Practical Exploratory Factor Analysis: An Overview

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Practical Exploratory Factor Analysis: An Overview

Introduction

- Factor Analysis is an important and widely used multivariate method.
- A number of techniques are referred to as "factor analysis methods," but experts currently concentrate primarily on two approaches, which we will refer to as *common factor analysis* and *principal component analysis*.
- People have been arguing the relative theoretical merits of these methods for many years, often with considerable ego-investment and with no shortage of vitriol.
- Interestingly, a significant percentage of the time the fundamental substantive conclusions from a common factor analysis and a component analysis will be very similar.

Introduction

- Proponents of the common factor model often present examples built around data sets (authentic or artificial) that fit the common factor model well, then expound on the fact that the solutions obtained by component analysis differs from that obtained by factor analysis.
- I'm not sure how seriously we should take these demonstrations, but I think we need to keep an open mind.
- In this lecture, we'll cover the basics of exploratory factor analysis.

Why Do an Exploratory Factor Analysis?

- Structural Exploration
- Structural Confirmation
- Data Reduction
- Attribute Scoring

Structural Exploration

- The two main factor analytic models can both be written

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{e} \tag{1}$$

- The *factor pattern* $\mathbf{F}$ is the set of linear weights carrying the (small number of) factors (or components) into the larger number of observed variables. If, for the $p \times m$ matrix $\mathbf{F}$, p is substantially larger then m, we may be able to attain several data analysis goals at once.
- First, the idea emerges that this compact $\mathbf{F}$ may be "a model for what underlies $\mathbf{y}$." A number of authors (most notably of late, the statistician and philosopher of science Michael Maraun) have severely criticized the ubiquity of this idea and its logical foundations. But if you accept the idea, then the form of $\mathbf{F}$ may reveal important aspects of the structure of the variance in a domain of content.
- In this course, we'll try to have our cake, and eat it too. We'll learn to interpret a factor pattern while remaining very cautious and skeptical about the factor analysis modeling enterprise.

Structural Confirmation

- In some situations, you expect to find a certain structure in your multivariate data.
- An exploratory factor analysis (depending on how it turns out) can confirm these expectations.
- Such expectations may be approached in a more formal way with so-called *confirmatory factor analysis* methods, which we'll examine in detail in a later lecture.

Data Reduction

- One important potential use for factor analytic methods is *data reduction*.
- If you have a large number of variables, and a relatively small number of observations, then it is quite likely that a number of dimensions in your data are redundant.
- Many variables may, essentially, be unreliable measures of the same construct.
- As we learn in courses on measurement, linear combination (even simple summing) of a number of unreliable measures of a construct can yield a more reliable measure of the construct, and reduce the number of variables in the analysis from several to one.
- This principle allows factor analysis methods to yield *scale scores* for those factors that appear to coincide with important attributes.

Steps in a Common Factor Analysis

- Design the Study
- Gather the Data
- Choose the Model
- Select m, the Number of Factors
- Rotate the Factors
- Interpret the Factors and Name Them
- Obtain Scale Scores

Design the Study

- Choose the domain of content
- Analyze its *facets*
- Try to have at least 3–4 items per facet
- Perform power and sample size analysis

Power and Sample Size Analysis

A factor analysis involves two fundamental kinds of power and sample size analysis.

One must have adequate sample size to accurately assess:

- Overall model fit
- Accuracy of parameter estimates
- Likelihood of Heywood cases (in the event common factor analysis is chosen)

Gather the Data

- It may take weeks, months, or even years to gather the data required to factor analyze a significant domain of behavior.
- False starts can be very costly.
- Always remember the problem of spurious correlations, and avoid combining males and females, or identifiable subgroups with significantly different mean or covariance structures.
- On the other hand, actively consider whether a convenience sample might produce false conclusions because it is not sufficiently representative of the population of interest.

Choose the Model

- Your fundamental choices are *common factor analysis* or *principal component analysis*.
- For onr detailed discussion of the issues regarding this choice, consult Fabrigar, Wegener, MacCallum, and Strahan (1999) available in our readings.
- One viewpoint is that smaller sample sizes or an emphasis on data reduction would favor principal component approaches, while situations where sample size is adequate and one "desires to go beyond the test space to discover new variables" favor a choice of common factor analysis.
- As we shall see when examining theoretical aspects of factor analysis, all such pronouncements have proven controversial.

Select the Number of Factors

- Selection of the number of factors is a key decision in a factor analysis.
- Either "under-factoring" or "over-factoring" can result in substantial differences in the ultimate result of the analysis.
- Several criteria are employed routinely in selection of the number of factors.
- Ironically, some of the more popular approaches have been strongly criticized.

Rotate the Factors

- In a system of the form $\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{e}$, when $m \geq 2$, there are infinitely many *empirically equivalent* representations for the factors.
- To see this, recall that $\mathbf{F}\mathbf{x} = \mathbf{F}^*\mathbf{x}^*$, where $\mathbf{F}^* = \mathbf{F}\mathbf{T}$, $\mathbf{x}^* = \mathbf{T}^{-1}\mathbf{x}$. If the original factors are orthogonal, i.e., $\text{Var}(\mathbf{x}) = \mathbf{I}$, then the "rotated" factors will have $\text{Var}(\mathbf{x}^*) = \mathbf{T}^{-1}\mathbf{T}'^{-1}$. Note that if $\mathbf{T}$ is an orthogonal matrix, that is, $\mathbf{T}\mathbf{T}' = \mathbf{I}$, then the factors will remain orthogonal after rotation.
- The goal of rotation is to obtain *simple structure*, a situation where there are lots of near-zero loadings, and most factors load only on a "coherent" subset of variables.
- Usually one tries orthogonal rotation first, then tries *oblique rotation* (allowing the factors to correlate) if the orthogonal solution is not sufficiently simple.

Interpret and Name the Factors

- Once simple structure has been obtained, one examines the pattern of loadings and attempts to interpret and understand the factors.
- The process involves examining the items that are common to a factor, and, relying on substantive knowledge of the domain of interest, conceptualizing a process common to these variables.