Factor Analysis: Exploratory

Robert Pruzek

Volume 2, pp. 606–617

in

Encyclopedia of Statistics in Behavioral Science

ISBN-13: 978-0-470-86080-9
ISBN-10: 0-470-86080-4

Editors

Brian S. Everitt & David C. Howell

Factor Analysis: Exploratory

Introduction

This year marks the one hundredth anniversary for exploratory factor analysis (EFA), a method introduced by **Charles Spearman** in 1904 [21]. It is testimony to the deep insights of Spearman as well as many who followed that EFA continues to be central to **multivariate analysis** so many years after its introduction. In a recent search of electronic sources, where I restricted attention to the psychological and social sciences (using PsychINFO), more than 20 000 articles and books were identified in which the term 'factor analysis' had been used in the summary, well over a thousand citations from the last decade alone.

EFA, as it is known today, was for many years called common factor analysis. The method is in some respects similar to another well-known method called **principal component analysis** (PCA) and because of various similarities, these methods are frequently confused. One of the purposes of this article will be to try to dispel at least some of the confusion.

The general methodology currently seen as an umbrella for both exploratory factor analysis and confirmatory factor analysis (*see* **Factor Analysis: Confirmatory**) is called **structural equation modeling** (**SEM**) Although EFA can be described as an exploratory or unrestricted structural equation model, it would be a shame to categorize EFA as nothing more than a SEM, as doing so does an injustice to its long history as the most used and most studied latent variable method in the social and behavioral sciences. This is somewhat like saying that **analysis of variance (ANOVA)** which has been on the scene for more than seventy-five years and which is prominently related to experimental design, is just a **multiple linear regression** model. There is some truth to each statement, but it is unfair to the rich histories of EFA and ANOVA to portray their boundaries so narrowly.

A deeper point about the relationships between EFA and SEM is that these methods appeal to very different operational philosophies of science. While SEMs are standardly seen as founded on rather strict hypothetico-deductive logic, EFAs are not. Rather, EFA generally invokes an exploratory search for structure that is open to new structures not imagined prior to analysis. Rozeboom [20] has carefully examined the logic of EFA, using the label *explanatory induction* to describe it; this term neatly summarizes EFA's reliance on data to induce hypotheses about structure, and its general concern for explanation.

Several recent books, excellent reviews, and constructive critiques of EFA have become available to help understand its long history and its potential for effective use in modern times [6, 8, 15, 16, 23, 25]. A key aim of this article is to provide guidance with respect to literature about factor analysis, as well as to software to aid applications.

Basic Ideas of EFA Illustrated

Given a matrix of correlations or covariances (*see* **Correlation and Covariance Matrices**) among a set of manifest or observed variables, EFA entails a model whose aim is to explain or account for correlations using a smaller number of 'underlying variables' called common factors. EFA postulates common factors as **latent variables** so they are unobservable in principle. Spearman's initial model, developed in the context of studying relations among psychological measurements, used a single common factor to account for all correlations among a battery of tests of intellectual ability. Starting in the 1930s, **Thurstone** generalized the 'two-factor' method of Spearman so that EFA became a multiple (common) factor method [22]. In so doing, Thurstone effectively broadened the range of prospective applications in science. The basic model for EFA today remains largely that of Thurstone. EFA entails an assumption that there exist uniqueness factors as well as common factors, and that these two kinds of factors complement one another in mutually orthogonal spaces. An example will help clarify the central ideas.

Table 1 below contains a correlation matrix for all pairs of five variables, the first four of which correspond to ratings by the seventeenth century art critic de Piles (using a 20 point scale) of 54 painters for whom data were complete [7]. Works of these painters were rated on four characteristics: composition, drawing, color, and expression. Moreover, each painter was associated with a particular 'School.' For current purposes, all information about Schools is

Table 1 Correlations among pairs of variables, painter data of [8]

	Composition	Drawing	Color	Expression	School D
Composition	1.00				
Drawing	0.42	1.00			
Color	−0.10	−0.52	1.00		
Expression	0.66	0.57	−0.20	1.00	
School D	−0.29	−0.36	0.53	−0.45	1.00

ignored except for distinguishing the most distinctive School D (Venetian) from the rest using a binary variable. For more details, see the file 'painters' in the Modern Applied Statistics with S (MASS) library in R or Splus software (*see* **Software for Statistical Analyses**), and note that the original data and several further analyses can be found in the MASS library [24].

Table 1 exhibits correlations among the painter variables, where upper triangle entries are ignored since the matrix is symmetric. Table 2 exhibits a common factor coefficients matrix (of order 5×2) that corresponds to the initial correlations, where entries of highest magnitude are in bold print. The final column of Table 2 is labeled h^2, the standard notation for variable communalities. Because these factor coefficients correspond to an orthogonal factor solution, that is, uncorrelated common factors, each communality can be reproduced as a (row) sum of squares of the two factor coefficients to its left; for example $(0.76)^2 + (-0.09)^2 = 0.59$. The columns labeled 1 and 2 are factor loadings, each of which is properly interpreted as a (product-moment) correlation between one of the original manifest variables (rows) and a derived common factor (columns). Post-multiplying the factor coefficient matrix by its transpose yields numbers that

Table 2 Factor loadings for 2-factor EFA solution, painter data

	Factor		
Variable name	1	2	h^2
Composition	**0.76**	−0.09	0.59
Drawing	**0.50**	**−0.56**	0.56
Color	−0.03	**0.80**	0.64
Expression	**0.81**	−0.26	0.72
School D	−0.30	**0.62**	0.47
Avg. Col. SS	0.31	0.28	0.60

approximate the corresponding entries in the correlation matrix. For example, the *inner product* of the rows for Composition and Drawing is $0.76 \times 0.50 + (-0.09) \times (-0.56) = 0.43$, which is close to 0.42, the observed correlation; so the corresponding residual equals -0.01. Pairwise products for all rows reproduce the observed correlations in Table 1 quite well as only one residual fit exceeds 0.05 in magnitude, and the mean residual is 0.01.

The final row of Table 2 contains the average sum of squares for the first two columns; the third entry is the average of the communalities in the final column, as well as the sum of the two average sums of squares to its left: $0.31 + 0.28 \approx 0.60$. These results demonstrate an additive decomposition of common variance in the solution matrix where 60 percent of the total variance is common among these five variables, and 40 percent is uniqueness variance.

Users of EFA have often confused communality with reliability, but these two concepts are quite distinct. Classical common factor and psychometric test theory entail the notion that the uniqueness is the sum of two (orthogonal) parts, specificity and error. Consequently, uniqueness variance is properly seen as an upper bound for error variance; alternatively, communality is in principle a lower bound for reliability. It might help to understand this by noting that each EFA entails analysis of just a sample of observed variables or measurements in some domain, and that the addition of more variables within the general domain will generally increase shared variance as well as individual communalities. As battery size is increased, individual communalities increase toward upper limits that are in principle close to variable reliabilities. See [15] for a more elaborate discussion.

To visualize results for my example, I plot the common factor coefficients in a plane, after making some modifications in signs for selected rows and the second column. Specifically, I reverse the signs of the 3rd and 5th rows, as well as in the second column, so that all values in the factor coefficients matrix become positive. Changes of this sort are always permissible, but we need to keep track of the changes, in this case by renaming the third variable to 'Color[−1]' and the final binary variable to 'School.D[−1]'. Plotting the revised coefficients by rows yields the five labeled points of Figure 1.

In addition to plotting points, I have inserted vectors to correspond to 'transformed' factors; the arrows show an 'Expression–Composition' factor and a second, correlated, 'Drawing–Color[−1]' factor. That the School.D variable also loads highly on this second factor, and is also related to, that is, not orthogonal to, the point for Expression, shows that mean ratings, especially for the Drawing, Expression, and Color variates (the latter in an opposite direction), are notably different between Venetian School artists and painters from the collection of other schools. This can be verified by examination of the correlations (sometimes called point biserials) between the School.D variable and all the ratings variables in Table 1; the skeptical reader can easily acquire these data and study details. In fact, one of the reasons for choosing this example was to show that EFA as an exploratory data analytic method can help in studies of relations among quantitative and categorical variables. Some connections of EFA with other methods will be discussed briefly in the final section.

In modern applications of factor analysis, investigators ordinarily try to name factors in terms of dimensions of individual difference variation, to

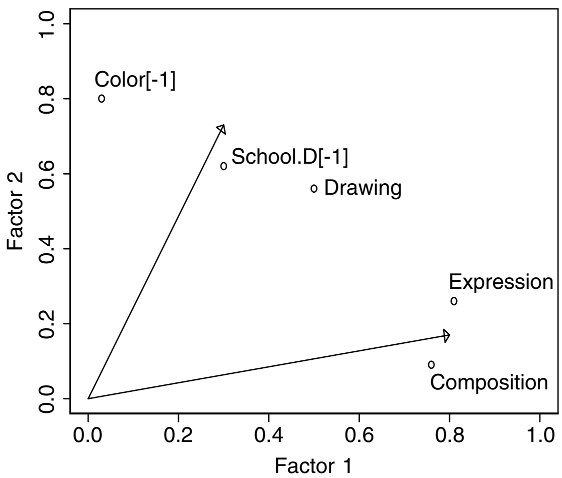

Figure 1 Plot of variables-as-points in 2-factor space, painter data

identify latent variables that in some sense appear to 'underlie' observed variables. In this case, my ignorance of the works of these classical painters, not to mention of the thinking of de Piles as related to his ratings, led to my literal, noninventive factor names.

Before going on, it should be made explicit that insertion of the factor-vectors into this plot, and the attempt to name factors, are best regarded as discretionary parts of the EFA enterprise. The key output of such an analysis is the identification of the subspace defined by the common factors, within which variables can be seen to have certain distinctive structural relationships with one another. In other words, it is the configuration of points in the derived space that provides the key information for interpreting factor results; a relatively low-dimensional subspace provides insights into structure, as well as quantification of how much variance variables have in common. Positioning or naming of factors is generally optional, however common. When the common number of derived factors exceeds two or three, factor transformation is an almost indispensable part of an EFA, regardless of whether attempts are made to name factors.

Communalities generally provide information as to how much variance variables have in common or share, and can sometimes be indicative of how highly predictable variables are from one another. In fact, the squared multiple correlation of each variable with all others in the battery is often recommended as an initial estimate of communality for each variable. Communalities can also signal (un)reliability, depending on the composition of the battery of variables, and the number of factors; recall the foregoing discussion on this matter.

Note that there are no assumptions that point configurations for variables must have any particular form. In this sense, EFA is more general than many of its counterparts. Its exploratory nature also means that prior structural information is usually not part of an EFA, although this idea will eventually be qualified in the context of reviewing factor transformations. Even so, clusters or hierarchies of either variables or entities may sometimes be identified in EFA solutions. In our example, application of the common factor method yields a relatively parsimonious model in the sense that two common factors account for all relationships among variables. However, EFA was, and is usually, antiparsimonious in another sense as there is one uniqueness factor for each variable as well as common factors to account for all entries in the correlation table.

Some Relationships Between EFA and PCA

As noted earlier, EFA is often confused with PCA. In fact, misunderstanding occurs so often in reports, published articles, and textbooks that it will be useful to describe how these methods compare, at least in a general way. More detailed or more technical discussions concerning such differences is available in [15].

As noted, the key aim of EFA is usually to derive a relatively small number of common factors to explain or account for (off-diagonal) covariances or correlations among a set of observed variables. However, despite being an exploratory method, EFA entails use of a falsifiable *model* at the level of manifest observations or correlations (covariances). For such a model to make sense, relationships among manifest variables should be approximately linear. When approximate linearity does not characterize relationships among variables, attempts can be made to transform (at least some of) the initial variables to 'remove bends' in their relationships with other variables, or perhaps to remove outliers. Use of square root, logarithmic, reciprocal, and other nonlinear transformations are often effective for such purposes. Some investigators question such steps, but rather than asking why nonlinear transformations should be considered, a better question usually is, 'Why should the analyst believe the metric used at the outset for particular variables should be expected to render relationships linear, without reexpressions or transformations?' Given at least approximate linearity among all pairs of variables – the inquiry about which is greatly facilitated by examining pairwise scatterplots among all pairs of variables – common factor analysis can often facilitate explorations of relationships among variables. The prospects for effective or productive applications of EFA are also dependent on thoughtful efforts at the stage of study design, a matter to be briefly examined below. With reference to our example, the pairwise relationships between the various pairs of de Pile's ratings of painters were found to be approximately linear.

In contrast to EFA, principal components analysis does *not* engage a model. PCA generally entails an algebraic decomposition of an initial data matrix into mutually orthogonal derived variables called *components*. Alternatively, PCA can be viewed as a linear transformation of the initial data vectors into uncorrelated variates with certain optimality properties. Data are usually centered at the outset by subtracting means for each variable and then scaled so that all variances are equal, after which the (rectangular) data matrix is resolved using a method called *singular value decomposition* (SVD). Components from a SVD are usually ordered so that the first component accounts for the largest amount of variance, the second the next largest amount, subject to the constraint that it be uncorrelated with the first, and so forth. The first few components will often summarize the majority of variation in the data, as these are principal components. When used in this way, PCA is justifiably called a *data reduction* method and it has often been successful in showing that a rather large number of variables can be summarized quite well using a relatively small number of derived components.

Conventional PCA can be completed by simply computing a table of correlations of each of the original variables with the chosen principal components; indeed doing so yields a PCA counterpart of the EFA coefficients matrix in Table 2 if two components are selected. Furthermore, sums of squares of correlations in this table, across variables, show the total variance each component explains. These component-level variances are the eigenvalues produced when the correlation matrix associated with the data matrix is resolved into eigenvalues and eigenvectors. Alternatively, given the original (centered and scaled) data matrix, and the eigenvalues and vectors of the associated correlation matrix, it is straightforward to compute principal components. As in EFA, derived PCA coefficient matrices can be rotated or transformed, and for purposes of interpretation this has become routine.

Given its algebraic nature, there is no particular reason for transforming variables at the outset so that their pairwise relationships are even approximately linear. This can be done, of course, but absent a model, or any particular justification for concentrating on pairwise *linear* relationships among variables, principal components analysis of correlation matrices is somewhat arbitrary. Because PCA is just an algebraic decomposition of data, it can be used for any kind of data; no constraints are made about the dimensionality of the data matrix, no constraints on data values, and no constraints on how many components to use in analyses. These points imply that for PCA, assumptions are also optional regarding statistical distributions, either individually or collectively. Accordingly, PCA is a highly general method, with potential for use for a wide range of data types or forms. Given their basic form, PCA methods provide little guidance for answering model-based questions, such as those central to EFA. For example, PCA generally offers little support for assessing how many components ('factors') to generate, or try to interpret; nor is there assistance for choosing samples or extrapolating beyond extant data for purposes of statistical or psychometric generalization. The latter concerns are generally better dealt with using models, and EFA provides what in certain respects is one of the most general classes of models available.

To make certain other central points about PCA more concrete, I return to the correlation matrix for the painter data. I also conducted a PCA with two components (but to save space I do not present the table of 'loadings').

That is, I constructed the first two principal component variables, and found their correlations with the initial variables. A plot (not shown) of the principal component loadings analogous to that of Figure 1 shows the variables to be configured similarly, but all points are further from the origin. The row sums of squares of the component loadings matrix were 0.81, 0.64, 0.86, 0.83, and 0.63, values that correspond to communality estimates in the third column of the common factor matrix in Table 2. Across all five variables, PCA row sums of squares (which should not be called communalities) range from 14 to 37 percent larger than the h^2 entries in Table 2, an average of 27 percent; this means that component loadings are substantially larger in magnitude than their EFA counterparts, as will be true quite generally. For any data system, given the same number of components as common factors, component solutions yield row sums of squares that tend to be at least somewhat, and often markedly, larger than corresponding communalities.

In fact, these differences between characteristics of the PCA loadings and common factor loadings signify a broad point worthy of discussion. Given that principal components are themselves linear combinations of the original data vectors, each of the data variables tends to be part of the linear combination with which it is correlated. The largest weights for each linear combination correspond to variables that most strongly define the corresponding linear combination, and so the corresponding correlations in the Principal Component (PC) loading matrix tend to be highest, and indeed to have spuriously high magnitudes. In other words, each PC coefficient in the matrix that constitutes the focal point for interpretation of results, tends to have a magnitude that is 'too large' because the corresponding variable is correlated partly with itself, the more so for variables that are largest parts of corresponding components. Also, this effect tends to be exacerbated when principal components are rotated. Contrastingly, common factors are latent variables, outside of the space of the data vectors, and common factor loadings are not similarly spurious. For example, EFA loadings in Table 2, being correlations of observed variables with latent variables, do not reflect self-correlations, as do their PCA counterparts.

The Central EFA Questions: How Many Factors? What Communalities?

Each application of EFA requires a decision about how many common factors to select. Since the common factor model is at best an approximation to the real situation, questions such as how many factors, or what communalities, are inevitably answered with some degree of uncertainty. Furthermore, particular features of given data can make formal fitting of an EFA model tenuous. My purpose here is to present EFA as a true exploratory method based on common factor principles with the understanding that formal 'fitting' of the EFA model is secondary to 'useful' results in applications; moreover, I accept that certain decisions made in contexts of real data analysis are inevitably somewhat arbitrary and that any given analysis will be incomplete. A wider perspective on relevant literature will be provided in the final section.

The history of EFA is replete with studies of how to select the number of factors; hundreds of both theoretical and empirical approaches have been suggested for the number of factors question, as this issue has been seen as basic for much of the past century. I shall summarize some of what I regard as the most enduring principles or methods, while trying to shed light on when particular methods are likely to work effectively, and how the better methods can be attuned to reveal relevant features of extant data.

Suppose scores have been obtained on some number of correlated variables, say p, for n entities, perhaps persons. To entertain a factor analysis (EFA) for these variables generally means to undertake an exploratory structural analysis of linear relations among the p variables by analyzing a p $\times$ p covariance or correlation matrix. Standard outputs of such an analysis are a factor loading matrix for orthogonal or correlated common factors as well as communality estimates, and perhaps factor score estimates. All such results are conditioned on the number, m, of common factors selected for analysis. I shall assume that in deciding to use EFA, there is at least some doubt, *a priori*, as to how many factors to retain, so extant data will be the key basis for deciding on the number of factors. (I shall also presume that the data have been properly prepared for analysis, appropriate nonlinear transformations made, and so on, with the understanding that even outwardly small changes in the data can affect criteria bearing on the number of factors, and more.)

The reader who is even casually familiar with EFA is likely to have learned that one way to select the number of factors is to see how many **eigenvalues** (of the correlation matrix; recall PCA) exceed a certain criterion. Indeed, the 'roots-greater-than-one' rule has become a default in many programs. Alas, rules of this sort are generally too rigid to serve reliably for their intended purpose; they can lead either to overestimates or underestimates of the number of common factors. Far better than using any fixed cutoff is to understand certain key principles and then learn some elementary methods and strategies for choosing m. In some cases, however, two or more values of m may be warranted, in different solutions, to serve distinctive purposes for different EFAs of the same data.

A second thing even a nonspecialist may have learned is to employ a 'scree' plot (SP) to choose the number of factors in EFA. An SP entails plotting eigenvalues, ordered from largest to smallest, against their ordinal position, 1, 2, ..., and so on. Ordinarily, the SP is based on eigenvalues of a correlation matrix [5]. While the usual SP sometimes works reasonably well for choosing m, there is a mismatch between such a standard SP, and another relevant fact: a tacit assumption of this method is that all p communalities are the same. But to assume equal communalities is usually to make a rather strong assumption, one quite possibly not supported by data in hand.

A better idea for SP entails computing the original correlation matrix, R, as well as its inverse R^{-1}. Then, denoting the *diagonal* of the inverse as D^2 (entries of which exceed unity), rescale the initial correlation matrix to DRD, and then compute eigenvalues of this rescaled correlation matrix. Since the largest entries in D^2 correspond to variables that are most predictable from all others, and vice versa, the effect is to weigh variables more if they are more predictable, less if they are less predictable from other variables in the battery. (The complement of the reciprocal of any D^2 entry is in fact the squared multiple correlation (SMC) of that variable with all others in the set.) An SP based on eigenvalues of DRD allows for variability of communalities, and is usually realistic in assuming that communalities are at least roughly proportional to SMC values.

Figure 2 provides illustrations of two scree plots based on DRD, as applied to *two simulated* random samples. Although real data were used as the starting point for each simulation, both samples are just simulation sets of (the same size as) the original data set, where four factors had consistently been identified as the 'best' number to interpret.

Each of these two samples yields a scree plot, and both are given in Figure 2 to provide some sense of sampling variation inherent in such data; in this case, each plot leads to breaks after four common factors – where the break is found by reading the plot from right to left. But the slope between four and five

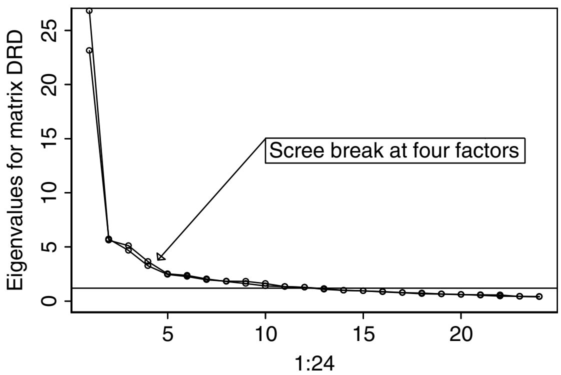

Figure 2 Two scree plots, for two simulated data sets, each n = 145

factors is somewhat greater for one sample than the other, so one sample identifies m as four with slightly more clarity than the other. In fact, for some other samples examined in preparing these scree plots, breaks came after three or five factors, not just four. Note that for smaller samples greater variation can be expected in the eigenvalues, and hence the scree breaks will generally be less reliable indicators of the number of common factors for smaller samples.

So what is the principle behind the scree method? The answer is that the *variance* of the $p - m$ smallest eigenvalues is closely related to the *variance* of residual correlations associated with fitting off-diagonals of the observed correlation matrix in successive choices for m, the number of common factors. When a break occurs in the eigenvalue plot, it signifies a notable drop in the sum of squares of residual correlations after fitting the common factor model to the observed correlation matrix for a particular value of m. I have constructed a horizontal line in Figure 2 to correspond to the *mean* of the 20 smallest eigenvalues (24–4) of DRD, to help see the variation these so-called '*rejected*' eigenvalues have around their mean. In general, it is the variation around such a mean of rejected eigenvalues that one seeks to reduce to a 'reasonable' level when choosing m in the EFA solution, since a 'good' EFA solution accounts well for the off-diagonals of the correlation matrix. Methods such as bootstrapping – wherein multiple versions of DRD are generated over a series of bootstrap samples of the original data matrix – can be used to get a clearer sense of sampling variation, and probably should become part of standard practice in EFA both at the level of selecting the number of factors, and assessing variation in various derived EFA results.

When covariances or correlations are well fit by some relatively small number of common factors, then scree plots often provide flexible, informative, and quite possibly persuasive evidence about the number of common factors. However, SPs alone can be misleading, and further examination of data may be helpful. The issue in selecting m vis-à-vis the SP concerns the nature or reliability of the information in **eigenvectors** associated with corresponding eigenvalues. Suppose some number m^* is seen as a possible underestimate for m; then deciding to add one more factor to have $m^* + 1$ factors, is to decide that the additional eigenvector adds useful or meaningful structural information to the EFA solution. It is possible that m^* is an 'underestimate' solely because a single correlation coefficient is poorly fit, and that adding a common factor merely reduces a single 'large' residual correlation. But especially if the use of $m^* + 1$ factors yields a factor loading matrix that upon rotation (see below) improves interpretability in general, there may be ex post facto evidence that m^* was indeed an underestimate. Similar reasoning may be applied when moving to $m^* + 2$ factors, etc. Note that sampling variation can also result in sample reordering of so-called *population eigenvectors* too.

An adjunct to an SP that is too rarely used is simply to plot the distribution of the residual correlations, either as a histogram, or in relation to the original correlations, for, say, $m, m + 1$, and $m + 2$ factors in the vicinity of the scree break; outliers or other anomalies in such plots can provide evidence that goes usefully beyond the SP when selecting m. Factor transformation(s) (see below) may be essential to one's final decision. Recall that it may be a folly even to think there is a single 'correct' value for m for some data sets.

Were one to use a different selection of variables to compose the data matrix for analysis, or perhaps make changes in the sample (deleting or adding cases), or try various different factoring algorithms, further modifications may be expected about the number of common factors. Finally, there is always the possibility that there are simply too many distinctive dimensions of individual difference variation, that is, common factors, for the EFA method to work effectively in some situations. It is not unusual that more variables, larger samples, or generally more investigative effort, are required to resolve some basic questions such as how many factors to use in analysis.

Given some choice for m, the next decision is usually that of deciding what factoring method to use. The foregoing idea of computing DRD, finding its eigenvalues, and producing an SP based on those, can be linked directly to an EFA method called *image factor analysis* (IFA) [13], which has probably been underused, in that several studies have found it to be a generally sound and effective method. IFA is a noniterative method that produces common factor coefficients and communalities directly. IFA is based on the m largest eigenvalues, say, the diagonal entries of Γ_m, and corresponding eigenvectors, say Q_m, of the matrix denoted DRD, above. Given a particular factor method, communality estimates follow directly from selection of the number of common factors.

The analysis usually commences from a correlation matrix, so communality estimates are simply row sums of squares of the (orthogonal) factor coefficients matrix that for m common factors is computed as $\Lambda_m = \mathrm{D}^{-1}\mathrm{Q}_m(\Gamma_m - \phi\ \mathrm{I})^{1/2}$, where ϕ is the average of the p – m smallest eigenvalues. IFA may be especially defensible for EFA when sample size is limited; more details are provided in [17], including a sensible way to modify the diagonal D^2 when the number of variables is a 'substantial fraction' of sample size.

A more commonly used EFA method is called *maximum likelihood factor analysis* (MLFA) for which algorithms and software are readily available, and generally well understood. The theory for this method has been studied perhaps more than any other and it tends to work effectively when the EFA problem has been well-defined and the data are 'well-behaved.' Specialists regularly advocate use of the MLFA method [1, 2, 16, 23], and it is often seen as the common factor method of choice when the sample is relatively large. Still, MLFA is an iterative method that can lead to poor solutions, so one must be alert in case it fails in some way. Maximum likelihood EFA methods generally call for large n's, using an assumption that the sample has been drawn randomly from a parent population for which multivariate normality (*see* **Catalogue of Probability Density Functions**) holds, at least approximately; when this assumption is violated seriously, or when sample size is not 'large,' MLFA may not serve its exploratory purpose well. Statistical tests may sometimes be helpful, but the sample size issue is vital if EFA is used for testing statistical hypotheses. There can be a mismatch between exploratory use of EFA and statistical testing because small samples may not be sufficiently informative to reject any factor model, while large samples may lead to rejection of every model in some domains of application. Generally scree methods for choosing the number of factors are superior to statistical testing procedures.

Given a choice of factoring methods – and of course there are many algorithms in addition to IFA and MLFA – the generation of communality estimates follows directly from the choice of m, the number of common factors. However, some EFA methods or algorithms can yield numerically unstable results, particularly if m is a substantial fraction of p, the number of variables, or when n is not large in relation to p. Choice of factor methods, like many other methodological decisions, is often best made in consultation with an expert.

Factor Transformations to Support EFA Interpretation

Given at least a tentative choice for m, EFA methods such as IFA or MLFA can be used straightforwardly to produce matrices of factor coefficients to account for structural relations among variables. However, attempts to interpret factor coefficient matrices without further efforts to transform factors usually fall short unless $m = 1$ or 2, as in our illustration. For larger values of m, factor transformation can bring order out of apparent chaos, with the understanding that order can take many forms. Factor transformation algorithms normally take one of three forms: Procrustes fitting to a prespecified target (*see* **Procrustes Analysis**), orthogonal simple structure, or oblique simple structure. All modern methods entail use of specialized algorithms. I shall begin with Procrustean methods and review each class of methods briefly.

Procrustean methods owe their name to a figure of ancient Greek mythology, Procrustes, who made a practice of robbing highway travelers, tying them up, and stretching them, or cutting off their feet to make them fit a rigid iron bed. In the context of EFA, Procrustes methods are more benign; they merely invite the investigator to prespecify his or her beliefs about structural relations among variables in the form of a target matrix, and then transform an initial factor coefficients matrix to put it in relatively close conformance with the target. Prespecification of configurations of points in m-space, preferably on the basis of hypothesized structural relations that are meaningful to the investigator, is a wise step for most EFAs even if Procrustes methods are not to be used explicitly for transformations. This is because explication of beliefs about structures can afford (one or more) reference system(s) for interpretation of empirical data structures however they were initially derived. It is a long-respected principle that prior information, specified independently of extant empirical data, generally helps to support scientific interpretations of many kinds, and EFA should be no exception. In recent times, however, methods such as confirmatory factor analysis (CFA), are usually seen as making Procrustean EFA methods obsolete because CFA methods offer generally more sophisticated numerical and statistical machinery to aid analyses. Still, as a matter of principle, it is useful to recognize that general methodology of EFA has for over sixty years permitted, and in some respects encouraged, incorporation of sharp prior questions in structural analyses.

Orthogonal rotation algorithms provide relatively simple ways for transforming factors and these have been available for nearly forty years. Most commonly, an 'orthomax' criterion is optimized, using methods that have been dubbed 'quartimax', 'varimax', or 'equamax.' Dispensing with quotations, we merely note that in general, equamax solutions tend to produce simple structure solutions for which different factors account for nearly equal amounts of common variance; quartimax, contrastingly, typically generates one broad or general factor followed by $m - 1$ 'smaller' ones; varimax produces results intermediate between these extremes. The last, varimax, is the most used of the orthogonal simple structure rotations, but choice of a solution should not be based too strongly on generic popularity, as particular features of a data set can make other methods more effective. Orthogonal solutions offer the appealing feature that squared common factor coefficients show directly how much of each variable's common variance is associated with each factor. This property is lost when factors are transformed obliquely. Also, the factor coefficients matrix alone is sufficient to interpret orthogonal factors; not so when derived factors are mutually correlated. Still, forcing factors to be uncorrelated can be a weakness when the constraint of orthogonality limits factor coefficient configurations unrealistically, and this is a common occurrence when several factors are under study.

Oblique transformation methods allow factors to be mutually correlated. For this reason, they are more complex and have a more complicated history. A problem for many years was that by allowing factors to be correlated, oblique transformation methods often allowed the m-factor space to collapse; successful methods avoided this unsatisfactory situation while tending to work well for wide varieties of data. While no methods are entirely acceptable by these standards, several, notably those of Jennrich and Sampson (direct quartimin) [12], Harris and Kaiser (obliquimax), Rozeboom (Hyball) [18], Yates (geomin) [25], and Hendrikson and White (promax) [9] are especially worthy of consideration for applications. Browne [2], in a recent overview of analytic rotation methods for EFA, stated that Jennrich and Sampson [12] 'solved' the problems of oblique rotation; however, he went on to note that '... we are not at a point where we can rely on mechanical exploratory rotation by a computer program if the complexity of most variables is not close to one [2, p. 145].' Methods such as Hyball [19] facilitate random starting positions in m-space of transformation algorithms to produce multiple solutions that can then be compared for interpretability. The promax method is notable not only because it often works well, but also because it combines elements of Procrustean logic with analytical orthogonal transformations. Yates' geomin [25] is also a particularly attractive method in that the author went back to Thurstone's basic ideas for achieving simple structure and developed ways for them to be played out in modern EFA applications. A special reason to favor simple structure transformations is provided in [10, 11] where the author noted that standard errors of factor loadings will often be substantially smaller when population structures are simple than when they are not; of course this calls attention to the design of the battery of variables.

Estimation of Factor Scores

It was noted earlier that latent variables, that is, common factors, are basic to any EFA model. A strong distinction is made between observable variates and the underlying latent variables seen in EFA as accounting for manifest correlations or covariances between all pairs of manifest variables. The latent variables are by definition never observed or observable in a real data analysis, and this is not related to the fact that we ordinarily see our data as a sample (of cases, or rows); latent variables are in principle not observable, either for statistically defined samples, or for their population counterparts. Nevertheless, it is not difficult to *estimate* the postulated latent variables, using linear combinations of the observed data. Indeed, many different kinds of factor score estimates have been devised over the years (*see* **Factor Score Estimation**).

Most methods for estimating factor scores are not worth mentioning because of one or another kind of technical weakness. But there are two methods that are worthy of consideration for practical applications in EFA where factor score estimates seem needed. These are called *regression estimates* and *Bartlett* (also, *maximum likelihood*) *estimates* of factor scores, and both are easily computed in the context of IFA. Recalling that D^2 was defined as the diagonal of the inverse of the correlation matrix, now suppose the initial data matrix has been centered and scaled as Z where Z'Z = R; then, using the notation given earlier in the discussion of IFA, Bartlett estimates of factor scores can be computed as $X_{m-\text{Bartlett}} = Z\ D\ Q_m\ (\Gamma_m - \phi\ I)^{-1/2}$. The discerning reader may recognize that these factor scores estimates can be further simplified using the singular value decomposition of matrix Z D; indeed, these score estimates are just rescaled versions of the first *m* principal components of Z D. Regression estimates, in turn, are further column rescalings of the same *m* columns in $X_{m-\text{Bartlett}}$. MLFA factor score estimates are easily computed, but to discuss them goes beyond our scope; see [15]. Rotated or transformed versions of factor score estimates are also not complicated; the reader can go to **factor score estimation** (**FSE**) for details.

EFA in Practice: Some Guidelines and Resources

Software packages such as CEFA [3], which implements MLFA as well as geomin among other methods, and Hyball [18], can be downloaded from the web without cost, and they facilitate use of most of the methods for factor extraction as well as factor transformation. These packages are based on modern methods, they are comprehensive, and they tend to offer advantages that most commercial software for EFA do not. What these methods lack, to some extent, is mechanisms to facilitate modern graphical displays. Splus and R software, the latter of which is also freely available from the web [r-project.org], provide excellent modern graphical methods as well as a number of functions to implement many of the methods available in CEFA, and several in Hyball. A small function for IFA is provided at the end of this article; it works in both R and Splus. In general, however, no one source provides all methods, mechanisms, and management capabilities for a fully operational EFA system – nor should this be expected since what one specialist means by 'fully operational' necessarily differs from that of others.

Nearly all real-life applications of EFA require decisions bearing on how and how many cases are selected, how variables are to be selected and transformed to help ensure approximate linearity between variates; next, choices about factoring algorithms or methods, the number(s) of common factors and factor transformation methods must be made. That there be no notably weak links in this chain is important if an EFA project is to be most informative. Virtually all questions are contextually bound, but the literature of EFA can provide guidance at every step.

Major references on EFA application, such as that of Carroll [4], point up many of the possibilities and a perspective on related issues. Carroll suggests that special value can come from side-by-side analyses of the same data using EFA methods and those based on structural equation modeling (SEM). McDonald [15] discusses EFA methods in relation to SEM. Several authors have made connections between EFA and other multivariate methods such as basic regression; see [14, 17] for examples.

– – an S function for Image Factor Analysis – –

```
'ifa'<-function(rr,mm) {
# routine is based on image factor
# analysis;
# it generates an unrotated common
# factor coefficients matrix & a scree
# plot; in R, follow w/ promax or
# varimax; in Splus follow w/ rotate.
# rr is taken to be symmetric matrix
# of correlations or covariances;
# mm is no. of factors. For additional
# functions or assistance, contact:
# rpruzek@uamail.albany.edu
     rinv <- solve(rr) #takes inverse
     # of R; so R must be nonsingular
     sm2i <- diag(rinv)
     smrt <- sqrt(sm2i)
     dsmrt <- diag(smrt)
     rsr <- dsmrt %*% rr %*% dsmrt
     reig <- eigen(rsr, sym = T)
     vlamd <- reig$va
     vlamdm <- vlamd[1:mm]
     qqm <- as.matrix(reig$ve[, 1:mm])
     theta <- mean(vlamd[(mm + 1)
     :nrow(qqm)])
     dg <- sqrt(vlamdm - theta)
     if(mm == 1)
         fac <- dg[1] * diag(1/smrt)
         %*% qqm
     else fac <- diag(1/smrt) %*% qqm
     %*% diag(dg)
```

```
    plot(1:nrow(rr), vlamd, type
    = "o")
    abline(h = theta, lty = 3)
    title("Scree plot for IFA")
    print("Common factor coefficients
    matrix is: fac")
    print(fac)
    list(vlamd = vlamd, theta = theta,
    fac = fac)
}
```

References

[1] Browne, M.W. (1968). A comparison of factor analytic techniques, *Psychometrika* **33**, 267–334.

[2] Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis, *Multivariate Behavioral Research* **36**, 111–150.

[3] Browne, M.W., Cudeck, R., Tateneni, K. & Mels, G. (1998). CEFA: Comprehensive Exploratory Factor Analysis (computer software and manual). [`http://quantrm2.psy.ohio-state.edu/browne/`]

[4] Carroll, J.B. (1993). *Human Cognitive Abilities: A Survey of Factor Analytic Studies*, Cambridge University Press, New York.

[5] Cattell, R.B. (1966). The scree test for the number of factors, *Multivariate Behavioral Research* **1**, 245–276.

[6] Darlington, R. (2000). Factor Analysis (Instructional Essay on Factor Analysis). [`http://comp9.psych.cornell.edu/Darlington/factor.htm`]

[7] Davenport, M. & Studdert-Kennedy, G. (1972). The statistical analysis of aesthetic judgement: an exploration, *Applied Statistics* **21**, 324–333.

[8] Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research, *Psychological Methods* **3**, 272–299.

[9] Hendrickson, A.E. & White, P.O. (1964). PROMAX: a quick method for transformation to simple structure, *Brit. Jour. of Statistical Psychology* **17**, 65–70.

[10] Jennrich, R.I. (1973). Standard errors for obliquely rotated factor loadings, *Psychometrika* **38**, 593–604.

[11] Jennrich, R.I. (1974). On the stability of rotated factor loadings: the Wexler phenomenon, *Brit. J. Math. Stat. Psychology* **26**, 167–176.

[12] Jennrich, R.I. & Sampson, P.F. (1966). Rotation for simple loadings, *Psychometrika* **31**, 313–323.

[13] Jöreskog, K.G. (1969). Efficient estimation in image factor analysis, *Psychometrika* **34**, 51–75.

[14] Lawley, D.N. & Maxwell, A.E. (1973). Regression and factor analysis, *Biometrika* **60**, 331–338.

[15] McDonald, R.P. (1984). *Factor Analysis and Related Methods*, Lawrence Erlbaum Associates, Hillsdale.

[16] Preacher, K.J. & MacCallum, R.C. (2003). Repairing Tom Swift's electric factor analysis machine, *Understanding Statistics* **2**, 13–43. [`http://www.geocities.com/Athens/Acropolis/8950/tomswift.pdf`]

[17] Pruzek, R.M. & Lepak, G.M. (1992). Weighted structural regression: a broad class of adaptive methods for improving linear prediction, *Multivariate Behavioral Research* **27**, 95–130.

[18] Rozeboom, W.W. (1991). HYBALL: a method for subspace-constrained oblique factor rotation, *Multivariate Behavioral Research* **26**, 163–177. [`http://web.psych.ualberta.ca/~rozeboom/`]

[19] Rozeboom, W.W. (1992). The glory of suboptimal factor rotation: why local minima in analytic optimization of simple structure are more blessing than curse, *Multivariate Behavioral Research* **27**, 585–599.

[20] Rozeboom, W.W. (1997). Good science is abductive, not hypothetico-deductive, in *What if there were no significance tests?*, Chapter 13, L.L. Harlow, S.A. Mulaik & J.H. Steiger, eds, Lawrence Erlbaum Associates, Hillsdale, NJ.

[21] Spearman, C. (1904). General intelligence objectively determined and measured, *American Jour. of Psychology* **15**, 201–293.

[22] Thurstone, L.L. (1947). *Multiple-factor Analysis: A Development and Expansion of the Vectors of Mind*, University of Chicago Press, Chicago.

[23] Tucker, L. & MacCallum, R.C. (1997). *Exploratory factor analysis*. [unpublished, but available: `http://www.unc.edu/~rcm/book/factornew.htm`]

[24] Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S*, Springer, New York.

[25] Yates, A. (1987). *Multivariate Exploratory Data Analysis: A Perspective on Exploratory Factor Analysis*, State University of New York Press, Albany.

ROBERT PRUZEK