INTEGRATIVE STATISTICS
Customized Statistical Analysis
and Survey Research

Guide to Decision-Making in Exploratory Factor Analysis

Roland B. Stark, June 2019

Introduction

Even 115 years after Charles Spearman's invention of factor analysis, this ingenious technique still presents the analyst with a bewildering array of decisions. This guide aims to help.

This article introduces *exploratory factor analysis* or EFA. This method is not to be confused with *confirmatory factor analysis*, which is arguably a more rigorous technique related to *structural equation modeling*. EFA can enable an analyst to identify patterns in data. It can help one characterize the unmeasured, *latent* dimensions (or themes or topics) that can account for the measured, observable variables in a study. In so doing EFA can also achieve *data reduction* – summarizing many scores or ratings into just a few.

The purposes and functions of EFA are actually many:

- To learn what essential themes are behind a set of traits, opinions or attitudes
- To see how large a role each such theme plays
- To uncover patterns of relationships not easily visible through correlations alone
- To summarize a large number of responses into a few valid and reliable scores
- To assess *construct validity*

EFA differs from many familiar statistical procedures in that significance testing plays little role. With the occasional exception, the analyst addresses the various decision points described below without recourse to *p*-values. One might describe EFA as "as much art as science," and at times it almost has the feel of a projective exercise like a Rorschach test. As you will see, there is seldom a single correct EFA solution, and any study for which it is claimed simply that "factor analysis shows …" should be treated skeptically. The analyst needs to back up such a claim with evidence as to why s/he made the various decisions s/he did. Hopefully this guide will help you both to make informed decisions with your own data and to critically evaluate the decisions made by others.

The decision points, which are interconnected:

A. Decide how to treat missing data
B. Assess data suitability for factor analysis
C. Choose an extraction method
D. Pick a rotation method
E. Decide how many factors to extract
F. Decide which items[1] to keep in the analysis
G. Choose a way of displaying results

This article explains each decision point and then ends with a cautionary tale.

A. Decide how to treat missing data

This is a step that many skip over, but it can greatly affect the validity of results. Suppose some cases (observations; typically, survey respondents) have blanks for certain variables. A default, perhaps unthinking approach is to include in the EFA only those cases with complete data. This is called *complete case analysis* or *listwise deletion*. One problem is that it may shrink the sample size too much to allow for precise estimates. Another problem is more insidious: with respect to the patterns to be uncovered by EFA, the cases left in the analysis may differ substantially from those left out. One would say the data, rather than being "*missing completely at random*" as we would like, are "*missing not at random.*"

One response to missing data is *pairwise deletion*: to include all data available, whether each case is complete or not. This carries its own risks. For one, it may make the analysis impossible if data are sparse enough. EFA software will produce errors indicating a "Heywood case" or "failure to converge." Secondly, like listwise deletion, pairwise deletion runs the risk of distorting findings if the data are missing in a non-random, biased way.

Since either listwise or pairwise deletion has the potential to bias results, many analysts consider filling in missing values via *imputation*. Imputation can be done in a variety of ways, from the crude (e.g., substituting the mean in place of any blanks) to the sophisticated (e.g., methods of *multiple imputation*). Whatever decision one makes with regard to missing data, a safe course is to run multiple analyses, each employing a different missing data strategy, and to compare results.[2]

[1] I follow convention in using "items" to indicate the variables used as inputs. Typically these will be survey items, i.e., fields representing responses to survey questions or statements.

[2] Recommended sources: 1. Allison, P. D. (2002). *Missing Data*. Newbury Park, California: Sage University Papers Series: Quantitative Applications in the Social Sciences. 2. Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annu. Rev. Psychol. 60*:549-576.

B. Assess data suitability for factor analysis

Factor analysis is based on correlation, and correlation most accurately describes relationships under two conditions. First, when distributions are close to *normal* (bell-curved or Gaussian); if they aren't, one may want to *transform*. Examples are taking the square root or the log of skewed data. Second, when relationships are essentially *linear*; if they follow some other curve, again one may want to transform.

> → *If I transform one item does that mean I have to transform all of them?* Try to steer clear of "have to" thinking in analysis generally, and in EFA in particular. Ask instead what results will likely occur from trying one course or another. In this case the biggest risk is probably that a particular audience for the results will object to treating different variables in different ways. But if pressed, would they be able to explain why this matters? EFA creates factor *loadings* (item-factor correlations) on a 0-to-1 scale, and *factor scores* on roughly a -3 to 3 scale; thus it performs its own transformations regardless of any pre-EFA transformation an analyst might make.

Another common question is

> → *If my data are nowhere near normally distributed, do I have to forego EFA?* Again, "do I have to?" is not as useful as "what is likely to happen?" Here and there you will find a respected source that recommends EFA for non-normal or even for binary data.[3] It is not out of the question. But it will limit the sensitivity of the procedure for identifying patterns in relationships. In writing up EFA results on such data I would be especially careful to avoid undue precision and so would reduce the number of decimal places I employed.

A specialized topic related to data suitability is *sampling adequacy*, as measured by the *Kaiser-Meyer-Olkin* statistic or KMO. For EFA items it's desirable to have a set of high *zero-order* ("vanilla") correlations relative to *partial correlations*. This will yield a solution where each item strongly relates to or loads on one and only one underlying factor, a happy condition known as *simple structure*. In contrast, a complex pattern of high partial correlations involves "muddier" relationships that make it difficult to cleanly associate each item with one factor. The KMO quantifies this ratio. A result above 0.8 shows that the data are very well suited to EFA; a result below about 0.6 would make one question that suitability and look for items to drop from the analysis. (See below. Choices to keep or drop can also be made using a more specialized set of statistics called *anti-image correlations*. The standards there are the same as for KMO.)

---

[3] E.g., Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

C. Choose an extraction method

Again, EFA has the dual functions of identifying latent variables and achieving data reduction. One can achieve these goals using several different methods of *extraction* – of creating from the data new, more abstract variables that could explain the patterns of relationships among the inputs. Different software will call these extraction methods by different names; the most common technique is *principal axis* extraction and the next is *maximum likelihood*.

Principal axis extraction will fit the needs of most factor analyses. Maximum likelihood (ML), on the other hand, has the advantage of more directly addressing the question of how many factors to extract. This is because it will yield a statistical significance test associated with a given solution, answering the question, "would more factors better account for these data?" This test only works well under specific conditions, however. It will be most accurate if item

> One extraction method is different enough that it bears calling out as <u>not</u> applicable to EFA, and that is *principal components*.
>
> Principal components analysis (PCA) is designed for data reduction, pure and simple. It is less effective than EFA for explaining underlying dimensions of subjective data. It is better suited to more objective data such as physical measurements or even economic indices. PCA will include in the derived dimensions (components) some information that is undesirable because it is either a) information unique to each item, or b) measurement error or noise.
>
> PCA, then, is applicable when analyzing objective measurements, when one seeks to use all information present in pure data reduction. It is not appropriate for revealing latent dimensions that meaningfully account for correlations among items. One might compare the PCA-EFA difference to the difference between running a predictive vs. an explanatory analysis. EFA's factors can be seen as *accounting for* the observed data, whereas PCA's components are best viewed as convenient *results* of those data.
>
> For more, see
>
> - Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*, 13-32,
> - Cudeck, R. (2000). Exploratory factor analysis. In Tinsley, H. E. A., and Brown, S. D., *Handbook of applied multivariate statistics and mathematical modeling* (San Diego: Harcourt Brace), 265-297.

distributions are fairly close to normal, i.e., skewness <2 and kurtosis <7, or, again, if one is willing to transform or drop non-normal items. ML extraction is also sensitive both to small and large sample sizes; it works best when N is a few hundred.

D. Pick a rotation method

In EFA, the algorithm will, at your discretion, manipulate extracted factors via *rotation* in order better to align and summarize the raw data. *Orthogonal* rotation (a common method is *varimax*) preserves each factor's independence (keeps them essentially uncorrelated), while *oblique* rotation allows for sizeable correlations among the factors. <u>Use oblique rotation as your default</u>. There is nothing to lose: if resulting correlations among factors are very low, you will have what amounts to a convenient, nearly orthogonal solution. If correlations are high, your solution will acknowledge the relatedness of your dimensions. But to force otherwise correlated would-be factors into an orthogonal relationship is to shoehorn the data into a preconceived arrangement and may undercut the validity of your findings.[4]

E. Decide how many factors to extract

With this decision, as with others, plan on multiple solutions. Rare is the factor analysis that is complete in the first iteration.

Unlike with PCA, with EFA we retain only those factors that are interpretable. This takes precedence over the considerations below.

Extraction involves *eigenvalues*: summaries of information. An eigenvalue of 3.0 contains as much noise-free information as three individual items do. Begin by extracting factors from all initial eigenvalues of at least 0.7 or 0.8. This is advice you won't often see elsewhere. A 50-year-old rule of thumb, the Kaiser-Guttman rule, suggests extracting eigenvalues >=1.0. But much research has shown that mechanically following this rule can under- or over-estimate the best (I will not say "correct") number of factors.

Once you see how solutions turn out, you can specify a criterion based either on a new minimum eigenvalue (probably somewhere between 0.7 and 3.0) or a specific number of factors. You can also take into account a *scree test*, which shows the few factors that tend to explain most of the explainable information while the rest, looking like boulders accumulating at the bottom of a hill, are less valuable.[5]

Factors tend to be most useful (again, I do not say most "correct") if they are reasonably distinct, i.e., if the correlations among them are weaker than, roughly, .6. With an r of .6, r^2 is .36; one factor can explain more than a third of the variance in the other. If correlations are higher, consider extracting a different number of factors if you seek more distinct dimensions. And if the largest initial eigenvalue is more than 8 or 10 times the size of the next, consider a one-factor solution. In such a case, you would treat a single, unidimensional scale as underlying the

[4] Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*, 13-32.

[5] Performed visually or numerically, the scree test is ideally conducted on extracted, not initial eigenvalues, although for some reason the initial are what you'll see displayed in the scree plots of most software.

responses to all the items in your analysis, and you'd conclude that it would be artificial to try to force your variables to fill out multiple dimensions.

Assuming between-factor correlations are not very high, a factor will be useful if the variance it can explain after rotation, as indicated by its eigenvalue, is at least 5 or 10% (>= 5% if there are about 20 or more inputs; >= 10% if about 10). If a factor explains, only, say, 3% of the variance in the whole set of items, its scores won't be very informative, and it probably won't replicate well.

Also, each factor will serve you best if it has at least 2 and probably 3 items with high loadings. The loading tells to what degree the item connects with, or can be explained by, the latent dimension represented by the factor. A "high" loading might be 0.5 in a data set where N~3,000, or 0.7 where N~300.[6] A factor with only one high loading might as well be represented instead by the original item in question; tacking on additional ones with small loadings will probably add more noise than signal.

For the advanced practitioner, there are two specialized methods worth investigating that can help in choosing the best number of factors. These are the *parallel analysis* method and the *minimum average partial correlation test*.

F. Decide which items to keep in the analysis

Items (inputs, or original variables) with weak loadings in one solution may behave quite differently in another. If after multiple attempts at a solution a given item still seems unrelated to the "story" told by the derived factors, it is perfectly fine to leave it out of the procedure. Leaving it in, with loadings of, say, < .3 will not dramatically undermine results -- after all, the *factor score* is weighted according to loading strength –, but will probably reduce reliability and validity.

As noted above, specialized analysis via anti-image correlations can also indicate items that will not fit in well with an EFA. Again, some commentators will object to treating different variables in different ways –"we must include the whole battery of questions or none" – and to that one might answer that in analysis it's advantageous, and customary, to try to maximize signal and minimize noise.

[6] This raises the question of how many cases are necessary to effectively run an EFA. Rules of thumb vary from the stringent (e.g., 20 cases per variable) to the liberal (20 cases per eventual factor.) Remember that, unlike with procedures such as regression, it's not as if the precision of *p*-values or standard errors will depend on having a sufficient sample size, since EFA usually produces no such statistics.

G. Choose a way of displaying results

Factor analysis yields many types of results, any of which might be of primary interest in a given situation. For example:

a. *How suitable the data are for such analysis.* Display KMO and/or anti-image results.

b. *To what extent a given number of factors can explain the information in the whole set of items.* Display a table of eigenvalues and explained variance.

c. *To what degree each item correlates with each factor.* Display loadings, as in a pattern matrix (see below).

d. *How the different factors relate to one another.* Display a factor correlation matrix and/or scatterplots detailing the position of each item in 2- or 3-dimensional space.

    i. To display two factors' relationship, one can actually use varimax rotation to plot the two clusters of items. This will show the factors' correlation, which will be indicated by the angle between them; specifically, *r* will equal the cosine of this angle. (See example below.)

e. *What factor scores characterize each person.* Either manually or automatically, one can compute for each person a number that indicates how high or low s/he scores on a given factor. These scores can be displayed via histograms, scatterplots, etc., and can be built into further analyses.

    i. To compute automatically: once you are satisfied with your FA solution, rerun EFA using the appropriate subcommand for saving scores. These will fill into new dataset columns and will usually range from -3 to 3.

    ii. To compute manually: one can devise an equation that weights each item according to its loading. However, it's sometimes sufficient to forego these weights and compute each factor score as the simple average of all variables with high loadings on that factor. The latter strategy has the advantage that such scores will match those obtained via a typical *Cronbach's alpha* test of internal consistency.

Below is a sample pattern matrix illustrating four dimensions that characterize clinical depression and showing the loadings of 12 items on those four dimensions. Here we see a relatively simple structure: all but two items have substantial loadings on only one factor. As is commonly done, loadings below a certain strength (in this case .3) have been left blank for easier interpretation of the table.

Table 1. Sample item loadings on factors of clinical depression

ITEM	MOOD	SELF-CONCEPT	COGNITIVE IMPAIRMENT	PHYSICAL SYMPTOMS
I feel sad	.9			
I feel hopeless	.8			
I can't experience joy	.8			
I feel like a failure		.8		
I don't like myself		.7		
I have trouble concentrating			.7	
I can't seem to make decisions			.7	
I feel as if I'm in a fog			.6	.4
I suffer from fatigue				.7
I am plagued by aches and pains				.7
My sleep pattern is disturbed				.6
I am preoccupied with my health			.4	.5

A cautionary tale

Once I analyzed data from a survey of parents' opinions about a certain group of schools. Are these schools academically strong? Are they too elitist? In the figure below,

▲ represents **Academic** items, e.g.,

- *These schools challenge students.*
- *These schools have excellent academic facilities.*

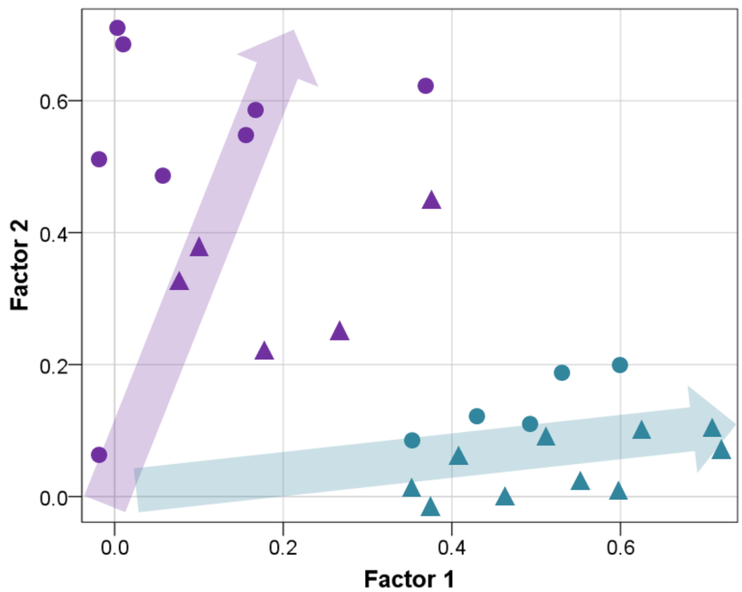
● represents **Socioeconomic** items, e.g.,

- *These schools are for rich people.*
- *These schools welcome students of diverse backgrounds.*

One would hope that Academic items would load mainly on one factor, while Socioeconomic, on another. The wrench in the works is that 28 such items were presented interspersed in two long grids in an online survey form. Long grids can exert a gravity of their own, blurring the connections a researcher expects to see by creating high, spurious correlations *within* each grid.

When the data were factor-analyzed, did all the triangles cluster together, and did all the circles cluster together? No! The factors derived were not based on content; they were mere artifacts of survey grid (represented by color). As such they said nothing pertinent about respondents' underlying opinions.

Figure 1. Factor plot in rotated factor space



All the items in **green** were from Grid 1; they all loaded most highly on derived Factor 1, the "Grid 1" factor. All those in **purple** were from Grid 2; they all loaded most highly on Factor 2, the "Grid 2" factor. We can see what harm this absurd result did to validity.

The lesson here is that, as with so many statistical methods, factor analysis will only be useful if the data on which it relies are obtained with careful attention paid to validity and reliability of measures.

Incidentally, as mentioned above, plots like these can show to what degree factors are correlated. Here, the angle between the purple and green arrows, about 65 degrees, translates via the cosine of 65 to an r of .42 and an R^2 of .18.

Additional Recommended Sources

Comrey, A. L., and Lee, H. B. (1992). *A first course on factor analysis* (2nd edition). Hillsdale, New Jersey: Lawrence Erlbaum.

Cudeck, R. (2000). Exploratory factor analysis. In Tinsley, H. E. A., and Brown, S. D., *Handbook of applied multivariate statistics and mathematical modeling* (San Diego: Harcourt Brace), 265-297.

Davis, J. D. (1985). *The logic of causal order*. Newbury Park, California: Sage University Papers Series: Quantitative Applications in the Social Sciences.

Dawis, R. V. (2000). Scale construction and psychometric considerations. In Tinsley, H. E. A., and Brown, S. D., *Handbook of applied multivariate statistics and mathematical modeling* (San Diego: Harcourt Brace), 65-94.

Doty, D. H., and Glick, W. (1998). Common methods bias: does common methods variance really bias results? *Organizational Research Methods*, *1*:4, 374-406.

Fabrigar, L. R., McCallum, R. C., Wegener, D. T. and Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*:3, 272-299.

Floyd, F. J. and Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*:3, 286-299.

Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

Spector, P. E. (1992). *Summated rating scale construction.* Newbury Park, California: State University Papers Series: Quantitative Applications in the Social Sciences.