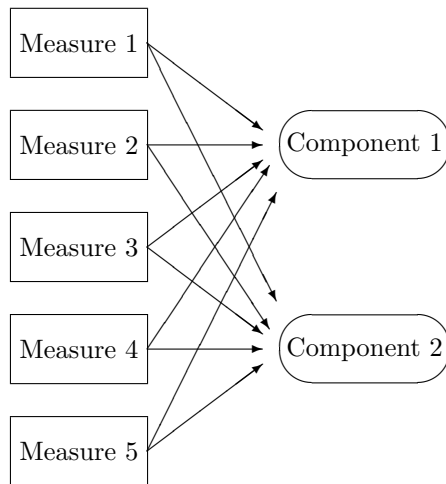


Figure 2.1: The model for principal components analysis.

## 2.4 Miscellaneous notes on EFA

- To have acceptable reliability in your parameter estimates it is best to have data from at least 10 subjects for every measured variable in the model. This number should be increased if you expect that the influence of the common factors is relatively weak. You should also have measurements from at least three variables for every factor that you want to include in your model.
- You should endeavor to have a wide variety of measurements for your EFA. The more accurately that your selection of measurements properly represents the “population” of measurements that could be taken, the more generality you will have in your findings.
- The procedures used to measure the fit of a model in CFA (described below) can also be used to test the fit of an EFA model. These tests reject the null hypothesis when the model does *not* fit, so you must be cautious interpreting the results. Even a very accurate model will generate a significant lack of fit statistic if the sample size is large.
- EFA can be performed in SAS using *proc factor*. Principal component analysis can be performed in SAS using *proc princomp*, while it can be performed in SPSS using the *Analyze/Data reduction/Factor analysis* menu selection. EFA cannot actually be performed in SPSS (despite the name of menu item used to perform PCA). You can, however, simply use component scores for the same purposes as you would typically use for EFA.

## Chapter 3

# Confirmatory Factor Analysis

### 3.1 Objectives

- The primary objective of a CFA is to determine the ability of a predefined factor model to fit an observed set of data.
- Some common uses of CFA are to
  - Establish the validity of a single factor model.
  - Compare the ability of two different models to account for the same set of data.
  - Test the significance of a specific factor loading.
  - Test the relationship between two or more factor loadings.
  - Test whether a set of factors are correlated or uncorrelated.
  - Assess the convergent and discriminant validity of a set of measures.

### 3.2 Performing CFA

There are six basic steps to performing an CFA:

1. **Define the factor model.** The first thing you need to do is to precisely define the model you wish to test. This involves selecting the number of factors, and defining the nature of the loadings between the factors and the measures. These loadings can be fixed at zero, fixed at another constant value, allowed to vary freely, or be allowed to vary under specified constraints (such as being equal to another loading in the model).
2. **Collect measurements.** You need to measure your variables on the same (or matched) experimental units.
3. **Obtain the correlation matrix.** You need to obtain the correlations (or covariances) between each of your variables.
4. **Fit the model to the data.** You will need to choose a method to obtain the estimates of factor loadings that were free to vary. The most common model-fitting procedure is *Maximum likelihood* estimation, which should probably be used unless your measures seriously lack multivariate normality. In this case you might wish to try using *Asymptotically distribution free* estimation.
5. **Evaluate model adequacy.** When the factor model is fit to the data, the factor loadings are chosen to minimize the discrepancy between the correlation matrix implied by the model and the actual observed matrix. The amount of discrepancy after the best parameters are chosen can be used as a measure of how consistent the model is with the data.

The most commonly used test of model adequacy is the  $\chi^2$  *goodness-of-fit* test. The null hypothesis for this test is that the model adequately accounts for the data, while the alternative is that there is a significant amount of discrepancy. Unfortunately, this test is highly sensitive to the size of your sample, such that tests involving large samples will generally lead to a rejection of the null hypothesis, even when the factor model is appropriate. Other statistics such as the *Tucker-Lewis index*, compare the fit of the proposed model to that of a “null model.” These statistics have been shown to be much less sensitive to sample size.

6. **Compare with other models.** If you want to compare two models, one of which is a reduced form of the other, you can just examine the difference between their  $\chi^2$  statistics, which will also have an approximately  $\chi^2$  distribution. Almost all tests of individual factor loadings can be made as comparisons of full and reduced factor models. In cases where you are not examining full and reduced models you can compare the *Root mean square error of approximation* (RMSEA), which is an estimate of discrepancy per degree of freedom in the model.

### 3.3 Miscellaneous notes on CFA

- CFA has strong links to structural equation modeling, a relatively nonstandard area of statistics. It is much more difficult to perform a CFA than it is to perform an EFA.
- A CFA requires a larger sample size than an EFA, basically because the CFA produces inferential statistics. The exact sample size necessary will vary heavily with the number of measures and factors in the model, but you can expect to require around 200 subjects for a standard model.
- As in EFA, you should have at least three measures for each factor in your model. Unlike EFA, however, you should choose measures that are strongly associated with the factors in your model (rather than those that would be a “random sample” of potential measures).
- CFA can be performed in SAS using *proc calis*, but cannot be performed in SPSS. However, SPSS does produce another software package called AMOS which will perform CFA. CFA are also commonly analyzed using LISREL.



## Chapter 4

# Combining Exploratory and Confirmatory Factor Analyses

- In general, you want to use EFA if you do not have strong theory about the constructs underlying responses to your measures and CFA if you do.
- It is reasonable to use an EFA to generate a theory about the constructs underlying your measures and then follow this up with a CFA, but this *must* be done using separate data sets. You are merely fitting the data (and not testing theoretical constructs) if you directly put the results of an EFA directly into a CFA on the same data. An acceptable procedure is to perform an EFA on one half of your data, and then test the generality of the extracted factors with a CFA on the second half of the data.
- If you perform a CFA and get a significant lack of fit, it is perfectly acceptable to follow this up with an EFA to try to locate inconsistencies between the data and your model. However, you should test any modifications you decide to make to your model on new data.

# References on Factor Analysis

- Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute Press.

This is an excellent resource for information about how to perform both exploratory and confirmatory factor analysis. It does a good job explaining both theoretical and applied issues.

- Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-014). Newbury Park, CA: Sage.

This manual details more advanced topics in factor analysis, such as the different extraction methods, rotation methods, and procedures for confirmatory factor analysis.

- Kim, J. O., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-013). Newbury Park, CA: Sage.

This manual provides an excellent overview to the theory behind factor analysis, including the common factor model. It also provides procedures for performing an exploratory factor analysis.

- Wegener, D. T., & Fabrigar, L. R. (in press). Analysis and design for nonexperimental data: Addressing causal and noncausal hypotheses. In R. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology*. London: Cambridge University Press.

This book chapter provides an overview of several nonexperimental statistical methods, including confirmatory and exploratory factor analysis, structural equation modeling, multidimensional scaling, cluster analysis, and hierarchical linear modeling.