

A Quick Primer on Exploratory Factor Analysis

Exploratory vs. Confirmatory Factor Analysis

Similarities

- Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are two statistical approaches used to examine the internal reliability of a measure.
- Both are used to investigate the theoretical constructs, or factors, that might be represented by a set of items.
- Either can assume the factors are uncorrelated, or *orthogonal*.
- Both are used to assess the quality of individual items.
- Both can be used for exploratory or confirmatory purposes.

Differences

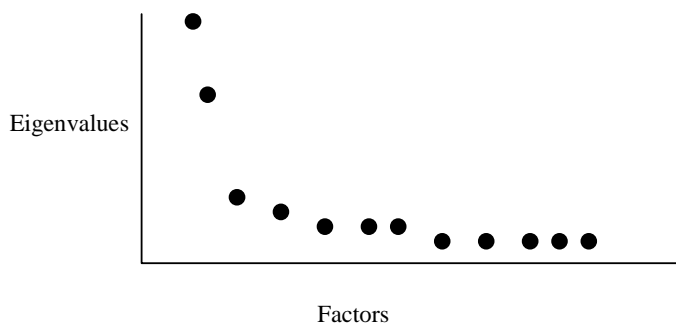
- With EFA, researchers usually decide on the number of factors by examining output from a principal components analysis (i.e., eigenvalues are used). With CFA, the researchers must specify the number of factors a priori.
- CFA requires that a particular factor structure be specified, in which the researcher indicates which items load on which factor. EFA allows all items to load on all factors.
- CFA provides a fit of the hypothesized factor structure to the observed data.
- Researchers typically use maximum likelihood to estimate factor loadings, whereas maximum likelihood is only one of a variety of estimators used with EFA.
- CFA allows the researchers to specify correlated measurement errors, constrain loadings or factor correlations to be equal to one another, perform statistical comparisons of alternative models, test second-order factor models, and statistically compare the factor structure of two or more groups.

Exploratory Factor Analysis: Purpose

Exploratory factor analysis (EFA) is generally used to discover the factor structure of a measure and to examine its internal reliability. EFA is often recommended when researchers have no hypotheses about the nature of the underlying factor structure of their measure. Exploratory factor analysis has three basic decision points: (1) decide the number of factors, (2) choosing an extraction method, (3) choosing a rotation method.

Exploratory Factor Analysis: Deciding the number of factors

The most common approach to deciding the number of factors is to generate a scree plot. The scree plot is a two dimensional graph with factors on the x-axis and *eigenvalues* on the y-axis. Eigenvalues are produced by a process called *principal components analysis* (PCA) and represent the variance accounted for by each underlying factor. They are not represented by percentages but scores that total to the number of items. A 12-item scale will theoretically have 12 possible underlying factors, each factor will have an eigenvalue that indicates the amount of variation in the items accounted for by each factor. If a the first factor has an eigenvalue of 3.0, it accounts for 25% of the variance ($3/12=.25$). The total of all the eigenvalues will be 12 if there are 12 items, so some factors will have smaller eigenvalues. They are typically arranged in a scree plot in decending order like the following:



From the scree plot you can see that the first couple of factors account for most of the variance, then the remaining factors all have small eigenvalues. The term “scree” is taken from the word for the rubble at the bottom of a mountain. A researcher might

examine this plot and decide there are 2 underlying factors and the remainder of factors are just “scree” or error variation. So, this approach to selecting the number of factors involves a certain amount of subjective judgment.

Another approach is called the Kaiser-Guttman rule and simply states that the number of factors are equal to the number of factors with eigenvalues greater than 1.0. I tend to recommend the scree plot approach because the Kaiser-Guttman approach seems to produce many factors.

Exploratory Factor Analysis: Factor Extraction

Once the number of factors are decided the researcher runs another factor analysis to get the loadings for each of the factors. To do this, one has to decided which mathematical solution to use to find the loadings. There are about five basic extraction methods (1) PCA, which is the default in most packages. PCA assumes there is no measurement error and is considered not to be a true exploratory factor analysis; (2) maximum likelihood (a.k.a. canonical factoring); (3) alpha factoring, (4) image factoring, (5) principal axis factoring with iterated communalities (a.k.a. least squares).

Without getting into the details of each of these, I think the best evidence supports the use of principal axis factoring and maximum likelihood approaches. I typically use the former. Gorsuch (1989) recommends the latter if only a few iterations are performed (not really possible in most packages). Snook and Gorsuch (1989) show that PCA can give poor estimates of the population loadings in small samples. With larger samples, most approaches will have similar results.

The extraction method will produce factor loadings for every item on every extracted factor. Researchers hope their results will show what is called *simple structure*, with most items having a large loading on one factor but small loadings on other factors.

Exploratory Factor Analysis: Rotation

Once an initial solution is obtained, the loadings are *rotated*. Rotation is a way of maximizing high loadings and minimizing low loadings so that the simplest possible structure is achieved. There are two basic types of rotation: *orthogonal* and *oblique*. Orthogonal means the factors are assumed to be uncorrelated with one another. This is the default setting in all statistical packages but is rarely a logical assumption about factors in the social sciences. Not all researchers using EFA realize that orthogonal rotations imply the assumption that they probably would not consciously make. Oblique rotation derives factor loadings based on the assumption that the factors are correlated, and this is probably most likely the case for most measures. So, oblique rotation gives the correlation between the factors in addition to the loadings. Here are some common algorithms for orthogonal and oblique rotation:

Orthogonal rotation: varimax, quartamax, equamax.

Oblique rotation: oblimin, promax, direct quartimin

I am not an expert on the advantages and disadvantages of each of these rotation algorithms, and they reportedly produce fairly similar results under most circumstances (although orthogonal and oblique rotations will be rather different). I tend to use promax rotation because it is known to be relatively efficient at achieving simple oblique structure.

References and Further Readings

Kim, J.-O., & Mueller, C.W. (1978). *Introduction to factor analysis: What it is and how to do it*. Newbury Park: Sage.

Kim, J.-O., & Mueller, C.W. (1978). *Factor analysis: Statistical methods and practical issues*. Newbury Park: Sage.

Gorsuch, R.L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, 25, 33-39.

Preacher, K.J., & MacCallum, R.C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2, 13-43.

Snook, S.C., & Gorsuch, R.L. (1989). Principal component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, 106, 148-154.

Tabachnick, B.G., & Fidell, L.S. (2000). *Using multivariate statistics* (4th Ed.). New York: Harper-Collins.